#160

***Notes for*:**

Meehl, P. E., & Yonce, L. J. (1994). Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*, *74*, 1059-1274. (Monograph Supplement 1-V74)

***For easier handling of this large monograph, main text and appendices are in two files*:**

160_MAMBAC_text_only.pdf (pp. 1059-1110 of the publication, including References)

160_MAMBAC_Appendices.pdf (pp. 1111-1274, Appendices A–J) .

TEXT (pp. 1059-1110)**: *Helpful information subsequent to the initial publication*:**

p. 1061 & 1110, "submitted" publication:

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

p. 1062, bottom lines, "A general treatment of the blind inductive scanning approach…" see:

Meehl, P. E. (1999). Clarifications about taxometric method. *Applied & Preventive Psychology*, *8*, 165-174. Reprinted, 2006, in *A Paul Meehl Reader: Essays on the practice of scientific psychology* (pp. 389-404). (N. G. Waller, L. J. Yonce, W. M. Grove, D. Faust, & M. F. Lenzenweger, Eds.). Mahwah, NJ: Erlbaum.

p. 1067, footnote 5: Monte Carlo samples are available online.

APPENDICES (pp. 1111-1274)**:**

LJY, October 2019

Psychological Reports, 1994, 74, 1059-1274. © Psychological Reports 1994
Monograph Supplement 1-V74

TAXOMETRIC ANALYSIS: I. DETECTING TAXONICITY WITH TWO QUANTITATIVE INDICATORS USING MEANS ABOVE AND BELOW A SLIDING CUT (MAMBAC PROCEDURE) [1, 2]

PAUL E. MEEHL AND LESLIE J. YONCE

University of Minnesota *Minneapolis, Minnesota*

Summary.—Given two quantitative indicators of a conjectured latent taxon, a statistical function defined as the difference between the observed means for cases of one indicator (designated for the procedure as the "output" indicator) falling above and below a sliding cut on the other indicator (designated as the "input" indicator) indicates whether the latent structure is taxonic or nontaxonic ("factorial," "dimensional"). If it is taxonic, latent parameters, e.g., base rate, hit rates, complement and taxon means, can be estimated. Graphs can be inspectionally sorted with very high accuracy, even by laypersons. MAMBAC (Mean Above Minus Below A Cut) is one of a related family of taxometric procedures in Meehl's Coherent Cut Kinetics Method.

CONTENTS

[1]Address correspondence to Paul E. Meehl, University of Minnesota, N218 Elliott Hall, 75 East River Road, Minneapolis, MN 55455.

[2]The authors are indebted to Robert R. Golden, who was involved in early stages of this work, and to William M. Grove, who has been a helpful consultant on many questions and who introduced the second author to the computer and to programming, generously sharing his expertise on countless occasions as we moved from crude hand-drawn graphs to the Monte Carlo evidence presented here. Without the help of these two people, it is very doubtful that this article would have been written. We are grateful to Golden, Grove, Niels Waller, and the anonymous referees whose helpful comments and corrections have enhanced the final product. Any infelicities that may remain are our responsibility.

Coherent cut kinetics is a system of procedures developed by Paul E. Meehl to determine whether the latent structure of a data set is taxonic and, if it is, to estimate the latent parameters associated with the taxon and complement (nontaxon) groups and to assign individual elements to membership in one or the other group.[3] The phrase ⌜coherent cut kinetics⌝ refers to the epistemology and mathematics of the approach: we move cuts on a designated input variable and study the statistical behavior of other (output) variables on cases in the region of the cut and 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 different variables and over different procedures. We say ⌜kinetics⌝ because the cuts move, ⌜coherent⌝ because the inferences should be consistent.

MAMBAC (Mean Above Minus Below A Cut) is one of the procedures used in the coherent cut kinetics method. The basic idea was first described by Meehl (1968, pp. 11ff) as an alternative procedure for locating the *hitmax cut* (that cut on an indicator which maximizes correct assignment of cases to the taxon or to the complement class). In fact, it does not locate the hitmax cut except in special circumstances, and the "quasi-proofs" in that technical report are unsound, probably due to poor approximation via differentials. What it does is to maximize the sum of *hit rates* above and below the cut. It was later presented as a consistency test for another procedure (the maximum covariance procedure, later called MAXCOV) and the expected shape of the graph for latently taxonic situations was described by Golden and Meehl (1973, pp. 15-16). It was used by Golden, Tyan, and Meehl (1974) as

[3]For discussion of the meaning, existence, and detection of taxa (= real, nonarbitrary categories, types, entities) in personology and psychopathology, see Meehl (1992), Meehl and Golden (1982), and methodological references cited in those papers.

a consistency hurdles test in culling items. It has been used by Waller, Putnam, and Carlson (submitted) to detect a taxon of pathological dissociation. In this article, it is shown that MAMBAC is capable of detecting taxonicity and estimating latent parameters. Although MAMBAC will be presented here as a stand-alone procedure, it is not optimally used that way. Rather it should be used as one of the battery of taxometric procedures and consistency tests in Meehl's coherent cut kinetics method, each of which will contribute indications of taxonicity (or lack of it) and produce parameter estimates. The other procedures should confirm MAMBAC results or give improved estimates in some cases for which MAMBAC is not the optimal procedure; likewise, MAMBAC will serve this function for the other procedures in the method. We will describe results to be expected using MAMBAC under both ideal and less than optimal conditions and present Monte Carlo evidence for the procedure.

This monograph is written with two audiences in mind. Because this is the first full presentation of MAMBAC, we have included extensive appendices documenting the statistical rationale for the procedure and results of our Monte Carlo tests. Researchers who merely want to apply MAMBAC to their data sets and to have a basic understanding of the procedure and how to interpret their results need only read the main text; the appendices are not necessary to grasping or using MAMBAC. Those with more curiosity about the statistical underpinnings and who want to see evidence from our Monte Carlo tests in greater detail will have the appendices readily available.

Selection of Indicators

MAMBAC requires two variables (scores, indicators), at least one of which must be continuously[4] distributed; if both variables are continuous, the analysis can be run bidirectionally. It is desirable to have additional continuous variables so that more curves and parameter estimates can be generated and so that other procedures, e.g., MAXCOV, which requires three variables (Meehl, 1973a; Meehl & Golden, 1982), can be run to provide consistency checks and confirmation of results obtained from MAMBAC.

Each indicator should be selected to give good separation between the complement and taxon groups, i.e., to have good validity, and also to be uncorrelated within either the taxon or the complement group with the other variable(s) being used, i.e., to have no nuisance covariance. *The selection of variables is in the context of discovery and is bootstrapped via the procedures.* The researcher cannot know the validity of an indicator beforehand. We must rely on clinical experience, nontaxometric research, e.g., fallible noso-

[4]Continuous is used here in the usual social science sense of *quantitative, numerical, dimensional* as contrasted with *qualitative* or *dichotomous*. Of course, no empirical numerical functions can be literally continuous in the mathematical sense, but, as in other sciences, we idealize that in the formalism, e.g., when we take derivatives to locate a maximum.

logic diagnosis, even theory, to suggest good candidate indicators. But we are not "relying" on these conjectures in the strong sense of "having to *assume*" them in order to justify using the indicators they suggest. The coherent cut kinetics procedures will help us to determine how good are the indicators that we have selected, and they will help us pick those that are better at discriminating an underlying taxon if one exists. That is, the coherent cut kinetics method (incorporating multiple procedures, MAMBAC being one of them) can be used as a "blind search" way of finding valid indicators for an underlying taxon.

We favor theoretically motivated selection of candidate indicators, where one has in mind a conjectured taxon, however loosely conceived. One of the eight defects of cluster analysis as employed by social scientists (Meehl, 1979) is the "blind" application of cluster algorithms to a mass of variables not so chosen. Despite this preference, based on historical, epistemological, and mathematical considerations to be elaborated elsewhere, we do not wish to be dogmatic on the issue. There may be circumstances in which an investigator has numerous quantitative scores on a large sample of subjects but has very weak faith in theoretical conjectures arising from clinical experience and conventional nontaxometric research. It may still be asked whether these data exhibit a pattern of relationships revealing a taxonic structure. In such situations we strongly advocate applying all of the several coherent cut kinetics procedures, because—lacking a "Gold Standard Criterion," relying on bootstrapped taxometric inferences—it is desirable to have as many epistemic paths to the inferred latent entity as possible. It is, of course, *possible* to employ MAMBAC alone in a blind inductive scanning of multiple candidate indicators. One would first compute conventional correlations between all indicators pairwise and eliminate from consideration pairs showing negligible correlations. MAMBAC is then run on the surviving pairs, and those pairs manifesting a taxonic curve shape (described below) are identified. Suppose we find that indicators *x, y, z, v* are pairwise taxonic, i.e., MAMBAC results are taxonic for all of the possible unordered pairwise combinations. Then we would be able to get $2\binom{4}{2} = 12$ estimates of the base rate, the latent means, and other parameters.

In a more complex situation, we might discover that indicators *x, y, z, u* are pairwise taxonic (indicating a taxon T_i), indicators *u, v, w* are pairwise taxonic (indicating a taxon T_j, possibly different from T_i, possibly not), but pairs *xv, xw, yv, yw, zv, zw* are not pairwise taxonic. We would eliminate the overlapping indicator *u* from the T_i set, retaining it for T_j, and then we would be able to get $2 \times 3 = 6$ estimates for the various T_i latent values and $2 \times 3 = 6$ estimates for T_j. A general treatment of the blind inductive scanning approach (TAXSCAN procedure which will also address the situation of multiple taxa) is planned for a subsequent article.

In the present article we have focused attention on taxonic separation of 2 *SD* (in latent units), with some Monte Carlo runs on smaller ones (more thorough investigation of larger and smaller separations is planned). It seems obvious that detection of a very "weak" taxon (one with low base rate and small indicator separations) will be more difficult. It also seems reasonable to anticipate decreased accuracy—either in the sense of increasing bias or from larger random sampling error—of estimates of the base rate and other parameters (although our Monte Carlo runs so far indicate that the detriment is not as great as might have been feared).

Some may consider 2 *SD* separations unduly optimistic and unrealistic in the field of psychopathology, but it is not unreasonable to expect separations at least as large as 1.5 *SD*. We would adopt a simple rule of thumb for selecting candidate indicators: when the study employs an equal number of patients and controls, the indicators should yield not less than 75% hits against crude, fallible, but relatively reliable diagnostic criteria of the usual concurrent or predictive validity sort. We could just as well have picked 66% or 80%, but let us explain why 75% seems a reasonable value. A taxometric indicator good enough for strong corroboration and accurate parameter estimates should, we think, perform at least as well as the old MMPI single scales do using a pre-DSM-III, generally unreliable psychiatric diagnosis as criterion. A refined psychometric or psychophysiological device in the present state of the art, bootstrapped from diagnoses based upon present-day research diagnostic practices and aiming at the inferred inner state (whether genetic, psychodynamic, or biochemical) rather than at the symptomatically dispersed forms of the symptoms and signs of mental disease, should, we argue, do better than the original MMPI did against the psychiatric diagnoses available in the early days of the derivation and validation of the MMPI as a diagnostic instrument. A hit rate of 75% with base rate $P = \frac{1}{2}$, assuming two overlapping normal distributions of equal variance to make the arithmetic easy, entails that the cutting score for that hit rate in the symmetrical case will be at the 75th percentile of the control distribution and at the 25th percentile of the taxon distribution. This percentile cut is at one probable error (*PE*, in the old terminology), which is at .6745 *SD*, or approximately two-thirds of a standard deviation above and below the two means of the complement and taxon groups, respectively. For this symmetrical case, that locates the hitmax cut midway between the latent means; thus the distance between the complement and taxon means is double the hitmax distance from each, or approximately $\frac{4}{3}$ *SD* (1.33σ) above the complement mean. In terms of a single scale on the old MMPI, this would require a separation such that the mean *T* score of the pathological group would be $\overline{T} = 63$, which is clearly worse than is found with any of the MMPI scales in respectable studies (the only scale which approaches that feebleness of discrimination is

Scale 6, Pa). Any clinician familiar with the MMPI research literature knows that almost all the scales considered singly, i.e., without regard to profile pattern, yield mean T scores, for reasonably carefully diagnosed pathological groups corresponding to the scale's name, in the region $T = 70—80$; some of the better, i.e., longer and more valid, scales come closer to 80 or even 85 (see, e.g., Lanyon, 1968). We think it not unreasonable to advise researchers to use indicators—whether psychometric, interview rating, biochemical, or physiological—that yield at least 75% (latent) correct classifications in the symmetrical case.

Adopting a rule of thumb "at least 75% hits" against chart diagnosis (concurrent validity, imperfect—no "Gold Standard Criterion") is intended to set a plausible lower bound on taxonic hit rate (construct validity for Omniscient Jones's perfect criterion, unavailable to the investigator until *inferred from* the taxometric results). It cannot provide a strictly safe lower bound for several reasons. Clinicians' diagnostic errors are likely to be correlated, vitiating a standard attenuation correction. A deeper (and intractable?) obfuscator is intrinsic deficiencies in the diagnostic *construct* itself (unreliability aside) which, of course, it is the aim of taxometrics to correct. Further, this "conceptual" nuisance influence may be correlated with the indicators, producing reliable but invalid components, tending spuriously to boost concurrent validity. Thus we have countervailing influences tending to inflate and depress concurrent diagnostic validity from the true latent taxonic validity, and no precise way to estimate their net effect (cf. Meehl, 1990e). *It is taxonic construct validity that is our regulative ideal in all these matters*, but that elusive number is known only indirectly and approximately, as the outcome of our taxometric labors.

Cut optimality can be illustrated with a simple hypothetical situation in organic medicine where symptoms and signs are specified by cuts on a pair of quantitative indicators and the results tallied in a 4-fold concordance table. Suppose the disease entity meningitis (a taxon) produces markedly elevated temperature, say, temperature $\geq 105°$ in the acutely ill, and—for the noncomatose—intense pain upon anteroflexion of the head, a symptom related to the "objective" Brudzinski sign. Let the meningitis patients be mingled with patients having other diseases, some of which produce fever (but of lower degree) and "stiff neck" (but less painful), and some of which produce only one or neither of these; and we will also include some cases free of any illness. If we define the *high fever sign* by a cut at the symptomatic temperature $\geq 105°$, and *neck sign* by the symptom excruciating pain upon anteroflexion, all of the meningitis cases will manifest both signs and symptoms, and none of the other sick or the well cases will do so. Hence the 4-fold table of sign/symptom concordance will exhibit no tallies in the discordant ($+/-$ or $-/+$) cells, and the ϕ coefficient will be 1. Let us now move one or both cuts downward, defining "fever" sign as temperature $> 100°$

and "stiff-neck" sign as any pain, mere discomfort, or resistance on bending. These lower fevers will now be found in patients with a variety of nonmeningitis diseases, some of whom will have stiff necks, others not. For instance, patients with minor muscle or joint stiffness due to bacterial infectious illness will usually be febrile, those with a viral infection usually afebrile. People can have a "stiff neck" from such conditions as arthritis, trauma, infectious disease, sleeping posture, chilling, or hysteria; some will be febrile, some not. These varying etiologies underlying the signs and symptoms will produce numerous cases tallied in the discordant cells of the 4-fold table, and the ϕ coefficient will be markedly reduced because our cuts are nonoptimal. This effect is, of course, not peculiar to signs and symptoms of disease or taxonicity generally; it is an instance of perhaps the most general of scientific epistemic principles: *specify concepts* (measure, classify, define, group, separate, aggregate) *so as to discern maximum order.* This powerful over-arching guideline applies whether we are researching inorganic chemistry, geology, animal species, personality, or types of mental illness.

Rationale for MAMBAC

The statistical procedure is motivated by a simple idea: the effect of *sorting* on a quantitative property that discriminates two categories. Suppose two groups, a taxon and its complement, i.e., all cases not members of that taxon, differ with respect to a quantitative variable y, so that, although the two distributions may overlap, their means are different, $\bar{y}_t \neq \bar{y}_c$. If cases are sorted perfectly into taxon and complement, the mean difference $\bar{d}_y$ between the two categories will be $\bar{y}_t - \bar{y}_c$. If now we randomly exchange some cases, "contaminating" the pure taxon group by introducing complement cases and conversely, the new $\bar{y}_t^*$ will be pulled down and the new $\bar{y}_c^*$ pulled up, decreasing the group difference $\bar{d}_y$ by some amount (ignoring rare sampling error). The more we scramble the groups, each becoming more "impure" by the presence of wrongly sorted members from the other, the greater will be the attenuation of the observed $\bar{d}_y$.

