#150

Factors and Taxa, Traits and Types, Differences of Degree and Differences in Kind

Paul E. Meehl
University of Minnesota

ABSTRACT A taxon is a nonarbitrary class whose existence is conjectured as an empirical question, not a mere semantic convenience. Numerous taxa are known to exist in nature and society (chemical elements, biological species, organic diseases, geological strata, kinds of stars, elementary particles, races, cultures, Mendelizing mental deficiencies, major psychoses, vocations, ideologies, religions). What personality types, if any, occur in the nonpathological population remains to be researched by sophisticated methods, and cannot be settled by fiat or "dimensional" preference. The intuitive concept of taxonicity is to be explicated by a combination of formal-numerical and causal criteria. Taxometric methods should include consistency tests that provide Popperian risk of strong discorroboration. In social science, latent class methods are probably more useful than cluster algorithms.

In the academic year 1939–40, I took a course in individual differences (regularly taken by *all* psychology majors at Minnesota in those days) from one of the "greats" of applied psychology, Donald G. Paterson, a founder of what was then called the "student personnel movement." A main theme of Paterson's lectures was that there are no *types* of persons, that categorical terminology (e.g., "introvert," "bright," "thin") is merely a convenient—and sometimes careless—way of demarcating rough regions on what are in reality quantitative traits, dimensions, or factors. He pointed out, for instance, that Jung had a typology of introverts and extroverts, but then had to add ambiverts,

Requests for reprints should be addressed to Paul E. Meehl, Department of Psychology, University of Minnesota, N218 Elliott Hall, 75 East River Road, Minneapolis, MN 55455-0344.

Journal of Personality 60, 117-74, 1992.

which is where most of us are on a bell-shaped curve of the dimension introversion-extroversion. Paterson said that there was a marked difference between American and European psychologists in this respect, the Europeans being fond of typology and the Americans, with some exceptions, suspicious of the concept. Big typological names in Europe were psychiatrists Kretschmer and Jung, and to some extent Freud; and psychologists such as Jaensch and Klages. Paterson did allow for exceptions to his doctrine, especially in the area of intelligence. There is an association between incidence of the higher levels of mental deficiency or borderline IQ and lower social class, whereas much lower IQ levels (in the idiot and imbecile ranges) are independent of parental SES. Typological language in the upper levels (e.g., 'moron,' 'borderline deficient,' 'dull-normal') he considered analogous to terms like 'introvert,' 'domineering,' or 'thin,' having no true typological or taxonomic significance. Moron, borderline, and dull-normal children he held to be simply the low end of the normal distribution of the polygenic determiners of *g*; whereas he thought the lowest IQ groups represented Mendelizing, karyotypic, or developmental anomalies, similar in that respect to the valid typologies of organic medicine. On this view the reason for the association of the nontypological mental deficiencies with social class was the transmission of low IQ polygenes from parents, plus a (slight, he thought) influence of poor environmental stimulation in the home.

These anti-typological views were almost universally held among the great American test builders and psychometricians such as Terman and Thurstone (but not Cattell), and are strongly represented today by Eysenck (1986). Eysenck properly emphasizes the importance of the shift from category thinking to dimensional thinking in the development of post-Galilean science, a contrast also stressed by Lewin (1931) in a classic paper. It is widely agreed by historians and philosophers of science that one of the respects in which post-Galilean science was superior to medieval science was the replacement of categorical, "essentialist" ways of conceptualizing the world by quantitative, dimensional modes of thought. Some consider this change as important as the invention of measuring instruments, the development of a powerful formalism (analytic geometry, calculus, probability theory), the liquidation of aesthetic and evaluative considerations, and the development of "corpuscularism" instead of essentialist phenomenally grouped properties (cf. Robert Boyle, 1667). Scientists replaced essences, substances, and occult qualities as the basic concepts with quantitative aspects (size, shape, local motion) of invisible

`

corpuscular matter, a way of thinking which is so natural to us from the general science we learned in secondary school that we hardly realize what a revolutionary change it represented.

However, one must be careful not to overdo this historical observation by turning it into a methodological dogma, closed to empirical evidence. Post-Galilean science retained a notion of categories in biology, the great contrib.-utor being Linnaeus. And while the medieval alchemists' occult qualities were liquidated, the post-Galilean chemical elements are obviously taxa, despite the fact that their indicators (e.g., density, valence, melting point) are quantitative. Even some of the quantitative properties go by discrete steps: atomic weight and atomic number are numerical values, but not continuous. In modern geology one learns about taxa such as the categories of igneous or sedimentary rocks, and the categorical labeling of geologic strata. In astronomy, nebulae (several species), clusters, stars, planets, comets, and moons are discrete taxa. Eysenck and others who totally reject the notion of taxa in psychopathology get too much methodological mileage from an admittedly important fact about the history of post-Galilean science. Whether or not the entities, properties, and processes of a particular domain (such as psychopathology, or vocational interest patterns) are purely dimensional, or are instead a mix of dimensional and taxonic relations, *is an empirical question*, not to be settled by a methodological dogma about "how science works." Among psychologists, particularly those who have a trade union animus against psychiatrists, and sometimes an extremely dogmatic treatment of the so called "medical model" (which few have thought deeply about, even for organic medicine, see Meehl, 1972a, pp. 20-24 [1973a, pp. 194-199], 1973a, pp. 284-289), one hears remarks such as, "Of course there are no entities, comparable to mumps or measles, in psychopathology." The *of course* is unwarranted.

The term 'taxonomic,' if used precisely, is a metaconcept referring to the theory or method of classification, hence I shall employ the neologism *taxonic* in the object language, designating the physical state of affairs that constitutes a real taxon. Since the question "Is the domain taxonic?" is an empirical question, how do we approach it? We cannot do so without some notion, however rough, of what taxonicity is. Unfortunately I cannot provide a rigorous definition of 'taxon.' I will, however, offer an explication that is contextual and will do for our purposes here (cf. Carnap, 1945). At this stage of taxometric analysis, we can get on with the mathematics and the empirical research without a rigorous definition of the *term* 'taxon,' just as we can carry on classical

psychometrics and factor analysis without a rigorous definition of a "real factor." (When is a mathematically identified factor an "artifactor"?) In the early stages of any science we usually can't define the core concepts precisely, contrary to the simplistic operationism taught in beginning psychology and sociology classes. For an illuminating quote by someone who was more methodologically sophisticated than some psychologists think, see the opening paragraph of Freud's (1915/1957) "Instincts and their Vicissitudes." What logicians and philosophers of science call *open concepts* are defined implicitly via their roles in the theoretical network, and since the network is incomplete—that's why we are engaging in research, if it were complete the job would be done!—one expects the concepts to have a certain fuzziness at the edges. We know that definitions are conventional, but that does not mean they should be whimsical. Our choice in defining or explicating an open concept depends upon how the world is. Of course the *word* used is unimportant (that's the truth of conventionalism and the old positivism); but the *concept* demarcated by the word is of great importance. It would be possible to adopt a convention that units of time are measured by the Pope's pulse, but it would be very inconvenient and make scientific physics impossible. One of the profound taxonomic insights of Linnaeus was the realization that the bat doesn't sort with the bird, nor the whale with the pickerel, but that, what would be surprising to a pre-Galilean scholar (e.g., Pliny), both the bat and the whale belong with the grizzly bear.

Readers unfamiliar with taxonomy should consult Blashfield (1984), Bolz (1972), Dahlstrom (1972), Gangestad and Snyder (1985, 1991), Meehl and Golden (1982), and references cited therein. I have attempted to hold down my overlap with these works to an unavoidable minimum. 'Taxon' comes from the Greek word for an arrangement or ordering. I will attempt to explicate the concept of taxonicity in three ways, without pre-judging their relationships (but conjecturing that they have some empirical correlation over domains).

Taxonicity Explicated Roughly by a Combination of Intuition, Common Sense, Synonyms, and Nonproblematic Examples

The easiest way, quite adequate for many purposes, as in organic medicine or botany, is to settle for a commonsense intuitive notion by providing synonyms and concrete examples. We say that a taxon is a "type," a "species," a "disease entity," a "non-arbitrary category," a "*natural kind*." We say with

`

Plato that one wants to "carve nature at its joints." We speak of "differences of kind rather than degree," although that should usually be "kind *as well as* degree" or "kind *as a source of* degree," because the indicators, even of a true and clearly specified taxon, are quantitative more often than not.

We supplement these rough verbal synonyms or clarifications with such examples as species in biology (there are chipmunks, there are gophers, but there are no gophmunks), disease entities in medicine (measles, mumps, scarlet fever), elements and compounds in chemistry, ideologies in politics, trade competencies in industrial psychology, tightly knit doctrinal positions in religion. It is important for social scientists suspicious of taxa to note that some socially learned behavioral taxa are as striking, clear-cut, and closely knit—in the sense of very high pairwise correlations of the indicators—as those in organic medicine or genetics. *Example*: As an undergraduate at the University of Minnesota I had a number of friends who were Communists (I was a Norman Thomas Socialist at the time), and they divided into Trotskyists and Stalinist Communists. Minneapolis in the 1930s was, along with New York and New Jersey, one of the centers for American Trotskyism. I quickly learned that there was a pair of beliefs that, taken jointly, were pathognomonic of the "Trotskyist syndrome." If a student opined that (a) the Soviet Union is a workers' state and must be defended at all costs against anybody, including the USA and (b) Stalin is a stupid counter-revolutionary bureaucrat, one could predict—not with 90% or 95% but with 100% accuracy—that the person would also hold a dozen or more other beliefs: that the doctrine of permanent revolution is a core thesis of the class struggle and was first enunciated by Leon Trotsky; that Trotsky was second to Lenin in leading the October Revolution; that the civil war against the Whites was won largely through the military genius of Trotsky; that the proper slogan for Socialists and Communists during the rise of Hitlerism in Germany should have been "No enemy to the left"; that the slogan "Socialism in one country" is unsound; that Leon Sedov was murdered; that the accused Bukharin of the 1938 Moscow trials was innocent; and so on. The statistical tightness of the facets of the Trotskyist syndrome was greater than any nosological entity in psychopathology, and in fact tighter than most clinical syndromes in organic medicine. This example, by the way, suffices to show that taxonicity in the behavior domain need not be a matter of biological agents (germ or gene), since Trotskyism is what Cattell would call an "environmental mold" type, and has no specific etiology.

In industrial psychology one can construct a trade test to detect who is a

lathe operator that can be as short as eight or ten items, which are almost perfectly Guttman scalable and generate a clear bimodality or a U-shaped distribution in the general population. If we consider such a social taxon as *bridge-player*, a person who knows the meaning of 'renege' or 'sluff' will be sure to know the meaning of the word 'vulnerable' (Meehl & Golden, 1982, p. 139). A person who knows how to examine or manipulate a bicycle tire will know about bicycle speeds, tire properties, and sprocket repair, but we do not assume that there is much topographical overlap in these behaviors, or in the subsystems of the brain that are involved in these different molar level activities. *Example*: One exhibition booth at the Minnesota State Fair has electronic apparatus which presents theological true/false queries, such as "People earn the right to grace by performing good works." Having made an intensive study of Christian theology in my youth, I can sample two or three items and "catch on" to which of the Conservative Protestant ideologies has prepared the test; that enables me to get a nearly perfect score on all of the remaining items, despite the fact that internal (intrinsic, logical) relations between the several doctrinal statements are absent or at best tenuous. The whole class of environmental mold traits and types illustrates nonbiological taxonicity (cf. Meehl, 1986b).

Perhaps the commonest way of explicating the taxon concept is to say that a genuine taxon is a *natural kind*, as contrasted with an *arbitrary class*. The connotation of 'natural kind' is that it would exist as a taxon in the perfect conceptual system of Omniscient Jones, that it is in some sense *really out there*, whether human scientists identify it or not. That way of saying it hinges upon taking a realist view of scientific concepts and theories, rather than a fictionist view as is common among psychologists; but I am a scientific realist, and I presuppose that position throughout this paper. The term 'arbitrary' does not imply that the classification is whimsical, pointless, or foolish. *Example*: A military personnel psychologist might be asked to do a computer search for all of the *short Swedish bakers* in the command, the general having in mind (rightly or wrongly) that Swedes tolerate the cold and social isolation of a certain kind of arctic duty better than Irish, Italians, or Jews, but that the installation is physically such that tall people are more likely to keep bumping their heads going through the low doorways. The conjunction of properties specified may be quite rational given certain administrative considerations, despite the fact that there may be no correlation in the army population between being a baker and being Swedish, and of course the correlation between

`

being short and Swedish is, in fact, negative. So the conjunctively defined class *short Swedish baker* is not in any sense a natural kind, does not claim to carve nature at its joints, and would not be of interest to a geneticist, ethnographer, or social psychologist. It is with respect to those disciplines an arbitrary class, but it is not a whimsical or useless one. When we demarcate an interval on a continuum or a volume region in a descriptor hyperspace of trait dimensions for some pragmatic purpose (juridical, industrial, educational, economic), its utility does not hinge on a claim that the class specified is a natural kind, that it would exist in the world as a real category if human psychologists had not had a purpose in sorting for it.

All taxa are classes, but not all classes are taxa. Those who criticize taxonomic concepts or statistical search methods from a conventionalist standpoint have failed to understand this. That all classification systems "are merely human conventions of convenience, not facts about the external world" is obviously not true of the chemical elements, nor of the elementary particles, nor of plant and animal species, nor of kinds of stones. There is no justification for appeal to that pseudosophisticated dogma when we are dealing with psychology rather than stones or chemicals. A class is defined by any conjunction or disjunction of properties as when the logician writes

$$x \in \mathrm{C} =_{\mathrm{def.}} \mathrm{Q}_1 x \,.\, \mathrm{Q}_2 x \,.\, (\mathrm{Q}_3 x \vee \mathrm{Q}_4 x)$$

and reads "Individual x is a member of class C if and only if x has properties Q_1 and Q_2 and either Q_3 or Q_4." The rubrics of *DSM-III* exemplify this way of writing it. Quantitative properties, which are what are usually involved when one looks closely enough, are always transformable into this format by writing, e.g., for the second property,

$$\mathrm{Q}_2 x \textit{ iff } a \leq f_2(x) \leq b$$

where f_2 is the logician's *functor*, so $f(x)$ does not denote a mathematical function of a numerical variable x, but rather a quantitative property (value) attributed to the individual x.

This rough, commonsense, intuitive, synonym-giving explication is all right for many purposes in practice and perhaps for some in theory. But when the taxa under investigation are loose, or the causality is murky, or when scientists or clinicians disagree about their existence or how to set them up, or when there are strong ideological influences impairing scientific objectivity, it is necessary to go past a commonsensical level (quite adequate for most

classifications of animals, stones, or organic diseases) to a more rigorous explication of the taxonicity concept. *Example*: It has been conjectured for many years that underlying the various subtypes of schizophrenia, there are only two basic taxa: the paranoids and "all the others" (schizophrenia simplex, catatonic, hebephrenic, chronic undifferentiated, whatever). This conjecture was suggested by several relations (stochastic, and some weaker than others) between such indicators as age of onset, degree of primary thought disorder, grossly inappropriate affect, schizophasic speech, weak heterosexual interest, and social withdrawal. There were also physical correlates noted by Sheldon, N. D. C. Lewis, and others (see discussion in Meehl, 1989, 1990a, 1990c). After two thirds of a century, the question of a basic taxonic dichotomy in schizophrenia is still unsettled. The trouble is that psychiatrists, and even psychologists who with their presumed psychometric and statistical sophistication should have known better, *have not recognized the need for a distinctively taxometric mathematics as a search procedure for investigating this conjecture*. If there is no taxon generating correlations between these various physical and psychological traits, but they have positive manifold, whether one gets an appearance of taxonicity by conventional *t* tests or chi squares or what—ever will depend simply upon the power function (Meehl, 1990b, 1990e). With traditional significance tests, a taxonic and a nontaxonic structure lead to the same predictions, because if the correlation of two traits is greater than zero and one selects patients who are high on trait X and contrasts them with patients who are low on that trait, one will of course get a significant difference on trait Y, regardless of whether the situation is taxonic or factorial. Millions of dollars of tax money have been wasted performing studies of this sort, studies which could not possibly answer the taxonic question (Meehl, 1990c, pp. 76-79).

