Meehl, P. E. (1999). Clarifications about taxometric method.
Applied & Preventive Psychology, 8, 165-174.

174

Clarifications about taxometric method

PAUL E. MEEHL
University of Minnesota

Abstract

Taxometrics is a statistical procedure for determining whether relationships among observables reflect the existence of a latent taxon (type, species, category, disease entity). A formal-numerical definition is needed because intuitive, commonsense notions of "carving nature at its joints" or "identifying natural kinds" cannot resolve disagreements as to taxonic reality for hard cases. Specific etiology (e.g., major gene, germ, traumatic event) is often unknown and is not appropriate in nonmedical domains. Lacking an infallible criterion, the taxonic inference relies on the internal configural relations among the conjectural fallible indicators. An essential feature is multiple consistency tests that will not be satisfied if the latent structure is not taxonic or the parameters are badly estimated. Common misconceptions are that the taxon must be "sharply" distinguished, quantitative indicators must be bimodal, the causal origin must be biological, emergence of a large dimensional factor refutes taxonicity, and adopting a taxon is a mere matter of convention or preference.

Key words: Diagnostic entities, Latent class, MAMBAC, MAXCOV, Psychometric method, Psychopathology, Taxometrics, Taxonomy, TAXSCAN, Types

I can highlight my theme with an anecdote that exemplifies how *not* to think about the topic. A highly capable postdoctoral student, someone who had thought deeply about trait theory and psychometric factors, a person as free of dogmatism and unquestioned assumptions as one can be in the "soft" areas of social science, was being considered for a job at a prestigious psychology department. After giving a first-rate talk on the well-corroborated eye-tracking anomaly in schizophrenics and their first-degree relatives, he thought he had a pretty good chance at the job. No offer was made, however; and a friendly faculty member consoled him, saying, "Oh, there's no doubt about your excellent qualifications, but, you see, we are dimension people, and you are a category person." The candidate concluded, given that indication of the level at which that clinical faculty functioned methodologically, he would be happier elsewhere.

There are two things wrong with the attitude of that department. First is the element of subjectivism, which increasingly infects contemporary social science; questions about the existence of categories (types, taxa, species, disease entities) are treated as a matter of taste rather than as factual questions involving causality and mathematics. It is as if, in espousing a theory of, say, schizophrenia, one was like a willful child contemplating the offerings in a candy store, "That one looks good," "I'll take two of those," "I hate chocolate," and the like. Second, aside from wallowing in subjectivity, we have a polarization in which one is *for* or *against* the use of certain concepts, rather than trying to understand the conceptual and statistical relations that might exist between them.

There is a dogmatism among many American psychologists that no taxa exist, or could *possibly* exist, in the realm of the mind. Are they unaware of the more than 150 Mendelizing mental deficiencies, to mention one obvious example? I have even heard it said that one should not think typologically in other areas of life, either, which would discount the several hundred infectious diseases known to medicine or the kinds of rocks in geology. Apart from theoretical science, the world of ordinary objects and events presents us with hundreds of taxa for which we have accepted labels and strong expectancies, different "natural kinds" as well as different artifactual kinds. We could not function in daily life if we were forbidden to use labels for substances and objects, being required to speak only a post-Galilean language of dimensions. When we empty the dishwasher, we sort utensils into knives, forks, and spoons, and we do not consider the possibility of coming across sporks.

This article is an expanded version of an invited address given on receipt of the Lifetime Achievement Award in Basic and Applied Research in Psychology from the American Association for Applied and Preventive Psychology, Washington DC, May 1997.

I am grateful to Leslie J. Yonce for programming and preparing the tables for this article.

Without knowing chemistry, my grandmother habitually used the terms 'baking soda,' 'salt,' and 'sugar' to designate taxa on the kitchen shelves. In the nonlife sciences, how could anybody who had taken a freshman inorganic chemistry course be unaware of the chemical elements, which are distinct natural kinds? In political psychology, a nonbiological social science, the taxonic term 'Trotskyist' denoted persons identified by two pathognomic beliefs—that the Soviet Union was a workers' state and that Stalin was a counterrevolutionary bureaucrat—from which one could predict a dozen other beliefs and attitudes with 95% accuracy. This ideological syndrome was more tightly knit than the great majority of organic diseases. It is of course appropriate for social scientists to bet on their favorite horses as to what future research will show, and I do not complain when others bet differently from me. But it is against the spirit of science to express dogmatic certainty about the nonexistence of entities of a certain sort merely on the grounds that sometimes putative natural kinds turn out to be illusory.

Although I've been interested in taxometric research for many years, I am neither a "procategory" nor an "anticategory" clinician or theoretician. I am persuaded by genetic and psychometric evidence that there are some genuine types of persons in psychopathology; but my main criticism of the *Diagnostic and Statistical Manual* (DSM-IV; American Psychiatric Press, 1994) is the proliferation of taxa when the great majority of clients or patients do not belong to any taxon but are simply deviates in a hyperspace of biological, psychological, and social dimensions, arousing clinical concern because their deviant pattern causes trouble. Further, for that minority of DSM rubrics that do denote real taxonic entities, the procedure for identifying them and the criteria for applying them lack an adequate scientific basis. I am convinced that schizophrenia is a taxonic entity. I am also convinced of the reality of unipolar and bipolar affective disorders, hardcore psychopathic deviate syndrome (not the mess called antisocial personality, but the Cleckley–Hathaway–Lykken psychopathic personality, whose Minnesota Multiphasic Personality Inventory [MMPI] [Hathaway & McKinley, 1940, 1983] profile is a pure 49′, having 2-sigma peaks on Pd [psychopathic deviate] and Ma [hypomania] scales, with no other elevations), textbook cases of compulsive-obsessive neurosis, panic disorder, and a variety of organic brain syndromes. About hysteria or histrionic personality, I have my doubts, although I have seen a few in the past. Hypochondriasis, I tend to think, with Bleuler (1924, 1950), is usually an atypical schizophrenia. I discount the psychophysiological disorders, because I take a Skinnerian view of them as based on respondent conditioning, not being motivated as in the striped muscle conditioning of Type R, and—whether it's in the skin, or the gut, or the bronchi—being largely determined by nonpsychological factors. I hope that small list makes clear that I am *not* an advocate of the view that there are numerous valid taxa. The American psychology tradition, which I learned from Donald G. Paterson in 1939, was strongly antitypology. We took it for granted that apparent types (e.g., 'introvert') were merely regions on a dimension. This viewpoint was partly based on evidence and partly biased, being a reaction against careless typologizing by nonquantitative European psychologists and psychiatrists.