Suppose there is a second variable x which also discriminates the taxon and complement members, but with overlap, and the variables are uncorrelated within categories (no xy nuisance covariance). Any cut on x (which we shall designate as the "input" variable) which classifies cases (into those above the x-cut and those below it) thereby determines a certain degree of scrambling of taxon and complement members, and hence produces a certain separation on y (the "output" variable). A "better" cut on x will tend to yield a larger observed separation on y. Hence the empirical function $\bar{d}_y(x) = \bar{y}_a(x) - \bar{y}_b(x)$ can be used to locate an optimum x-cut (derivation of the latent formula for MAMBAC is given in Appendix A, pp. 1111-1119). The initial intuition was that maximizing $\bar{d}_y(x)$ would serve as a way of locating the hitmax cut on x (as in the MAXCOV-HITMAX procedure; see

Meehl, 1973a; Meehl & Golden, 1982), but this is not the case. Maximizing $\bar{d}_y(x)$ does not maximize total hits, i.e., minimize misclassifications, but rather it maximizes the sum of hit *rates* above and below the x-cut, that is, it maximizes the latent quantity $h = h_a + h_b$ where

$$h_a = \frac{H_a}{N_a} \text{ and } h_b = \frac{H_b}{N_b}$$

H_a and H_b being the "raw" hits, i.e., frequency of correctly classified cases above and below, respectively. The denominators N_a and N_b are the cases falling above and below the cut rather than correctly classified cases as in the valid positive rates

$$p_t = \frac{H_a}{N_t} \text{ and } p_c = \frac{H_b}{N_c}$$

Although the Monte Carlo tests reported in this article are based on Gaussian distributions, normality is not a required assumption for MAMBAC. The analytical theorems on which MAMBAC is based do *not* postulate normality or equality of variance. The basic theorem $\bar{d}_y(x) = (h_a + h_b - 1)(\bar{y}_t - \bar{y}_c)$ is an algebraic identity. That $h_b \to Q$ as $x \to \infty$ holds for any pair of smooth distributions such that the taxon density function $f_t(x)$ exceeds the complement density function $f_c(x)$ for all values $x > K$ (large enough) and both densities are asymptotic to the baseline (Fisher's "high contact") as $x \to \infty$. For empirical data, absent gross measurement or clerical error, this yields a cut on x above which literally no complement cases fall. The only further assumption is unimodality of the latent curves.

Additional Monte Carlo work is needed to cover non-Gaussian latent distributions and to investigate unequal variances and nuisance correlations more thoroughly than has been done in this article. We encourage other researchers to participate in this large-scale task and also to apply MAMBAC to nonartificial empirical data that depart markedly from the idealizations of our Monte Carlo studies to date.

The Monte Carlo Samples

In order to see how MAMBAC performs for different situations a researcher might encounter, we generated samples for various configurations, e.g., with different base rates, sample sizes, separations of the latent complement and taxon classes, presence of nuisance covariance within the latent classes. We could not, of course, cover every possible situation, but we tried to choose configurations that would illustrate those variations researchers would most expect to encounter. The different configurations will be introduced in later sections as they become relevant.

Because there would be random error associated with any particular sample, we generated 25 separate samples for each of the configurations.

Thus, for a configuration having a particular base rate, a particular separation between the complement and taxon classes, and a particular amount of nuisance covariance, there will be 25 samples. We present data from some of the samples (usually the first 10 samples) in the text; data for all 25 samples are given in appendices for readers who wish to see more examples for each configuration. Each sample has a coded name, consisting of a configuration code with the particular sample number (1 through 25) appended to it. We have tried to present all material in such a way that *it is not necessary to learn the configuration code names*; they merely identify particular Monte Carlo samples for readers who may want that amount of detail.

Each Monte Carlo sample has four continuously distributed variables (because another procedure in the coherent cut kinetics method requires that many). MAMBAC requires only two variables. When there are more variables, MAMBAC can be used on all possible two-variable combinations.

Our Monte Carlo samples were all constructed using normally distributed random deviates. Thus a nontaxonic sample will have an expected mean = 0 and an expected standard deviation = 1.0. A researcher may or may not standardize scores before using MAMBAC on a data set; *it will not affect the interpretation of MAMBAC results.* (A researcher who has more than two variables may want to standardize them if other coherent cut kinetics procedures are going to be used; for instance, there may be advantages for interpreting MAXCOV results.)

Here is a more detailed description of the way we generated our Monte Carlo samples, for readers with a special interest. Construction of a sample began with a unique file of random deviates.[5] To generate four indicators per subject, we used five random numbers per subject; hence we started with a file of $5 \times N$ random deviates for each sample. For a given subject$_i$, the first random deviate in a sequence was multiplied by a factor loading (predetermined by the desired configuration) of each indicator in turn, to make four expected scores for the subject. If the sample was to be a taxonic one with no nuisance covariance, i.e., no indicator correlation within the taxon or within the complement, this factor loading was .001, to approximate zero without crashing the computer program; the factor loadings for each indicator were higher when we wanted to create nuisance covariance within the taxon and nontaxon classes. If the sample was to be a nontaxonic one, factor

[5]This set of normally distributed random numbers was generated by a program written in C by William M. Grove. The algorithm consists of two steps: Generate uniform random numbers, and use these to generate normal random numbers. The uniform generator is a straight multiplicative congruential r.n.g. of modulus $2^{31} - 1$, described by Fishman and Moore (1982) and using their multiplier VIII. This multiplier passes tests of independence and equidistribution on univariate through trivariate spaces. It also passes the lattice test. The critical parts of the program are in 80 x 87 assembly language code for speed and for consistent 64-bit precision (80 bits are carried on intermediate results). The method of generating normals is the "convenient" method of Marsaglia and Bray, which is documented by Kennedy and Gentle (1980). The Monte Carlo samples are available upon request.

loadings were set to create a desired amount of expected correlation between the indicators. The next four random deviates were used to add error components to each of those expected scores; if the subject was to be a taxon member, the assumed amount of taxonic separation was also added to each variable. Then the following sequence of five random deviates from the input file was used to create the four scores for the next subject$_{i+1}$, and so on for the rest of the subjects in the sample. Nontaxonic comparison samples were generated by using factor loadings such as to match the manifest correlations between variables that resulted from different taxonic separations and nuisance covariance in taxonic samples.

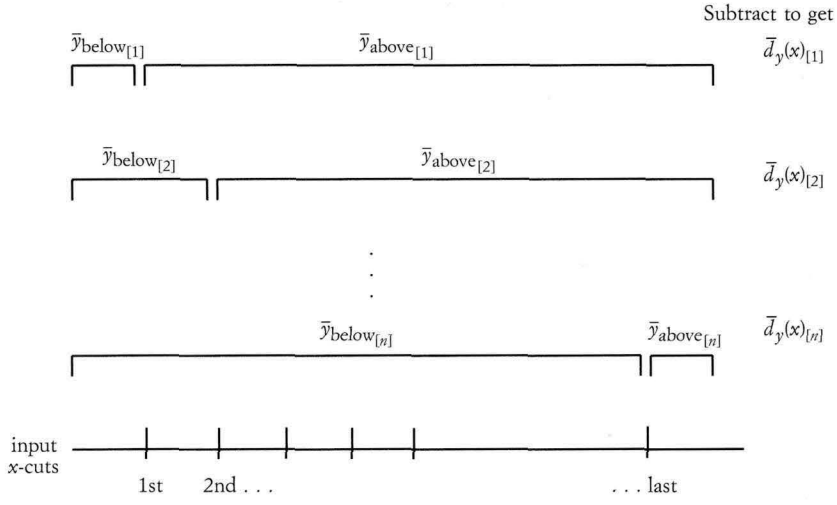
Calculation of MAMBAC Values

The first step is to draw a graph of differences between means calculated from output variable values for cases located above and below successive cuts along a continuous input variable. Restating this: Successive cuts are defined along the input variable. At each cut we calculate a mean on the output variable for the cases located above the input cut, and another mean for the cases below the input cut; we subtract (mean above minus mean below) and the resulting difference is one point on our MAMBAC graph. The input variable is used to determine whether a subject falls above or below the cut; the subject's output scores are used to calculate the mean differences. The number of points on the MAMBAC curve will be equal to the number of cuts we have defined on the input variable.

Although the procedure is quite simple, it is an unfamiliar way of manipulating data, so we will describe the algorithm once more, this time with a diagram. Let us suppose we have N subjects, and each subject has a score on x and a score on y. To help in explication, let us say that we sort the subjects from low to high according to their x-scores (for a computer program, an explicit sorting is not necessary; the program need only ask whether a subject's x-score is above or below a particular cut). We choose an initial cut near the bottom of the x-distribution. For the subjects falling below this cut, we calculate the mean of their y-scores, and we do the same for the subjects falling above the cut. Subtracting the y-mean for those cases below the x-cut from the y-mean for subjects above the cut (Mean Above Minus Below A Cut), we get a difference $\bar{d}_y(x)_{[1]}$. Then we move the cut to a second point on the x-distribution and get a second mean difference, $\bar{d}_y(x)_{[2]}$. We repeat this procedure for successive cuts, getting a $\bar{d}_y(x)$ value each time. When we plot these $\bar{d}_y(x)$ values, the curve shape will tell us whether the underlying structure is taxonic or not.

When both variables are continuously distributed, we reverse the roles of input and output and get a second $\bar{d}_x(y)$ MAMBAC curve, with y as the input variable and x as output. Additional variables are used in all possible combinations. In our Monte Carlo tests, we have used four continuous vari-

ables (x, y, z, v), which give $\binom{4}{2} \times 2 = 12$ possible input/output combinations, resulting in 12 MAMBAC graphs for a given sample.



The cuts along the input variable can be made on the basis of either intervals on the abscissa, e.g., .25 standard deviation intervals along the distribution of the input variable, or number of cases, e.g., deciles on the sorted input distribution. Whichever of those ways we choose, two things are relevant in deciding how coarse to make the cuts. Because the *shape of the graph* of the mean differences is what we are interested in, there should be enough cuts, i.e., enough points on the MAMBAC graph, to make the shape of the MAMBAC curve clear. Going solely by this criterion, one would think the more points (cuts), the better. But, of course, we want to include enough cases (from the output distribution of scores) per cut (on the input distribution of scores) to keep the progression of the curve fairly stable at either end.

Most of the Monte Carlo runs included here used intervals on the abscissa. Steps at -2.50 *SD*, -2.25 *SD*, . . . $+2.50$ *SD* were defined around the observed mean of the input distribution. We checked both ends of the input distribution and accumulated cases (moving forward from the low end, backward from the high end) until there were at least 15 cases below (or above) the defined cut. This located the beginning and ending cut locations for a single MAMBAC graph. Fifteen was chosen (more or less arbitrarily, based on observations from initial Monte Carlo runs) as a minimal n for obtaining fairly orderly results. Because the method uses all the cases below

and above each successive cut, this guaranteed that any cut would have at least 15 cases above and below it. The result is usually about a dozen cuts along each input variable for samples with $N = 300$.

Pseudocode for calculating MAMBAC using four continuously distributed variables and cutting on the basis of abscissa intervals and of deciles is shown in Fig. 1. This was extracted from the Modula-2 (LOGITECH, Inc., Version 3.0) program used for the Monte Carlo runs reported here.

Pseudocode from MAMBAC program used in Monte Carlo runs

```
combinations of variables: xy, xz, xv, yz, yv, zv
FOR each combination of variables DO
  Begin with the first variable as input, second variable as output;
    then repeat the procedure using the second as input, first as output

Get d̄output(input) at successive cuts of .25 SD from the mean of the input variable
  Determine possible cuts around the mean of the input variable: -2.50 SD, -2.25 SD,
    -2.00 SD . . . 2.50 SD
  Check each end of the input distribution and move inward until there are at least 15
    cases below (or above) a predetermined cut; this defines the first and last cut at which
    d̄ will be calculated

FOR each cut c DO
  FOR all cases i in the sample DO
    IF the input score[i] < the cut THEN
      sumBelow : = sumBelow + output score[i];
      nBelow : = nBelow + 1;
    ELSIF the input score[i] ≥ the cut THEN
      sumAbove : = sumAbove + output score[i];
      nAbove : = nAbove + 1;
    END (* if *);
  END (* for i *);
  outMeanBelow[c] : = sumBelow / nBelow[c];
  outMeanAbove[c] : = sumAbove / nAbove[c];
  MAMBAC[c] : = outMeanAbove[c] - outMeanBelow[c];
END (* for each cut *)

Plot obtained MAMBAC values over cuts (Smooth curve if necessary)

Get d̄output(input) by decile cuts on input
  Sort scores on input variable, keeping output variable scores properly associated with each
  Determine the number of cases in each decile (depends on N)
  Proceed as above using cuts at successive deciles on the input variable

END (* for each combination of variables *)

(* It will be helpful later to have saved the MAMBAC[c], nBelow[c], and nAbove[c] values
   at each cut in an output file that can be read for making parameter estimates *)
```

FIG. 1. Pseudocode for MAMBAC

With more sophisticated graphical analysis software which has become available since we began our Monte Carlo work, it may be possible to slide the cut by any arbitrary amount, even one case at a time, and then impose a smoothing procedure on the final curve. An example of such code, including the resultant MAMBAC curves, is given in Fig. 2. The code shown there is *S-Plus* (Statistical Sciences, 1992).

Example of MAMBAC code and resultant curves using the *S-Plus* language

```
# start with a data vector "xy" composed of variable pairs x and y arranged in two columns

# using x as input, y as output to get d̄y(x), sort on x, assumed to be in the first column
#   of "xy"
    sortedx <- xy[sort.list(xy[,1]),]
    N <- length(sortedx[,1])                    # define N as the number of pairs,
                                                #   600 in this sample
    MAMBACxy <- vector("numeric",N)     # define a vector to hold the MAMBAC values

# at successive cases on x, get the difference between ȳ above and ȳ below
# ignore 15 cases at either end for improved stability, move the cut by one case each time
    for (i in seq(15,585,1) ) {
        MAMBACxy[i] <- mean(sortedx[i:N,2]) - mean(sortedx[1:i,2]) }

# create a window, plot the MAMBAC values, and overlay the points with a lowess
#   smoothed curve
    win.graph( )
    plot(MAMBACxy[15:585])
    lines(lowess(MAMBACxy[15:585]) )
```

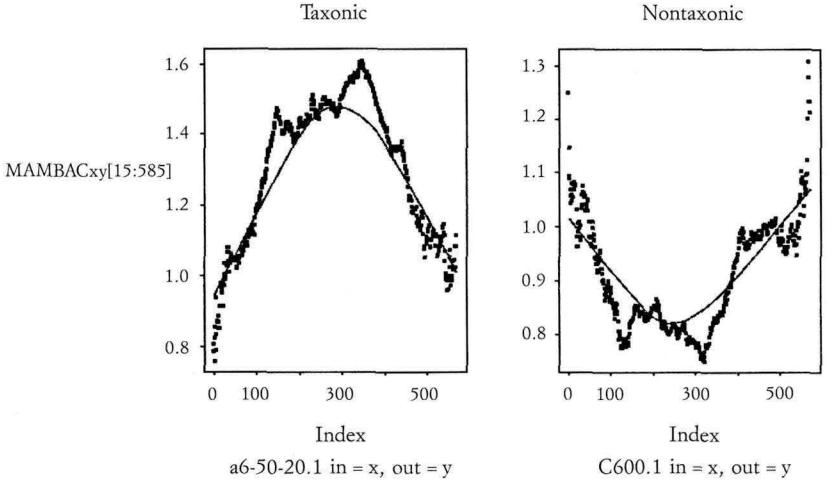


FIG. 2. *S-Plus* code for MAMBAC

While a computer program is handy for large samples or for doing MAMBAC repeatedly, it is not required. A researcher with only a single, not too large data set can easily do MAMBAC with paper and pencil. An optimal way to go about it would be to record each subject's scores for x and y on a 3- × 5-in. slip and sort the slips in ascending order for x. Set up columns on a tabulation sheet: Successive cuts on x, sum of y-scores above cut, n cases (= slips) above cut, $\bar{y}$ above (= sum of y-scores above/n above), sum of y-scores below cut, n cases below cut, $\bar{y}$ below (= sum of y-scores below/n below), $\bar{d}_y(x)$ (= $\bar{y}$ above − $\bar{y}$ below). Add, divide, and subtract to get the values at each cut, then plot the $\bar{d}_y(x)$ values on graph paper. To get a second MAMBAC curve, $\bar{d}_x(y)$ for the data set, sort the slips according to scores on y and repeat the process.

Smoothing may make the MAMBAC curve more aesthetic and easier to interpret at a glance. We have experimented with different curve-smoothing techniques—weighted means, polynomial fit, lowess (Cleveland, 1979), and repeated medians (described by Tukey, 1977, pp. 212-213). It does not seem to matter much which technique is used; under some circumstances, e.g., base rate $P = .50$, good indicator validity, no nuisance covariance, $N \geq 300$, and cuts made on the basis of deciles, smoothing may not be needed at all. With base rates less than .50, some smoothing techniques show the underlying curve shape better (see below). All Monte Carlo curves presented here were drawn using *S-Plus* (Statistical Sciences, 1992); data points are plotted and superimposed with smoothed curves obtained by Tukey's procedure as implemented by *S-Plus*.