When Hathaway and I were writing the *Atlas* (Hathaway & Meehl, 1951) applying his code system to the analysis of MMPI profiles, we were struck by the great variation in the frequencies of various two-digit codes such as 49′, 27′, and 31′; their incidence was sometimes an order of magnitude larger than certain highly infrequent patterns (such as 19′). Some rough computations (never published) showed that the incidence of these pairs of deviant scores vastly exceeded what could be predicted from the zero-order Pearson *r* in either the normal or psychiatric population. Some of this was attributable to differences in scale length (not compensated for by the linear T transformation, as some mistakenly believe), but most of it could not be so interpreted.

`

This fits our clinical experience nonpsychometrically, that there are such *types* as the hard-core solid-gold psychopath (49′), the hysteric (31′), or the anxious and depressed dysthymic syndrome (27′). *Example*: When Robert R. Sears (1943) and others reported significant correlations between the three aspects of Freud's conjectured anal character (orderliness, stubborness, and parsimony), critics argued that the correlations, while statistically significant, were not very large and hence did not support Freud. This line of criticism is defective because it treats a taxonomic problem like a factorial one, ignoring the expected size of correlations generated by a latent taxonic situation. The initial sentence of Freud's classic paper reads: "Among those whom we try to help by our psycho-analytic efforts we often come across a type of person who is marked by the possession of a certain set of character-traits, while at the same time our attention is drawn to the behavior in his childhood of one of his bodily functions and the organ concerned in it" (Freud, 1908/1959, p. 169). Freud does not say, or imply, that the three traits—orderliness, stubborness, parsimony—seen as quantitative dimensions are markedly correlated among persons in general; what he says is that we often come across a *type of person* marked by (extreme degrees of) these traits. Assume that there is an anal taxon that has a base rate $P = .10$ in the general population, and that the three traits—orderliness, stubborness, parsimony—treated quantitatively each have a 1.75 *SD* separation of the anal taxon from the "normal" complement class. What is the expected value of the Pearson *r* in an unselected population of college students? It turns out to be .28,[1] lower than the values reported by Sears.

Some features of a single manifest frequency distribution are suggestive of latent taxonicity but cannot be treated as clearly demonstrating it. Bimodality is rather strongly suggestive (but see Murphy, 1964), bitangentiality (two points having identical derivatives) less so, and even skewness may indicate taxonicity, unless some other influences (e.g., an unfortunate distribution of item difficulties and inter-item phi coefficients) provide an obvious artifactual explanation. Any clear "clumping" of individuals in the descriptor space

---

[1] This predicted *r* is obtained from the General Covariance Mixture Theorem (see Meehl & Golden, 1982), a distribution-free algebraic identity that expresses the manifest covariance of two indicators in a mixed group in terms of the latent class separations and the category proportions, $\mathrm{cov}(yz) = P\,\mathrm{cov}_t(yz) + Q\,\mathrm{cov}_c(yz) + PQ(\bar{y}_t - \bar{y}_c)(\bar{z}_t - \bar{z}_c)$, where the subscripts 't' and 'c' refer to the taxon and complement groups. Setting the within class covariances at zero—assuming the (*yz*)-relation to be generated solely by taxonicity—we obtain from the third term: $r = (.90)(.10)(1.75)(1.75) = .28$.

is suggestive. The distribution of inter-person distances in the traditional cluster algorithms attempts to objectify this notion of clumping (see below, formal-numerical definition of taxonicity).

Taxonicity Defined by Causality

A second way of explicating the taxon concept is by reference to causality. Most of us are relatively comfortable with the idea of a taxon such as a disease entity in organic medicine that has been empirically shown to arise from a specific identifiable germ, major gene, or event. A germ merely moves the problem back to the species problem in microbiology, but since that situation is much easier to resolve, the movement helps. In organic medicine a disease entity is specified by a *conjunction* of pathology and etiology (see Meehl, 1973a, the diagram on page 287 and associated text). If we have distinguishable etiologies producing substantially the same pathology, we consider them separate diseases; likewise, we consider them separate diseases if a single etiological agent gives rise through divergent causal pathways to very different pathologies and sometimes nonoverlapping symptomatologies. The symptoms of a syphilitic gumma mimic brain tumor and do not overlap with the symptoms of tertiary lues of the CNS parenchyma; and the symptoms of tabes are in turn very different from those of paresis (same specific etiology, different pathology). If we did not have the history of luetic infection and the biochemical tests to inform us, we would have no reason for grouping these three diseases together. The symptoms of different pneumonias are similar, as are the symptoms of the various meningitides. Whether the patient has dyphtheritic meningitis or tuberculous meningitis or meningococcal meningitis, the important thing is the neurological consequence of inflamatory process in the meninges, whatever germ is responsible. But pneumococcal meningitis is a different disease from a tuberculous meningitis, the division being on the basis of the specific etiology. Potts' disease is very different from pulmonary tuberculosis, and both are very different from tuberculous meningitis, despite the specific etiological agent being Koch's bacillus in all three instances. For a more detailed discussion of these lines of thought see Meehl (1972a, 1977) and Meehl and Golden (1982). In the 1977 paper I present a dozen meanings of strong causal influence ranging from very strong to so weak that nobody would consider them examples of specific etiology. Scholarly clinicians I have polled

`

allow only the strongest three influences as clear cases of specific etiology: (1) an intrinsically dichotomous qualitative causal factor necessary and sufficient for the disease (e.g., dominant mutation → Huntington disease); (2) a sine qua non but not sufficient factor (e.g., schizogene → schizophrenia on my theory); or (3) a quantitative factor of which disease probability is a quasi-step-function with p(Disease) = 0 below the step (e.g., ascorbic acid deficiency → scurvy), but the probability to the right of the step may rise or remain flat with an increase in the etiological factor.

Open-minded empirical exploration of taxonic possibilities will, I am sure, lead to some interesting surprises. *Example*Ì: We do not usually consider *severity* of a specific organic disease as taxonic, rather we view severity as involving one or more quantitative components within a taxon. But Hoagland (1947), in a fascinating study, showed—relying on the chemist's Arrhenius equation for reaction rate as a function of temperature—that log α EEG frequency plotted against T° during diathermy identified early, intermediate, and advanced paresis as discrete categories. The conjectured explanation is in terms of a step-function effect due to different cerebral enzymes being inactivated at different stages of the paretic process's progression.

Formal-Numerical Definition of Taxonicity

A third general way of explicating taxonicity may be labeled *formal-numerical*. It relies upon *implicit* (*contextual*) *definition*, probably the commonest kind of "definition" in empirical science, as it is in the formal sciences of pure logic and mathematics. Metatheoretical considerations and history of science show that explicit definitions via observables (the "operational definitions" still peddled as mandatory in some beginning psychology and sociology courses) are in fact rare. Explicit definition via other theoretical terms takes place by reduction in the Comte Pyramid of the Sciences (Meehl, 1990d), but usually this happens only in advanced states of scientific knowledge; and of course the problem then recurs for the definitions of the defining terms in that theoretical reduction. Very common is implicit definition via the mathematics; but the mathematics, the formalism, must be supplemented by an embedding text, and the latter can be divided into *operational text* and *interpretive text* (Meehl, 1990d, 1990e). I call this explication of taxonicity 'formal-numerical' because the formalism asserts postulates regarding the latent situation that are not themselves theorems of

general mathematics, and they could be false in the physical world when interpreted by the embedding text; the numerical values of the variables are given by the empirical facts which, when substituted appropriately into the formalism, will either fit the structure or not.

Ultimately the somewhat vague word 'taxon' will be explicated by a combination of causal and formal criteria. This joint causal-formal explication, which future research may succeed in making precise, will presumably permit a definition in the old fashioned rigorous sense of the concept *define*, which will be seen to correspond moderately well with the intuitive, commonsense semantics above. It cannot be required to correspond exactly, because the present intuitive semantics of 'taxon' are themselves fuzzy at the edges. Taxonicity is a nice example of the logician's *open concepts* (Pap, 1953, 1958, pp. 302-360, 1962, chapter 3; Meehl, 1972a, references on p. 21 [1973, p. 195]). As Carnap (1945) pointed out in his original discussion of an explication in semantics, we normally anticipate that there will be borderline cases, usually but not invariably discerned as such under the aegis of our intuitive preanalytic notion of a concept, so that one's intuitions are not badly offended by a classification in either direction ("in" or "out") by the more articulate explication. A good explication, in addition to assorting cases that would previously have been considered borderline, or on which intuitions disagreed, may even reclassify some instances in which our prior intuitions were strong and consensual. Such reclassification is not fatal, because if we had thought the intuitions were perfect we wouldn't have needed to explicate them, only to spell them out carefully. However, in these new counterintuitive reclassifications a good theoretical understanding should enable us to grasp *why* the old intuitions were "mistaken." There is no necessity, and no point, scientifically, to become involved in a semantic dispute along these lines. What is important is the development of theoretical understanding. If we know with high confidence what the epistemological and causal situation is, the words do not matter, except for scientists motivated by tendentious ideology.

The French mathematician d'Alembert, when other mathematicians fretted about Bishop Berkeley's attack upon the calculus (roughly, "How can an 'infinitesimal' be an infinitely small chunk and still be real, still exist as a chunk?", not satisfactorily answered until Cauchy, Weierstrass, Dedekind and company over a century later!), said reassuringly, "Go forward, and faith will come to you." This is good advice. Physical scientists did not suspend the process of taking derivatives and integrating functions pending rigorous treat-

`

ment of the infinitesimal. To have done so because of Berkeley's searching criticisms, which Newton's defenders did not satisfactorily answer, would have retarded the development of astronomy, physics, and chemistry by many years. We will explicate 'taxon' by conducting taxometrics on a large number of real empirical research problems distributed over a variety of domains, together with mathematical and epistemological analysis.

By *taxometrics* I mean any formal (mechanical, mathematical, algorithmic) method for (1) identifying taxa and, having done that, (2) classifying individual entities into or out of the taxa. The term 'classification' covers both of these procedures, but the first is taxonomic and the second diagnostic (cf. Meehl & Golden, 1982). While the conceptual distinction between the *detection of the taxon* and the *assignment of individuals* to it (or to the complement class) is logically important, taxometric methods differ as to whether they carry out these operations sequentially or simultaneously. I prefer the neologism 'taxometrics' to the more familiar 'numerical taxonomy' for several reasons. First, a minor point, it's shorter. Second, it lends itself to adjectival form whereas the commoner terminology does not (similarly to why *schizophrenia* is handier than the older *dementia praecox*). Third, 'taxometrics' is verbally analogous to other one-word creations in which the first half of the term refers to the empirical domain and the second half indicates that it is done "metrically," that is, by measurement (e.g., psychometrics, econometrics, scientometrics, anthropometrics, cliometrics, biometrics, technometrics). Fourth—the most important consideration—because of the major contributions and high visibility of Sokal and Sneath (1963; Sneath & Sokal, 1973), whose treatise may be considered the ground breaking and now standard treatment, their view of the matter is strongly identified with the term 'numerical taxonomy.' In the controversy among biologists concerning the value of numerical taxonomy, which as yet receives allegiance from only a small minority of taxonomists in zoology and botany (cf. Hull 1988, personal communication, August 6, 1990), one bone of contention is the strong thesis of numerical taxonomy as viewed by Sneath and Sokal, to wit, that all of the respects in which entities differ appreciably from one another should be thrown into the statistical pot without regard for a priori theoretical considerations which might lead to excluding most variables as of little taxonomic relevance. In that view, it is not merely that theoretical considerations are unnecessary as a basis for selecting the set of attributes to be treated statistically in the taxonomic search process, but the employment of theory for that purpose, such

as an initial assigning of weights to possible indicators, is actively discouraged. For scientists having a nodding acquaintance with the numerical taxonomy domain, the phrase has a strong connotation of the Sneath-Sokal position, a position which I do not share for psychopathology. (I express no opinion as to its validity in other domains, e.g., entomology, geology.) Therefore I prefer the more neutral term 'taxometrics.'

There are two general approaches to taxometrics: cluster methods and latent class methods. Cluster methods have conventionally been divided into agglutinative and divisive, the former by far the preponderant and more researched to date. The essential feature of cluster methods is that one locates the individuals (honey bees, flowers, stones, mental patients, politicians) in a descriptor hyperspace, computes some appropriate distance measure (or its complement, a similarity measure) between the individuals, and then operates upon this matrix of inter-individual distances with an algorithm that sorts the individuals into groups. The latent class methods do not begin by asking how similar two individuals are when compared on a set of attributes, but employ some other statistical consequence of postulates regarding the latent situation. Cluster algorithms have not proved themselves as powerful as was hoped, and it is probably fair to say that there is no domain of empirical science whose practitioners would give a majority vote to cluster methods as the preferred way of parsing the facts. While there are over 150 cluster methods to be found in the literature, differing as to the metric for separation and then the algorithm applied to the matrix of such separation measures, over 95% of research studies rely upon a favorite half dozen (Blashfield, 1976, 1984; Golden & Meehl, 1980).

Two grave defects in the cluster methods are that they always yield clusters, their "stopping" rules not being effective (Grove, 1991a), and that they do not provide multiple "risky" consistency tests to reassure one as to the taxonic structure and the accuracy of inferences about latent values. We need stronger methods deriving from a mathematical formulation of the latent structure. The rationale of a search method should spring from one's explication of the taxonicity concept, however loose and approximate it must be at this time. It is a weakness of the six conventional cluster algorithms that they do not do this, although Ward's method, which on available evidence appears to be the best or one of the two best, comes closest to having a mathematical structural rationale. The cluster methods are best viewed as "plausible" numerifications of the intuitive notion of clumping in the phenotypic descriptor hyperspace.

`

It is not my purpose here to analyze the alleged deficiencies of cluster methods, or engage in a debate with their advocates. I only want to say that it is premature for psychologists working in personality and psychopathology to reject taxometrics generically because of what they perceive as deficiencies of cluster algorithms, since that is only one way of going about the taxometric task, and perhaps not the best one. It is well known that different cluster methods applied to the same data set (real data or artificial data in Monte Carlo runs) do not agree with one another as much as we should hope if we believe that they are all "different ways of getting at the same thing" (see, e.g., Blashfield, 1976; Golden & Meehl, 1980; but see Borgen, 1983). It must be admitted as a historical fact, whatever the mathematical or philosophical arguments may be, that no psychiatric entity or organic medical disease has been first discovered by means of cluster algorithms (Meehl, 1979). It is also a social fact that cluster methods have not succeeded in converting the majority of biologists, or the majority of physicians, to their use. Since I am acutely conscious of the extent to which psychologists in my own field (clinical psychology) can persist in a biased refusal to accept clear facts and strong arguments over more than a generation of accumulating consistent evidence (cf. Dawes, Faust, & Meehl, 1989; Meehl, 1986c), I do not attach as much weight to this social fact as some might. I only say that a busy scholar in the personality domain could hardly dismiss it as totally irrelevant that cluster methods have not demonstrated such clear cut advantages in biology and medicine as to command the assent of most workers in those fields. But it is not my aim to attack or defend cluster versus latent class methods, despite my own preference for the latter. "The proof of the puddingÛa2Û\5Ûa2," and personologists will not obtain proof one way or the other if we refrain from trying taxometric methods because of anti-taxonomic dogmas (see current objections to taxonicity, below).