Among the half-dozen great changes in thought mode that characterized the rise of Galilean science, historians and philosophers of science agree that a shift of emphasis from looking for qualitative Aristotelean *essences* to causally relevant *dimensions* played a crucial role. The development of instrumentation to quantify those dimensions accurately and the development of analytic geometry, calculus, and probability theory to theorize about them was a major breakthrough. I consider Lewin's classic work on the contrast between Galilean and Aristotelean modes of thought (Lewin, 1935) an important metatheoretical contribution. When quantified measures of intelligence, special abilities, and personality traits were rendered more theoretically interesting and powerful by Thurstone's invention of factor analysis, the dimensional orientation became firmly ensconced. But very few psychologists remember something Thurstone (1935) wrote about this in *The Vectors of Mind*, and even I, with my taxometric interest, had forgotten it until my coauthor quoted it in Waller and Meehl (1998):

> It is conceivable, and not improbable, that some reference abilities will be found to be sufficiently elemental that they can be declared to be either present or absent in each individual without intermediate gradations in amount or degree of presence…. *If only two numerical values occur in the population for the standard scores in a primary ability, then the primary ability is a unitary ability*. This is a genetic interpretation of factors. (1935, pp. 51–52)

> A special case of the positive manifold is that in which each factor is either completely present or entirely absent in each member of the experimental population. Each individual member of the population has, then, one of only two possible standard scores—one positive, which represents the presence of the trait, and the other negative, which represents the complete absence of the trait. The numerical values of these two possible standard scores are determined by the proportion of the population that has the trait. (Thurstone, 1947, p. 343)

To infer from the emergence of factors in a correlational study that there are *only* factors and no taxonic distributions of factors is, in addition to being dogmatic, a mistake in mathematical reasoning. Doubtless the common psychometric situation involves one or more quantitative latent factors, each having a noncomposite distribution, *but not always*. Sometimes a factor's distribution is a mixture, perhaps because of a second-order factor that is dichotomous. The ideal formalism leaves open the taxonic question, and when Thurstone's classical model is generalized so as to allow factors with 0,1 distribution, interesting theorems are derivable that

prescribe powerful taxon-detecting techniques. This achievement by Niels Waller is presented in Waller and Meehl (1998).

I would like to offer a general rigorous definition of the word "taxon," but after trying for many years, I have been unable to generate one. Fortunately, we need not define all terms explicitly or operationally, as is misleadingly told to beginning psychology students. When teaching, I provide a formal-numerical implicit definition and call it "taxon$_{PM}$," meaning taxon as defined by Paul Meehl's method. If a population (patients, honeybees, stones, daffodils) is characterized by several quantitative observables such that statistical relations among these observables satisfy certain taxometric criteria derivable from a postulated latent structural model, then the situation is taxonic. In the case of a disease entity, one can combine that condition with a plausible specific etiology (e.g., a germ, a gene, a vitamin deficiency, a trauma). The largest collection of taxa that we find in human psychology is in the *Dictionary of Occupational Titles* (U.S. Department of Labor, 1977). The only specific etiology you can give for "brain surgeon" is that it's a physician who has had a residency in brain surgery, and hence possesses a package of perceptual and motor skills, verbal knowledge, and certain attitudinal and self-concept characteristics. The entries—roughly 22,000 of them—in the DOT comprise one of the biggest collections of taxa that exist, being exceeded only by the number of species in biology. If we delete such environmental mold taxa in the human psychosocial domain as being a special case, we could reiterate the request for a generic definition; but I still wouldn't know how to specify one. A verbal definition of taxonicity that would be sufficiently general, cutting across various domains of science, would have to be a very long intricate paragraph clumsily stating in words what the mathematics does rigorously and succinctly. Analogously, suppose one were asked to define the mathematician's term, "Jacobian," without using the symbolism for a determinant or a partial derivative. It would take more than a paragraph, and would be completely incomprehensible except to the mathematician, who would recognize it as a useless verbal rendition of powerful mathematical notation.

Although it is intellectually unsatisfying to lack a highly general cross-domain conceptual definition of "taxon" (see Meehl, 1992, especially the list of taxa on p. 163), I offer a consoling thought. Would it help us much to possess such a "verbal-conceptual" explication supplementing our formal-numerical one? I doubt it. Suppose a researcher in political science wonders if *right-wing extremist* is taxonic, as is, say, *Trotskyist*, and asks whether taxometrics can help decide that. Do we need a generic definition to advise the scientist? Rather, I would inquire concerning the research problem, the nontaxometric facts known, the theories in contemplation, and the indicators available. Searching queries would be, "What, roughly, do *you* have in mind in speaking of a type, class, real kind? What are some analogous examples in political science that you are convinced are, and are not, types? Why is mainstream liberal not a type but Trotskyist is? What facts lead you to suspect right-wing extremist is a type? Do you have any conjectures as to how such a type could arise? What sorts of evidence would tend to confirm its existence? What would refute it? Why couldn't it merely be a subgroup of individuals who deviate importantly on several dimensions? Is your interest pragmatic (e.g., to predict something useful) or theoretical (to understand the causal or compositional structure)? What would it mean for the supposed type to be illusory, an artifact? How would you proceed differently, in research or practical application, if it were taxonic or not?" I believe the replies to these questions would usually reveal whether or not the taxometrician has a contribution to make.