Detection of Taxonicity with MAMBAC

If the underlying structure is taxonic, MAMBAC graphs tend to be convex upward, with the location of the peak depending on the latent base rate and the taxonic separation. Theoretically, the peak will be near the hitmax cut, i.e., that cut on the input indicator which minimizes misclassifications. As stated above, what is literally maximized at the peak value, however, is *not total hits* but the sum of *hit rates* above and below the cut, $h = h_a + h_b$ (see proof in Appendix A, pp. 1111-1119). A nontaxonic latent structure results in graphs with a fairly symmetrical dish shape (concave upward), lower in the middle range and higher on the ends. Fig. 3 shows the ideal MAMBAC curves (based on error-free normal-curve values, not Monte Carlo data) for a good taxonic situation and for the nontaxonic situation.

MAMBAC curves from Monte Carlo samples are shown in Fig. 4. The panels in the top row are from 10 taxonic samples with $N = 600$, a base rate $P = .50$, and 2 *SD* separation on each of four variables; cuts on the input variable were made on the basis of .25 *SD* units. Each panel represents the 12 curves (offset for visual display) generated by four variables for a single sample; the input/output ordering of the variables is shown on the left. Su-

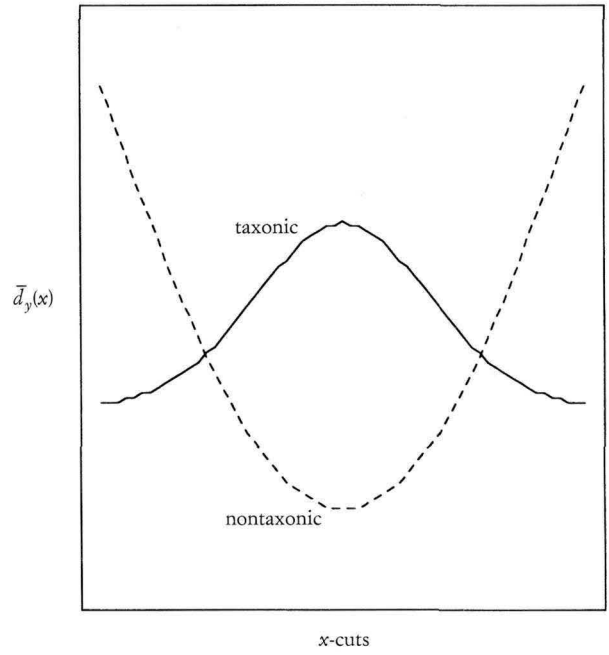


FIG. 3. MAMBAC error-free curve shape for $P = .50$, 2σ separation on each variable, and no nuisance covariance. Solid line is the taxonic situation; dashed line is the nontaxonic situation when $r_{xy} = .50$.

perimposed on the points are smoothed curves using Tukey's 4(3RSR)2H twice method (Tukey, 1977, Chapters 7 and 16). MAMBAC curves from nontaxonic samples are shown in the lower row of panels. Because taxonic separation generates correlation between the variables, we imposed factor loadings on the variables to generate a comparable correlation when we created the nontaxonic samples. In the taxonic samples shown in Fig. 4 the expected $r_{ij} = .50$ ($P = .50$, 2 *SD* separation on each variable, and no nuisance covariance; see Appendix B, pp. 1120-1121). In the nontaxonic comparison samples, factor loadings of .707 on each variable generate the same expected $r_{ij} = .50$.

For exposition, we have shown curves from only 10 samples of each configuration (taxonic and nontaxonic) in Fig. 4. They are unselected, i.e., they are merely the first 10 samples from each of these two Monte Carlo configurations. To see the difference between curves generated by a taxonic versus a nontaxonic configuration, any panel in the top row could be compared with any panel in the bottom row. Curves from all 25 samples that have been generated for each configuration may be found in Appendix C (pp. 1122-1151; the first 10 panels there will be identical to those in Fig. 4).

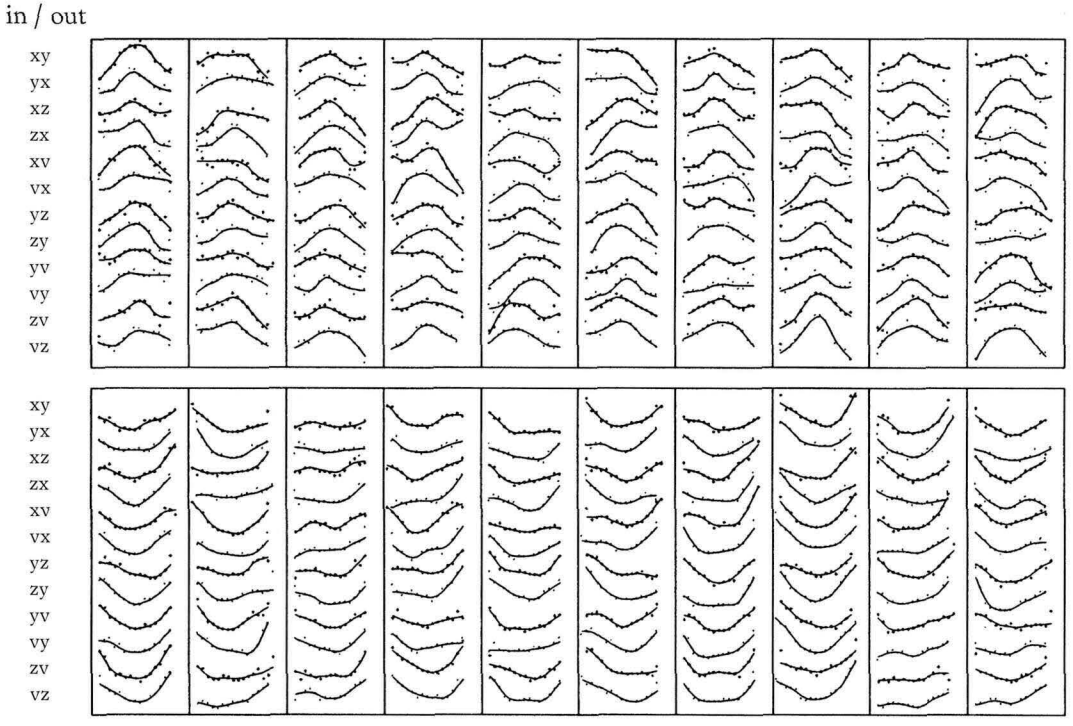


FIG. 4. Detection of taxonicity with MAMBAC cutting at .25 *SD* intervals on the input variable. Top panels are from taxonic samples: $N = 600$, $P = .50$, 2 *SD* separation on each variable, no nuisance covariance, $r_{ij} = .50$ (because of complement-taxon mixture). Bottom panels are from non-taxonic samples: $N = 600$, $r_{ij} = .50$ (from factor loadings of .707 on each variable). Data from these samples are also shown in Fig. 5.

Clearly MAMBAC detects taxonicity for the samples shown here. Curves from taxonic samples tend to be arched upward in the middle. The nontaxonic curves tend to be symmetrically dish-shaped. In fact, the curve shapes from taxonic and nontaxonic samples are almost mirror images of each other for this taxonic configuration. Although not all curves within every sample show the expected shape, at least one curve (usually more) in each sample clearly indicates either a taxonic or nontaxonic shape. With this amount of correlation between the variables, there are no curves in our nontaxonic Monte Carlo samples that would mislead a researcher into thinking the sample is taxonic when it is not (see Appendix C, configuration C600, p. 1131).[6]

Within the taxonic samples, we sometimes see individual curves that are not clearly taxonic. This points up the value of *multiple continuously distributed variables*: Often when one input/output combination of variables does not give a clearly taxonic curve, reversing their input/output status does. For example, in sample 23 of this configuration (Appendix C, A6-50-20, p. 1130), the curve for variables y (as input) and v (as output) is clearly taxonic, while the v/y input/output combination gives a curve that is not clearly anything. A researcher with only these two variables could rightly conclude that the sample was taxonic. But this advantage with curve pairs is not guaranteed. In sample 18 (Appendix C, A6-50-20, p. 1130), values obtained using x and z in both input/output combinations yield curves neither of which looks taxonic when smoothed. Thus, with only two variables, it is possible to miss detecting a taxonic situation.

Because we do not yet have an algorithm for testing whether a MAMBAC curve is taxonic or not, we asked people to sort graphs by inspection. For this task we mixed nontaxonic curves and taxonic curves from samples with base rates of .50, .25, and .10 (which yield different curve shapes; see below). Twelve people with no special training in psychology or other areas relevant to the task sorted 150 panels (12 curves per panel) with 98% over-all accuracy. Four psychologists (one of them the first author) and a

[6] Under what *latent* conditions can a nontaxonic quantitative factor generate a MAMBAC curve that looks clearly taxonic? Assuming the classic psychometric factor model, the following necessary and sufficient conditions hold for a test score defined as summed dichotomous (1 or 0) items: Steep item-characteristic ogives closely located *iff* [if and only if] high interitem ϕ-coefficients *iff* a U-shaped frequency distribution of the input variable *iff* ipso-MAMBAC (our term to designate the graph of $\bar{d}_x(x)$, the MAMBAC function computed on the values of the input variable itself, rather than for y upon x) hump on the input variable *iff* a MAMBAC hump on the output variable. The parametric questions "how steep" and "how closely located" await Monte Carlo investigation. But, it is clear that *very* high ϕ-coefficients are required to countervail the powerful Central Limit Theorem; and if the 4-fold table has disparate marginals the ϕ-coefficients are severely constrained, e.g., difficulty levels of .90 versus .50 for an item pair impose an upper bound $\phi_{ij} \leq .30$ even if each item were a step-function of the latent factor score. As an empirical fact, such item properties are almost never found even when one works hard to achieve them. To our knowledge, the only test domain that presents U-shaped frequency distributions is trade tests; but these results are not spurious as a skilled trade is a genuine taxon of environmental mould origin.

FIG. 5. Detection of taxonicity with MAMBAC *cutting at decile intervals* on the input variable. Top panels are from taxonic samples: $N = 600$, $P = .50$, 2 *SD* separation on each variable, no nuisance covariance, $r_{ij} = .50$ (because of complement-taxon mixture). Bottom panels are from non-taxonic samples: $N = 600$, $r_{ij} = .50$ (from factor loadings of .707 on each variable). These are the same samples as in Fig. 4.

graduate student in clinical psychology sorted them with accuracy rates of 99.7% for taxonic samples and 99.4% for nontaxonic samples. To test how well a researcher could sort curves if only two variables were available (hence only two curves from a sample instead of 12), three of the psychologists who had sorted the panels later sorted 900 pairs of curves. All three had accuracy rates of 99%. Details of these sortings are given in Appendix D (pp. 1152-1156).

Effect of Different Methods of Cutting on the Input Distribution

The variability in the curves is reduced when cuts on the input variable are made according to deciles. When MAMBAC curves are plotted both ways on the same axes, the decile curves will look dampened compared to those generated on the basis of abscissa cuts. This is because there are fewer points to be plotted with the decile curves and they are stretched along the *x*-axis (with very large samples, one might use finer cuts than deciles and thereby increase the number of points). When axes are adjusted, decile curves also indicate taxonicity or nontaxonicity and may present a "cleaner" looking and somewhat more orderly graph, such that smoothing may seem superfluous with a large sample. Otherwise, for detection of taxonicity there seems to be no particular advantage or disadvantage to either way of making the cuts. Fig. 5 shows the same samples as in Fig. 4 but with deciles used as cutting points on the input variable. Notice the tendency for the same curves, i.e., same sample and same input/output combination, to look more or less peaked (or dish-shaped in the nontaxonic samples) and to have generally the same orientation vis à vis other curves within the sample whether abscissa units or deciles are used as the basis for the MAMBAC cuts. To avoid needless duplication, we will show only curves based on abscissa cuts in the rest of this article.

Effect of Sample Size

MAMBAC can detect taxonicity with smaller samples, but large samples are highly recommended, especially if the base rate is suspected to be less than .50. When the base rate is low, there simply may not be enough taxon members at the high end of the input distribution to generate a taxonic MAMBAC curve. The stability of the curves increases with larger sample sizes, and there is less chance to be misled. With smaller samples the number of intervals, hence the number of points for plotting, is reduced when cuts are based on the abscissa. Taxonic curves from different sample sizes are shown in Fig. 6.

Fig. 7 shows nontaxonic curves from samples of different sizes. With small samples, it is particularly important to have more than two continuous variables so that more MAMBAC curves can be plotted.

Effect of Base Rate

With base rates less than .50, the peak of the MAMBAC curves moves

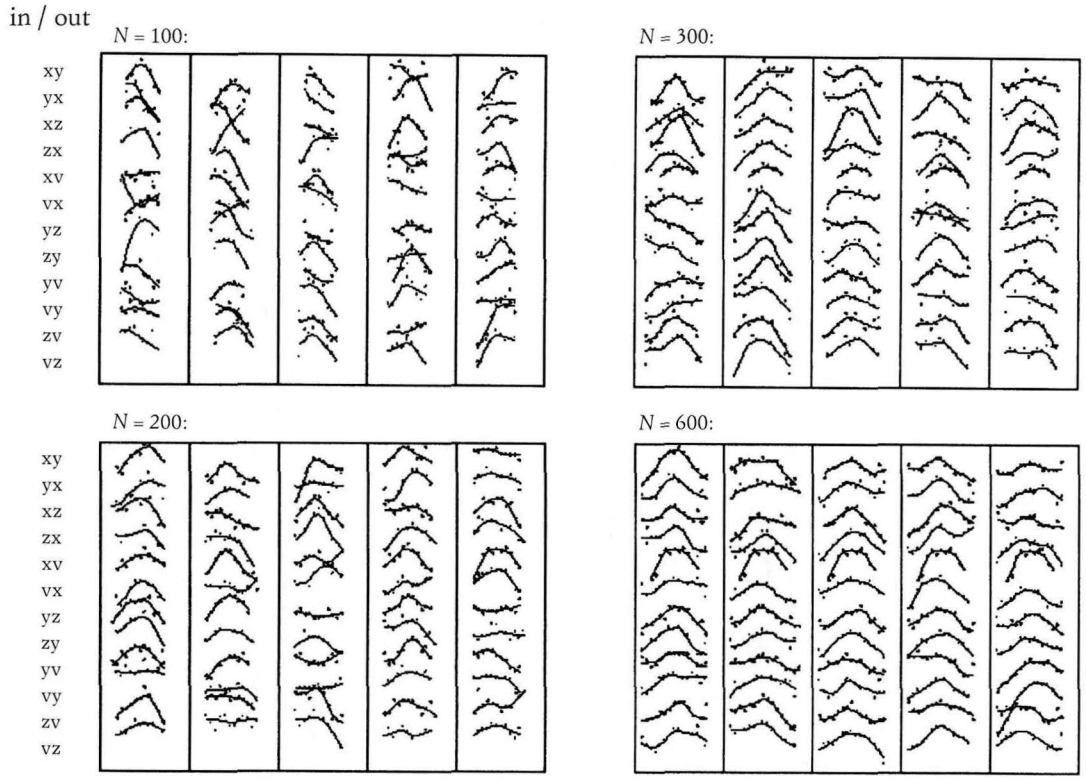


FIG. 6. Effect of sample size, $N = 100, 200, 300, 600$. MAMBAC cuts at .25 *SD* intervals on the input variable. All samples shown here are taxonic: $P = .50$, 2 *SD* separation on each variable, no nuisance covariance, $r_{ij} = .50$ (because of complement-taxon mixture). (Curves from all 25 Monte Carlo samples for each sample size may be found in Appendix C, pp. 1122-1151.)