A Better Method for Detecting Taxonicity

In classroom instruction I often say that "a taxon is a class whose quantitative indicators behave taxonically in the sense of Meehl's taxometric procedures." If this personalization bothers some students, I tell them they can adopt Korzybski's convention and use the word 'taxon' with personalized subscripts, thus 'taxon_{PM}.' For this to be appropriate I must show: (*a*) the several Meehl procedures tend strongly to agree as to whether a latent structure is taxonic and, within statistical tolerances, as to the latent parameters; (*b*) those classes

that are generally agreed by scientists in various areas to be taxonic (e.g. biological species, organic diseases with a specific etiology, phenotypic syndrome produced by a major gene) are detected as such by the procedures; (*c*) Monte Carlo results generated by artificial data samples with either taxonic or nontaxonic (factorial, dimensional) structure are properly distinguished with high trustworthiness.

My approach conjectures a latent class situation in which a set of fallible quantitative indicators each discriminates the latent taxon from the complement class. We have two unimodal overlapping distributions, the taxon class and the complement class, so the probability of taxon membership for an individual *i* is a monotone increasing function of his score x_i on the indicator. We make as few additional assumptions as possible. For example, we assume nothing as to the factorial purity, scalability, or Cronbach α of the *x*-indicator. As regards α, given a fixed set of item difficulties $\{p_i\}$ for the taxon (mean difficulty = $\overline{p}_t$) and complement (mean difficulty = $\overline{p}_c$), the mean score difference on the *m*-item scale is $\overline{d} = m(\overline{p}_t - \overline{p}_c)$ and the score variance within each class (distribution-free algebraic identity!) is

$$\sigma^2 = \sum^{m} p_i q_i + 2 \Sigma \, \mathrm{cov}_{ij}$$

where cov_{ij} = numerator of an inter-item ϕ-coefficient. Hence the derived taxon separation is increased (group overlap decreased) by *reducing* the mean ϕ-coefficient within each class, of which α is a monotone function. *Moral*: Taxometric research sometimes requires one to think differently from classical psychometrics about the same statistical formalism. (A referee for this paper reminds us that classical psychometrics also requires that there be no correlation between items in a subgroup of individuals with identical true scores.) We proceed to derive as many nonequivalent theorems as possible. (How they can be "nonequivalent" and yet be valid consequences of the postulates characterizing the latent situation as taxonic will be discussed below.)

My approach differs greatly from the conventional one of psychostatisticians in the broadly "Fisherian" tradition, in that I lay heavy emphasis upon approximate empirical agreement between numerical inferences made from different procedures and indicators, rather than being concerned with a kind of mathematical precision and optimality that is in any case illusory in the social sciences. If asked to defend this emphasis, I could do it both mathematically

`

and philosophically, but the short answer is from the history of science. Astronomy, physics, chemistry, and portions of geology and physiology were in an advanced scientific state of quantification—more advanced than almost any area of social science is today—before R. A. Fisher was born. As I read the scientific record, the most persuasive evidence for theories concerning theoretical entities, or events and processes observable in principle but not observed because of spatial or temporal remoteness from the scientist, lay in finding *consistent numerical results via qualitatively disparate epistemic paths* (cf. Meehl, 1978; 1990e). While there may perhaps be legitimate philosophical disagreements about this stance (I myself doubt even that), there cannot be any disagreement about the plain historical fact.

Taking convergence of evidence as fundamental to solving the taxometric problem, several kinds of consistency tests can be applied to a data set in an effort to satisfy ourselves whether the latent structure is taxonic or not. We can take estimates of the same latent numerical values (I shall call them "parameters" as is customary, although I'm not happy with the application of that term in this context) reached by different computational procedures which are formally nonequivalent. We can also apply a particular procedure to different conjectured indicators of the taxon. For example, the taxon base rate *P*, the first "big important parameter" in any taxometric study, should be the same (within tolerances) when inferred from different indicators, despite the fact that the indicators may vary greatly in their taxonic separations. Combining these two approaches, the results from different theorems applied to different indicator sets should agree within tolerances. The tolerances are set by a combination of analytical derivation and Monte Carlo runs. While I have no objection to showing that a certain taxometric estimator is a MLE, I am not compulsive on that score, and would be much more distressed if it turned out that an estimator has a bias which we cannot adequately correct for.

Specifically, I advocate a system of *coherent cut kinetics* (Meehl & Yonce, in preparation), which uses a succession of procedures to detect taxonicity, estimate latent parameters, and assign individual elements to either the taxon or the complement class. The phrase *coherent cut kinetics* refers to the epistemology and the mathematics of the approach; one moves cuts along an "input" variable and examines the statistical behavior of other ("output") variables on cases in the region of the cut or in regions demarcated by the cut. Inferring latent parameters (base rates, means, valid and false positive rates),

we test the model and the numerical values for consistency over (a) different variables and (b) different procedures. Hence, 'kinetics' because the cuts move, 'coherent' because the inferences should be consistent.

For example, in the MAXCOV (MAXimum COVariance) procedure, we look at the subjects in successive intervals along an input variable x. We get the covariance of two other scores y and z for the cases within each of the x-intervals and plot those covariances. If the underlying structure is taxonic, the curve will be peaked in the middle (for a base rate of $P = .50$; the peak moves to the right as the base rate becomes less); if the latent structure is not taxonic, the curve will be flat. Relying on the General Covariance Mixture Theorem (see footnote 1), we solve quadratic equations in the x-intervals to get the taxon frequencies per interval, and with further manipulations we infer the base rate P, hitmax cut, valid and false positive rates, etc. If we have only one psychometric indicator, item-pairs can be used as output, with covariances $p_{ij} - p_i p_j$ (numerator of ϕ-coefficient).

Using another procedure—MAMBAC ("Mean Above Minus Mean Below A Cut")—we take successive cuts on an input variable x; and at each cut we calculate for an output variable y the mean for subjects above the cut minus the mean for subjects below the cut. When these differences $\overline{d}_y(x) = \overline{y}_a(x) - \overline{y}_b(x)$ are plotted, a latent taxonic structure with a base rate of P .50 gives us a curve bowed in the middle (or shifted to the right if the base rate is lower), like an inverted saucer; if the latent structure is not taxonic, the curve will be dish-shaped. Calculations on the values of $\overline{d}_y(x)$ at the ends of the input distribution yield estimates of the base rate and latent means.

To use the coherent cut kinetics procedures *optimally* requires measures on at least four continuously distributed indicators. Each variable (or indicator) should be selected (1) to give good separation between the complement and taxon groups (good validity) and (2) to be uncorrelated within either the taxon or complement group with the other variables being used (no nuisance covariance). The selection of variables is in the context of discovery (Reichenbach, 1938) and is bootstrapsed via the procedures. The researcher cannot know the validity of an indicator beforehand. We must rely on clinical experience, nontaxonic research (e.g., fallible nosologic diagnosis), even theory, to suggest good candidate indicators. But we are not "relying" on these things in the strong sense of "having to *assume*" them in order to justify using the indicators they suggest. The procedures themselves will help us to determine how good the indicators are that we have selected, and they will

`

help us pick those that are better at discriminating an underlying taxon if one exists. (For more information about particular procedures see Golden & Meehl, 1973b, 1978; Golden, Tyan, & Meehl, 1974a, 1974b, 1974c; Meehl, 1965, 1968, 1973b, 1989, 1995; Meehl & Golden, 1982; Meehl & Yonce, in preparation). Variants of the basic procedures have been applied in real data/real problem or real data/pseudo-problem situations to detect: biological sex (Golden & Meehl, 1973a, ~~1980~~; Meehl, Lykken, Burdick, & Schoener, 1969); borderline personality disorder (Trull, Widiger, & Guthrie, 1990); childhood nephrosis (Golden & Freeman, 1983); children at risk for schizophrenia (Erlenmeyer-Kimling, Golden, & Cornblatt, 1989); dementia (Golden, 1982); depressive syndrome (Grove, Andreasen, Young, Keller, Endicott, Lavori, Hirschfeld, & Reich, 1987); neonatal brain dysfunction (Golden, Vaughan, Kurtzberg, & McCarton, 1988); schizotypy (Golden & Meehl, 1974, 1979; Lenzenweger & Korfine, 1991; Lowrie & Raulin, 1990); self-monitoring (Gangestad & Snyder, 1985); tardive dyskinesia (Golden, Campbell, & Perry, 1987); Type A syndrome (Strube, 1989).

One respect in which my taxometric approach differs from familiar ways of thinking is my emphasis that in the empirical testing (subjecting to danger of falsification) of a substantive theory, one makes use of *auxiliary conjectures* without which the logical or mathematical derivation of numerical predictions does not go through (Meehl, 1978, 1990b, 1990d, 1990e). In conventional statistics classes, psychology students learn to call auxiliary conjectures *assumptions*, the standard terminology of statisticians. From a neo-Popperian standpoint this language is not optimal, and I therefore do not employ it. In most statistical reasoning there are two kinds of "assumptions," and my conception of an auxiliary conjecture fits neither one. The first kind of statistician's assumption is an ancillary statement that one requires in order for a certain statistical procedure, say a *t* test, to be appropriate; and that ancillary assumption is subject to fairly direct test. For example, one says that the Pearson *r* is not an optimal descriptive statistic if the bivariate distribution is not linear, homoscedastic, and normal within arrays and in the marginals. (Of course *r* is computable, and will provide the correct value of the residual sum of squares, even if those assumptions are not met; but if none of them is met, it's a poor descriptive statistic.) All three of these assumptions are directly testable, with the usual setting up of confidences or significance levels, by examination of the data at the manifest level. The second kind of assumption is one that we cannot directly test but which we must simply postulate, in the

sense that we hope to God it's true, because if it isn't, what we are going to compute next will not be warranted and our substantive conclusions, based upon results we get from what we do statistically, will be invalid. For example, if I get a skewed distribution on some kind of cognitive function, I may be willing to "assume" that the underlying ability tapped by my test is essentially Gaussian, and that the skewness is attributable to some unfortunate combination of the distribution of item difficulties and the inter-item phi-coefficients. In that case, if my aim is to estimate the relation between the inferred ability dimension and some other variable, I make a nonlinear (Gaussian integral) transformation of the raw scores. I am not criticizing such procedures, but merely pointing out that neither is what I mean by 'auxiliary taxometric conjectures.' My approach considers that the *conjunction* of the main substantive conjecture of interest (e.g., that there is a latent taxon underlying the schizophrenia scale of the MMPI) with auxiliary conjectures about normality, linearity, independence within the classes, or whatever, leads to certain risky numerical predictions. If those predictions fail, we may have a problem deciding where the trouble lies, but at least we know that we cannot rely upon the derived numbers, because something is wrong with the latent model we conjectured. Instead of mere statistical significance, we make numerical point predictions. If the theory is too weak for that, we can at least make predictions about *one* set of numerical values from *another* set of numerical values *within the data*. If the tolerances (derived analytically and from Monte Carlo results) are narrow enough to make the antecedent probability of "hitting it right" quite unlikely conditional upon falsity of the conjunction, then passing the test is a corroborator of both the main theory and the auxiliaries employed (Meehl & Golden, 1982). Simply put: We hypothesize a certain latent structure underlying our data (i.e., that there is a taxon group and a complement group), and we know that if we are correct certain quantitative relations in the data are entailed by others. We examine these several relations to see if they obtain, as they will if our conjecture about the latent structure is correct.

An important feature of the reasoning in my methods is that the first step is to decide whether the latent (Cattell's "source trait") structure generating the pattern of observed results among the indicators is taxonic or factorial (quantitative). The taxon base rate P that I try to infer is the proportion of taxon members *in my particular sample*. If one knows within some tolerance ΔP what the taxon proportion is in a particular sample, then it is a simple problem in con-

`

ventional statistics to set up confidence intervals for the population value, provided the sample was drawn randomly from a specifiable physical population. In most social science research (e.g., the psychophysiology of schizophrenia) no such random sampling claim is made, which is fortunate for those research programs because, if made, it could not be plausibly defended. In my research methodology seminar I have for several years asked graduate students whether their thesis research involved a true randomization procedure in the sense specified by Fisher, and nobody's hand has been raised yet, a zero percentage in a sample of over 150 doctoral candidates. But if one *had* sampled a group of schizophrenes randomly from, let us say, all patients so diagnosed in private or public hospitals in the State of Minnesota (already an extremely biased and highly selected sample of all persons who are in fact clinically schizophrenic), then setting up a confidence interval for the base rate of some property (e.g., conjecturing that there are several distinct taxa in schizophrenia due to genetic heterogeneity) is not a complex statistical problem. It is, I must add, not a problem about which I can work up any great scientific interest. Surely the important issue here is whether there is taxonic structure indicating heterogeneity among schizophrenes carefully diagnosed. Having made that inference with confidence, the question of whether a certain subtype has a base rate of .30 or .35 in the State of Minnesota hospital system is of little interest, except maybe to an insurance actuary.

How can theorems validly derivable from the taxonic postulates concerning the conjectured latent structure be nonequivalent? The short answer is that one normally has only unidirectional derivability, at least at some stages, so that you can't go from the theorems up to the postulates and then back to the other theorems; this is a general characteristic of scientific theories, not peculiar to taxometrics. Consider a pair of expressions whose numerical equality is deriveable in my system. Given the empirical definitions of the numerified terms, some of these will be summary statistics on observables, others latent values inferred in a more complicated way from the taxonic model. Then one cannot make a *horizontal* derivation from one expression to the other. In mathematical language, we are dealing with the distinctions between an *equation*, an *equality*, and an *identity*. The theorems relating numerical values are derivable given the structural (taxonic) postulates. They are not identities of algebra, probability theory, or calculus. A philosopher of science would handle this question by making Carnap's (1939) distinction between the *general calculus* (arithmetic,

algebra, analytic geometry, probability theory, calculus, theory of functions, topology, set theory) and the *special calculus* which is obtained by adjoining certain postulates, *expressible in* the formalism but not themselves *theorems of* the formalism, intended to describe the physical state of affairs in a particular scientific theory (cf. Meehl, 1990d, 1990e).

How We Decide Whether to Trust a Taxometric Method

A researcher studying a taxonomic problem is entitled to require of a taxometric method that it bring credentials: "Before I rely on Meehl's MAXCOV-HITMAX procedure for detecting a conjectured latent taxon, what are the grounds for believing that it *works*?" For a taxometric procedure to "work" it must detect real taxa and not concoct nonexistent ones, provide accurate estimates of the taxon base rate P and the taxonic separations ($M_t - M_c$), locate optimal cutting scores and the latent hit rates p_t and p_c achieved by them, and diagnose individuals as accurately as the indicator validities permit, using Bayes's Formula on indicator patterns. Four kinds of considerations are brought to bear on the validity of a taxometric procedure, and I believe that jointly they provide as much assurance of validity as one customarily demands in the life sciences. I shall present them briefly (and hence some—what dogmatically), leaving extended discussion, with answers to objections, for another place.

First, the mathematics speaks for itself. Absent formal error, if a theorem about certain latent-manifest relations is deduced from a postulated latent taxonic structure, we know that a set of physical entities that is a model of the postulates must satisfy the theorems. But because such deducibility is usually only one-directional, we demand multiple evidentiary paths to the latent situation, such that close agreement between several nonredundant epistemic bases would be antecedently improbable—a Salmonian "damn strange coincidence"—if the postulates were grossly in error (Meehl, 1978, 1990d, 1990e; Salmon 1984, personal communication, June, 1980). For example, two nonequivalent algorithms applied to the same indicator set should yield the same estimate of P, within tolerance; the same algorithm applied to different indicator sets should give consistent results; and cross-algorithm-cross-indicator estimates should agree. Another kind of consistency test employs theorems relating some latent parameters to others, and to observed values. I cannot emphasize too

`

strongly that, in my view, *multiple consistency tests are a desideratum*. They are not merely something nice if available, frosting on the statistical cake, handy when you can get them. *Since taxometrics* (*like factor analysis*) *is a bootstraps procedure, forced to rely on the internal pattern of relations among fallible indicators* (*having no external defining criterion*), *any taxometric method that lacks consistency tests is radically defective*. Happily, the vexing rotation problem of factor analysis has no close analogue in taxometrics *when done by coherent cut kinetics*; other methods (e.g., cluster analysis) do have problems analogous to the rotation problem in factor analysis.