For some reason, it seems difficult to think clearly about the taxonic problem, even for someone who has had a lot of practice at it. I, who have been reflecting on this for years and have more awareness of the philosophical aspects (as distinguished from the mathematics) than most psychologists, sometimes fall into stupid mistakes. Recently, I found myself referring to the "false-positive rate" when discussing a latent nontaxonic structure, which obviously can't have a false-positive rate, because there is no taxon. The late Eliot Slater, one of the clearest thinkers that psychiatry ever had, wrote an article on hysteria (Slater, 1976) in which he criticized another clinician's list of hysterical traits on the ground that each one of them was dimensional and that all of us have at least a little bit of each. This is an astonishing blooper for a psychiatrist with Slater's methodological knowledge and sophistication. If signs, symptoms, or traits being quantifiable and showing sizable individual differences among normals as well as among the pathological group were a valid basis for rejecting a diagnostic concept, most of the organic diseases, which Slater and everybody would consider as clearly taxa—measles and Huntington's disease and phenylketonuria are taxonic entities if anything is—would be rejected. Most symptoms, signs, and laboratory tests used in organic medicine are quantitative when scrutinized closely and almost always show overlap between the sick and the well. I have a mildly elevated blood sugar and am technically classified as "glucose intolerant." (The books say it is bad tactics to label a patient like myself as a diabetic.) I have diabetic ancestry and, therefore, probably some genes for this condition. I have had numerous glucose-tolerance curves drawn over the years, and almost every one is anomalous. There are many nonpatients whose blood sugar is simply at the high end of the "nondiabetic" distribution. Do we conclude from all of this that diabetes is a spurious entity, that no such taxon exists? Of course not.

Despite efforts by me and others over many years to clarify these elementary points, psychologists continue to make similar reasoning mistakes. For example, we continue to see the statement that if the latent structure is categorical, then quantitative indicators of the conjectured latent taxon should be bimodally distributed. A classic article by epidemiologist

Murphy refuted this in 1964, and I have cited that article repeatedly for a quarter-century with negligible effect. With equal variances, two normal curves must be separated by at least two standard deviations to give a discernible manifest bimodality, and this will not appear if the base rate of the taxon is less than .50. What happens to the distribution of the manifest indicator with changes in the base rate is somewhat complicated. Starting with a certain separation and base rate .50, we observe a symmetrical curve with pronounced platykurtosis. Holding the separation fixed, as the base rate declines, we get increased skewness and the peakedness changes to leptokurtosis. A proper search principle is not a *rule* but a *rough guideline*; namely, extreme skewness or marked platykurtosis, lacking any obvious distortion of the metric or bias in the sampling, are *suggestive* of latent taxonicity. Although a markedly nonnormal manifest distribution cannot be trusted as an inclusion test, it is at least evidentiary. When departure from normality is bimodal, most persons would agree (despite Murphy's mathematical point) that it is fairly strong evidence for taxonicity. But the reverse argument does not hold even weakly; that is, unimodality cannot function even as a weak refuter of a taxonic conjecture. Psychologists should familiarize themselves with the informed literature on this subject and quit propagating the mistake.

Another error is requiring a latent category to be "sharp," a term that does not even have an exact mathematical meaning, unless one is talking about a step-function or a nonarbitrary cut on some quantitative dimension. Most organic diseases and genetic syndromes are not "sharp" as regards their quantitative indicators, or as a weighted composite of them. If a sharpness exists, it lies not in the syndrome but in the specific etiology (e.g., a major gene rather than a polygenic system).

Another common error is to think that if factor analysis reveals a big quantitative factor, that refutes the categorical hypothesis. It doesn't, nor does it even argue against it. On the contrary, if a set of indicators have appreciable validity for a latent taxon, then a conventional factor analysis must necessarily reveal a big factor. Some of my writings may have contributed to the confusion on this point because I treat the manifest indicators as quantitative and draw their distributions with respect to the taxon. This is not incorrect; but perhaps I should have reminded readers of a rather obvious point: psychometric indicator variables that discriminate latent categories usually do so because they are loaded with a latent quantitative factor, *which in turn* has two partially overlapping latent distributions.

Some think that a taxon must necessarily have arisen from a specific etiology in the medical or genetic sense; that is, it must have a dichotomous causal factor such as a germ or a gene. (Even here, note that *how many germs* is quantitatively relevant to probability of sickness, as is how many carcinogenic mutations, or how severe a vitamin deficiency—all these operate causally as quasi-step functions, i.e., graphs steeper in one region.) Although these are the obvious examples and easy to explain, there is nothing about the etiological concept of a taxon, or about the mathematics of detecting one, that says anything so restrictive about substantive causation. A specific environmental event, such as postulated in Freud's (1896/1962) theory of the etiologies of hysteria versus obsessional neurosis, constitutes a good specific etiology of a social kind. If a polygenic system has a quasi-step function, where in a certain narrow region there is a rise in probability of illness from zero to some substantial value (e.g., as in Irving Gottesman's [1991; Gottesman & Shields, 1982] theory of schizophrenia), this is, of course, compatible with formal taxonicity.