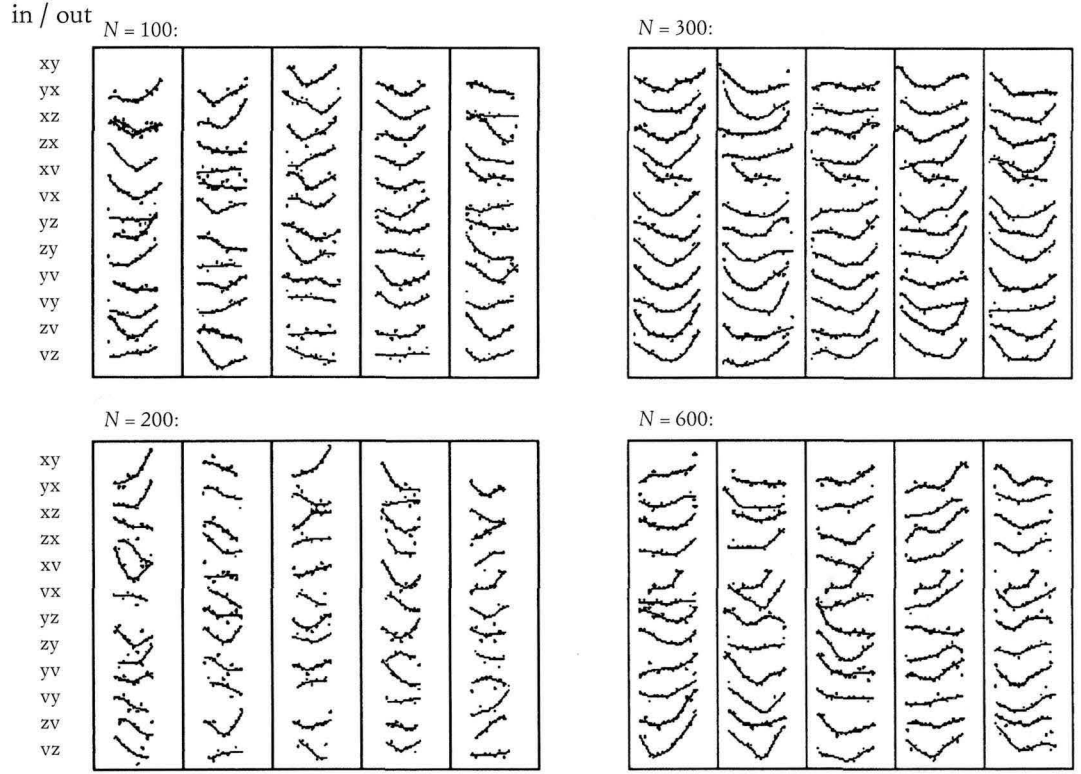


FIG. 7. Effect of sample size, $N = 100, 200, 300, 600$. MAMBAC cuts at .25 *SD* intervals on the input variable. Samples shown here are non-taxonic: expected $r_{ij} = .50$ (factor loadings = .707 on each variable).

to the right. Fig. 8 shows the effect of different base rates on error-free MAMBAC curves with 2 *SD* separation and no nuisance covariance for either variable. The dish-shaped curve is the nontaxonic situation with r_{xy} = .50. Larger samples become more important with lower base rates, but curves from samples of $N = 300$ look different when $P = .50$, .25, or .10, other factors being favorable, e.g., multiple variables with 2 *SD* separation on each, no nuisance correlation. For base rates greater than .50, the MAMBAC peak shifts to the left of center. Thus for $P = .75$, MAMBAC curves would look like curves for $P = .25$ but with the peak on the left. Fig. 9 shows MAMBAC curves for Monte Carlo samples with different base rates. The taxonic panels are stacked vertically to make it easier to compare the effect of different base rates. The nontaxonic comparison curves are presented on the right. Of course, a difference in base rate produces a difference in the expected

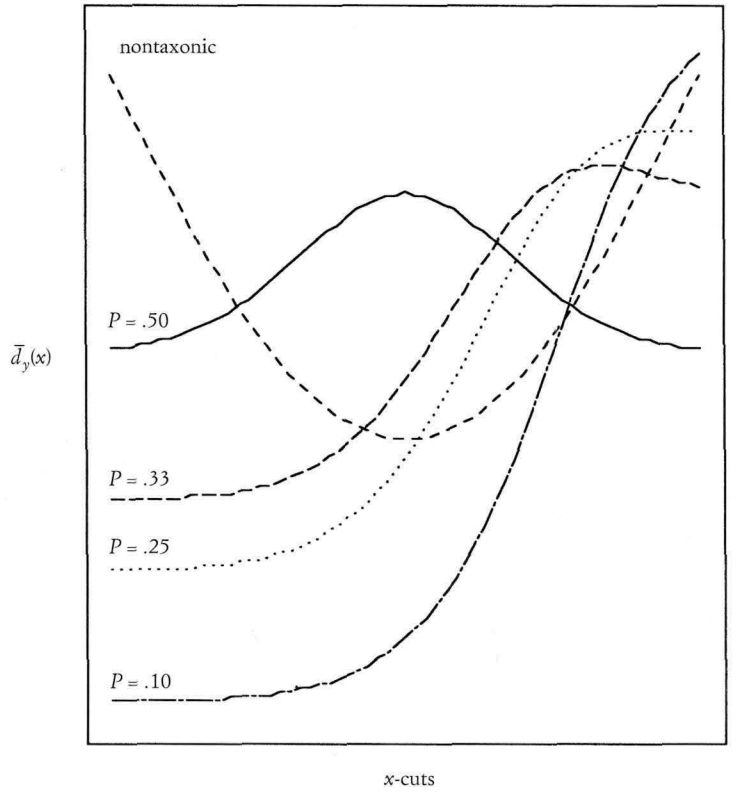


FIG. 8. MAMBAC error-free curve shapes for different base rates (no nuisance covariance, and 2σ separation on each variable). The dish-shaped dashed line is the nontaxonic situation with $r_{xy} = .50$.

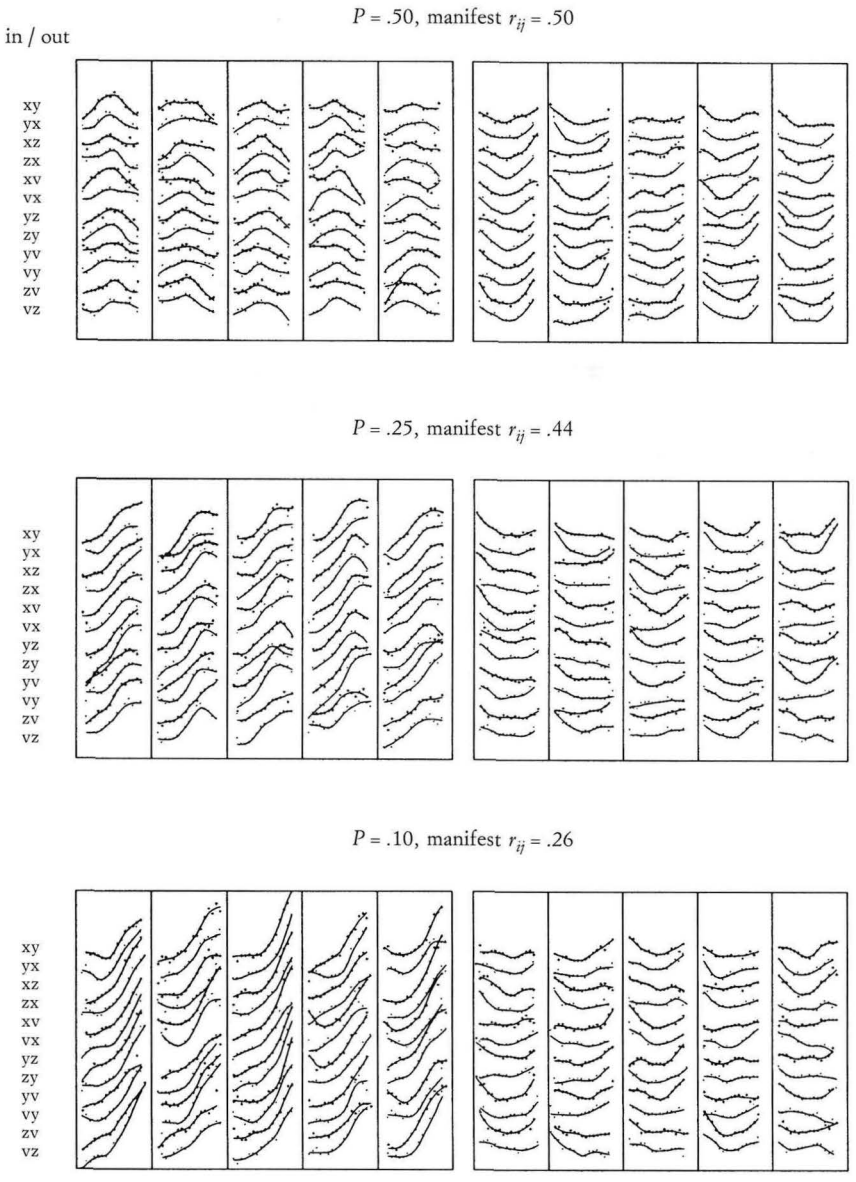


FIG. 9. Effect of base rate. MAMBAC cuts at .25 *SD* intervals on the input variable. Samples on the left are taxonic: $N = 600$, 2 *SD* separation on each variable, no nuisance covariance. Samples on the right are nontaxonic with expected r_{ij} matching that of the taxonic samples in each row.

correlation between any two variables (see Appendix B, pp. 1120-1121). Adjustments were made via factor loadings so nontaxonic comparison samples would have comparable expected r_{ij} values. Notice that the differences in the expected r_{ij} values shown here have no obvious effect on the nontaxonic curves.

Effect of Taxon Validity

The effect of indicator separation on an error-free MAMBAC curve with $P = .50$ and no nuisance covariance for either variable may be seen in Fig. 10. These taxonic curves have been centered to make it easier to see the attenuation of the taxonic curve peak with lower validity; if they were plotted on a fixed x-axis, their peaks would move to the right with increased separation. Taxonic curves become more peaked with increased separation and the location of the peak shifts with the mean of the taxonic distribution, e.g., with a separation of 3 *SD* the peak is at 1.50. Thus the peak of

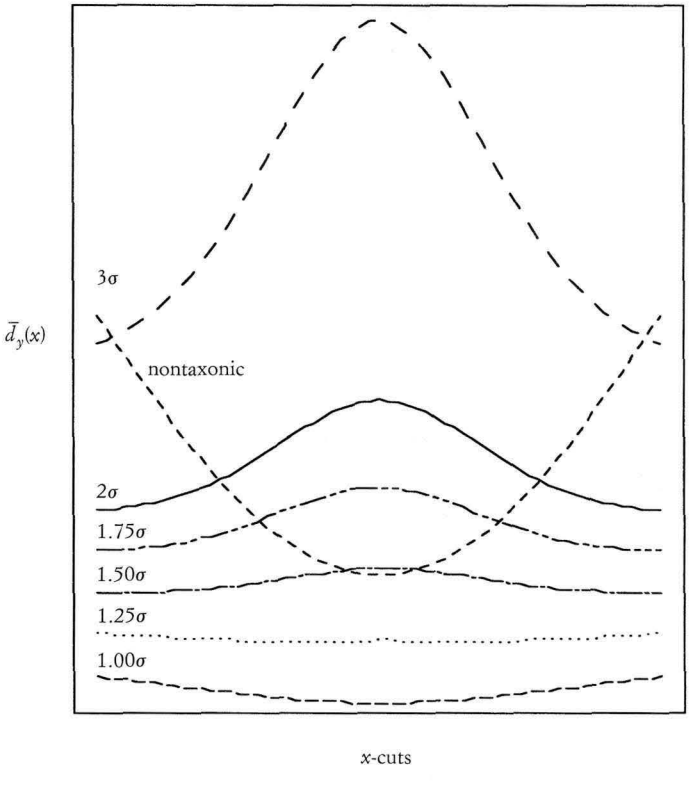


FIG. 10. MAMBAC error-free curves for different amounts of separation (validity). The dish-shaped dashed line is for a nontaxonic situation with $r_{ij} = .50$.

the taxonic curve is influenced by indicator validity as well as by the base rate.

Fig. 11 shows Monte Carlo graphs when taxonic separation is reduced to 1.5 *SD* on each of the four variables ($N = 600$, $P = .50$, no nuisance correlation, expected $r_{ij} = .36$). (See also Fig. 14, which shows samples with various validities for the four variables in addition to nuisance covariance.) The nontaxonic curves with comparable correlation are clearly nontaxonic, but curves in taxonic samples may also look nontaxonic, e.g., the second curve in the fifth taxonic panel. Obviously, having more variables is helpful when validities are reduced, and all other conditions will have to be optimal if latent taxonicity is to be detected with low separations.

We recommend choosing variables with more validity rather than relying on the other factors being optimal; however, we plan to investigate the use of consistency tests to help in cases of low validity. For instance, it may be possible to use estimates derived from the data to predict the correlation due to taxonicity (when there is no nuisance covariance) and compare that with the observed correlation. It may be that this will help distinguish the taxonic from nontaxonic situation (because the estimates used to predict r_{ij} will be invalid if the underlying structure is nontaxonic). Additionally, we hope eventually to rigorize the decision process, perhaps by fitting a polynomial, e.g., a quartic, to the data and setting up criteria of taxonicity based on the polynomial's coefficients.

Effect of Nuisance Covariance

Nuisance covariance is correlation between the variables within the complement, the taxon, or both groups. Ideally, we want variables that are uncorrelated within each subgroup (although, of course, they will be correlated for the total, combined group). Increasing nuisance covariance flattens the MAMBAC curve progressively. Fig. 12 shows the effect of nuisance covariance on error-free MAMBAC curves with $P = .50$ and 2 *SD* separation on both variables.

Monte Carlo samples were generated with factor loadings added to variables in the taxon and complement groups to produce nuisance covariance. MAMBAC curves from those samples are shown in Fig. 13. Nuisance covariance (in the amount used here) is often detrimental to the taxonic MAMBAC curves. Here again, it is very helpful to have more than two variables so that more curves can be obtained. On the other hand, the nontaxonic samples with comparable expected r_{ij} values still produce clearly nontaxonic MAMBAC curves; the investigator should not be misled into thinking the underlying situation is taxonic when it is dimensional.

When several indicators are available, it may be feasible to estimate nuisance covariance "directly" by identifying high-confidence cases, e.g., employing three indicators x, y, z dichotomously, finding the + + + and

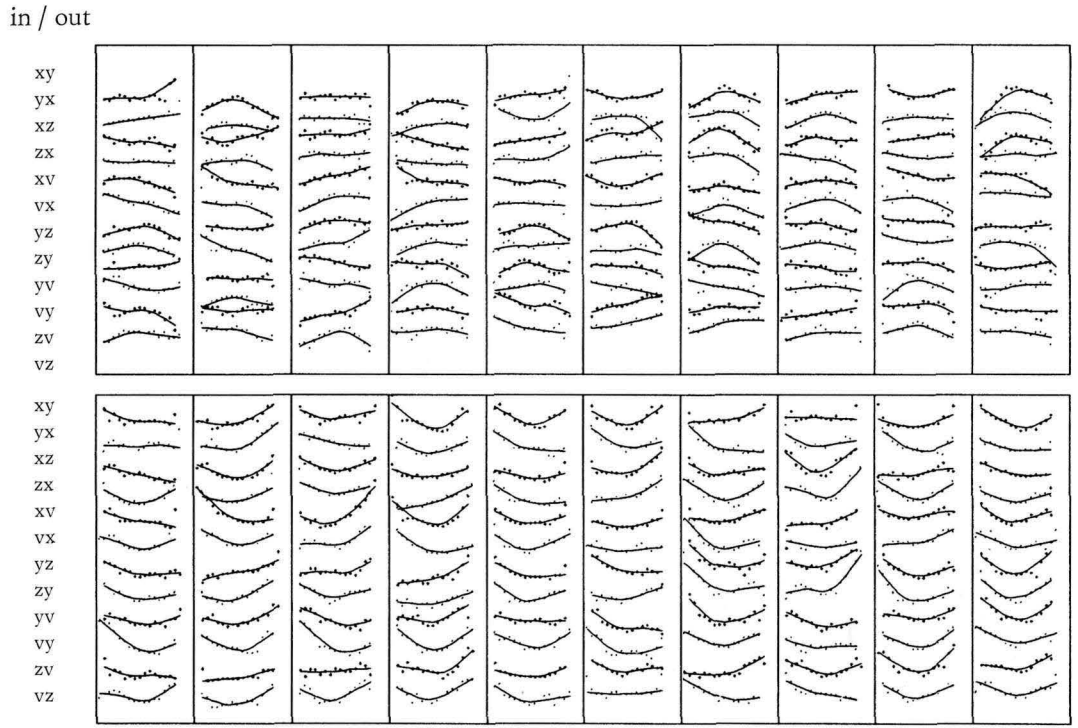


FIG. 11. Effect of reduced indicator validity on MAMBAC curves. Cuts are at .25 *SD* intervals on the input variable. Top panels are from taxonic samples: $N = 600$, $P = .50$, 1.5 *SD* separation on each variable, no nuisance covariance, $r_{ij} = .36$ (because of complement-taxon mixture). Bottom panels are from nontaxonic samples: $N = 600$, $r_{ij} = .36$ (from factor loadings of .60 on each variable).