An important kind of *negative consistency test* discorroborates the conjecture that an apparent taxonicity has been produced artifactually by an unfortunate combination of item difficulty levels and inter-item correlations, and that the underlying source of indicator behavior is a strong quantitative factor rather than a taxon. This contribution to taxometrics was made by my (skeptical) colleague Auke Tellegen, in discussion with Gangestad, who was struggling with the pseudo-taxon problem. Briefly, the negative test (a "factorial control" Monte Carlo run) creates artificial items matched with the real data items as to difficulty level and inter-item phi-coefficients, but the item responses of individuals are assigned by a factor score model rather than a taxonic one. So far the control works beautifully, as seen in the very different graphs generated by taxonic and factorial sources (Gangestad & Snyder, 1985; Strube, 1989; Trull, Widiger, & Guthrie, 1990).

The second kind of validation of the taxonic method is Monte Carlo. Since the mathematics, if validly derived, speaks for itself, why do we need anything more to warrant trusting a taxometric procedure? Here the problem is similar to that in most other inferential statistics. The formalism idealizes the physical situation (e.g., we take derivatives to locate maxima or minima, assuming a continuity that we know does not literally obtain in the physical world). Also, the math may be intractable (or we are not clever enough) for answering certain questions formally. And we sometimes make auxiliary conjectures that are unavoidable in doing the mathematics but which are not literally true, giving rise to robustness problems. For these reasons we have recourse to Monte Carlo studies of artificial data.

A third validation method is what I call *real data, pseudo-problem* (RDPP) where we apply the procedures to a well-corroborated taxon with a defining criterion (spirochete in brain) or quasi-infallible indicator (positive CSF Wassermann plus first-zone colloidal gold curve) but delete these "gold

standard" criteria from the taxometric data base. Does the method deliver the goods? A favorite RDPP with Golden and myself is biological sex, a clear taxon, with strong but fallible MMPI indicators (of "psychological femininity"). Since we know that the MMPI items can discriminate biological sex with around 90% accuracy, if a taxometric method operating on such data cannot tell the boys from the girls, it is a pretty poor method.

I have previously viewed RDPP trials as a major validation method and one unaccountably neglected. My notion was that we should do RDPP studies in a variety of substantive domains (internal medicine, botanical classification, Mendelizing mental deficiencies, political ideologies) despite the massive amount of work involved. Robyn Mason Dawes (personal communication, 1980) and William M. Grove (personal communication, 1990) have argued against this emphasis, and so persuasively that I do not know where I stand as of this writing. They point out that any worrisome property (e.g., skewness, discontinuity, nuisance covariance) known to be found in real data with appreciable frequency can be mimicked Monte Carlo, and large numbers of samples drawn to get statistically stable relationships. "The math speaks for itself, and what it cannot answer the Monte Carlo runs will." My only rejoinder is that we do not know how to distribute artificial situations in the parameter space (rectangular? normal?); but to this Dawes and Grove point out that, without a horrendous amount of work, one cannot accurately infer the parametric densification in real world taxa either. Nor could we safely sample only a few real taxa from various research domains, because it is surely unlikely that the parameter spaces of psychiatry, social psychology, genetics, neurology, zoology, internal medicine, political science, vocational interest, religion, etc., would be densified in the same way. For pedagogical purposes, a few clear examples of successfully bootstrapsing real taxa that the skeptical clinician already knows a lot about seems sufficiently persuasive.

Finally, the long term fate of a taxometric method will depend on its ability to solve *real data real problems*. This may sound like a vicious circularity ("how can you evaluate a *method* if you are as yet in doubt about the *substantive answer*?"), but it is not. As Feyerabend says—C. I. Lewis (1929) said it first—there's nothing wrong with arguing in a circle if it's a big enough circle. Less metaphorically, empirical science is rarely a matter of linear argumentation, it is a coherency matter with certain statements (protocols) conventionally privileged although defeasible. Deciding whether the taxometric results have

`

"solved the problem" is no different from that question in nontaxometric science, and always involves a variety of considerations of differing (and, alas, somewhat subjective) weight. Suppose we are researching a conjectured taxon in psychopathology. Starting with a batch of putative indicators (suggested by clinical experience, conventional nontaxometric research, perhaps a favorite theory, cf. Meehl, 1990c, p. 45f) we conduct our taxometric analysis and conclude that there is a taxon, that its base rate in our population is P, that indicators x, y, z are good separators but candidates u, v are nondiscriminating, and so on. We can now study the subjects diagnosed "taxon members" versus the "complement class members" in various ways, looking at properties and relations not employed in the taxometrics (e.g., family members? MMPI profile types? cumulative school records? course? response to therapy? Additional physiological, neurological, anatomical features?). We can examine the taxon and complement cases as to indicator correlations, provided we are careful not to use any that are numerically forced by the taxometric procedure itself. Even the "theoretical comprehension" afforded by contemplating the indicators need not be a totally subjective matter ("now I see how they sort out, it makes psychological sense to me"). If there are enough valid indicators available, we can have one group of skilled clinicians name (or briefly characterize) the taxon by reflecting on one indicator set, and then see whether a new group of experts can match the first group's taxonic concept with the second indicator set, as in the RIT method I devised years ago for use in factor analytic interpretation (Meehl, Lykken, Schofield, & Tellegen, 1971).

The Super Bootstraps Theorem

There is a way to locate an infallible indicator by means of fallible ones. I proved the Super Bootstraps Theorem many years ago, and it has subsequently been presented in various papers (Golden & Meehl, 1978; Meehl, 1965, 1973b, 1989, 1990a, p. 303 [where the final equation should read $p_t - p_c = 1$, not zero], 1990c; Meehl & Golden, 1982); but this important theorem has been unaccountably neglected by both geneticists and social scientists. I will not present the proof (see Meehl, 1990a, Appendix, p. 303) but only summarize the theorem here. Assume a set of fallible indicators shows clear evidence of taxonicity, and consistency tests indicate that the inferred latent values (base rate, means, optimal cut, hit rates at that cut, negligible nuisance covariance)

are fairly accurate. One can define a set of "descriptor cells" by taking all configurations of the fallible indicators substituted into Bayes's Theorem, such that for every cell one has an estimate of the proportion of taxon members in it. For instance, suppose we have three fallible indicators, x, y, and z, dichotomized at the hitmax cut of each. Having computed and checked the latent valid and false positive rates achieved by each cut, we can assign all subjects to one of eight cells, depending on their indicator sign patterns, $x^+ y^+ z^+$, $x^+ y^+ z^-$, … $x^- y^- z^-$. Suppose there is a new indicator, v, that did not enter into any of this analysis, such that a suitable cut on v classifying every subject as either a v^+ or a v^- leads to a v^+ proportion $p(v^+)$ in each cell that matches the inferred taxon rate of that cell within tolerance. For this to be true, it is necessary and sufficient that when optimally cut, variable v is infallible. Hence we can locate a sign and prove that it is a perfect sign, pathognomonic (perfectly valid both as an inclusion and exclusion test, sensitivity = 1 and specificity = 1), relying on its relationship to a set of fallible indicators which are fallible both individually and collectively. Since high validity markers in the study of the behavior disorders (or loose genetic syndromes generally) are difficult to come by, I have been surprised that the Super Bootstraps Theorem has been so neglected. I conjecture that it is so counter-intuitive that people have trouble believing it. Generalizing the theorem to a fallible new sign, its validity can be estimated and may be shown to exceed the validity of an original set considered jointly (Dawes & Meehl, 1966).

"Pseudo-Taxa"

In factor analysis one ask whether a mathematically identified factor corresponds to a meaningful psychological or social dimension. If that question were meaningless, there would be no reason to be concerned with the rotation problem. In taxometrics one has the problem of pseudo-taxa, of data sets that may behave taxonically when examined by whatever taxometric method, but that are in some sense spurious, artifactual, not "real entities." This concern was raised most forcibly on the local scene by Auke Tellegen and has therefore come to be designated in Minnesota circles as the *Tellegen Case*. His original example was in the field of developmental psychopathology, but for expository purposes I will use one that is easier to explain and I think less controversial. Suppose a federal judge in his infinite wisdom determines that mentally retarded children should be mainstreamed, and to heighten the intended effect

`

specifies that they shall attend the public school in Edina, an affluent Minneapolis suburb. A psychologist might carry out a taxometric analysis, whether by the familiar cluster algorithms or by a latent class method such as mine, and would undoubtedly detect a taxonic structure in the pattern of relationships of the eleven subtests of the WISC. If the range of retardation did not include developmental, Mendelizing, or karyotypic cases but merely the low end of heritable and environmental components of *g*, this taxon identified taxometrically from the subtests would in some sense be not a "real taxon," not a natural kind, since it merely represents a judicial selection of children in the low tail of the normal curve, all administratively concentrated in one school. Is this an indictment of taxometric methods? I believe not. *No statistical method should ever be applied blindly, unthinkingly, paying no attention to anything else about the situation than the way the numbers behave*. A statistical procedure is not an automatic, mechanical truth-generating machine for producing or verifying substantive causal theories. Of course we all know that, as an abstract proposition; but psychologists are tempted to forget it in practice. (I conjecture the temptation has become stronger due to modern computers, whereby an investigator may understand a statistical procedure only enough to instruct an R.A. or computer lab personnel to "factor analyze these data.")

Would a psychologist be seriously in danger of being misled in this special situation? Would she fail to ascertain that the source of the statistically identified taxon was an administrative action rather than a biological species or an organic disease? Surely not. Finding this strange taxon at the low end of the curve, preventing the distribution from being even approximately Gaussian, she would go to the records and discover that the home addresses of almost all taxon members identified psychometrically were not in Edina but in other Twin City areas or perhaps from around the whole state. Noting this, she would question the school principal, who would tell her about the federal judge's order. Even the statistics alone could provide a strong hint, because in Minnesota an IQ < 75 is a necessary but insufficient condition for adjudication as retarded; this would produce a manifest distribution decomposable into a quasi-Gaussian component (Edina normals) and a sharply truncated tail component (the retarded).

Second, a more interesting point theoretically, whether something is a "real taxon" depends upon where we are operating in Comte's "Pyramid of the Sciences." The social scientist is not committed to identifying "real types"

as only those due to a gene or a germ. A critical environmental event will do. A highly stylized environmental mold (e.g., Trotskyism) will do. The Canadian Hutterites would show up as a statistical taxon in a heterogeneous sample of Canadians if a researcher looked at indicators of tendencies to schizophrenia and tendencies to manic-depressive disease; it would be due to the founder effect, all of the Canadian Hutterites having come from a gene pool in which the cyclothymia genome was present in at least one person who came over on the ship with the original small group; but apparently nobody in that original group of Hutterites carried the schizogene. Do we want to call this a pseudo-taxon? Well, it depends partly upon what discipline you are studying. If you are a population geneticist, an epidemiologist, or a sociologist, it is a genuine taxon, of great theoretical interest for several reasons. Two non-interbreeding populations separated by water, mountain ranges, or religion (e.g., 13th century Catholics and Jews), perhaps combined with small-sample genetic drift effects, can show statistical taxonicity of blood group indicators, anatomical features, or disease proneness (e.g., Tay-Sachs), and I see no compelling reason to label such "historically accidental" taxa as *spurious*. Using this kind of argument with our earlier example, a professor of jurisprudence or political science—perhaps even a researcher in the field of educational administration?—might consider a group defined by the complex social property "mandated by judicial order to be mainstreamed despite low IQ" as a meaningful taxon.

More generally, it is obvious without any fancy mathematics that if any attribute, whether intrinsically dichotomous (e.g., passing a particular item on an intelligence test) or made dichotomous by establishing a cutting score on a quantitative variate, is strongly correlated with each of a set of quantitative traits, then partitioning a population on the basis of this dichotomy will yield sets of partly overlapping distributions, even if the correlations between the indicators and between each indicator and the dichotomous item are generated by a latent factorial rather than taxonic structure. However, my coherent cut kinetics system insures that if you do not have such a dichotomous basis for sorting individuals into two baskets initially, you will not get an apparent taxonicity in the statistics. One of my specific procedures, called SCUD ("Sign Count U Distribution"), discriminates taxonicity from factorial origin on the basis of whether the graph of sign counts of dichotomous indicators is convex upward or downward. But is easy to show that, with very high factor loadings

`

in a nontaxonic situation, the SCUD graph will resemble the one generated taxonically. Therefore, in a domain where it is plausible to consider the possibility of very high factor loadings (they have to be as high as we get in the cognitive domain with good intelligence tests), one would be ill-advised to rely upon the SCUD procedure alone. Since in my approach one *never* relies on one of the procedures taken singly, that presents no serious problem.

The circumstances in which a nontaxonic latent structure, such as a set of factor loadings on a powerful quantitative factor, can produce an appearance of taxonicity (not via the kind of truncation or selection involved in the Tellegen case) are not clear at this time. Analytical work and Monte Carlo runs by Golden (personal communication, 1990) have shown that if all the indicators are dichotomous, there are adverse circumstances which can give an appearance of taxonicity even though the artificial data are generated by a fundamentally nontaxonic model. Consider the extreme case of a quantitative dimension, say an ability, for which it is possible to construct dichotomous test items each of which has a steep item discrimination ogive when plotted against the underlying ability score. Let us suppose further that they are closely clustered at the same difficulty level. This will result in high pairwise phi-coefficients, which in turn will mean that practically every "able" subject attains a perfect or near perfect score, or if he is below that threshold on the ability, a score of zero or near zero. This U-shaped distribution of scores is merely a psychometric artifact; it would lead one to infer the existence of a taxon where there really is none. The extent to which this is a major worry when the indicators are quantitative rather than dichotomous is currently a matter of discussion and disagreement between those involved in work on my taxometric methods (Robert Golden, William Grove, and myself). My current view is that some particular procedures, such as SCUD, could yield a misleading result, but the factor loadings have to be very high—higher than we get in most personality domains, but not (perhaps) higher than we get in some cognitive domains. I believe that some of my other procedures—e.g., MAXCOV, or MAMBAC—do not suffer from a serious danger here, but that stands in need of further investigation as of this writing.

Obviously the "interesting taxa" *to a psychologist* are not those concocted by dichotomizing a psychological or social impact property, which we would be inclined to look upon as the "output" side of the causal situation. At the psychologist's level of analysis, we look for a dichotomous fact that is plausibly assigned a *causal role* with respect to the dispositions and social niche of the

individual, such as a life history event, a certain kind of developmental milieu (e.g., schizophrenogenic mother), or a dichotomous feature of the endophenotype due to a major gene.

Contemplating the situation involving only a single indicator variable (which as a last resort can be examined taxometrically but the results should be treated with caution), an interesting possibility is to define formally useful taxa by reference to Karl Pearson's system of frequency curves (Craig, 1936; Cramér, 1946; Elderton & Johnson, 1969; Johnson & Kotz, 1970; Kenney, 1939; Pearson, 1894, 1895; Rietz, 1927; Stigler, 1986). Just as it is impossible to obtain a Gaussian distribution by summing two Gaussians (except for the trivial case of identical parameters where you are merely inflating area), it is impossible to assign proportions P to a taxon and Q to a complement class which, when multiplied by any two Pearsonian curves, yields a composite curve that is Pearsonian. A manifest frequency distribution not fittable by any of Pearson's 12 curve types, but (quasi-)decomposable into two of them (with weights P, Q assigned as required for the fit), suggests taxonicity. We should consider the possibility of giving a highly general explication of statistical taxonicity in terms of the generating conditions for members of Pearson's curve family, but I am unable to say any more about that at this time. For comments along those lines with reference to the schizotypal taxon, see Meehl (1990a).