In behavior genetics, some think that taxometric analysis is unimportant because when the neurochemist finds, say, the "purple spot" (as Seymour Kety called it, referring to chromatography) that is specific for schizophrenia, simple pedigree analyses, using psychopathology only to locate pedigrees at risk, will tell us all we want to know. Some psychologists are troubled by this, worrying that the geneticists will "scoop" us here. That is a possibility, but—trade union competition aside—I refuse to worry about it. At the phenotypic level—including psychometrics, soft neurology, and psychophysiology—schizophrenia, or, what I emphasize, *schizotypy*, and even more, the basic neurological aberration *schizotaxia*, is what logicians call an *open concept*. To offer a so-called "operational definition" which lists necesssary and sufficient criteria as logical disjunctions and conjunctions of signs and symptoms is self-deceptive and pretentious, considering the statistical model we face. Implementing DSM IV by a structured interview (e.g., Schedule for Affective Disorders and Schizophrenia [SADS]; Endicott & Spitzer, 1978), it can be inferred that we will not be more than 90% accurate when we label someone as a schizophrene (Spitzer, Endicott, & Robins, 1978). Most clinicians, including those engaged in research, will run lower than that. On my dominant gene theory of schizophrenia (Meehl, 1990), around 80% of correctly diagnosed schizophrenes are schizotaxic (what the gene determines). If everybody cooperates, and the biochemists test the blood, or cerebrospinal fluid, or whatever is "Kety-infallible," we expect 36% of first-degree relatives to show the purple spot. Will such a finding be a clincher for everybody? Hardly. A conventional nontaxometric showing that 36% "differs significantly from controls" (e.g., sibs of nonschizophrenic psychiatric patients = 18%, sibs of "normals" = 10%) merely corroborates a genetic influence, which we already know. It does not help us in choosing among genetic models. Many would see it as tending to refute my dominant gene model; whereas the proper taxometric methodology recognizes 36% ± ΔP (small tolerance) as a strong corroborator, passing a Popperian "risky test." And how about an obtained 31% or 40%? The point is that a clear showing that something micro-anatomic or biochemical is the specific etiology of a schizotaxic brain that underlies the development of the schizotypal personality, only a minority of whom develop

clinical schizophrenia, hinges on a good quantification of the three theoretical entities (i.e., schizotaxia, schizotypy, and schizophrenia). It is sufficient, although not optimal (for such genetic model appraisals), to know with some précision the *proportion* of the inferred taxon in a specified population at risk, without knowing with high confidence *which individuals* belong to the taxon versus the complement class.

The merit of a taxometric method may be studied in five ways, some better than others. First, the mathematics speaks for itself, a point that some psychologists seem not to appreciate. One does not "validate the mathematics" by empirical facts; a set of physical entities that satisfies the postulates of a formal system (i.e., is a model of the calculus) necessarily satisfies the theorems. The only way a mathematically derived search procedure can fail to do its job is by its formal-numerical conjectures being erroneous, or because *N* is too small to avoid gross random sampling errors. Second, Monte Carlo runs on artificial data are powerful support for a method's accuracy, provided that the parameter space is widely sampled. We know what we put into the computer, and we find out whether the procedure delivers the correct answer. Third, we apply the method to real data pseudo-problems where the truth is known with high confidence by direct epistemological paths rather than taxometrics (e.g., pathologist reports in organic disease, DNA or biochemical path in genetic syndromes, experimental manipulation of qualitatively distinct problem-solving strategies, self- and group-labeled membership in religious sects or political parties). For example, given our clear knowledge that the MMPI item pool is *capable* of psychometrically identifying the sexes with 85%–90% accuracy, a taxometric method that falls far short of this hit-rate in telling males from females is not praiseworthy. Fourth, the taxometric method should agree with other mathematically independent methods of detecting latent classes. This epistemic avenue involves complex issues in sequential versus concurrent "validities" that I cannot elaborate here; but in a suitably hedged sense, two search methods (*each having its independent formal rationale*) can, in principle, corroborate one another by showing high agreement. Fifth, a taxometric method should have a good track record in solving empirical problems. Here again, lacking direct access to theoretical truth, the logician's Total Evidence Rule is all we can rely on—the taxometric result makes "theoretical sense," facilitates technological efficacy, enables us to sharpen measures, suggests fruitful next research steps, and so on. *Coherence* of distinct, nonredundant procedures within the method as well as with nontaxometric information is the guiding principle.

There are several methods of detecting conjectured latent taxa: mixture models, cluster analysis, inverse factor analysis, latent class analysis, and my taxometric method and extensions of it by Robert R. Golden, William M. Grove, and Niels G. Waller. A taxometric search algorithm should be rationalized by a postulated theoretical model rather than by roughly quantifying a semi-intuitive notion of some sort of closeness in the descriptor space. My *method* involves several distinct, minimally redundant *procedures* which, if the taxon exists and the inferred latent values are fairly accurate, must agree with one another within tolerance. Agreement of these distinct procedures already constitutes a kind of consistency or coherency test. But in addition, we have developed several further consistency tests that will not be satisfied if the taxon is illusory or if an unfortunate sampling error or gross distortion of the metric is at work. Any adequate taxometric method must contain internal tests that tell us whether or not it is doing its job, because if we had acceptable external criteria we wouldn't be doing taxometrics in the first place. It is not enough that a taxometric search method often delivers correct results, it should provide red warning flags on the occasions when it misleads. Mine is the only taxometric method that provides these internal consistency tests. (Perhaps Lazarsfeld's [1959] higher order "accounting equations" are a kind of consistency test, but his latent structure analysis is not taxometrics, the indicators being dichotomous and the latent variable being quantitative—reverse of the taxometric case.)