– – – patterns, and then correlating other indicators, e.g., u, v, within those groups. Generalizing MAMBAC for situations of nonnegligible nuisance covariance is planned for presentation in a subsequent publication. Psychologists may be reassured by noting how hard we usually work in the "soft" areas, e.g., psychopathology, in an effort to *raise* correlations to levels above those for which the MAMBAC procedure is fairly robust under departures from the idealization. When candidate indicators are carefully selected

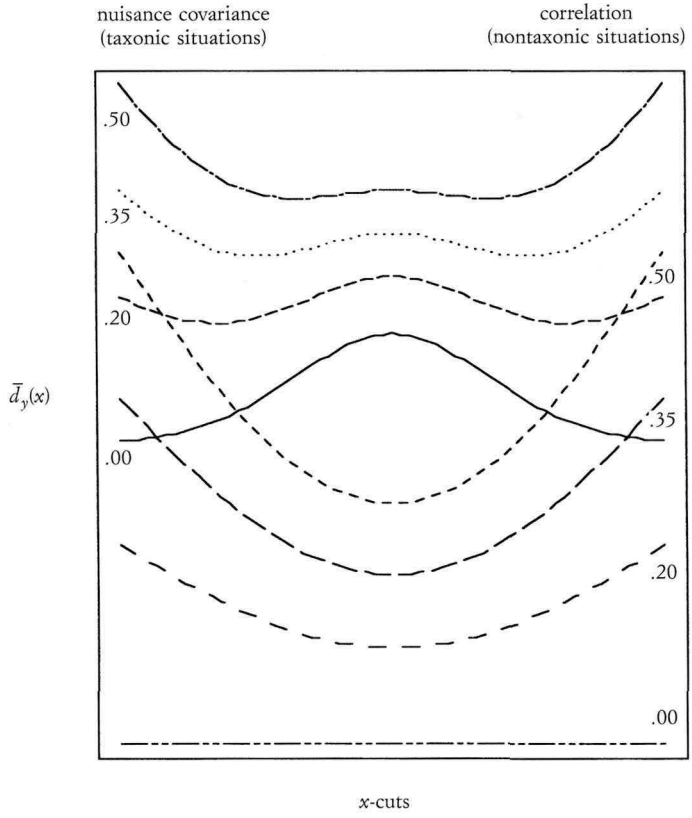


FIG. 12. MAMBAC error-free curves for various amounts of nuisance covariance (labeled on the left). The dish-shaped curves for nontaxonic situations with comparable amounts of correction between the variables are labeled on the right.

for qualitative diversity, e.g., an MMPI schizotypal score, a SADS interview, the SPEM eye-tracking anomaly, dysdiadochokinesia—a diversity strongly desirable on *theoretical* grounds, apart from the auxiliary statistical conjecture of MAMBAC—there is no reason why they should be appreciably correlated within the complement class ("normals") or markedly within the taxon (al-

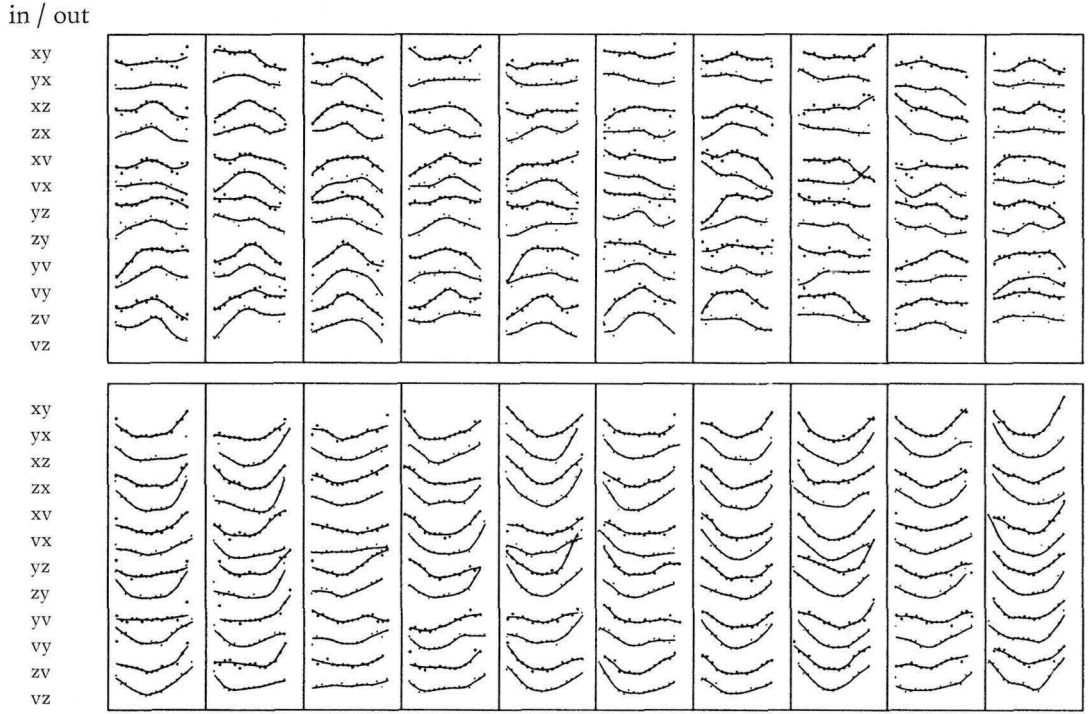


FIG. 13. Effect of nuisance covariance on MAMBAC curves. Cuts are at .25 *SD* intervals on the input variable. Top panels are from taxonic samples: $N = 600$, $P = .50$, 2.0 *SD* separation on each variable. Factor loadings added to the taxonic samples: $x = .70$, $y = .50$, $z = .40$, $v = .20$. Expected correlations (due to the complement-taxon mixture and the factor loadings on the variables): $r_{xy} = .68$, $r_{xz} = .64$, $r_{xv} = .57$, $r_{yz} = .60$, $r_{yv} = .55$, $r_{zv} = .54$. Bottom panels are from nontaxonic samples: $N = 600$, expected r_{ij} values equal to those in the taxonic samples by imposing factor loadings on: $x = .84$, $y = .79$, $z = .77$, and $v = .69$.

though, of course, they will be correlated within the mixed distribution). Even different aspects of "the same" basic phenomenon may be only weakly correlated, e.g., intrusive saccades and low pursuit on SPEM correlate $r \leq .20$ in schizophrenia. While nuisance covariance is often troublesome when the measures are psychometric, e.g., two MMPI scores, here the investigator has item-analytic procedures available to reduce it, sacrificing some scale length (and hence some validity) to hold down nuisance correlation. It should usually be possible to set reasonable bounds on within-category correlations by correlating within crude fallible diagnostic groups, e.g., diagnosed schizophrenes in remission, compensated MZ twins of schizophrenes, "normals" with no family history or even borderline MMPI profile.[7]

Fairly direct estimates of nuisance correlation are usually available via conventional nontaxometric methods despite the lack of a Gold Standard Criterion. *Example*: Suppose the conjectured latent taxon is schizotaxia, a subtle neurological disorder predisposing to schizophrenia but leading to the florid diagnosable disease picture in only a minority of cases (Meehl, 1962, 1989, 1990c, 1990d). We compute pairwise correlations for candidate indicators, e.g., MMPI schizotypy score, a good SPEM anomaly measure, interview rating, P50 evoked potential, dysdiadochokinesia, among high-certainty schizophrenes in remission and among their (clinically well) MZ twins. Controls are chosen for clear family history, no mental disorder (any diagnosis!), and no MMPI elevations. We look for negligible correlations between the candidate indicators in all these groups.

Suppose such careful inclusion/exclusion screening is unfeasible. Setting safe upper bounds on the taxon rate among crudely defined "normals" and on the false positive rate among "presumed schizotypes," we enter the table (Appendix B, pp. 1120-1121) to estimate how much correlation can be nonnuisance-generated by taxonic mixture. The point about such rough estimates is that they can provide a near guarantee that the nuisance correlation—*not* precisely estimated—is small enough to rely on robustness under departures from the idealization. "Small enough is good enough" for our purposes. We rely on philosopher Herbert Feigl's advice to social scientists, "Don't cut butter with a razor."

Combined Effects of Reduced Validities and Nuisance Covariance

The combined effects of nuisance covariance and various validities (all 2 *SD* separations or less) for four variables may be seen in Fig. 14. The taxonic samples used in these tables were generated to provide a deliberately difficult test for detecting taxonicity. Again, the value of multiple variables and large

[7] *Caveat*: For this context of discovery purpose, it is *not* safe to treat "miscellaneous nonschizotypal psychiatric patients" as the complement group, because there are good theoretical and empirical grounds for expecting that a nonnegligible fraction of these patients are unrecognized schizotypes (cf. Meehl, 1973a).

FIG. 14. Combined effects of nuisance covariance and various validities of 2 *SD* separation or less on MAMBAC curves. Cuts are at .25 *SD* intervals on the input variable. Top panels are from taxonic samples: $N = 600$ and $P = .50$. Factor loadings added to the taxonic samples: $x = .70$, $y = .50$, $z = .40$, $v = .20$; separations: $x = 2.00$, $y = 1.75$, $z = 1.50$, $v = 1.25$. Expected correlations (due to the complement-taxon mixture and the factor loadings on the variables): $r_{xy} = .65$, $r_{xz} = .58$, $r_{xv} = .46$, $r_{yz} = .52$, $r_{yv} = .41$, $r_{zv} = .37$. Bottom panels are from nontaxonic samples: $N = 600$, expected r_{ij} values equal to those in the taxonic samples by imposing factor loadings on: $x = .85$, $y = .76$, $z = .68$, and $v = .54$.

samples is obvious. Curves from the taxonic samples are not always clearly taxonic, and some individual curves look clearly nontaxonic; but with four continuous variables, $N = 600$, and $P = .50$, none of the nontaxonic samples we have examined (see also Appendix C, pp. 1122-1151) should mislead an investigator.

Dichotomous Output Variables

Some very preliminary tests (looking at only one output curve per sample) have been done using dichotomized output variables. While MAMBAC usually produced peaked curves for the taxonic situations and dish-shaped curves for nontaxonic samples, there were a few misleading curves for each latent situation. We have not yet done enough testing to recommend confidently this procedure with dichotomous variables.

Estimating the Base Rate with MAMBAC

Theoretically, as the cut moves upward, demarcating the extreme high tail of the input distribution, the hit rate above, i.e., h_a, the proportion of cases that we label as taxon members who truly are taxon members, approaches 1. With large samples of real data, it will actually equal 1, since there is a cut above which no complement cases lie. (If, in rare instances, one or two outliers from the complement group should happen to fall above the taxon group, no harm will be done.) Asymptotically, the proportion of hits below the cut, i.e., h_b, the proportion of cases that we label as complement members who truly are complement members, approaches the proportion of the complement class, Q, in the entire group. Thus, for a cut at the high end we can infer the asymptotic values $h_a \rightarrow 1$, $h_b \rightarrow Q$, and write

$$\mathrm{Hi}[\bar{d}_y(x)] \simeq (h_b + h_a - 1) \cdot \mathrm{separation}_y \simeq Q \cdot (\bar{y}_t - \bar{y}_c)$$

Similarly, for a cut at the low end, as $h_b \rightarrow 1$, $h_a \rightarrow P$, we can write

$$\mathrm{Lo}[\bar{d}_y(x)] \simeq (h_b + h_a - 1) \cdot \mathrm{separation}_y \simeq P \cdot (\bar{y}_t - \bar{y}_c)$$

(see Appendix A, pp. 1111-1119, for proofs of these theorems).

When the MAMBAC graphs indicate that the underlying data structure is taxonic, the taxon base rate P can be estimated using the ratio $R_{\mathrm{Hi/Lo}}$ of $\bar{d}_y(x)$ values computed at the low and high ends of the distribution of an input variable. For smaller samples, the fact that there are a few cases above the cut at the high end and below the cut at the low end may make a difference when we use the MAMBAC values as estimates of P and Q, so we multiply by the appropriate numbers of cases to correct for that. (More detailed explanation is given in Appendix A, pp. 1111-1119.) Thus we estimate the base rate P by

$$R = \frac{\mathrm{Hi}N_b}{\mathrm{Lo}N_a} \cdot \frac{\mathrm{Hi}[\bar{d}_y(x)]}{\mathrm{Lo}[\bar{d}_y(x)]} = \frac{Q}{P}$$

where ${}_{\text{Lo}}[\bar{d}_y(x)]$ is the MAMBAC value we have calculated at the low end of the input curve (at the first cut),[8] ${}_{\text{Lo}}N_a$ is the number of cases above that first cut, ${}_{\text{Hi}}[\bar{d}_y(x)]$ is the MAMBAC value calculated at the high end of the input curve (at the last cut), and ${}_{\text{Hi}}N_b$ is the number of cases below that last cut. (Note that the other MAMBAC values used to plot the graph are ignored; they are not used in these calculations.) Then

$$R = \frac{1 - P}{P}$$

and our base-rate estimate is

$$\hat{P} = \frac{1}{R + 1} \text{ and } \hat{Q} = 1 - \hat{P}$$

```
PROCEDURE EstimateBaseRate

    (* It is helpful to have recorded the numbers of cases above and below each cut when
       the MAMBAC values were calculated; these numbers could be saved in an output
       file on a computer or simply recorded on a tabulation sheet *)

  FOR each input/output combination

      Read MAMBAC_low; Read nAbove at this low-end cut;
      Read MAMBAC_high; Read nBelow at this high-end cut;

      R : = (nBelow[high] / nAbove[low]) * (MAMBAC[high] / MAMBAC[low]);
                  (* i.e., R = nBelow at the high cut / nAbove at the low cut
                                × MAMBAC at high end / MAMBAC at low end *)

      estimate_P : = 1 / (R + 1);
  END (* for each input/output combination*);

  estimate_P : = average over estimates from all in/out combinations;
END EstimateBaseRate
```

FIG. 15. Pseudocode for MAMBAC base-rate estimate

With four quantitative variables, we can get 12 $[= 2 \binom{4}{2}$, each indicator pair used bidirectionally] estimates of P from a sample, one for each input/output combination. We average these 12 estimates to get a mean $\hat{P}$ for the sample. The $\hat{P}$ values in Table 1 were based on the lowest and highest

[8] If MAMBAC cuts were made on the basis of one case at a time, e.g., as illustrated in Fig. 2., the "first" and "last" MAMBAC values should be ones chosen about 15 cases in from the bottom and top of the input curve. In our experience, that should provide enough stability to get an accurate base-rate estimate.

abscissa cuts on the input distributions.[9] Notice that it is possible to get negative estimates when the base rate is small (Table 1, $P = .10$, $N = 300$, sample 24). This happens more frequently for estimates based on single input/output combinations; this may be seen in tables for the individual samples in Appendix E (see especially configurations A3-10-20 and A6-10-20, pp. 1170 and 1172). We did not exclude negative or extreme parameter estimates throughout this article, but it might be rational for a researcher to do so. Although the averaged base-rate estimates in Table 1 may seem not so bad, estimates of subsequent parameters depend on the base-rate estimate and on each other, the result being propagation of error when clearly impossible values are carried along. (To see how the averaged $\hat{P}$ may improve, the interested reader may recalculate means across input/output combinations for Monte Carlo samples, e.g., sample A3-10-20.24 in Appendix E, p. 1171, omitting the negative estimates.)

When the base rate is small, there may be a slight bias toward overestimation of the base rate. This effect was greater when we used decile cuts to make estimations (for true $P = .10$ and $N = 300$, mean $\hat{P} = .14$, $SD = .05$ over 25 samples; for $P = .10$ and $N = 600$, mean $\hat{P} = .16$, $SD = .03$). We surmise it is due to the impurity of the cases at the top cut on a distribution; they are not exclusively taxon members as is required for a correct base-rate estimate. The bias is less, if there at all, when abscissa cuts are used, probably because that way of cutting takes one farther out on the distribution, hence the cases being used are more likely to be all taxon members. (We are considering trying an iterative procedure on taxon rates of inferred latent intervals to see whether the slight upward bias of $\hat{P}$ for low base rates is correctable. A simpler crude solution may be to reduce all low $\hat{P}$ values by a fixed "expected" bias ΔP. We have investigated various combinations, such as the relationship between error and standard deviation of the 12 $\hat{P}$ estimates, but have as yet been unable to discern any orderliness that would be helpful to researchers assessing the bias of their estimates.)

It is important to note that with a base rate of .50 the MAMBAC $\hat{P}$ by itself does not say anything about taxonicity; *it is imperative to look at the MAMBAC graphs as well.* A nontaxonic sample routinely gives a "base-rate" estimate of about .50 (see Appendix F, p. 1186); but if the latent structure is dimensional, good MAMBAC curves will be symmetrically dish-shaped. Only if the MAMBAC graph is peaked near the middle *and* the estimated base rate is about .50 may one infer a taxon with that base rate.

In our Monte Carlo tests, neither nuisance covariance nor less separation on the variables seemed to affect the MAMBAC base-rate estimations (it would be interesting to correlate the goodness of a taxonic curve with the

[9] The base-rate estimations for all of the input/output combinations for all samples may be found in Appendix E (pp. 1157-1185).