We may ask whether a taxon, say, one in some strong meaning of the term, such as the phenotypic pattern in psychopathology produced by a dominant gene, or by a specific bacterium in organic medicine, can appear as a *factor* if one does a factor analysis, and the answer is definitely yes. This presents no hard problem. If a particular statistical method will not answer a question, the obvious thing is to employ a different statistical method that will do so. I take the same position with regard to current controversies about the genetics of schizophrenia. For example, in a recent exchange among a half-dozen of the ablest geneticists in the world on the merits and weaknesses of linkage approach to the schizophrenia genome problem (Risch, 1990a, 1990b, 1990c), we find persisting disagreement both on fundamental questions regarding the appropriate mathematical statistics and, setting those aside, on the degree to which a statistical answer can corroborate or refute a particular genetic model. My position on this matter is unconventional, but I believe it is based upon a sound philosophy of science. When there is that much disagreement among scientists both on the mathematics itself and on the genetic inferences warranted, then perhaps (at least in the immediate future) that's not the best

`

way to go about it. I hold that we have gone as far as we can go by statistical fiddling with formal diagnosis as the dependent variable, and I argue that no material advance in deciding (on that basis) among competing genetic theories of schizophrenia has been made in 20 years, since the first book by Gottesman and Shields (1972). One should employ formal diagnosis of schizophrenia solely for the purpose of locating families in which the conjectured genome is likely to be present—only *likely* because of genetic heterogeneity. The measures obtained on the relatives of these schizophrenic probands should be of the psychophysiological, soft neurological, and perhaps anatomical kind, with a bit of psychometrics and perceptual functioning included in the indicator list. Suppose it has been clearly shown that the unaffected MZ twins of schizophrenes deviate from normal controls quite markedly on these indicators, as do schizophrenes in remission, but not, for instance, manic-depressives in remission. Let p_s and p_n be the valid and false positive rates of a dichotomous schizotaxic sign, inferred from taxometric analysis (here using 's' and 'n' to denote the schizotaxic and nonschizotaxic groups). Then we can derive a strong prediction from the dominant gene theory, namely, that if such validated indicators are studied in the parents of carefully diagnosed schizophrenes, first screening out the probands who (although schizophrenic) do not show those indicators themselves when in remission, the expected covariance of each single indicator in the parent (father/mother) pairs will be $\text{cov}_{fm} = 1/4(p_s - p_n)^2$ (Golden & Meehl, 1978). I do not believe it possible to derive this consequence plausibly from a polygenic theory. For one thing, we would have to assume very high negative assortative mating for indicators such as SPEM and P50, whereas the data show there is no high negative assortative mating for either diagnosable schizophrenia or for scores on MMPI scale 8, surely more socially visible traits. The numerical prediction, if false, would constitute a strong potential falsifier of the dominant gene theory. Hence, thinking Popperian, if it turns out as predicted it's a strong corroborator for the dominant gene theory (Meehl, 1978, 1990b, 1990c, 1990e; Popper, 1959, 1962).

Environmental Mold Taxa

One finds some resistance to the notion of any really strong and interesting taxa in the field of personality and social psychology when we deal with the nonpathological range of individual differences. I don't think we know to what

extent this skepticism is rationally based, or arises simply from the unquestioning acceptance of the Paterson "no types, only dimensions" dogma cited at the beginning of this paper. One tends to think that true taxa or types should be like disease entities in organic medicine, attributable to the influence of a qualitatively unique and distinct etiological agent such as a germ or a gene. Somewhat reluctantly, we also permit a highly specific pathogenic environmental event such as childhood seduction, or "being dropped on your head when you were a baby," as the saying goes. But the unquestioned existence of such highly cohesive and dynamically effective taxa as Trotskyism, Baptist Fundamentalism, or Frenzied Egalitarianism—one could exemplify with a variety of political, economic, religious, and even esthetic types—should suffice to persuade us that strong and important taxa in the personality domain need not originate in germs, genes, or single dramatic environmental happenings. When we consider the possibility that such a trait as self-monitoring could be taxonic, as argued by Gangestad and Snyder (1985; Snyder & Gangestad, 1986), it does strike most psychologists, myself included, as a bit much to accept. I shared the skepticism of my colleagues when I sat as an examiner on Gangestad's oral, and I still don't know exactly what to make of the findings. I am not impressed by the criticism of Miller and Thayer (1989), who apparently did not take the trouble to understand the taxometric theory and methods involved (see Gangestad & Snyder, 1991). For example, in taxometric separations, the factorial purity, alpha-reliability, or unidimensionality of a psychometric instrument are irrelevant. All that matters is that the probability of taxon membership is a monotone increasing function of the score, plus internal tests against psychometric artifactual taxonicity. An extreme case—mathematically ideal for maximal taxonic separation—would be an item set with all pairwise inter-item correlations $\phi_{ij} = 0$, hence zero scale "reliability" (= internal consistency) within the taxon and complement classes, but each item discriminating the categories. (One learns in elementary statistics class to prefer negligible inter-predictor correlations in multiple regression with quantitative predictor variables, but that idea does not easily generalize to the scale-to-item situation, being paradoxical there [Loevinger, 1954; and see Humphreys, 1956].) Psychologists have trouble imagining such situations, but physicians do not. (A referee points out that we do want indicator reliability, which constrains taxonic validity. Since single-item reliabilities are often rather poor, we may look for item packets which possess high internal consist-

`

ency, but still aim for the several packets—distinct taxon indicators—to have negligible pairwise correlations, as the text supra argues.)

I have the impression that when a socially developed nonbiological taxon can be easily understood in terms of environmental molding, psychologists accept it more readily. It might be a complicated matter to analyze why somebody joined the Socialist Workers party, although even there it might be due to a single event, such as marrying a Trotskyist. But once the person joins, it is easy to understand how the environment molds the attitudes and beliefs into an ideological syndrome that is statistically tighter, and often more resistant to change, than many diseases recognized as taxa in internal medicine. The existence of diverse causal paths to one's membership in a "normal range" personality or social taxon precludes a methodological stipulation linking formal-numerical taxonicity to a unique cause, as in specific etiology (medicine) or a major locus (genetics).

An environmental mold taxon emerges because persons subjected to certain (formal or informal) learning experiences—precepts, models, and reinforcement schedules—acquire motives, cathexes, cognitions, and act dispositions that the social group "teachers" tend to transmit together, at least stochastically. People with certain interests, perhaps in considerable part genetic in origin (Arvy, Bouchard, Segal, & Abraham, 1989; Bouchard, Lykken, McGue, Segal, & Tellegen, 1990; Carter, 1932; Nichols, 1966, 1978; Pedersen, Plomin, McClearn, & Friberg, 1988; Tellegen, Lykken, Bouchard, Wilcox, Segal, & Rich, 1988; Vandenberg & Kelly, 1964; Vandenberg & Stafford, 1967), tend to associate with others of similar interests, so that a whole set of attitudes, values, and verbal habits become reinforced by the social environment in addition to the skills and cognitions that define the particular occupation. If you major in subject S, or join political party P, or convert to religion R, or cultivate hobby H, you will learn to *want, value, perceive, believe*, *say*, and *do* such and such things with a higher probability than those not so educated.

This strong dependence of covariant dispositions on statistically clustered experiences does not, however, imply a complete absence of *intrinsic* covariance between the elements of an environmental mold taxon. Reflecting on the two main kinds of environmental mold taxa, *vocational* and *ideological*, one can discern three sorts of intrinsic relations among learned contents that facilitate taxonic molding and help protect taxa, once formed, from erosion. These three intrinsic content relations I shall call *instrumental*, *cognitive*, and *psychodynamic*, without implying sharp divisions or non-compresence. A

vocational taxon (e.g., lathe operator) exists because some persons learn that skilled trade while most of us do not. The molding environment is the vocational school. But that such a mold exists is not, so to say, a "historical accident." The verbal items making up a trade test for lathe operator are semantically *about* a specially arranged set of objects (parts of the machine, the piece worked on) and sequential actions performed on these, with instrumental meaning (e.g., "turning a newel post"). In the case of ideology (political, religious, social), there are semantic or purportedly causal relations between the attitudes, sentiments, and beliefs that comprise the cluster. These internal conceptual relations are part of the reason why an ideological mold comes into being in a society in the first place, but the theorist cannot safely insist upon that being invariably operative. A social psychologist may turn to sociology, political science, or history for a proximal explanation of an ideological taxon, and such an account is not vitiated by a finding that those disciplines provide no accepted explanation of the mold at their level of analysis. (If proximal explanations were considered defective because not ultimate, would there be any science except physics in Comte's Pyramid?) Freud (1914/1957b), beginning with a quote from Adler, has a helpful comment on this problem:

> "If you ask where repression comes from, you are told, 'from civilization'; but if you go on to ask where civilization comes from, you are told 'from repression'. So you see it is all simply playing with words." A tithe of the acuteness and ingenuity with which Adler has unmasked the defensive devices of the "nervous character" would have been enough to show him the way out of this pettifogging argument. What is meant is simply that civilization is based on the repressions effected by former generations, and that each fresh generation is required to maintain this civilization by effecting the same repressions. I once heard of a child who thought people were laughing at him, and began to cry, because when he asked where eggs come from he was told "from hens," and when he went on to ask where hens come from he was told "from eggs." But they were not playing with words; on the contrary, they were telling him the truth. (pp. 56-57)

Example: Right-wing ideology is so familiar a cluster of beliefs and attitudes that we easily þëleave unnoticed the puzzle of its content relations. When we do reflect on these, one plausible conjecture is that a certain view of "human nature"—not always verbalized explicitly in social scientists' measures of the

`

rightwing factor—could explain a large portion of the covariances (McClosky & Meehl, in preparation). Roughly, the "core notion" is that humans are a pretty bad lot, full of wicked impulses, irrational, and just barely restrained, so that it is necessary to impose tight, strict, pervasive, and punitive controls all along the line (families, church, school, police, etc.), otherwise society will come apart at the seams. Authoritarian child rearing, subjection of women, harsh criminology, persecution of deviates, censorship of books and movies, religious and patriotic dogmatism, dislike of social science, disapproval of questioners and skeptics, suspiciousness of psychotherapy, "keeping everyone in his place," and a kind of diffuse anti-intellectualism can all be understood in terms of this basic fear of the human impulse life.

When cognitive coherence, a certain "strain toward consistency," is combined with psychodynamics as a source of covariation, matters can become pretty complicated and causal analysis murky. Sometimes cognitive consistency seems to facilitate environmental molding. But sometimes the facets of a taxon seem not only lacking in coherence but even contradictory. *Example*: Passionate anti-abortion persons and groups *tend*, with numerous exceptions, to have supported the Vietnam War and to favor capital punishment. If one takes their "pro-life" self-label in a Buddha-Schweitzer "reverence for life" sense, these views seem not to mesh well. But a given component of ideology can usually be subsumed under more than one conceptual rubric, and these available rubrics may differ as to the quality (or dimension) they make relevant, or the *level* of description (in a hierarchical system) at which they draw the line. Since generic reverence for life doesn't fit, we can try religious and political conservatism. If a right winger classifies what occurred in Vietnam not as a case of *napalming babies* but as *fighting godless communism*, the paradox vanishes. And if a core (not the sole) element of right-wingism is the negative and fearsome view of mankind described above, a person holding that view subsumes capital punishment under the heading of necessary firm social control of wickedness.

But psychodynamics might be employed here with minimal emphasis on the "wicked human nature" concept as source trait. Suppose the right winger is typically a repressed, character-armored person with a rigid defense system that permits little guilt-free gratification of drives, especially the societally dreaded impulses of aggression and sexuality. He projects his forbidden wishes (hence the dark view of human nature, and the diffuse fear of other people's

impulses). But a partial gratification is available, since the resulting condemnatory-punitive belief system permits a socially acceptable expression of aggression (against the "bad people," and even against those societal agents that fail to punish severely). It is well known how thinly veiled is the aim-inhibited eroticism of puritans and censors such as Anthony Comstock. Some such psychodynamics is suggested by the statistical finding that these right wing attitudes are correlated with low self-esteem and low life-satisfaction (McClosky & Meehl, in preparation).

Taxometrics and Causal Inference

These complexities lead to a fundamental question, "What can taxometrics tell us about causality?" In answering this metaquestion, two opposite errors must be avoided. The first is the temptation to treat a taxometric method as a kind of automatic truth-generating machine, which would apply a taxometric algorithm to a data set and yield a computer printout with a definite "Yes" or "No" to any causal-theoretic question, at any stage of knowledge. Of course no one would officially hold that view, but psychologists could be tempted to act *as if* they held it, as has happened with other statistical procedures (factor analysis, *t* tests, multidimensional scaling, analysis of covariance). The prophylactic principle against such mindless taxometricizing could read: *No statistical procedure should be treated as a mechanical truth generator*. (By the same token, a good taxometric method should not be faulted because some social scientist employed it blindly or inappropriately, defying this principle.)

The opposite error is more likely in the present climate of opinion about taxa in social science. It is the error of holding that a taxometric analysis cannot *bear strongly* on a causal conjecture. If a causal conjecture substantively entails the existence of a taxon specifying (on theoretical grounds) its observable indicators, a clearcut nontaxonic result of taxometric data analysis discorroborates the causal conjecture. But any strong empirical test capable of discorroborating a theory is, in principle, capable of corroborating it, namely, being a high-risk prediction that, by succeeding, does not refute. The metaprinciple here reads: *The numerical results of a taxometric analysis are evidentiary with respect to a causal conjecture, as are any other kinds of empirical results; the corroborative strength of this evidence depends, as usual, on the antecedent probabilities and the predictions of competing causal conjectures*.

When an environmental mold shaping an ideology is not discernible, the

`

explanatory problem becomes more complicated. One big help is to keep in mind Langmuir's important but oft-neglected distinction between *convergent* and *divergent causality* (Langmuir, 1943; London, 1946; Meehl, 1954, p. 61; 1973b, p. 218; 1978, p. 809). In convergent causality the structure of the causal system, the initial and boundary conditions, and the kind of causal laws involved are such that slight deviations in initial conditions tend to average or cancel out. In divergent causality slight variations in initial conditions cumulate and amplify, moving the whole system in one direction rather than another (cf. Bak & Chen, 1991; Maruyama, 1963). Analogous is the "butterfly effect" in meteorology, so called from the (imagined) dramatic example of a butterfly's wing movements in South America playing a critical role in a causal chain eventuating in a devastating tropical storm (Monastersky, 1990; cf. Amato, 1990; Peterson, 1990). A frightening example of divergent causality is that a quantum uncertain event stimulating a photoreceptor in the retina of a French sniper may have determined that Hitler escaped death when he was a runner in World War I, the end result of which was the Holocaust and World War II (cf. Meehl, 1970, pp. 395-396, fn 18). I have considered the discordance for schizophrenia in a pair of MZ twins from this standpoint (Meehl, 1972b, pp. 404-405), and I believe the notion of "luck" is one of the most important factors in human life, both for individuals and for groups. It is a factor that social scientists tend to ignore, for the obvious reason that it is not something we can plug into our equations or use to enhance our theoretical understanding (Meehl, 1978, p. 118). One can think of the life history of a person with a genetic potential for schizophrenia as a random walk, and the first attack and resulting hospitalization, with all its attendant consequences for labelling and self-concept and the like, as having the properties of an absorbing barrier in random walk theory (Kemeny, Snell, & Thompson, 1956). The MZ co-twin writes schizy poetry and receives the Pulitzer Prize. I am not attempting to persuade by these examples, but only to suggest the possibilities so that we will be open-minded and look to the facts rather than to a methodological dogma that says there *must not be* any "interesting, real taxa" in the area of socially acquired and maintained dispositions, with the exception of those that are ideological or vocational molds.