Over the last 30 years, my colleagues and I have concocted eight distinct, largely nonredundant search procedures for identifying a latent taxon and five modified versions of these, for a total of 13 procedures. This may seem to be methodological overkill, but it isn't. Absent a Silver Standard external criterion (we never have a Gold Standard), which would obviate using taxometrics in the first place, the evidence to support the inferred latent structure consists of the coherence of numerical values inferred from the various theorems that can be derived from the postulates that implicitly define the latent structure. The term "postulate" here designates a *theoretical conjecture*—subjected to indirect but strong empirical tests—not an "assumption" in the usual (dangerous) sense. Taxometrics is not different from other kinds of science in this respect; it is more apparent here, because of psychologists' historical reliance on *definatory* external criteria of a test's validity. Taxometrics, like factor analysis and multidimensional scaling, is construct validity par excellence. Only two of the main procedures, MAMBAC (acronym from "Mean Above Minus Below A Cut"; Meehl & Yonce, 1994) and MAXCOV (MAXimum COVariance; Meehl, 1973, 1995b; Meehl & Yonce, 1996), have been published in detail (for a summary of them, see Meehl, 1995a).

I call my method *coherent cut kinetics*, referring to both the epistemology and the mathematics of the approach. We move cuts on a designated input variable and study the statistical behavior of other (output) variables on cases in regions demarcated by the cuts. Inferring latent parameters (base rates, means, valid and false positive rates), we look for numerical consistency over different variables and over different procedures. We say *kinetic* because the cuts move, *coherent* because the inferences should be consistent.

Formulating the two procedures briefly, the MAMBAC procedure stems from a simple intuition: Suppose one could

classify subjects accurately into two groups by the use of an observable indicator *x*; then the classified groups will differ on another valid indicator *y*. Other things being equal, the more accurate the classification (that is, the more hits and fewer misses achieved by a classificatory cut on *x*), the larger will be the two groups' mean difference on *y*. In MAMBAC, we determine cuts on *x*, calculate the mean *y* for the cases above and below each *x* cut, get the differences between these *y* means, and graph the values as the *x*-cut moves. If the latent situation is taxonic, this graph will tend to be humped, the height and location of its high point depending jointly on the indicator separations ("validities") and the taxon base rate. With a base rate around .50, the maximum is near the center; as base rates become smaller, the peak moves to the right, becoming merely a high point at the end for very small base rates. If one has four quantitative candidate indicators (the more, the better), there are six ways to pick a pair; and because MAMBAC can be run bidirectionally (interchanging each input-output pair), we have 12 graphs to look at. If the latent structure is nontaxonic, with, say, a single quantitative factor unimodally distributed (it need not be normal) and generating an ordinary correlation between *x* and *y*, the MAMBAC graph is dish-shaped. The depth of the dish depends on the correlation and hence on the factor loadings of the indicators *x* and *y*. Although once in a while a single MAMBAC graph may leave one in doubt, a panel of 12 is easy to sort inspectionally, even by persons without statistical or psychological education.

The MAXCOV procedure requires three indicator variables and relies on the mathematical fact that the covariance of two indicators in a subset of cases is a function of the subset's taxonic mix. That mix is maximum in the interval of a third variable (which we call the *input*, not in a causal but only a statistical sense) where the latent frequency functions of the input variable intersect (i.e., $p = q$). An *x*-cut that maximizes hits is called the hitmax cut and the interval containing it the hitmax interval. As with MAMBAC, if the latent structure is taxonic the MAXCOV graphs are peaked, the peak location and height depending on the base rate and latent validities. Nontaxonic factorial structure yields a flat graph. Inspectional sorting of 450 Monte Carlo 12-graph panels by two psychologists and two laypersons yielded zero misclassifications.

Each procedure permits a calculation of the taxonic base rate and other latent values with good accuracy. The procedures are almost wholly independent mathematically. MAMBAC deals with means, and MAXCOV with covariances; MAMBAC means are computed for entire regions above and below cuts, and MAXCOV covariances for sliding small intervals. Good numerical agreement between the procedures corroborates both the taxonic conjecture and the trustworthiness of the latent values.

These two procedures have been applied to various research problems in psychopathology and personality theory some two dozen times around the world, and other studies are in progress by psychologists, psychiatrists, and geneticists. A nice example of taxometrics definitively answering a vexed question in psychopathology is Waller, Putnam, and Carlson's (1996) use of the MAXCOV-HITMAX procedure to clarify the controversial concept of dissociative disorder. Starting with Bernstein-Carlson and Putnam's (1986) 28-item Dissociative Experiences Scale (DES), they showed the existence of a dissociative taxon, not merely a continuum of propensity to dissociative experiences. They further found that only one third of patients labeled "dissociative disorder" belonged to that taxon, and only eight of the DES items were good taxon discriminators. The item content clearly reflects the difference between "pathological" dissociative experiences and "normal range" individual differences in absorption, absentmindedness, and other nonpathological personality dispositions. The striking U-shaped distribution of diagnostic probabilities (Waller, Putnam, & Carlson, 1996, Fig. 3, p. 313) is alone sufficient to convince one of taxonicity.

Despite the Minnesota tradition of so-called "dust-bowl empiricism," in which I was educated, I strongly advocate a theory-motivated choice of candidate indicators as preferable to an exploratory search of a large heterogenous collection of signs, symptoms, traits, and test scores. Some psychologists prefer to proceed atheoretically in an exploratory way, however, and, for that reason, we have conducted preliminary Monte Carlo studies of a "blind, empirical" search procedure that merely conjectures there may be multiple taxa underlying a batch of indicators, some of which may have been chosen for theoretical plausibility and others just being available in the clinic files. Details of this TAXSCAN procedure will be presented in a subsequent publication, but here are initial results for two simple situations.