TABLE 1
MAMBAC ESTIMATES OF BASE RATE

	Sample Configuration																									
True Base Rate:											$P = .50$										$P = .25$				$P = .10$	
A[1]1[2]-50[3]-20[4]			A2-50-20		A3-50-20		A6-50-20		A3-50-15		A6-50-15		N3-50-20		N6-50-20		D3-50-v1		D6-50-v1		A3-25-20		A6-25-20		A3-10-20		A6-10-20	
	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*
1	.51	.08	.52	.04	.51	.08	.49	.06	.49	.12	.52	.13	.52	.05	.49	.05	.51	.03	.52	.07	.27	.08	.26	.10	.18	.13	.05	.10
2	.51	.14	.53	.05	.48	.08	.53	.06	.52	.10	.56	.09	.51	.04	.51	.05	.48	.09	.49	.05	.24	.05	.25	.09	.05	.18	.13	.13
3	.55	.06	.50	.08	.52	.06	.52	.05	.49	.09	.45	.10	.47	.06	.51	.05	.47	.07	.51	.04	.27	.11	.28	.06	.07	.16	.12	.10
4	.49	.04	.50	.04	.54	.07	.51	.07	.51	.08	.50	.15	.46	.03	.50	.05	.55	.07	.51	.05	.21	.06	.25	.08	.17	.09	.14	.17
5	.48	.08	.51	.05	.51	.06	.47	.05	.50	.11	.49	.11	.46	.04	.47	.05	.55	.09	.49	.04	.17	.15	.22	.11	.13	.10	.15	.10
6	.50	.07	.47	.09	.48	.05	.52	.09	.45	.11	.53	.11	.53	.05	.51	.06	.54	.06	.52	.04	.23	.10	.28	.08	.15	.13	.11	.14
7	.51	.08	.50	.06	.47	.07	.52	.09	.48	.05	.56	.08	.54	.06	.51	.06	.49	.07	.48	.05	.20	.09	.24	.07	.07	.16	.11	.13
8	.49	.07	.57	.08	.49	.07	.50	.09	.51	.11	.50	.08	.48	.06	.52	.05	.52	.07	.46	.12	.26	.13	.29	.06	.08	.11	.13	.05
9	.45	.06	.50	.08	.56	.11	.51	.06	.49	.09	.56	.08	.50	.07	.53	.05	.46	.10	.48	.05	.26	.08	.25	.10	.12	.18	.10	.07
10	.46	.08	.50	.04	.50	.06	.51	.07	.44	.08	.54	.14	.51	.04	.49	.04	.45	.06	.48	.08	.26	.05	.27	.10	.24	.15	.14	.13
11	.51	.09	.44	.10	.49	.06	.51	.05	.53	.10	.56	.07	.49	.05	.48	.06	.47	.06	.54	.05	.29	.06	.27	.05	.15	.10	.12	.10
12	.49	.05	.51	.08	.50	.05	.53	.04	.49	.14	.50	.08	.53	.06	.50	.07	.50	.05	.51	.06	.25	.07	.30	.08	.11	.13	.14	.19
13	.46	.07	.49	.07	.51	.07	.47	.07	.56	.11	.49	.11	.49	.06	.50	.06	.52	.04	.54	.07	.25	.09	.25	.08	.07	.11	.11	.08
14	.52	.04	.54	.06	.54	.11	.52	.05	.52	.07	.50	.13	.53	.09	.50	.05	.47	.08	.51	.09	.22	.06	.25	.10	.05	.12	.09	.10
15	.56	.15	.48	.08	.52	.06	.48	.07	.47	.20	.55	.08	.50	.07	.53	.04	.51	.06	.50	.08	.22	.09	.25	.06	.07	.14	.08	.15
16	.54	.06	.51	.09	.49	.07	.52	.06	.51	.14	.48	.11	.51	.04	.49	.05	.52	.06	.49	.07	.26	.05	.26	.05	.22	.16	.10	.11
17	.46	.07	.41	.10	.54	.09	.51	.08	.53	.09	.56	.13	.52	.04	.52	.04	.48	.05	.46	.11	.20	.07	.23	.06	.09	.12	.16	.13
18	.45	.07	.50	.05	.51	.07	.56	.06	.52	.11	.53	.08	.52	.05	.53	.04	.46	.09	.55	.05	.29	.07	.24	.06	.07	.11	.11	.08
19	.52	.04	.51	.09	.47	.05	.51	.05	.54	.11	.49	.12	.45	.06	.47	.05	.51	.07	.48	.08	.18	.07	.26	.09	.21	.11	.12	.12
20	.55	.09	.51	.05	.49	.07	.53	.11	.46	.11	.52	.07	.54	.07	.50	.07	.44	.07	.49	.05	.29	.09	.31	.08	.15	.13	.07	.14

(continued on next page)

[1]'A' samples have no nuisance covariance; 'N' and 'D' samples have nuisance covariance. [2]This number multiplied by 100 equals the sample size: 100, 200, 300, or 600. [3]Multiplied by .01 gives the true base rate of the sample (also indicated in the line above as *P* values): .50, .25, or .10. [4]Usually, multiplied by .1 gives the amount of separation on each variable: 2.0 *SD* or 1.5 *SD*; 'v1' samples have different separations on the four variables.

TABLE 1 (CONT'D)
MAMBAC ESTIMATES OF BASE RATE

True Base Rate:	Sample Configuration																									
	$P = .50$																				$P = .25$				$P = .10$			
A[1]1[2]-50[3]-20[4]	A2-50-20		A3-50-20		A6-50-20		A3-50-15		A6-50-15		N3-50-20		N6-50-20		D3-50-v1		D6-50-v1		A3-25-20		A6-25-20		A3-10-20		A6-10-20	
	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*	$\hat{P}$	*SD*
21	.55	.09	.50	.06	.52	.08	.53	.05	.48	.08	.54	.07	.50	.03	.47	.05	.57	.08	.46	.06	.26	.12	.22	.08	.10	.15	.14	.07
22	.53	.09	.51	.06	.53	.07	.48	.08	.53	.13	.54	.18	.51	.07	.51	.04	.54	.07	.55	.04	.27	.09	.24	.16	.12	.11	.11	.10
23	.48	.06	.51	.10	.57	.10	.50	.07	.57	.09	.51	.08	.48	.05	.51	.04	.47	.08	.59	.11	.32	.07	.27	.08	.02	.37	.15	.11
24	.50	.06	.49	.07	.47	.08	.47	.07	.56	.09	.52	.09	.49	.03	.50	.05	.52	.09	.48	.09	.31	.07	.26	.07	-.03	.30	.18	.09
25	.49	.04	.49	.03	.45	.06	.49	.06	.45	.09	.52	.09	.48	.06	.50	.03	.48	.06	.49	.07	.28	.11	.26	.07	.04	.17	.09	.10
Means and standard deviations over 25 samples per configuration:																										
$\hat{P}$	.50	.03	.50	.03	.51	.03	.51	.02	.50	.03	.52	.03	.50	.02	.50	.02	.50	.03	.50	.03	.25	.04	.26	.02	.11	.06	.12	.03
$\lvert\hat{P}\text{-}P\rvert$	.03	.02	.02	.02	.02	.02	.02	.01	.03	.02	.03	.02	.02	.01	.01	.01	.03	.02	.02	.02	.03	.02	.02	.01	.05	.04	.03	.02
$\hat{P}\text{-}P$	.00	.03	.00	.03	.01	.03	.01	.02	.00	.03	.02	.03	.00	.02	.00	.02	.00	.03	.00	.03	.00	.04	.01	.02	.01	.06	.02	.03

[1]'A' samples have no nuisance covariance; 'N' and 'D' samples have nuisance covariance. [2]This number multiplied by 100 equals the sample size: 100, 200, 300, or 600. [3]Multiplied by .01 gives the true base rate of the sample (also indicated in the line above as *P* values): .50, .25, or .10. [4]Usually, multiplied by .1 gives the amount of separation on each variable: 2.0 *SD* or 1.5 *SD*; 'v1' samples have different separations on the four variables.

accuracy of the base-rate estimate). When the graphs are unclear and $\hat{P} \simeq .50$, it may not be possible to answer the taxonic/nontaxonic question using only those two factors.

If the estimated base rate is notably less than about .40, that fact by itself corroborates[10] a conjecture of taxonicity. Fig. 16 shows the distributions of pooled $\hat{P}$ values (over 12 input/output combinations per sample) for all the Monte Carlo samples. The distributions are rather similar for nontaxonic samples and all the taxonic samples with a true base rate of .50; samples with true base rates of .25 or .10 give distinctly different $\hat{P}$ values.

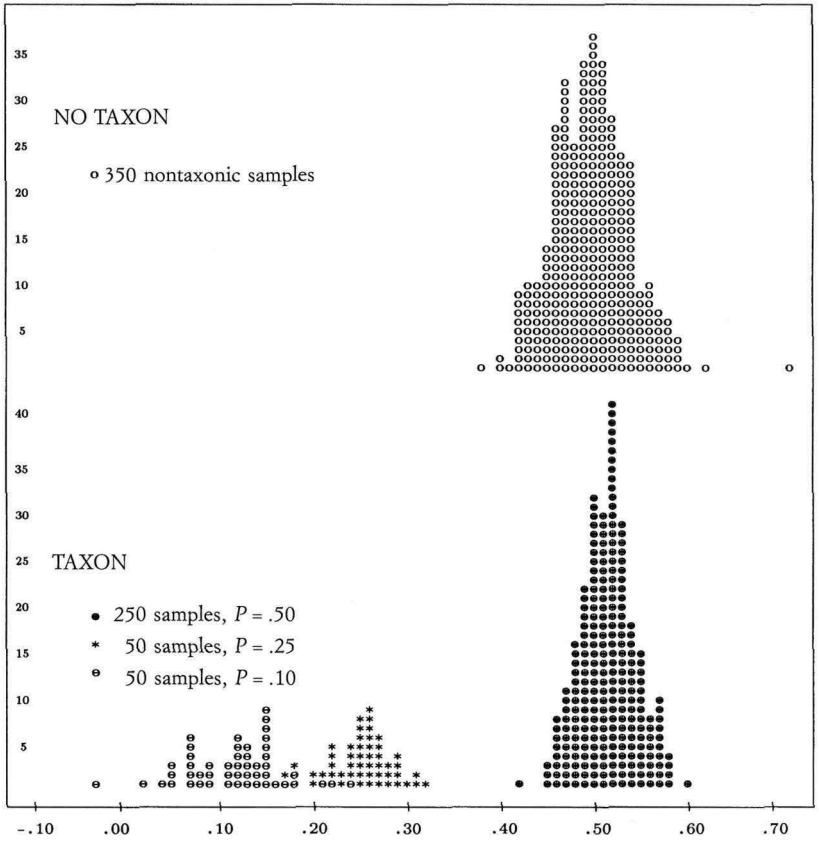


FIG. 16. Distributions of MAMBAC base-rate estimates (based on first and last abscissa cuts) for all of the Monte Carlo samples

[10] We use ⌜corroborate⌝ as does Popper, "pass a (risky) test." Inductivists may substitute ⌜confirm⌝, ⌜support⌝, ⌜tend to prove⌝. Of course these epistemological terms cannot be taken to mean *deduce* (*necessarily*) in empirical research.

Estimating Validity (Separation)

Once we have $\hat{P}$ (it being a pooled estimate based on as many variables as we have available),[11] we can solve for the taxonic separation on the output variable. For each pair of variables that is available, x (considered as the input variable in the following equations) and y (considered as the output variable), we already have

observed		latent		latent	
${}_{\text{Hi}}[\bar{d}_y(x)]$	=	$Q \cdot$ separation on y	=	$Q \cdot (\bar{y}_t - \bar{y}_c)$	[1]
${}_{\text{Lo}}[\bar{d}_y(x)]$	=	$P \cdot$ separation on y	=	$P \cdot (\bar{y}_t - \bar{y}_c)$	[2]

which we used (with the addition of a bias adjustment) to obtain $\hat{P}$. Now we can solve for the separation, doing it at both the high and low ends of an input distribution,

$$\widehat{\text{sep}}_y = \frac{{}_{\text{Hi}}[\bar{d}_y(x)]}{\hat{Q}} \qquad \text{Using the high end } (\hat{Q} = 1 - \hat{P}) \qquad [3]$$

$$\widehat{\text{sep}}_y = \frac{{}_{\text{Lo}}[\bar{d}_y(x)]}{\hat{P}} \qquad \text{Using the low end} \qquad [4]$$

Again, we make an adjustment for bias when using the MAMBAC values as estimates of P and Q, and the corrected separation estimates, computed independently at the two extremes, are

$${}_{\text{Hi}}[\bar{d}_y(x)] = {}_{\text{Hi}}b_{\text{b}}\ \text{sep}_y = \frac{NQ}{{}_{\text{Hi}}N_{\text{b}}}\ \text{sep}_y$$

at the high end:

$${}_{\text{Hi}}\widehat{\text{sep}}_y = \frac{{}_{\text{Hi}}N_{\text{b}}}{NQ}\ {}_{\text{Hi}}[\bar{d}_y(x)]$$

$${}_{\text{Lo}}[\bar{d}_y(x)] = {}_{\text{Lo}}b_{\text{a}}\ \text{sep}_y = \frac{NP}{{}_{\text{Lo}}N_{\text{a}}}\ \text{sep}_y$$

at the low end:

$${}_{\text{Lo}}\widehat{\text{sep}}_y = \frac{{}_{\text{Lo}}N_{\text{a}}}{NP}\ {}_{\text{Lo}}[\bar{d}_y(x)]$$

[11] We have not yet done Monte Carlo runs to see which $\hat{P}$ is better to use in calculating estimates of the separation when multiple indicators are used. There is one $\hat{P}$ estimated for a given input/output combination; if MAMBAC is run bidirectionally on those indicators, one might use the average of the two $\hat{P}$ estimates obtained; or one might use $\hat{P}$ pooled over estimates from all the variables run bidirectionally. It seems nearly certain that the last would be the most accurate estimate, hence the best one to use. That is what was used for Monte Carlo tests reported here; we used the average of the 12 $\hat{P}$ estimates we got from running MAMBAC bidirectionally on four continuous indicators.

These estimate the same latent value, so we average them,

$$\widehat{sep}_y = (\widehat{\bar{y}_t - \bar{y}_c}) = \tfrac{1}{2}({}_{Hi}\widehat{sep}_y + {}_{Lo}\widehat{sep}_y) \quad [5]$$

As one would expect, taxonic separation is estimated better with larger sample sizes, larger base rates, and no nuisance covariance. Monte Carlo re-

```
PROCEDURE EstimateSeparation

    (* It is helpful to have an output file of MAMBAC values and the numbers of cases
       above and below the cut when each was calculated; this file could be created
       when MAMBAC values are obtained for plotting the MAMBAC curve *)

  FOR each input/output combination

      Read MAMBAC_low; Read nAbove at this low-end cut;
      Read MAMBAC_high; Read nBelow at this high-end cut;
      (* estimate_P has been calculated previously *)

      sep_high : = nBelow[high] / (N * (1 - estimate_P) * MAMBAC[high];
                  (* i.e., nBelow at the high cut / (N x Q) x MAMBAC at high end *)

      sep_low : = nAbove[low] / (N * estimate_P) * MAMBAC[low];
                  (* i.e., nAbove at the low cut / (N x P) x MAMBAC at low end *)

      estimate_sep : = (sep_high + sep_low) / 2;

END (* for each input/output combination *);
      END EstimateSeparation.
```

FIG. 17. Pseudocode for MAMBAC validity estimation

sults for 14 sample configurations are shown in Table 2 (estimates for all input/output combinations for each taxonic sample may be found in Appendix G, pp. 1187-1215). There is variability in the accuracy of the estimates of separation for all the parametric configurations presented in Table 2, but two generalizations may be made: nuisance covariance produces overestimation of taxonic separation, and a low base rate leads to underestimation of the separation. Development of procedures to assess and correct bias in separation estimates due to these two influences has yet to be done.

Unless the MAMBAC curve suggests taxonicity, one would not reach this step. But, for the curious, the result of trying to estimate taxonic separation when the true latent situation is factorial depends on the amount of correlation between variables (see Appendix H, p. 1216). Estimated "separations" (averaged over 25 samples each) ranged from 1.01 to 2.17 for our nontaxonic samples. The correlation between those estimates and the expected amount of correlation between variables is .89 (the comparable correlation for taxonic samples is .83).