Some Current Objections to Taxonicity

The extent to which taxonicity in the behavioral area, despite such striking examples as the Mendelizing mental deficiencies and political and religious

syndromes, is ruled out by psychologists is indicated by the frequency with which one hears certain objections that are so careless and naive that one knows bright people wouldn't voice them unless they had a bias. Probably the worst is the argument that a certain trait or symptom "exists in all degrees," that there are persons who have large amounts of it but short of diagnosable illness, hence, there cannot be a taxon involved. This fallacious reasoning is based upon an elementary confusion between the mathematical character of the indicator and the mathematical character of the latent structure. Distinguishing latent and manifest ("source" and "surface" traits), all four combinations of *quantitative* and *qualitative* can exist. We can have qualitative (dichotomous) indicators of a latent factor or dimension, quantitative indicators of a latent factor, qualitative indicators of a latent taxon, or quantitative indicators of a latent taxon. If it were true that a symptom that is intrinsically a matter of degree could not be indicative of a taxon, then none of the infectious diseases for which a symptom like fever is important in diagnosis could exist as genuine taxa, since temperature is a matter of degree. Over the years I have collected pseudosophisticated objections to the taxon schizophrenia, and will content myself here with quoting a brief passage from a previous paper dealing with that subject:

> In almost any discussion of research strategy or data interpretation, one will hear plausible statements like the following: "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 non-symptomatic 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 three decades ago (Carnap 1936-37, pp. 461-463)—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, 1972a, pp. 21-22 [1973a, p. 196])

In clinical psychology a common objection to taxa stems from the erroneous cliché that "psychiatric nosology has been shown to be completely unreliable." To examine the empirical side of that argument would take us far beyond the scope and space constraints of this paper, so I shall say merely that it is not an accurate statement of the current diagnostic situation, at least for the major categories of mental illness (e.g., schizophrenia). Formalization of descriptive nosological criteria, especially as implemented by structured interviews, achieves reliabilities that are quite respectable by customary standards of social science, including psychometrics (cf. Grove, 1987 and references cited therein). But even older studies do not warrant the usual assertion as regards major psychopathological taxa. For example, the classic study of Schmidt and Fonda (1956) is often cited for this purpose; but while that research showed grave unreliability in diagnosing many of the old rubrics (and especially their subdivisions), for the clinically important dichotomy schizophrenia/nonschizophrenia the interclinician reliability was .95, better than any psychometric instrument or normal range rating scale I could name. Each diagnostic category should be separately studied, employing skilled clinicians who have been trained in careful diagnosis and "believe in it" as something valid and worth doing. A kappa computed over numerous diagnoses does not assist us appreciably in thinking about the taxonicity and genetics of schizophrenia. For sophisticated discussions of the problems of diagnostic

reliability, see Carey and Gottesman (1978); Grove (1987); Grove, Andreasen, McDonald-Scott, Keller, and Shapiro (1981); and Andreasen, Grove, Shapiro, Keller, Hirschfield, and McDonald-Scott (1981).

Empirical reliabilities aside, the argument requires much closer methodological scrutiny than it usually receives. The intrinsic validity (empirical meaningfulness) of a diagnostic construct cannot be dismissed *ipso facto* on grounds of poor average clinician agreement. Admittedly, for practical (e.g., forensic, drug selection, prognostic) purposes, we have not usually calibrated each clinician as to net attenuated construct validity. That being so, a replicable finding that *interclinician agreement* (a better word here than the generic 'reliability,' with its several overtones from classical psychometrics) is modest or poor is discouraging—although I must point out that for some clinical purposes (e.g., which drug to try first) even unreliable diagnoses are preferable to random trial. In organic medicine this is taken for granted. When we move from clinical tasks to the research context, matters become more complicated, both conceptually and mathematically. High average interjudge agreement is *neither necessary nor sufficient for construct validity*. The point can be made by considering an extreme case. Clinician A is inexperienced, poorly trained, anti-nosological, uninterested in schizophrenia, and a careless interviewer. Clinician B (= Bleuler) is bright, careful, has examined thousands of patients (followed over many years at Burghölzli), has had a special interest in schizophrenia since his sister fell ill with it, etc. If a "true case of schizophrenia" were *theoretically* defined as such-and-such psychisms caused by a dominant schizogene (Meehl, 1962, 1972b, 1989, 1990a, 1990c), then Bleuler might be diagnosing 95% correctly and A diagnosing at a nearly chance level; hence their agreement would be very poor. Of course one cannot simply *assume* one of two disagreeing clinicians is highly accurate; the methodological point is that their disagreement is not a dispositive argument against the *existence* of the conjectured taxon. *The correlation of selected clinicians' judgments with other conjectured indicators which are embedded in a network of indicator and family correlations is more important in theoretical research than unselected clinicians' agreements with one another*. Even inexplicable "clinical intuition" is sometimes acceptable, however frustrating it may be to one's drive toward methodological purity, provided it "works" (cf. my intuitive spotting of psychopathic deviates, Meehl, 1959, pp. 104-105 [1973a, p. 93]; Reichenbach's clairvoyant, 1938, pp. 358-359; or philosopher of science Feigl's "diagnostic dog," which mysteriously barked in the presence

`

of carcinoma, personal communication, 1941). Some clinicians may conceivably be better at diagnosing schizophrenia because they have a certain olfactory gene analogous to the PTC (phenothiocarbamide) "taster" gene (Smith & Moody, 1961; Smith & Sines, 1960; Smith, Thompson, & Koster, 1969). Since human judgments and ratings have been shown to behave like test items in that they follow the Spearman-Brown prophecy formula, one way to include putative indicators of a conjectured taxon that are not readily objectified is by pooling interviewer ratings. (cf. Gottesman & Shields, 1972, pp. 215-216, where the estimated construct validity of their six "schizophrenia experts" when combined was .95. It is worth noting that "conservative" psychiatrist Slater and "liberal" psychologist Meehl still agreed 84% of the time; family-dynamics psychiatrist Mosher and genetics-oriented psychologist Meehl correlated $r = .89$ in rating severity of psychopathology. I permit myself the conjecture that for some kinds of clinical judgments, competence, motivation, and conscientiousness matter more than theoretical disagreements.) A final point is that when the *existence* of a construct taxon is at issue, and reliability studies vary widely as to interjudge agreement found, the high-agreement results—absent artifact (e.g., contamination) or sheer sampling error in small *N*—tend to corroborate the taxon's existence more than the poor-agreement results tend to refute it.

The conventional approach, stemming from our laudable interest in objectivity as a mark of science, goes like this: "We can't be scientific about a putative diagnostic taxon unless clinicians and researchers agree on its operational definition, so that we know which patients 'have it' and which don't. But clinicians employ different lists, or at least weight various elements differently. So we can't even find out whether the putative taxon has reality." This customary reasoning is plausible, but it is deeply wrong-headed. To detect a latent taxon it is not necessary to "agree—a matter of taste, of convention—on the criterion." In fact, two researchers need not even agree completely on the *candidate list* of indicators, to get the taxometric process going. Psychologists should accustom themselves to thinking about taxa as classical psychometric theory deals with factors, where the equations explicitly include uniqueness (specificity + error) along with communality as components of a test's variance. The distribution of different judges' agreements and disagreements as to a diagnosis, if treated in a mathematically sophisticated way, is itself a valid source of information. A nice example of this is the study of schizophrenia by Young, Tanner, and Meltzer (1982), where latent class

analysis strongly corroborates the taxonic conjecture ("there is such an entity as schizophrenia, which explains the pattern of diagnostic agreements and disagreements"), and estimates the specificity and sensitivity of the four competing sign-sets. For an excellent methodological and mathematical treatment of the general case, see Uebersax and Grove (1990).

One stumbling block for hyperoperational psychologists is a fear of scientific sin when dealing with inferred entities. It is in the intrisic nature of theoretical constructs in the life sciences—an empirical fact of the biological world rather than a careless scientist's tolerance for conceptual sloppiness—that their causal linkage to indicators is typically stochastic and, hence, their epistemic indicator function is (casewise) imperfect. In this respect taxometrics is no different from the psychometrics of quantitative factors, but somehow—I have been unable to figure out why—it makes psychologists more nervous. A taxonic conjecture that is well corroborated by coherency tests is analogous to a confirmatory factor analysis, if not stronger. But despite the confidence we may properly repose in the structural and parametric inferences (taxon base rate P, or factor loading a_{ij}) the classification of individuals is of course probabilistic, as is the estimate of a subject's factor score on *g* from his pattern of WAIS subtest scores. It is odd that psychologists (and even some psychiatrists, since the ascendancy of *DSM-III*) should be unclear about this data-to-construct relation, given its ubiquity in life science explanations. We cannot theorize or experiment on latent learning, intelligence testing, psychodynamics, or genetics without distinguishing between "latent" and "manifest" (MacCorquodale & Meehl, 1948; Meehl, 1986a, 1986b). Without allowing that elementary distinction, one cannot even explain what a 'recessive gene' is. In this respect, the useful "operationalizing" of psychopathology as in *DSM-III* criteria has sometimes been simplistically interpreted in a way quite unlike the so-called "medical model" of organic medicine, where physicians are routinely prepared to learn from the pathologist (Cabot Method) that an error of omission arose because the disease (e.g., staghorn kidney, silent CNS tumor) was asymptomatic. I tell clinical psychology students to read the courageous statistical summary by Minneapolis internist Thomas Peppard (1949) of his diagnostic mistakes; 29% of his errors of omission were clearly attributable (despite stringent self-criticism) to *symptoms and signs not found*. For a powerful and clarifying critique of *DSM-III* "strict empiricism" see Faust and Miner (1986). The value of this critique does not depend on one's

`

accepting their amount of Kuhnian "theory-ladenness of observations," which I do not (cf. Meehl, 1983, pp. 389-395).

The anti-taxon bias leads some to rely on buzzwords instead of incisive mathematical and metatheoretical analysis (e.g., "an individual cannot be sorted into a simplistic category or slot." Whence the adjective 'simplistic'? Does this tendentious language help clarify anything?). Objections are made that, while sometimes qualitatively sound, apply to the identification and measurement of dimensions as well as to taxa, and whose adverse import is a matter of degree, harmful or harmless depending on the parameters. It is argued that since personality variables are distributed continuously (*petitio principii*!) and subject to consensual definition, purported types are bound to be arbitrary constructions. Exactly the same complaint can of course be made against quantitative phenotypic traits, and against psychometric factors inferred from them. All scientific investigation begins with selecting a set of quasi-observable properties or dimensions for study. (As Popper says, the blanket instruction "Observe" is not a helpful one.) What keeps the scientist's initial choice of "what to look at" from *remaining* personal, subjective, arbitrary, "purely conventional," is the next step: statisticizing the covariations. I refer the reader to Jensen's (1969) clarifying discussion of how the broad, vague, pre-analytic notion of *cognitive task* develops into the scientific concept of *g*. Given the ramified network of relations among subtests of an omnibus intelligence test, and the hundreds of correlations between measures of *g* and extra-test variables, we have a rich (but still partial) interpretation of the "general intelligence" construct. If someone wants now to put 2-point threshold into the list of indicators (Binet and Simon tried and abandoned it), that is no longer a permissible arbitrary choice, because this candidate indicator just won't fit in empirically. The objection conflates a truism about all empirical knowledge—"you have to start somewhere, have to get your foot in the door"—with "there can be no rational, intersubjective procedures for finding out what goes with what," a false statement, not validly inferable from the truism. Pliny sorted the bat with the chickadee, and the whale with the pickerel, both for pretty obvious reasons. Linnaeus, seeing—better, *thinking*—deeper, classified both bat and whale with grizzly bear. We consider this reclassification a stroke of taxonomic genius, we do not say it is merely a matter of whether one's classificatory tastes match an 18th century Swede's or an ancient Roman's.

Developing and validating *strong tests* (O'Hear, 1980; Popper, 1959, 1962,

1983; Schilpp, 1974) of a conjectured taxon may be a task of considerable difficulty, both as to the metaconcepts and the mathematics. The same is true for initial search techniques (absent a conjecture) in the context of discovery. I have no wish to minimize the difficulties. What I object to is pseudosophisticated dogmas about arbitrariness, conventionality, and ineradicable subjectivity. Whatever crude, intuitive, pre-analytic considerations suggest a candidate list of indicators, whether they cohere taxonically (or factorially) is a matter of empirical fact. A theorist's personal predilection for 2-point threshold or speed of tapping cannot force these measures to be highly correlated with vocabulary and Kohs blocks; nor can my clinical impression about spidery, spatulate, knobby fingers (Meehl, 1990c, p. 47) produce an empirical correlation with schizotypal thought disorder, blunted affect, and aberrant eye tracking (Clementz & Sweeney, 1990). I suspect that the psychologist's antinosology bias is here assisted by remnants of (misunderstood) logical positivist doctrine concerning the stipulative ("conventional") character of definitions. Even 1930 vintage Vienna did not say anything so foolish as "one meaning stipulation is just as good as another." Against this vulgar error, Gustav Bergmann spoke often of the *Bergmann Index*, giving a perfectly "operational" definition: B.I. = (IQ) ÷ (Weight)$^{1/3}$—totally useless for science.

The unreliability sometimes found for the individual elements of a type is often mentioned, here again ignoring the fact that this problem can arise for quantitative traits also (e.g., a single MMPI or WAIS item; a single playground episode of aggression; an occasion of being late or punctual). Psychology has evolved a set of procedures for increasing reliability. In ratings we can improve the format in several well known ways (e.g., asymmetric spread of anchoring adjectives, finer scales, reversing "good-bad" direction, separating similar contents, forcing dispersions). We delete the poorest items. We select and train raters carefully. We provide feedback. We enlarge the behavior sample. If these joint efforts still leave us with unsatisfactory interjudge agreement, we pool judgments. A pairwise inter-rater reliability coefficient $r = .60$, representative for many traits, tells us we need 7 raters to get a reliability $r^* = .90$, via the inverse form of the Spearman-Brown prophecy formula $n = r^*(1 - r)/r(1 - r^*)$. Finally, the effect of indicator unreliability is to attenuate taxonic separation, not to prevent taxon identification (except via statistical instability, countered by increasing N).

This last brings me to a metatheoretical comment of great importance, not

`

confined to taxometrics. Many (most?) psychologists have a "lazy" attitude about theory testing, stemming from the Fisherian emphasis on "exact" small sample statistical inference. In a taxometric problem, our statistical questions can be about (1) significance, (2) estimation, or (3) structure. The last is the big threshold question, the first is of little importance, and in my coherent cut kinetics (described above) the second is the epistemic path to the third. Now it is a truism of statistical inference that *problems of estimation cannot be solved by small samples*. It is useless to know that a statistic is, say, an MLE, or unbiased, if it can deviate hugely from the parameter, given a small *N*. Everyone learns this in undergraduate statistics class; but if you tell people, "this taxometric procedure should probably not be used with an $N < 300$," they tend to feel somehow put upon, as if it's "not fair" to demand such a thing. Physicists, chemists, astronomers, epidemiologists, geneticists take it for granted that some empirical questions cannot be studied satisfactorily except by making very large numbers of observations. Some current astrophysical puzzles (e.g., the deficiency of solar neutrinos) are being researched by making millions of observations of rare nuclear events over a period of years. *If a taxometric question cannot be answered without more time, money, and subjects than your resources or motivation allow, then you should not try to study it*.

What Difference Does It Make?

A referee asks (as does my colleague David Lykken, personal communications, March, 1991) what difference does it make? Why should we *care* about taxonicity anyway? Does it—or should it—make a difference in how we proceed? I offer four replies, any one of which suffices to warrant raising the taxometric question. First, if there *are* real taxa in a domain, theoretical science should come to know them. I assert this as a scientific realist, recognizing that fictionists and instrumentalists may think otherwise.

Second, in constructing assessment devices (tests, rating scales, checklists, episode sampling procedures, work products), the psychometric strategy is very different, the distinct technological aims being *assignment of individuals to a category* versus *location of individuals on a dimension*. For the latter task, item difficulties and correlations should be chosen so as to disperse scores widely and discriminate effectively in all regions of the dimension (latent or manifest), a very different function from *sorting* at a best cut so as to minimize

"in/out" misclassifications. (Critics of the MMPI have pointed out that it was never clear from the start which scales were to do which kind of job, or some of both? MMPI-2 has not clarified this.) Construction and selection of items to compose a quantitative indicator cannot optimize both its dimensional and taxonic power.