Consider a case in which there are four taxa with base rates .40, .30, .20, and .10, each subject belonging to one of the taxa. Here the *complement class* of each taxon is composed of the three other taxa. Each taxon is discriminated by three quantitative indicators that do not differentiate among the three taxa in the complement class. The first step is to calculate an ordinary Pearson correlation between all the indicator pairs. If one has a very large number of indicators, as is typically true in such atheoretical exploratory scannings, this reduces the huge number of MAMBAC graphs to be inspected (e.g., 9900 for 100 indicators). We first identify indicator pairs whose Pearson correlation exceeds a value that would be generated by the weakest taxon one hopes to recover, say, a base rate of .10 and small separations. For a Monte Carlo study of 25 samples, Table 1 shows the four taxa clearly by triads of correlations ranging from .27 to .49, the rest of the correlations in the Table being negative. (Lest the reader worry that individual Monte Carlo samples may not look like this summary matrix, the standard deviations of the means in Table 1 range from .03 to .05.) These negative correlations, which I call *parataxonic*, arise from the mathematical fact that each taxon belongs to the complement class of each other taxon. If there is no within-class correlation, a negative correlation between indicators that discriminate two different

Table 1. *Pearson r Matrix* (E*ach Cell is Averaged Over 25 Monte Carlo Samples, N = 600*) *for the Situation in Which There are 4 Taxa* (*base rates P = .40, .30, .20, .10*) *and Each Individual Belongs to Only One Taxon.*.

	P = .40			*P* = .30			*P* = .20			*P* = .10		
	1	2	3	4	5	6	7	8	9	10	11	12
1		.35	.38	-.15	-.16	-.18	-.13	-.13	-.15	-.07	-.08	-.09
2			.49	-.20	-.22	-.24	-.16	-.17	-.19	-.08	-.09	-.08
3				-.22	-.25	-.26	-.16	-.18	-.21	-.10	-.10	-.11
4					.36	.38	-.10	-.11	-.12	-.07	-.06	-.07
5						.41	-.12	-.14	-.13	-.08	-.07	-.07
6							-.13	-.14	-.15	-.09	-.07	-.08
7								.35	.39	-.05	-.06	-.05
8									.42	-.05	-.06	-.05
9										-.07	-.06	-.06
10											.28	.27
11												.27

Variables that discriminate the same taxon are correlated with each other. Parataxonic correlations are negative

taxa follows as a consequence. If candidate indicators show a clear MAMBAC panel (typically six taxonic graphs), we estimate the base rate with each possible pairing and pool these six estimates to get one $\hat{P}$ for each Monte Carlo sample. The accuracy of these base rate estimates obtained by MAMBAC is shown in Table 2, where the $\hat{P}$ values averaged over all 25 Monte Carlo samples are given. Having found evidence of taxonicity with MAMBAC, we would then subject the indicator triads to the MAXCOV procedure, whence we would get three graphs to look at and more estimates of the base rate and other parameters. A more complicated situation occurs when there are indicators that discriminate the taxa but there are also nontaxonic measures that are correlated by virtue of nontaxonic factor loadings on some underlying quantitative factor. Suppose the factor loadings for these complement class cases have values that generate Pearson correlations in roughly the same ranges as those taxonically generated. A second Monte Carlo study considers the case of two taxa having base rates .30 and .15, so that the majority of our subject sample do not belong to either of the taxa and the Pearson correlations they show are caused by nontaxonic latent quantitative factors. Table 3 shows the averaged correlation matrix (standard deviations for these means range from .02 to .05), which, of course, includes several correlations that could be generated taxonically or not. This is highly representative of the individual sample matrices.

Table 2. *Base Rate Estimates* (*Using MAMBAC Procedure*) *for the First Monte Carlo Study.*

True *P*	$\hat{P}$	*SD*	Mean ($\hat{P} - P$)	Mean | $\hat{P} - P$ |
.40	.42	.04	.02	.04
.30	.31	.03	.01	.03
.20	.21	.03	.01	.03
.10	.12	.06	.02	.05

The parataxonic correlations are negative; the intrataxonic and nontaxonic (factorially generated) correlations are positive, and the remaining correlations are negligible. Usually an investigator's single matrix will not be pre-grouped by taxonic and factorial indicators as this one is. We must apply taxometric procedures (first MAMBAC to find variables that are pair-wise taxonic, then MAXCOV to the triads) to all the positively correlated indicators to identify (e.g., in this case) the two taxonic triads as distinct from the other correlations. When the taxa have been identified, the indicators may be rearranged in the correlation matrix, grouping those that identify taxa and the nontaxonic ones, to reveal the clusters of taxonic, parataxonic, and nontaxonic correlations. The accuracy of the taxon base rate estimates obtained with MAMBAC (averaged over the 25 Monte Carlo samples) is shown in Table 4. If it seems surprising that the nontaxonic factors do not tend to deceive us, the reason is that—unlike cluster algorithm methods—the presence of correlations in the matrix that are nontaxonically generated has no influence on those that are. The influence of nontaxonic latent quantitative factors is only "noise" when we are first looking at the correlation matrix; it does not function as noise, masking taxonic signal, in the next two steps.

How clearly taxonic or nontaxonic are the MAMBAC and MAXCOV curves in these two situations? In a visual sorting of 200 panels of MAMBAC graphs (six curves per panel) generated by each cluster of three highly correlated variables in the matrices in Tables 1 and 3 (including the nontaxonic clusters in Table 3), I (seeing these particular curves for the first time) misclassified only one panel. Visual sorting of the MAXCOV curves alone was slightly less accurate; 96% of the 200 panels were correctly classified. When both the MAMBAC and MAXCOV panels were provided for the previously missorted samples (randomly mixed with a control subsample of panels initially sorted correctly), classification was correct in all cases except one. A researcher is in the latter

Table 3. *Pearson r Matrix* (E*ach Cell is Averaged Over 25 Monte Carlo Samples, N = 600*) *for the Situation in Which There are 2 Taxa* (*base rates P =.30,.15*) *and Variables That Do Not Discriminate a Taxon But Have Factor Loadings* (*Ranging from .50 to* .70) *on a Nontaxonic Quantitative Factor.*.