Estimating Latent Means

Given that we have reason to infer underlying taxonicity from the MAMBAC curve(s) and now have estimates of the base rate and the separations, we can estimate the means of the complement and taxon groups that make up the manifest distribution. We have another linear equation:

$$\underline{\text{observed}} \qquad \underline{\text{latent}}$$

$$\bar{y} \quad = \quad P\bar{y}_t + Q\bar{y}_c \qquad [6]$$

P and Q are known by our estimation of P, and the separation has been estimated as described above. Now we can solve the system of equations,

$$\underline{\text{latent}} \qquad \underline{\text{known}}$$

$$\bar{y}_t - \bar{y}_c \quad = \quad \widehat{\text{sep}}_y \qquad \text{From [5], inferred} \qquad [7]$$

$$\underline{\text{latent}} \qquad \underline{\text{known}}$$

$$P\bar{y}_t + Q\bar{y}_c \quad = \quad \bar{y} \qquad \text{From [6], observed} \qquad [8]$$

to get $\bar{y}_t$ and $\bar{y}_c$.

```
PROCEDURE EstimateComplement&TaxonMeans
  FOR each input/output combination
    (* estimate_P has been calculated previously *)
    (* estimate_sep has been calculated previously *)
    variableMean : = observed mean of output variable distribution;
    estimate_CompMean : = variableMean - estimate_P * estimate_sep;
    estimate_TaxonMean : = estimate_CompMean + estimate_sep;
  END (* for each input/output combination *);
END EstimateComplement&TaxonMeans.
```

FIG. 18. Pseudocode for MAMBAC estimation of complement and taxon means

Estimated means for Monte Carlo samples are shown in Table 3 (estimations for the complement group) and Table 4 (estimations for the taxon group). Estimates for all the input/output combinations for the taxonic samples may be found in Appendix I (pp. 1217-1273). Smaller samples ($N = 300$ here) with low base rates ($P = .10$ here) estimate complement means adequately, but they may do a poor job of estimating taxon means (see Table 4, configuration A3-10-20). There is great variability in results from the individual samples with that configuration (see Appendix I, pp. 1217-1273).

TABLE 2
MAMBAC Estimates of Separation

													Sample	
Expected separation:*			2.00 *SD*						1.50 *SD*				2.00 *SD*	
	A[1]1[2]-50[3]-20[4]		A2-50-20		A3-50-20		A6-50-20		A3-50-15		A6-50-15		N3-50-20	
Sample	$\hat{sep}$	*SD*	$\hat{sep}$	*SD*	$\hat{sep}$	*SD*	$\hat{sep}$	*SD*	$\hat{sep}$	*SD*	$\hat{sep}$	*SD*	$\hat{sep}$	*SD*
1	2.01	.26	2.02	.23	1.83	.37	2.06	.29	1.42	.29	1.38	.29	2.59	.28
2	1.48	.43	1.97	.35	1.82	.24	2.06	.20	1.40	.18	1.24	.40	2.71	.35
3	1.75	.22	2.14	.33	2.09	.22	2.02	.28	1.44	.22	1.47	.33	2.73	.33
4	1.89	.30	2.23	.13	1.93	.36	2.04	.27	1.44	.20	1.21	.34	2.63	.51
5	1.85	.12	2.26	.20	1.96	.21	1.91	.29	1.32	.31	1.54	.36	2.80	.40
6	1.94	.19	1.93	.33	1.97	.26	2.00	.24	1.63	.18	1.46	.37	2.82	.38
7	1.71	.36	2.21	.27	2.07	.29	2.06	.25	1.61	.31	1.32	.34	2.80	.38
8	1.97	.23	1.99	.30	1.96	.35	1.97	.23	1.26	.43	1.33	.20	2.56	.35
9	1.84	.25	1.92	.24	1.83	.36	1.95	.19	1.51	.32	1.56	.30	2.48	.36
10	1.89	.22	2.10	.34	2.07	.20	1.87	.33	1.34	.23	1.33	.34	2.62	.64
11	1.78	.20	1.92	.27	1.91	.27	2.05	.26	1.24	.23	1.69	.25	2.26	.34
12	2.10	.20	1.97	.46	1.85	.26	2.20	.40	1.33	.25	1.37	.35	2.62	.39
13	2.15	.19	2.03	.26	1.92	.20	2.00	.23	1.38	.22	1.27	.16	2.59	.40
14	2.07	.33	1.78	.18	1.84	.29	2.12	.32	1.42	.24	1.63	.25	2.41	.34
15	1.54	.27	1.96	.35	2.03	.41	2.05	.32	1.36	.49	1.51	.29	2.65	.39
16	1.90	.37	2.00	.20	2.16	.37	2.03	.27	1.25	.26	1.45	.28	2.45	.33
17	1.90	.36	1.71	.27	2.01	.27	1.97	.28	1.38	.30	1.35	.47	2.57	.32
18	1.95	.40	1.84	.23	1.93	.36	2.18	.21	1.31	.34	1.48	.38	2.72	.37
19	1.66	.22	1.96	.34	2.01	.23	2.03	.37	1.44	.28	1.49	.27	2.59	.43
20	1.56	.28	2.01	.23	2.24	.25	1.99	.27	1.52	.28	1.33	.26	2.60	.39
21	1.62	.41	1.85	.37	1.94	.33	1.98	.23	1.60	.39	1.52	.19	2.92	.42
22	1.80	.23	1.98	.32	2.07	.34	1.93	.22	1.25	.26	1.46	.27	2.58	.58
23	1.86	.25	1.86	.28	1.96	.35	1.94	.28	1.66	.37	1.42	.22	2.79	.38
24	1.95	.30	1.95	.25	1.93	.37	1.90	.27	1.32	.36	1.39	.25	2.46	.51
25	2.10	.22	1.96	.32	2.09	.31	2.08	.18	1.47	.33	1.33	.28	2.76	.30
Means and standard deviations over 25 samples per configuration:														
$\hat{sep}$	1.85	.18	1.98	.13	1.98	.11	2.02	.08	1.41	.12	1.42	.11	2.63	.15
$\lvert\hat{sep}$-sep$\rvert$	.18	.14	.10	.09	.09	.06	.06	.05	.13	.07	.12	.08	.63	.15
$\hat{sep}$-sep	−.15	.18	−.02	.13	−.02	.11	−.02	.08	−.09	.12	−.08	.12	.63	.15

[1]'A' samples have no nuisance covariance; 'N' and 'D' samples have nuisance covariance. [2]This gives the true base rate of the sample: .50, .25, or .10. [4]Usually, multiplied by .1 gives the are $x = 2.00$, $y = 1.75$, $z = 1.50$, $v = 1.25$; the average of these (1.625) was substituted for "true used in the generation of the Monte Carlo samples; in fact, the true separation varied from

This is due in part to propagation of error from allowing impossible values such as negative base rate and negative estimates of separation to remain in our calculations. Pilot runs indicate substantial correlations (ranging from .45 to .78 in the A3-10-20 samples) between absolute error in estimates of the taxon means and the ratio of the standard deviation of the base-rate estimates to the mean $\hat{P}$ for each sample. Defining a "bad sample" as one in which at least one of the four taxon means is mal-estimated with an absolute

TABLE 2 (CONT'D)
MAMBAC ESTIMATES OF SEPARATION

Configuration													
2.00 *SD*		*x*: 2.00 *SD*, *y*: 1.75 *SD* *z*: 1.50 *SD*, *v*: 1.25 *SD*				2.00 *SD*							
N6-50-20		D3-50-v1		D6-50-v1		A3-25-20		A6-25-20		A3-10-20		A6-10-20	
sêp	*SD*	sêp	*SD*	sêp	*SD*	sêp	*SD*	sêp	*SD*	sêp	*SD*	sêp	*SD*
2.65	.43	2.09	.52	2.07	.63	1.70	.21	2.02	.48	1.75	.58	1.81	1.68
2.72	.32	2.27	.77	2.27	.51	1.93	.32	1.78	.41	1.44	1.49	1.88	1.04
2.42	.57	2.33	.76	2.31	.66	1.77	.32	1.86	.34	1.42	1.38	1.81	.77
2.73	.47	2.15	.56	2.48	.52	1.90	.44	1.91	.35	1.69	.57	1.81	1.06
2.69	.52	2.49	.76	2.40	.77	1.86	.82	1.80	.52	1.53	.77	1.99	.68
2.99	.49	2.40	.70	2.29	.69	1.89	.60	2.01	.37	1.47	.58	1.58	1.09
2.87	.47	2.23	.54	2.20	.79	1.95	.53	1.87	.40	1.45	1.43	1.70	.86
2.93	.49	2.22	.64	2.11	.80	1.78	.60	2.03	.38	1.68	1.20	1.70	.39
2.66	.32	2.21	.57	2.38	.71	1.85	.36	1.76	.39	1.54	.98	1.84	.94
2.73	.41	1.84	.61	2.38	.74	2.04	.36	1.87	.30	1.39	.35	1.56	.62
2.71	.54	2.20	.47	2.37	.66	2.10	.29	1.96	.23	1.49	.54	1.62	.75
2.63	.50	1.99	.68	2.64	.79	1.68	.37	1.99	.34	1.75	1.26	1.87	1.20
2.87	.37	2.27	.54	2.38	.67	1.93	.46	2.02	.33	1.63	1.35	1.48	.59
2.77	.39	2.17	.56	2.23	.64	1.82	.27	2.00	.54	1.42	1.38	1.98	1.19
2.70	.46	2.29	.62	2.22	.77	1.90	.53	2.05	.32	1.65	1.49	1.30	1.05
2.83	.47	2.23	.59	2.35	.67	2.12	.39	1.94	.21	1.21	.41	1.68	.84
2.87	.45	2.33	.50	2.34	.91	1.85	.58	1.94	.30	1.52	.92	1.88	.94
2.61	.40	2.23	.53	2.23	.59	1.91	.25	1.96	.38	1.60	1.32	1.77	.58
2.76	.41	2.09	.42	2.54	.74	1.81	.44	1.84	.44	1.42	.40	1.87	.97
3.03	.44	2.40	.65	2.17	.77	1.95	.31	1.99	.37	1.33	.65	2.18	2.18
2.50	.64	2.11	.53	2.18	.64	1.60	.33	1.70	.49	1.50	1.25	1.67	.44
2.66	.46	1.94	.55	2.45	.54	1.76	.35	1.85	.54	1.68	.79	2.00	.98
2.67	.46	1.77	.60	2.36	.61	1.94	.28	1.91	.30	4.38	8.09	1.75	.78
2.64	.48	2.05	.61	2.45	.55	1.73	.36	1.90	.28	−0.25	3.74	1.76	.30
2.67	.32	2.20	.47	2.11	.52	1.83	.62	2.00	.31	1.80	3.23	1.61	.98
2.73	.14	2.18	.17	2.32	.14	1.86	.12	1.92	.09	1.58	.68	1.76	.18
.73	.14	.55	.17	.69	.14	.16	.09	.09	.08	.61	.52	.25	.16
.73	.14	.55	.17	.69	.14	−.14	.12	−.08	.09	− .42	.68	−.24	.18

number multiplied by 100 equals the sample size: 100, 200, 300, or 600. [3]Multiplied by .01 amount of separation on each variable: 2.0 *SD* or 1.5 *SD*. In 'v1' samples expected separations separation" to get values for |sêp–sep| and sêp–sep. *The expected amount of separation was sample to sample, and those true separations may be found in Appendix G, pp. 1187-1215.

error ≥ 1 *SD*, cutting the $\hat{P}$-dispersion consistency test at $\frac{SD\hat{p}}{\hat{p}} \geq 1.50$ identifies 85% of the bad samples at the expense of only 8% of acceptable ones, an encouraging preliminary result.

Another possibility for estimating taxon means relies on the fact that for small *P* and a cusp at the high extreme of the MAMBAC graph a sizeable subsample can be defined (say, all cases below the input variable's median)

TABLE 3
MAMBAC Estimations of Complement Means

Configuration	Pooled Estimates								Algebraic Error								Absolute Error							
	$\bar{x}$	*SD*	$\bar{y}$	*SD*	$\bar{z}$	*SD*	$\bar{v}$	*SD*	$\bar{x}$	*SD*	$\bar{y}$	*SD*	$\bar{z}$	*SD*	$\bar{v}$	*SD*	$\bar{x}$	*SD*	$\bar{y}$	*SD*	$\bar{z}$	*SD*	$\bar{v}$	*SD*
Expected Mean:	**.00**		**.00**		**.00**		**.00**																	
A1-50-20[1]	**.05**	.14	**.07**	.15	**.05**	.12	**.04**	.14	**.02**	.14	**.04**	.13	**.07**	.07	**.00**	.13	**.12**	.06	**.12**	.07	**.08**	.06	**.11**	.08
A2-50-20	**.00**	.12	**−.04**	.13	**.02**	.17	**.01**	.15	**−.01**	.12	**−.02**	.14	**.04**	.12	**.01**	.14	**.10**	.07	**.10**	.09	**.10**	.08	**.12**	.09
A3-50-20	**−.01**	.12	**−.04**	.14	**−.01**	.11	**−.02**	.09	**−.04**	.13	**−.06**	.13	**.02**	.09	**−.01**	.09	**.12**	.08	**.11**	.09	**.07**	.06	**.08**	.05
A6-50-20	**−.03**	.13	**−.01**	.10	**.00**	.12	**−.02**	.11	**−.02**	.09	**−.04**	.11	**−.02**	.11	**−.03**	.10	**.07**	.06	**.09**	.07	**.08**	.08	**.08**	.06
Base Rates = .25 or .10																								
A3-25-20	**.06**	.13	**.05**	.13	**.05**	.12	**.05**	.12	**.05**	.13	**.06**	.14	**.02**	.09	**.04**	.16	**.10**	.09	**.13**	.09	**.07**	.06	**.13**	.09
A6-25-20	**.00**	.10	**.04**	.07	**.01**	.10	**.01**	.09	**−.01**	.09	**.03**	.08	**−.01**	.07	**.00**	.11	**.08**	.06	**.07**	.05	**.06**	.04	**.08**	.07
A3-10-20	**.04**	.13	**.02**	.12	**.04**	.12	**.03**	.12	**.03**	.14	**.02**	.14	**.03**	.11	**.03**	.14	**.11**	.09	**.11**	.09	**.09**	.06	**.11**	.09
A6-10-20	**−.02**	.10	**.01**	.10	**.01**	.10	**−.01**	.11	**−.02**	.10	**.00**	.11	**−.01**	.08	**−.02**	.13	**.07**	.06	**.08**	.07	**.06**	.05	**.10**	.08
Nuisance Covariance																								
Factor loadings on $x = .70$, $y = .50$, $z = .40$, $v = .20$																								
N3-50-20	**−.39**	.16	**−.34**	.15	**−.26**	.13	**−.19**	.13	**−.42**	.16	**−.36**	.17	**−.29**	.10	**−.19**	.16	**.42**	.16	**.36**	.17	**.29**	.10	**.20**	.15
N6-50-20	**−.49**	.10	**−.41**	.11	**−.34**	.11	**−.17**	.12	**−.48**	.11	**−.40**	.12	**−.35**	.09	**−.18**	.12	**.48**	.11	**.40**	.12	**.35**	.09	**.18**	.10
Taxonic Separation = 1.5 *SD*																								
A3-50-15	**.03**	.14	**.06**	.12	**.03**	.12	**.06**	.10	**.02**	.15	**.05**	.13	**.03**	.09	**.06**	.11	**.12**	.09	**.11**	.08	**.08**	.05	**.10**	.07
A6-50-15	**.03**	.11	**.05**	.13	**.00**	.12	**.01**	.11	**.03**	.11	**.03**	.12	**−.01**	.09	**.01**	.11	**.09**	.07	**.10**	.08	**.07**	.06	**.09**	.07
Nuisance Covariance + Taxonic Separation																								
Factor loadings on $x = .70$, $y = .50$, $z = .40$, $v = .20$																								
Separations $x = 2.00$, $y = 1.75$, $z = 1.50$, $v = 1.25$																								
D3-50-v1	**−.62**	.15	**−.42**	.16	**−.26**	.16	**−.05**	.11	**−.61**	.16	**−.44**	.15	**−.26**	.13	**−.07**	.10	**.61**	.16	**.44**	.15	**.26**	.13	**.10**	.07
D6-50-v1	**−.73**	.15	**−.52**	.11	**−.36**	.14	**−.04**	.15	**−.72**	.16	**−.54**	.11	**−.36**	.14	**−.07**	.13	**.72**	.16	**.54**	.11	**.36**	.14	**.11**	.09

[1]See Table 1 notes for explanation of sample configuration codes.