Third, in clinical work, should prognosis and treatment choice be mediated by *classifying* the patient or client, as in organic medicine? Physicians understandably assume that, but many psychologists object to the rubrics of *DSM-III* because they presuppose a categorical model, analogous to disease entities in organic medicine. (The odd use of 'axis' by the system's builders suggests that they were somewhat muddled about the taxon/dimension question.) Given a set of fallible "input" indicators (symptoms, signs, biochemical and psychological tests, life history facts, family data) and a set of "outcome" dispositions (drug of choice, second drug to try, suicide risk, response to group therapy, relapse probability, vocational competence), under what conditions is it worthwhile to mediate (input → output) epistemic transitions via an inferred latent taxon? I once used philosopher Feigl's argument that dealing with $(m + n)$ correlations is simpler than $(m \times n)$, plus the need to extrapolate research findings from one clinic to another when a single installation cannot study everything at once (Meehl, 1959). Grove (1991b) has examined the first of those arguments analytically and Monte Carlo, with surprising and discouraging results. The parametric conditions required for taxon-mediated (input → output) inference to improve on a direct dimensional (linear regression) approach are more limited than the conventional wisdom, relying on the medical practice of category diagnosis, had supposed. Grove's path-finding paper will, I trust, stimulate mathematical and empirical studies of this important but long-ignored problem.

Fourth, *causally* oriented research will often proceed differently if a taxonic conjecture has initially been taxometrically corroborated. *Example*: Suppose *psychopathic deviate* as conceived at Minnesota (*not* the sociologically specified "sociopathic personality" of *DSM-III*, a nonpsychological hotchpotch, but the solid-gold Cleckley-Lykken syndrome with pure 49′ MMPI code, free of neurotic and psychotic elements) turns out to be a real taxon when a mixed battery of psychometric, EEG, ratings, and life history indicators are taxometrically analyzed. That would give a major gene locus etiology higher antecedent probability than if the psychopathic syndrome were

`

merely a matter of being "far out" on a set of independent dimensions (e.g., low anxiety parameter, egocentricity, mesomorphic toughness, impulsivity). *Example*: If schizotypy is a taxon, continuing a dogged search for the biochemist's "purple spot," despite the history of nonreplicable artifacts, would be more rational in the context of discovery than if the statistical structure of phenotypic descriptors were found to be factorial rather than taxonic.

Lykken has pressed me for a generic but "substantive" explication of taxonicity, a conceptual definition bridging the gap between an exemplifying list of concrete taxa and the taxometric formalism. While I understand (and sympathize with) this desire, I am not hopeful that it can be satisfied. Despite sharing it, I have been unable, during 25 years of reflection, to produce any such, nor has anyone else; *and no meta-proof exists that such a thing is possible*. What would Lykken's gap-bridging explication look like? He uses the term 'structure' for what he has in mind. Consider the following list of categories, which informed persons agree are taxa, and which would satisfy my formal-numerical taxometric criteria:

chipmunk	hurricane	ablative absolute
daffodil	PKU	protein
igneous rock	neutron	Danish pastry
schizotype	stroke	operetta
bridge player	scurvy	Baptist
Fascist	gout	potassium
surgeon	measles	sonnet
comet	Gothic cathedral	"big" forest fire
revolution	bureaucracy	conic section

It is hard to discern a "structure" (static or dynamic) common to all items in that list. If we parsed the list in a meta-taxonomy of taxa, such as organic diseases, ideologies, major gene syndromes, biological species, personality types, vocations, chemical substances, heavenly bodies, etc., perhaps common structures could be specified for the taxa within each *sub*group, although I am not fully confident even about that. (This, it turns out, is what Lykken has in mind.) Or, focusing on causality rather than scientific domains or outcome types, we might distinguish several causal structures, e.g., specific etiology (in any of the three "strong influence" forms described above), environmental mold, Langmuir divergence, autocatalysis, catastrophe, positive feedback, random walk with absorbing barrier. But I can see no further level of

abstraction subsuming these. In biological classification, the higher taxa (e.Ûa2g., phylum, class, order) can be specified by shared "structural" criteria. But divisions at lower levels (e.g., two gopher species, or even two genera of order *Rodentia*) are based upon quantitative distances, structure being shared between taxa. When we consider environmental mold taxa (e.g., Trotskyist, C.P.A.) the structure concept does nothing for us; it's all a matter of quantitative separation. Focusing on the taxometric search methods, we might try: "There exists a latent or historical property A (or conjunctive property A . B . C …), such that individuals having the property are distributed unimodally on manifest quantitative indicators x, y, z, … with frequency functions $f(x)$, $g(y)$, $h(z)$, …; while individuals lacking the property distribute unimodally as different frequency functions $F(x)$, $G(y)$, $H(z)$, …." But this is surely too "empty" to meet Lykken's desideratum, being *so* abstract that it is hardly more than a verbal restatement of the formalism. It may be fruitful to conceive of the set of individuals whose terminal values of indicator variables are distributed in a broadly specified way (e.g., as one of Pearson's twelve curve types, cf. discussion in Meehl, 1990a, pp. 258–263) as ensembles of temporally successive states, and to inquire what distributions of initial values, combined with what class of transitional generating functions—not a Markov process!—are necessary and sufficient to yield one, versus two, unimodal outcome distributions. But such an investigation demands mathematical competence beyond mine. Meanwhile, the lack of a Lykken—generic explication of taxonicity, while intellectually frustrating, is not an impediment to the research enterprise. The taxometrist asks what the scientist's aim is, what is conjectured about the domain, what question is to be answered, and what would be acceptable as an answer. Given those substantive specifications as to a particular research problem, we will, I believe, be able to decide whether our family of taxometric procedures are adequate to the task.

Personologists investigating taxa usually think of search procedures involving (a) complex statistical treatment of (b) individual differences in (c) quantitative indicators. There is nothing wrong with that, but it is salutary meanwhile to reflect on the variety of taxa, and ways of discovering them, that abound in nonbehavioral sciences. It may turn out that nontaxometric (e.g., experimental) approaches are better in some psychological or sociological domains. Quasi-dichotomous indicators may cohere so tightly that taxometric methods are pointless. Sometimes the term 'indicator' is inappropriate, the full

`

meaning of the taxon being contained in the manifest pattern. *Example*: The list above includes *big forest fire* as an undisputed taxon. 'Big' denotes not size of area burned, but a set of present-or-absent properties that cohere perfectly and are *constitutive* of the concept. Among other things, a big forest fire creates its own windstorm; ignites objects distant from its front via superheated air, no flame needed; and generates upward draft that lifts large burning pieces that fall ahead and start "spot fires." No statistics were required to discern that big forest fires are a clear taxon, and there is no *inferred latent* construct "fallibly indicated" by the defining properties (Carroll & Raiter, 1990, p. 11). *Example*: Rutherford's analysis of radioactive radiation proceeded experimentally with a single "output" measure.

> …Rutherford showed that the radiation from uranium, though complex, consists of two entirely different types—and he achieved this with a convincing simplicity. He simply covered his uranium with thin foils of aluminium, gradually increasing the number of foils. For the first three layers of foil the radiation escaping from the uranium decreased progressively in such a way as to suggest an ordinary law of absorption—i.e., that the thicker the layer of aluminium the less radiation penetrated to ionise the air. More thicknesses of aluminium, however, had little further effect in reducing the radiation at first, but eventually the intensity of radiation began to diminish again as even more foils were added. "These experiments show that the uranium radiation is complex and that there are present at least two distinct types of radiation—one that is very readily absorbed, which will be termed for convenience the alpha-radiation, and the other of more penetrative character which will be termed the beta-radiation."
>
> We now know after eighty more years of investigation that these two different forms of radiation are caused by two different forces which are among the most fundamental features of the physical world. It is an extraordinary tribute to Rutherford's imagination, and his sheer power of measuring things previously unmeasured, that he was able to distinguish between them in his very first investigation of the subject. (Wilson, 1983, p. 126)

In psychology there are doubtless many examples of experimentally detected *qualitative* differences analogous to Rutherford's α- and *β*-rays. For example, cognitive psychology finds that subjects adopt either a "confirmatory" or "falsifying" strategy in problem solving, and for some tasks this

qualitative dichotomy produces a large quantitative difference (even zero overlap) in their performance. Whether we would label that a 'taxon,' absent other strong correlates, is a semantic question, but one rationally discussable in the taxonomic metalanguage. This does not mean, contrary to some quick-and-easy solutions, that the whole question is "merely semantic." *Given* a stipulation as to conditions for applying the term 'taxon,' the taxometric *facts* are what they are, and so decide the answer.

Psychological classification poses a conceptual problem for those who are methodologically sophisticated, and its resolution is of wider interest than taxometrics, or even psychometrics generally. When a simplistic operationism is replaced by recognizing the legitimacy and unavoidability of open concepts, and the feeble testing of theories through H_0-refutation (Meehl, 1967, 1978) is replaced by demand for strong Popperian risks (or Salmonian coincidences, see Meehl, 1990b, 1990d, 1990e), we have a metatheoretical puzzle. How reconcile these two methodological desiderata? Two of the three kinds of concept "open-ness" (Meehl, 1977)—extensibility of the list of indicators and stochastic character of the postulated relations—would seem to prevent tight derivation chains running from taxonic postulates to relations between observables. Roughly, if a concept is "loose," if there's a lot of "play" in the system, how can we tack it down to hard facts? The generic answer, while requiring some ingenuity and mathematics in application, is simple: *To achieve strong tests of open concepts, we must statisticize the open-ness*. *Example*: It is hard to assess a dominant gene theory of the disposition to schizophrenia given low clinical penetrance, numerous correlated nuisance factors, and disagreement about the indicator list (Meehl, 1989, 1990a, 1990c). Lacking an indicator of schizotaxia sufficiently valid to serve as a genetic marker (I use 'marker' in the strict sense of a pleiotropic effect, not chromosomal linkage or mere breeding population correlate), statistics on the *patterns* of indicator correlations can be used to refute the dominant theory (and, hence, to corroborate it if a high risk falsifier does not occur). For examples of taxometric theorems usable in this way, see Golden and Meehl (1978), Meehl and Golden (1982), and Meehl (1990c). The basic metatheoretical approach is no different from that of other sciences that treat probabilistic domains rigorously, where we do not know the truth about each individual entity, state, property, or event, but theory is nevertheless strong enough to derive theorems concerning proportions, means, variances, covariances, etc., of groups (e.g., quantum mechanics, genetics, epidemiology, meteorology).

`

CONCLUSION

In conclusion, my thesis is that the existence of interesting taxa in the nonpathological range of individual differences that are studied by personality and social psychologists should be viewed as a question for empirical determination by appropriate taxometric methods, and that even unlikely candidates for taxonicity should not be dismissed on a priori grounds. If we cannot understand how a certain kind of personality or social impact syndrome could have become so strongly crystallized as to yield a statistical taxon, if we find this theoretically puzzling, the proper stance is not to reject the facts, but to revise the theory. I do not myself anticipate the discovery of numerous strong taxa within the nonpathological range of personality differences, except for some clearcut ideological ones and the ones we have become accustomed to in the field of vocations. I do, however, expect that taxometric analysis applied to a variety of traits in the normal range will turn up more taxa than current views in American psychology lead us to expect.

REFERENCES

Amato, I. (1990). Solutions to crystal-growth mysteries. *Science News, 138,* 116.

Andreasen, N. C., Grove, W. M., Shapiro, R. W., Keller, M. B., Hirschfield, M. A., & McDonald-Scott, P. (1981). Reliability of lifetime diagnosis. *Archives of GeneralPsychiatry*, *38*, 400-405.

Arvy, R. D., Bouchard, T. J., Jr., Segal, N. L., & Abraham, L. M. (1989). Job satisfaction: Environmental and genetic components. *Journal of Applied Psychology*, *74*, 187-192.

Bak, P., & Chen, K. (1991). Self-organized criticality. *Scientific American*, *264* (1), 46-53.

Blashfield, R. K. (1976). Mixture model tests of cluster analysis: Accuracy of four hierarchical agglomerative methods. *Psychological Bulletin*, *83*, 377-388.

Blashfield, R. K. (1984). *The classification of psychopathology: Neo-Kraepelinian and quantitative approaches*. New York: Plenum Press.

Bolz, C. R. (1972). Types of personality. In R. M. Dreger (Ed.), *Multivariate personality research: Contributions to the understanding of personality in honor of Raymond B. Cattell* (pp. 161-260). Baton Rouge, LA: Claitor's Publishing Division.

Borgen, F. H. (1983). Cluster analytic detection of biological sex: A second look. *Multivariate Behavioral Research*, *18*, 343-360.

Bouchard, T. J., Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science*, *250*, 223-228.

Boyle, R. (1979). The origine of formes and qualitiesÌ. Reprinted in M. A. Steward (Ed.), *Selected philosophical papers of Robert Boyle* (pp. 1-96). New York: Barnes & Noble. (Original work published 1667)

Carey, G., & Gottesman, I. I. (1978). Reliability and validity in binary ratings. *Archives of General Psychiatry*, *35*, 1454-1459.

Carnap, R. (1936-37). 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. Reprinted in H. Feigl & M. Broadbeck (Eds.), *Readings in the philosophy of science* (pp. 47-92). New York: Appleton-Century-Crofts, 1953.

Carnap, R. (1939). Foundations of logic and mathematics. *International Encyclopedia of Unified Science*, *1* (3).

Carnap, R. (1945). The two concepts of probability. *Philosophy and Phenomenological Research*, *5*, 513-532. Reprinted in H. Feigl & W. Sellars (Eds.), *Readings in philosophical analysis* (pp. 330-348). New York: Appleton-Century-Crofts, 1949. Reprinted in H. Feigl & M. Broadbeck (Eds.), *Readings in the philosophy of science* (pp. 438-455). New York: Appleton-Century-Crofts, 1953.

Carroll, F. M., & Raiter, F. R. (1990). *The fires of autumn: The Cloquet–Moose Lake Disaster of 1918*. St. Paul, MN: Minnesota Historical Society.

Carter, H. D. (1932). Twin similarities in occupational interests. *Journal of Educational Psychology*, *23*, 641-655.

Clementz, B. A., & Sweeney, J. A. (1990). Is eye movement dysfunction a biological marker for schizophrenia? A methodological review. *Psychological Bulletin*, *108*, 77-92.

Craig, C. C. (1936). A new exposition and chart for the Pearson system of frequency curves. *The Annals of Mathematical Statistics*, *7*, 16-28.

Cramér, H. (1946). *Mathematical methods of statistics*. Princeton, NJ: Princeton University Press.

Dahlstrom, W. G. (1972). *Personality systematics and the problem of types*. Morristown, NJ: General Learning Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668-1674.

Dawes, R. M., & Meehl, P. E. (1966). Mixed group validation: A method for determining the validity of diagnostic signs without using criterion groups. *Psychological Bulletin*, *66*, 63-67. Reprinted in P. E. Meehl, *Psychodiagnosis: Selected papers* (pp. 156-164). Minneapolis: University of Minnesota Press, 1973.

Elderton, W. P., & Johnson, N. L. (1969). *Systems of frequency curves*. London: Cambridge University Press.

Erlenmeyer-Kimling, L., Golden, R. R., & Cornblatt, B. A. (1989). A taxometric analysis of cognitive and neuromotor variables in children at risk for schizophrenia. *Journal of Abnormal Psychology*, *98*, 203-208.

Eysenck, H. J. (1986). A critique of contemporary classification and diagnosis. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology* (pp. 73-78). New York: Guilford.

Faust, D., & Miner, R. A. (1986). The empiricist and his new clothes: DSM-III in perspective. *American Journal of Psychiatry*, *143*, 962-967.

Freud, S. (1957a). Instincts and their vicissitudes. In J. Strachey (Ed. and Trans.), *Standard edition of the complete psychological works of Sigmund Freud* (Vol. 14, pp. 117-140). London: Hogarth Press. (Original work published 1915)

`

Freud, S. (1957b). On the history of the psychoanalytic movement. In J. Strachey (Ed. and Trans.), *Standard edition of the complete psychological Works of Sigmund Freud* (Vol. 14, pp. 7-66). London: Hogarth Press. (Original work published 1914).