	P = .30			*P* = .15			*Nontaxonic variables*					
	1	2	3	4	5	6	7	8	9	10	11	12
1		.36	.40	-.10	-.10	-.10	.01	.00	.00	.01	.01	-.00
2			.42	-.10	-.10	-.10	-.00	-.00	-.01	-.02	-.01	-.00
3				-.11	-.10	-.10	.01	.01	.00	.01	-.01	.00
4					.33	.33	-.01	-.00	.00	-.01	-.01	.01
5						.34	-.00	.01	-.00	.00	-.01	.01
6							-.01	-.00	.00	.00	.01	.01
7								.47	.42	.42	.38	.35
8									.39	.38	.36	.33
9										.36	.32	.30
10											.32	.29
11												.27

Variables that identify the same taxon are correlated with each other, as are nontaxonic variables that have high factor loadings. The nine parataxonic correlations are negative and noticeably different from zero. Correlations between taxon discriminators and variables that do not discriminate taxa are close to zero.

(favorable) situation, having identified candidate MAXCOV triads by first discerning overlapping MAMBAC pairs, so curves from both procedures are known. In that sense, the triads' MAXCOV-taxonic behavior functions as a consistency test.

One common objection to my method is that the auxiliary conjecture (I never use the usual word 'assumption' in this context; Meehl, 1992; Meehl & Golden, 1982) of negligible nuisance correlation is unplausible. In fields like psychopathology we often strain mightily to get correlations *up to* .30 or .40, so it should not be difficult to hold them below that. For psychometric instruments, such as MMPI keys, one can use factor analysis or old-fashioned biserial item analysis to reduce the correlations between candidate indicator keys. For nonpsychometric measures, a suitable nonlinear transformation may reduce the correlation to safe size. One can estimate the within-class nuisance correlation with sufficient accuracy by studying carefully selected groups in which an external criterion, while imperfect, has a safe upper bound of taxon and complement class contamination. Extensive Monte Carlo runs have shown that the method is surprisingly robust under departures from the idealizations. Finally, I have presented a procedure for taking the nuisance correlation into explicit account (Meehl, 1995b), although that has not yet been adequately investigated.

Table 4. *Base Rate Estimates* (*Using MAMBAC Procedure*) *for the Second Monte Carlo Study.*

True *P*	$\hat{P}$	*SD*	Mean ($\hat{P} - P$)	Mean | $\hat{P} - P$ |
.30	.31	.03	.01	.03
.15	.16	.03	.01	.03

What about very small base rates? The method is not intended for miniscule base rates, as found, for instance, in the various Mendelizing mental deficiencies (for them, taxometric methods are seldom required). But consider a conjectured taxon that previous nontaxometric research suggests has a base rate of perhaps 3% in a clinical population. Starting with a sample of 1,000 clinic cases, one can add standard scores on, say, four candidate indicators and truncate out perhaps the bottom 700, which with low-to-moderate indicator validities will assure that nearly all taxon cases are in the remaining set. The single indicator variables will be somewhat skew as a result, but that is harmless, as my method never assumes normality. The remaining sample of 300 will have a taxon base rate of 10%, which is sufficient for the method's use. It is appropriate to comment here that psychologists should get accustomed to demanding large samples for studying certain kinds of questions, as do astronomers, physicists, chemists, geneticists, and epidemiologists. The two worst influences of Fisherian statisticians on psychology are the glorification of significance testing and the contentment with small samples. If the sample size is not large enough to investigate a taxonic problem, then the researcher should investigate something else.

Some complain that the formal numerical definition of taxonicity fails to illuminate them as to the psychological nature of the taxon identified. Of course it doesn't; neither does factor analysis or multidimensional scaling or, for that matter, an ordinary correlation coefficient or multiple regression equation or *t*-test. *No statistic is self-interpreting*, and taxometrics is not different from any other psychometric or statistical method in that regard. Here, as always, the guiding light is the logicians' Total Evidence Rule.

Although Meehl and Golden (1982) studied separations as small as .5 *SD*, we have worked mostly with separations of

1.5 to 2 standard deviations. Separations of 1 *SD* or greater are often considered excessively optimistic, but I do not understand why. I have examined typical separations in psychopathology, in organic medicine, and in various other areas of psychology, such as vocational interests, intelligence in relation to occupation, scores on oral trade tests, and the differentiation of selected groups on other tests such as the California Personality Inventory (Gough, 1987). Without having done a formal meta-analysis, I can say that separations of better than 1.5 *SD* are common; and many separations, even in soft psychology, go to 2.5 or 3 standard deviations.