Freud, S. (1959). Character and anal erotism. In J. Strachey (Ed. and Trans.), *Standard edition of the complete psychological works of Sigmund Freud* (Vol. 9, pp. 167-175). London: Hogarth Press. (Original work published 1908).

Gangestad, S., & Snyder, M. (1985). "To carve nature at its joints": On the existence of discrete classes in personality. *Psychological Review*, *92*, 317-349.

Gangestad, S. W., & Snyder, M. (1991). Taxonomic analysis redux: Some statistical considerations for testing a latent class model. *Journal of Personality and Social Psychology*, *61*, 141-146.

Golden, R. R. (1982). A taxometric model for detection of a conjectured latent taxon. *Multivariate Behavioral Research*, *17*, 389-416.

Golden, R. R., Campbell, M., & Perry, R. (1987). A taxometric method for diagnosis of tardive dyskinesia. *Journal of Psychiatric Research*, *21*, 101-109.

Golden, R. R., & Freeman, K. D. (1983). *Taxometric diagnosis of a latent disease without use of a criterion variable*. Columbia University Reports, Division of Biostatistics, School of Public Health, Columbia University, Technical Report No. B-34.

Golden, R., & Meehl, P. E. (1973a). *Detecting latent clinical taxa, IV: Empirical study of the maximum covariance method and the normal minimum chi-square method, using three MMPI keys to identify the sexes* (Report No. PR-73-2). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., & Meehl, P. E. (1973b). *Detecting latent clinical taxa, V: A Monte Carlo study of the maximum covariance method and associated consistency tests* (Report No. PR-73-3). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., & Meehl, P. E. (1974). *Detecting latent clinical taxa, VIII: A preliminary study in the detection of the schizoid taxon using MMPI items as indicators* (Report No. PR-74-6). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., & Meehl, P. E. (1978). Testing a single dominant gene theory without an accepted criterion variable. *Annals of Human Genetics London*, *41*, 507-514.

Golden, R., & Meehl, P. E. (1979). Detection of the schizoid taxon with MMPI indicators. *Journal of Abnormal Psychology*, *88*, 217-233.

Golden, R., & Meehl, P. E. (1980). Detection of biological sex: An empirical test of cluster methods. *Multivariate Behavioral Research*, *15*, 475-496.

Golden, R., Tyan, S., & Meehl, P. E. (1974a). *Detecting latent clinical taxa, VI: Analytical development and empirical trials of the consistency hurdles theory* (Report No. PR-74-4). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., Tyan, S., & Meehl, P. E. (1974b). *Detecting latent clinical taxa, VII: Maximum likelihood solution and empirical and artificial data trials of the multi-indicator multi-taxonomic class normal theory* (Report No. PR-74-5). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., Tyan, S., & Meehl, P. E. (1974c). *Detecting latent clinical taxa, IX: A Monte Carlo method for testing taxometric theories* (Report No. PR-74-7). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R. R., Vaughan, H. G., Jr., Kurtzberg, D., & McCarton, C. M. (1988). Detection of neonatal brain dysfunction without the use of a criterion variable: Analysis of the statistical problem with an illustrative example. In P. Vietze & H. G. Vaughan, Jr. (Eds.), *Early identification of infants at risk for mental retardation* (pp. 71-95). Orlando, FL: Grune and Stratton.

Gottesman, I. I., & Shields, J. (1972). *Schizophrenia and genetics, A twin study vantage point*. New York: Academic Press.

Grove, W. M. (1987). The reliability of psychiatric diagnosis. In C. G. Last and M. Hersen (Eds.), *Issues in diagnostic research* (pp. 99-119). New York: Plenum.

Grove, W. M. (1991a). The validity of cluster analysis stopping rules as detectors of taxa. In D. Cicchetti and W. M. Grove (Eds.), *Thinking clearly about psychology. Vol. 1: Matters of public interest* (pp. 313-329). Minneapolis: University of Minnesota Press.

Grove, W. M. (1991b). When is a diagnosis worth making? A comparison of two statistical prediction strategies. *Psychological Reports*, 68, 3-17.

Grove, W. M., Andreasen, N. C., McDonald-Scott, P., Keller, M. B., & Shapiro, R. W. (1981). Reliability studies of psychiatric diagnosis. *Archives of General Psychiatry*, *38*, 408-413.

Grove, W. M., Andreasen, N. C., Young, M. A., Keller, M. B., Endicott, J., Lavori, P. W., Hirschfeld, R. M. A., & Reich, T. (1987). Isolation and characterization of a nuclear depressive syndrome. *Psychological Medicine*, *11*, 471-484.

Hathaway, S. R., & Meehl, P. E. (1951). *An atlas for the clinical use of the MMPI*. Minneapolis, MN: University of Minnesota Press.

Hoagland, H. (1947). Enzyme kinetics and the dynamics of behavior. *Journal of Comparative and Physiological Psychology*, ¤40Ì, 107-127.

Hull, D. L. (1988). *Science as process*. Chicago: University of Chicago Press.

Humphreys, L. G. (1956). The normal curve and the attenuation paradox in test theory. *Psychological Bulletin*, *53*, 472-476.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, *39*, 1-123.

Johnson, N. L., & Kotz, S. (1970). *Distributions in statistics: Continuous univariate distributions-I*. New York: John Wiley and Sons.

Kemeny, J. G., Snell, J. L., & Thompson, G. L. (1956). *Introduction to finite mathematics*. Englewood Cliffs, NJ: Prentice-Hall.

Kenney, J. F. (1939). *Mathematics of statistics, Part II*. New York: D. Van Nostrand.

Langmuir, I. (1943). Science, common sense and decency. *Science*, *97*, 1-7.

Lenzenweger, M. F., & Korfine, L. (1991). Confirming the latent structure and base rate of schizotypy: A taxonomic analysis. Manuscript submitted for publication.

Lewin, K. (1931). The conflict between Aristotelian and Galilean modes of thought in contemporary psychology. *Journal of General Psychology*, *5*, 141-777. Reprinted in *A dynamic theory of personality: Selected papers* (pp. 1-42). New York: McGraw-Hill, 1935.

Lewis, C. I. (1929). *Mind and the world-order*. New York: Scribner's.

`

Loevinger, J. (1954). The attenuation paradox in text theory. *Psychological Bulletin*, *51*, 493-504.

London, I. D. (1946). Some consequences for history and psychology of Langmuir's concept of convergence and divergence of phenomena. *Psychological Review*, *53*, 170-188.

Lowrie, G. S., & Raulin, M. L. (1990, March-April). Search for schizotypic and nonschizotypic taxonomies in a college population. Paper presented at the Sixty-First Annual Convention of the Eastern Psychological Association, Philadelphia, PA.

MacCorquodale. K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*, 95-107.

Maruyama, M. (1963). The second cybernetics: Deviation-amplifying mutual causal processes. *American Scientist*, *51*, 164-179.

McClosky, H., & Meehl, P. E. (in preparation). *Ideologies in conflict*.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.

Meehl, P. E. (1959). Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology*, *13*, 102-128. Reprinted in *Psychodiagnosis: Selected papers* (pp. 90-116). Minneapolis: University of Minnesota Press, 1973.

Meehl, P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, *17*, 827-838. Reprinted in *Psychodiagnosis: Selected papers* (pp. 135-155). Minneapolis: University of Minnesota Press, 1973.

Meehl, P. E. (1965). *Detecting latent clinical taxa by fallible quantitative indicators lacking an accepted criterion* (Report No. PR-65-2). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

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: Aldine, 1970.

Meehl, P. E. (1968). *Detecting latent clinical taxa, II: A simplified procedure, some additional hitmax cut locators, a single-indicator method, and miscellaneous theorems* (Report No. PR-68-2). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Meehl, P. E. (1970). Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Minnesota studies in the philosophy of science: Vol. IV. Analyses of theories and methods of physics and psychology* (pp. 373-402). Minneapolis: University of Minnesota Press.

Meehl. P. E. (1972a). Specific genetic etiology, psychodynamics and therapeutic nihilism. *International Journal of Mental Health*, *1*, 10-27. Reprinted in *Psychodiagnosis: Selected papers* (pp. 182-199). Minneapolis: University of Minnesota Press, 1973.

Meehl, P. E. (1972b). A critical afterword. In I. I. Gottesman & J. Shields, *Schizophrenia and genetics: A twin study vantage point* (pp. 367-416). New York: Academic Press.

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

Meehl, P. E. (1973b). MAXCOV-HITMAX: A taxonomic search method for loose genetic syndromes. In *Psychodiagnosis: Selected papers* (pp. 200-224). Minneapolis: 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.

Meehl, P. E. (1979). A funny thing happened to us on the way to the latent entities. *Journal of Personality Assessment*, *43*, 563-581.

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: Vol. X, Testing scientific theories* (pp. 349-411). Minneapolis: University of Minnesota Press.

Meehl, P. E. (1986a). Diagnostic taxa as open concepts: Metatheoretical and statistical questions about reliability and construct validity in the grand strategy of nosological revision. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology* (pp. 215-231). New York: Guilford Press.

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

Meehl, P. E. (1986c). Causes and effects of my disturbing little book. *Journal of Personality Assessment*, *50*, 370-375.

Meehl, P. E. (1989). Schizotaxia revisited. *Archives of General Psychiatry*, *46*, 935-944.

Meehl, P. E. (1990a). Schizotaxia as an open concept. In A. I. Rabin, R. Zucker, R. Emmons, & S. Frank (Eds.), *Studying persons and lives* (pp. 248-303). New York: Springer.

Meehl, P. E. (1990b). 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

Meehl, P. E. (1990c). Toward an integrated theory of schizotaxia, schizotypy and schizophrenia. *Journal of Personality Disorders*, *4*, 1-99.

Meehl, P. E. (1990d). *Corroboration and verisimilitude: Against Lakatos' "sheer leap of faith"* (Working Paper, MCPS-90-01). Minneapolis: University of Minnesota, Center for Philosophy of Science.

Meehl, P. E. (1990e). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant using it. *Psychological Inquiry*, *1*, 108-141, 173-180.

Meehl, P. E. (1995). Extension of the MAXCOV-HITMAX taxometric procedure to situations of sizeable nuisance covariance. In D. Lubinski and R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings*. Minneapolis: University of Minnesota Press.

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.

`

Meehl, P. E., Lykken, D. T., Burdick, M. R., & Schoener, G. R. (1969). *Identifying latent clinical taxa, III. An empirical trial of the normal single-indicator method, using MMPI Scale 5 to identify the sexes* (Report No. PR-69-1). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Meehl, P. E., Lykken, D. T., Schofield, W., & Tellegen, A. (1971). Recaptured-item technique (RIT): A method for reducing somewhat the subjective element in factor-naming. *Journal of Experimental Research in Personality*, *5*, 171-190.

Meehl, P. E., & Yonce, L. J. (~~in preparation~~). *Taxometric analysis by the method of coherent cut kinetics*.

Miller, M. L., & Thayer, J. F. (1989). On the existence of discrete classes in personality: Is self-monitoring the correct joint to carve? *Journal of Personality and Social Psychology*, *57*, 143-155.

Monastersky, R. (1990). Forecasting into chaos. *Science News*, *137*, 280-282.

Murphy, E. A. (1964). One cause? Many causes? The argument from the bimodal distribution. *Journal of Chronic Disease*, *17*, 301-324.

Nichols, R. C. (1966). The resemblance of twins in personality and interest. *National Merit Scholarship Reports*, *2* (Whole No. 8).

Nichols, R. C. (1978). Twin studies of ability, personality, and interest. *Homo*, *29*, 158-173.

O'Hear, A. (1980). *Karl Popper*. Boston: Routledge & Kegan Paul.

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: The Free Press.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London* (*A*), *185*, 71-110.

Pearson, K. (1895). Contributions to the mathematical theory of evolution, II: Skew variation. *Philosophical Transactions of the Royal Society of London* (*A*), *186*, 43-414.

Pedersen, N. L., Plomin, R., McClearn, G. E., & Friberg, L. (1988). Neuroticism, extraversion, and related traits in adult twins reared apart and together. *Journal of Personality and Social Psychology*, *55*, 950-957.

Peppard, T. A. (1949). Mistakes in diagnosis. *Minnesota Medicine*, *32*, 510-511.

Peterson, I. (1990). Rock and roll bridge. *Science News*, *137*, 344-346.

Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books. (Original work published 1935)

Popper, K. R. (1962). *Conjectures and refutations*. New York: Basic Books.

Popper, K. R. (1983). *Postscript* (Volume I): *Realism and the aim of science*. Totowa, NJ: Rowman and Littlefield.

Reichenbach, H. (1938). *Experience and prediction*. Chicago: University of Chicago Press.

Rietz, H. L. (1927). *Mathematical statistics*. Chicago: Open Court. [Carus Monograph No. 3]

Risch, N. (1990a). Linkage strategies for genetically complex traits. I. Multilocus models. *American Journal of Human Genetics*, *46*, 222-228.

Risch, N. (1990b). Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *American Journal of Human Genetics*, *46*, 229-241.

Risch, N. (1990c). Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *American Journal of Human Genetics*, *46*, 242-253.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Schilpp, P. A. (Ed.). (1974). *The philosophy of Karl Popper*. LaSalle, IL: Open Court.

Schmidt, H. O., & Fonda, C. P. (1956). The reliability of psychiatric diagnosis: A new look. *Journal of Abnormal and Social Psychology*, *52*, 262-267.

Sears, R. R. (1943). Survey of objective studies of psychoanalytic concepts. *Social Science Research Council Bulletin*, **51**.

Smith, K., & Moody, A. C. (1961). The effect of meprobamate (Miltown®), RO 1-9569/12 (Nitoman®), and SCH-6673 (Tindal®) on the odor of schizophrenic sweat. *American Journal of Psychiatry*, **117**, 1034-1035.

Smith, K., & Sines, J. O. (1960). Demonstration of a peculiar odor in the sweat of schizophrenic patients. *Archives of General Psychiatry*, **2**, 184-188.

Smith, K., Thompson, G. F., & Koster, H. D. (1969). Sweat in schizophrenic patients: Identification of the odorous substance. *Science*, **166**, 398-399.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy*. San Francisco, CA: Freeman.

Snyder, M., & Gangestad, S. (1986). On the nature of self-monitoring: Matters of assessment, matters of validity. *Journal of Personality & Social Psychology*, **51**, 125-137.

Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of numerical taxonomy*. San Francisco: Freeman.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge: Harvard University Press.

Strube, M. J. (1989). Evidence for the type in Type A behavior: A taxometric analysis. *Journal of Personality and Social Psychology*, **56**, 972-987.

Tellegen, A., Lykken, D. T., Bouchard, T. J., Jr., Wilcox, K. J., Segal, N. L., & Rich, S. (1988). Personality similarity of twins reared apart and together. *Journal of Personality and Social Psychology*, **250**, 223-228.

Trull, T. J., Widiger, T. A., & Guthrie, P. (1990). Categorical versus dimensional status of borderline personality disorder. *Journal of Abnormal Psychology*, **99**, 40-48.

Uebersax, J. S., & Grove, W. M. (1990). Latent class analysis of diagnostic agreement. *Statistics in Medicine*, **9**, 559-572.

Vandenberg, S. G., & Kelley, L. (1964). Hereditary components in vocational preferences. *Acta Geneticae Medicae et Gemellologiae*, **13**, 266-274.

Vandenberg, S. G., & Stafford, R. E. (1967). Hereditary influences on vocational preferences as shown by scores of twins on the Minnesota Vocational Interest Inventory. *Journal of Applied Psychology*, **51**, 17-19.

Wilson, D. (1983). *Rutherford: Simple genius*. Cambridge, MA: MIT Press.

Young, M. A., Tanner, M. A., & Meltzer, H. Y. (1982). Operational definitions of schizophrenia: What do they identify? *Journal of Nervous and Mental Disease*, **170**, 443-447.

Manuscript received February 23, 1991; revised August 21, 1991

`