There is much work yet to be done. An important issue requiring investigation by analytic, Monte Carlo, and real-data study is which consistency tests are sensitive to spurious taxonicity and which ones mainly respond to inaccurate parameter estimates despite the basic model being correct. We need to explore further the effect of extreme nonnormality of latent distributions. None of my analytical derivations hinge on normality or equal variance, but most of our Monte Carlo samples have had those features, as a matter of convenience. The extent to which nontaxonic, nonlinear dependencies present a grave danger of pseudo-taxonicity is unknown, although preliminary investigations (Cleland & Haslam, 1996; Haslam & Cleland, 1996) suggest the method is robust in this regard. We plan to investigate how badly the method works in highly unfavorable scenarios, such as small separations, extreme skewness, small base rate, excessive nuisance correlations, and markedly different variances. Both MAMBAC and MAXCOV can be employed using only a single quantitative input indicator, with the output means or covariances being dichotomous items. This has been done in several encouraging studies, but it remains to be investigated whether there is a greater danger of getting pseudo-taxonic curves with dichotomous output indicators. (I have seen no evidence of this.) The relationship between my method and other widely used methods such as mixture analysis of single variables and cluster algorithms should be studied; but what data I have suggest that when they don't agree, Meehl's taxometrics is more likely to be right (Golden & Meehl, 1980; Grove & Meehl, 1993; Waller & Meehl, 1998). Performance of taxometrics in the identification of organic disease, such as in internal medicine or neurology, when we have a Gold Standard criterion from the pathologist should be investigated.

Relying on extensive Monte Carlo runs from eight laboratories involving several thousands of single Monte Carlo samples, and the few empirical applications to date, I am prepared to say, perhaps rashly, that I have solved the taxometric problem, at least for the great majority of latent situations. An *American Psychologist* action editor once asked me to delete such a strong claim, but I pointed out that such an objection was probably because we are so accustomed, in the field of psychopathology, that nothing ever is solved. I was allowed to retain my claim provided I said I *conjectured* that I had solved it. If you are correct in a theory that there exists a latent taxon, and you have some good candidate indicators (some of them may be invalid but that fact will easily be discovered by the procedures), then applying the coherent cut kinetics method will allow you to identify the taxon, estimate its base rate accurately, locate the optimal cutting score on each indicator, estimate the valid- and false-positive rates achieved by that hitmax cut, and classify individuals via Bayes' theorem as accurately as you could with, say, Fisher's discriminant function (which requires a Gold Standard criterion to begin with). You will also have multiple consistency tests to reassure that you are not getting a pseudo-taxon. If your conjecture is incorrect, the panel of graphs will appear clearly nontaxonic and the consistency tests will be failed.

References

American Psychiatric Press. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Press.

Bernstein-Carlson, E. B., & Putnam, F. W. (1986). Development, relaibility, and validity of a dissociation scale. *Journal of Nervous and Mental Disease*, *174*, 727–735.

Bleuler, E. (1924). *Textbook of psychiatry* (A. A. Brill, Ed. and Trans.). New York: Macmillan.

Bleuler, E. (1950). *Dementia praecox* (J. Zinkin, Trans.). New York: International Universities Press. (Original work published 1911)

Cleland, C., & Haslam, N. (1996). Robustness of taxometric analysis with skewed indicators: I. A Monte Carlo study of the MAMBAC procedure. *Psychological Reports*, *79*, 243–248.

Endicott, J., & Spitzer, R. L. (1978). A diagnostic interview: The schedule for affective disorders and schizophrenia. *Archives of General Psychiatry*, *35*, 837–844.

Freud, S. (1962). The aetiology of hysteria. In J. Strachey (Ed. & Trans.), *Standard edition of the complete psychological works of Sigmund Freud*, Vol. 3 (pp. 191–221). London, Eng.: Hogarth. (Original work published 1896)

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

Gottesman, I. I. (1991). *Schizophrenia genesis*: *The origins of madness*. New York: Freeman.

Gottesman, I. I., & Shields, J. (1982). *Schizophrenia, The epigenetic puzzle*. New York: Cambridge University Press.

Gough, H. G. (1987). *CPI, Administrator's Guide*. Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (1987). *CPI, Administrator's Guide*. Palo Alto, CA: Consulting Psychologists Press.

Grove, W. M. & Meehl, P. E. (1993). Simple regression-based procedures for taxometric investigations. *Psychological Reports*, *73*, 707–737.

Haslam, N., & Cleland, C. (1996). Robustness of taxometric analysis with skewed indicators: II. A Monte Carlo study of the MAXCOV procedure. *Psychological Reports*, *79*, 1035–1039.

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, *10*, 249–254.

Hathaway, S. R., & McKinley, J. C. (1983). *The Minnesota Multiphasic Personality Inventory manual*. New York: Psychological Corporation.

Lazarsfeld, P. F. (1959). Latent structure analysis. In S. Koch (Ed.), *Psychology*: *A study of a science. Volume 3. Formulations of the person and the social context* (pp. 476–543). New York: McGraw-Hill.

Lewin, K. (1935), *A dynamic theory of personality*. New York: McGraw-Hill.

Meehl, P. E. (1973). 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. (1990). Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. *Journal of Personality Disorders*, *4*, 1–99.

Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality*, *60*, 117–174.

Meehl, P. E. (1995a). Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist*, *50*, 266–275.

Meehl, P. E. (1995b). Extension of the MAXCOV-HITMAX taxometric procedure to situations of sizeable nuisance covariance. In D. Lubinski & R. V. Dawis (Eds.), *Assessing individual differences in human behavior*: *New concepts, methods, and findings* (pp. 81–92). Palo Alto, CA: Davies-Black.

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., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*, *74*, 1059–1274.

Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). *Psychological Reports*, *78*, 1091–1227.

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

Slater, E. (1976). What is hysteria? *New Psychiatry*, *3*, 714–715.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry*, *35*, 773–782.

Thurstone, L. L. (1935). *The vectors of mind*. Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.

U. S. Department of Labor, Employment and Training Administration, Employment Service. (1977). *Dictionary of occupational titles* (4th ed.). Washington, DC: Government Printing Office.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures*: *Distinguishing types from continua*. Newbury Park, CA: Sage.

Waller, N. G., Putnam, F. W., & Carlson, E. B. (1996). Types of dissociation and dissociative types: A taxometric analysis of dissociative experiences. *Psychological Methods*, *1*, 300–321.

pdf by ljy May 2000