TABLE 4
MAMBAC Estimations of Taxon Means

Configuration	Pooled Estimates								Algebraic Error								Absolute Error							
	$\bar{x}$	*SD*	$\bar{y}$	*SD*	$\bar{z}$	*SD*	$\bar{v}$	*SD*	$\bar{x}$	*SD*	$\bar{y}$	*SD*	$\bar{z}$	*SD*	$\bar{v}$	*SD*	$\bar{x}$	*SD*	$\bar{y}$	*SD*	$\bar{z}$	*SD*	$\bar{v}$	*SD*
Expected Mean:	**2.00**		**2.00**		**2.00**		**2.00**																	
A1-50-20[1]	**1.91**	.21	**1.89**	.18	**1.91**	.14	**1.91**	.14	**−.09**	.19	**−.08**	.15	**−.08**	.11	**−.11**	.11	**.17**	.13	**.15**	.09	**.10**	.08	**.13**	.08
A2-50-20	**1.98**	.15	**2.02**	.14	**1.95**	.13	**1.97**	.13	**−.05**	.13	**.02**	.14	**−.04**	.11	**−.02**	.14	**.10**	.10	**.12**	.07	**.09**	.08	**.10**	.10
A3-50-20	**1.94**	.16	**1.98**	.14	**1.95**	.12	**1.95**	.13	**−.07**	.13	**−.05**	.11	**−.04**	.11	**−.03**	.15	**.13**	.08	**.10**	.07	**.09**	.07	**.13**	.09
A6-50-20	**2.01**	.11	**2.00**	.07	**1.98**	.08	**2.00**	.09	**.01**	.10	**−.02**	.09	**−.01**	.08	**−.02**	.09	**.08**	.06	**.07**	.05	**.06**	.05	**.08**	.06
Base Rates = .25 or .10																								
A3-25-20	**1.89**	.17	**1.92**	.25	**1.91**	.21	**1.93**	.23	**−.12**	.19	**−.07**	.22	**−.09**	.16	**−.06**	.22	**.18**	.13	**.18**	.14	**.14**	.11	**.17**	.16
A6-25-20	**1.96**	.21	**1.86**	.19	**1.93**	.13	**1.96**	.16	**−.04**	.18	**−.16**	.18	**−.04**	.13	**−.04**	.13	**.15**	.11	**.20**	.14	**.12**	.07	**.12**	.06
A3-10-20	**1.27**	1.41	**1.74**	.97	**1.54**	1.49	**1.81**	1.88	**−.70**	1.40	**−.37**	1.02	**−.46**	1.50	**−.15**	1.88	**.89**	1.27	**.74**	.79	**1.04**	1.17	**1.08**	1.54
A6-10-20	**1.93**	.61	**1.66**	.47	**1.64**	.39	**1.76**	.48	**−.08**	.57	**−.36**	.49	**−.35**	.39	**−.27**	.42	**.44**	.36	**.51**	.34	**.42**	.31	**.41**	.28
Nuisance Covariance																								
Factor loadings on $x = .70$, $y = .50$, $z = .40$, $v = .20$																								
N3-50-20	**2.43**	.14	**2.38**	.12	**2.30**	.14	**2.23**	.15	**.42**	.13	**.37**	.10	**.29**	.10	**.21**	.16	**.42**	.13	**.37**	.10	**.29**	.10	**.22**	.15
N6-50-20	**2.52**	.11	**2.44**	.09	**2.37**	.10	**2.20**	.11	**.50**	.12	**.42**	.10	**.34**	.09	**.18**	.10	**.50**	.12	**.42**	.10	**.34**	.09	**.19**	.09
Taxonic Separation = 1.5 *SD*																								
Expected Mean:	**1.50**		**1.50**		**1.50**		**1.50**																	
A3-50-15	**1.47**	.11	**1.44**	.13	**1.47**	.12	**1.44**	.10	**−.04**	.10	**−.06**	.11	**−.04**	.10	**−.08**	.12	**.10**	.05	**.10**	.07	**.09**	.07	**.11**	.09
A6-50-15	**1.44**	.10	**1.42**	.12	**1.46**	.07	**1.45**	.11	**−.07**	.12	**−.10**	.12	**−.05**	.07	**−.06**	.12	**.11**	.08	**.11**	.10	**.07**	.05	**.10**	.09
Nuisance Covariance + Taxonic Separation																								
Factor loadings on $x = .70$, $y = .50$, $z = .40$, $v = .20$																								
Separations $x = 2.00$, $y = 1.75$, $z = 1.50$, $v = 1.25$																								
Expected Mean:	**2.00**		**1.75**		**1.50**		**1.25**																	
D3-50-v1	**2.13**	.15	**1.92**	.14	**1.76**	.16	**1.55**	.12	**.09**	.11	**.18**	.11	**.27**	.13	**.28**	.16	**.12**	.08	**.19**	.09	**.27**	.13	**.28**	.16
D6-50-v1	**2.22**	.14	**2.01**	.12	**1.86**	.13	**1.53**	.10	**.21**	.14	**.23**	.13	**.34**	.11	**.26**	.12	**.21**	.13	**.24**	.12	**.34**	.11	**.26**	.12

[1]See Table 1 notes for explanation of sample configuration codes.

which will be negligibly contaminated with taxon cases and hence provide a trustworthy estimate of the complement mean. (In this case it would be estimated by the mean of the output variable cases in that subsample.) We can then estimate the taxon mean using Equation [8] alone, bypassing Equation [7]. Equation [8] is precise since it is a distribution-free set-theoretical identity (for the sample). The adequacy of this and other approaches requires further Monte Carlo study. But we cannot emphasize strongly enough that taxometric research requires large samples, particularly when base rates are small. Since the graph shape is trustworthy as a taxon detector, when its maximum is a cusp at the upper extreme of the input variable, strongly suggesting a low base rate, the researcher with a small sample is alerted to be cautious about taxon mean estimates.

Nuisance covariance has the biggest predictable effect on accuracy of estimations. It leads to overestimation of taxonic means and underestimation of the complement means. In Equations [7] and [8], the numerator determinants solving for unknowns $\bar{y}_t$ and $\bar{y}_c$ expand as $(Q\ \hat{\text{sep}}_y + \bar{y})$ and $(\bar{y} - P\ \hat{\text{sep}}_y)$, respectively. Since $\hat{\text{sep}}_y$ is inflated due to nuisance correlation, the former is spuriously raised and the latter spuriously lowered.

There appears at first to be an anomaly in estimates of the taxon mean for x in nuisance covariance samples ('N' configurations in Table 4) and in samples with both nuisance covariance and various separations ('D' configurations in Table 4); in the latter, x has the same factor loading (.70) and the same expected separation (2 *SD*), yet the taxon mean is not overestimated to the same extent. This is explained when we remember that the estimates involve x in combination with y, z, and v, all of which have smaller taxonic separations which reduce the estimates of separation and, hence, of the taxon mean for x. Likewise, the larger separations on x, y, and z increase estimates for v. Interactions between nuisance covariance and validity remain to be explored.

Although the largest errors (.40s and .50s) in Table 3 are irksome, they are not as "bad" as they seem. The metric here is expressed in standard scores of the latent distributions, whereas the researcher's units will be based on the observed standard deviations (the only thing the researcher can know!), and the latter will be considerably larger. For example, in the configuration with nuisance covariance and $N = 600$ (N6-50-20 in Table 3), the $\bar{x}$-error in the manifest standard score metric is $\frac{.44}{\sqrt{2}} = .31$, less than ⅓ *SD*. This error is only slightly larger than the rule of thumb for "coarse grouping" (in the old precomputer days) that Karl Pearson showed loses 10% of the information.

Inferring Latent Frequency Curves, Locating the Hitmax Cut, and Estimating the Over-all Hit Rate and Valid and False Positive Rates

When there is no nuisance covariance, if there were no sampling error in the taxon and complement means within defined intervals along the input distribution, i.e., if the y-mean of taxon members within any x-interval = the taxon mean, the observed mean y in an x-interval depends on the ratio of taxon members to complement members in the interval. We know this is literally false (due to random fluctuation), but it is the best estimate. Thus we can get the proportions of taxon and complement members in each interval from

$$\text{In interval } x_i\text{:} \qquad \underbrace{p_{ti}\bar{y}_t + q_{ci}\bar{y}_c = p_{ti}\bar{y}_t + (1 - p_{ti})\bar{y}_c}_{\text{observed}} = \bar{y}(x_i)$$

We solve for p_i in each x-interval, then multiply it by n_i (the frequency in the interval) to get the taxon frequency n_{ti}.

This allows us to draw the inferred latent x-curves, and from them we can locate the hitmax cut (which will be the place where the complement and taxon curves intersect), compute valid and false positive rates achieved by that cut, compute over-all hit rate, get estimates of the variances and additional estimates of the means (which will provide consistency tests of those previously estimated), and assign subjects to complement or taxon membership. Monte Carlo study of these estimates is planned for future publications.

Consistency Tests and the Importance of Coherence

A core feature of Meehl's coherent cut kinetics method (going back to the first technical report, Meehl, 1965) is emphasis on the fundamental importance of consistency tests in taxometric analysis. If the latent structure is as conjectured and the inferred numerical values are correct within acceptable limits of error, then there are formal procedures in which mathematical relations between two or more quantitative variables, latent or manifest, observed or inferred, may be expected to obtain, within allowed tolerances. There is a tendency for psychologists and statisticians to view such consistency tests as a pleasant adjunct, nice to have if you can get them, merely "icing on the psychometric cake." This attitude is mistaken; consistency tests are absolutely essential in taxometric analysis in all empirical domains.

Why is this? The necessity for consistency tests arises from the problem of construct validity (Cronbach & Meehl, 1955; Loevinger, 1957; Campbell & Fiske, 1959). In simple clear-cut cases of predictive validity, we have available what some psychometrists have called a "Gold Standard Criterion," and the "validity" of a predictor variable (or a composite of such, as in multiple regression, linear discriminant function, or actuarial table) is problematic

only in the sense of random sampling error, because it does not involve any theoretical inference to unobserved states, structures, or events. For example, if an aviation psychologist wants to predict which candidates for flight training will succeed, a stanine score that combines measures of dial reading, general intelligence, and emotional stability "does just as well as it does" (which in World War II turned out to be very well indeed). If an educational psychologist counts spelling errors in a sample of several thousand sentences written by a person in a representative sample of settings, this may be taken as a Gold Standard Criterion of real-life spelling ability; and the correlation of a multiple-choice hundred-item spelling test with the count from that huge sample is a simple case of concurrent validity, the only problem being random sampling fluctuation on the criterion side. In organic disease, the pathology and etiology jointly define a disease entity, and a sign or symptom has a validity commonly expressed in terms of sensitivity and specificity (valid and false positive rates).

One can divide Gold Standard Criteria into two rough subcategories, worthwhile distinguishing for conceptual clearness but not differing appreciably in their function as validators of a proposed fallible indicator. Sometimes a Gold Standard Criterion is *definitory*, that is, the explicit literal *meaning* of the dimension or category is exemplified by the criterion property. For example, general paresis is explicitly defined as a certain kind of cerebral pathology produced by parenchymal infection with *Treponema pallidum*. Certain pathognomonic brain changes (such as gross disarrangement of cellular layers and free iron in the ganglion cells) conjoined with presence of the spirochete in the cortical tissue constitute the explicit operational definition of paretic brain state. In organic disease, the nosological entity is defined *conjunctively*, by pathology *and* etiology, if the latter is known (cf. Meehl, 1973b, pp. 285-288; 1992, pp. 126-127). For a nonmedical example, if an insurance company wants to know whether an insurance agent will sell a lot of insurance, then the remarkably high correlation of the insurance salesman key of the Strong Vocational Interest Blank with dollars sold constitutes concurrent (and, for selection of personnel in the future, predictive) validity against the Gold Standard Criterion of sales. In the second subset, more common in medicine, indicators are not definitory but are nevertheless two-way pathognomonic, i.e., perfectly valid as both an inclusion and an exclusion test, with sensitivity = specificity = 1. Thus, a positive spinal fluid Wasserman and first zone colloidal gold curve, while not definitory of paresis, are jointly pathognomonic and therefore this two-sign configuration counts as a Gold Standard Criterion of the disease entity.

Moving to the use of statistical methods for testing substantive causal theories, e.g., genetics of schizophrenia, we no longer have a Gold Standard Criterion available to us. If we have understood the comparative feebleness

of refutation of the null hypothesis in appraising scientific theories (Morrison & Henkel, 1970; Meehl, 1978, 1990a), we are looking for stronger tests against which competing theories may be judged. Lacking a Gold Standard Criterion for the validity of an indicator of some theoretical construct, how is it possible to arrive at a "risky prediction" (or "difficult hurdle" or "dangerous potential falsifier," to use Karl Popper's terminology)? Insistence upon consistency tests in taxometrics does not hinge upon acceptance of Popper's metatheoretical views. A non-Popperian, e.g., an inductivist, might rather express this riskiness feature as does Wesley Salmon, in terms of deriving certain relationships among the facts from a theory such that, if the theory had negligible verisimilitude, this successful derivation would constitute a "damned strange coincidence" (Salmon, 1984, pp. 213-227; and see discussion in Meehl, 1990a, pp. 116-121; 1990b, pp. 39-42).

The reasoning here is simple and ineluctable. A sufficiently strong theory may permit a numerical prediction directly, as sometimes happens in physics, chemistry, or astronomy; it happens less commonly in the life sciences but can occur in fields like genetics, e.g., on Meehl's dominant gene theory of schizotaxia, the incidence of schizotaxia in the first-degree relatives of schizophrenes should be one-half. In most of the life sciences, and in almost all of the behavioral life sciences, theories are not sufficiently strong to make numerical predictions but can, at most, by specifying the conjectured causal or compositional structure, *derive theorems within that structure* that relate some parameters to others.[12] This possibility gives rise to several kinds of consistency tests.

Suppose the latent structure is conjectured to be taxonic, and in fact it is. Then when a given taxometric procedure, e.g., MAMBAC, is applied to different sets of indicators, estimates of a latent parameter, say, the taxon base rate P, should agree within tolerance.[13] Second, a given set of indicators analyzed by different nonredundant procedures, e.g., MAMBAC and MAXCOV, should lead to the same value of P. Third, a more severe and hence more strongly corroborative test, applying two or more different taxometric

[12] To these strong theoretical considerations from mathematics and epistemology may be added an observation from the history of science, that convergence of different lines of evidence upon a theoretical entity has constituted one of the most powerful modes of theory appraisal. The classic example, emphasized by philosopher Wesley Salmon (1984, pp. 213-227; Nye, 1972; and see discussion in Meehl, 1990a, pp. 116-121; 1990b, pp. 39-42), is the clinching of the reality of molecules by the agreement of 13 independent ways of estimating Avogadro's Number (number of molecules in a mole), which all came out with an order of magnitude 10^{23}. Those unfamiliar with history of the developed sciences may ask, if consistency tests are so important, why do not chemists, astronomers, even physiologists, talk about them? The answer is that the concept is taken so much for granted in the advanced sciences, no special terminology is needed for it!

[13] That MAMBAC estimates of the "base rate" may "agree," i.e., they tend to be .50, when the latent structure is nontaxonic does not weaken this argument; if the initial MAMBAC curves are nontaxonic, one has no reason to attempt to estimate base rates, and the weakness of this consistency test arises only when $\hat{P} \simeq .50$.

procedures, e.g., MAMBAC, MAXCOV, to partially or wholly disjoint sets of indicators should lead to the same value of P within tolerance. These three approaches all examine the agreement via different epistemic paths to the numerical estimate of an inferred latent quantity such as base rate, separation of means, regression slope, nuisance correlation, standard deviation of the latent distributions, or whatever. A fourth kind of consistency test is more complicated and consists of deriving a theorem (from the postulated latent taxonic structure) in which two or more different numerical values, e.g., a manifest correlation or mean difference, and some inferred latent value, e.g., a slope, base rate, or hit rate, are related (see, e.g., Golden & Meehl, 1973; Meehl, 1973a, p. 215, 1979; Meehl & Golden, 1982).

It should be clear from this discussion that coherent cut kinetics relies on no single index for a conclusion of taxonicity or dimensionality. Rather, multiple procedures are used as consistency tests for one another, and, if the data set is large enough, some procedures (MAMBAC being one of them) may be fruitfully used on subsets selected by the initial application. For instance, if a large data set looked taxonic with a small base rate after applying MAMBAC, the researcher might construct a new, smaller sample composed of the taxon members (identified by using MAMBAC on the entire data set) plus an equal number drawn randomly from cases initially assigned to the complement group. Applying MAMBAC to this new subset, the researcher would expect to get MAMBAC curves that look taxonic, base-rate estimates close to .50, separations and estimates of complement and taxon means close to those found with the complete data set, and the same assignment of individuals to the complement and taxon membership. A researcher might be left in doubt by a single procedure but should never be misled by the coherent cut kinetics *method*.

As we said at the beginning, MAMBAC has been presented here by itself; its use with consistency tests (and as such a test for other procedures) is planned for a subsequent publication. But we may briefly illustrate the importance and usefulness of this underlying coherence of the formalism by considering the 19 pairs of MAMBAC curves (out of a total of 900 pairs) that were incorrectly sorted by visual inspection (see Appendix D, pp. 1152-1156). In a research situation, data analysis would not stop with simple visual inspection of the curves. Some pertinent additional sources of information are: MAMBAC curves generated by other variables from the same samples, base-rate estimates, kurtosis of the distributions of the variables individually, and, because we have available more than two variables in these Monte Carlo samples, the MAXCOV procedure.

Because these Monte Carlo samples have four variables, there are six pairs of MAMBAC curves for each sample. In some instances two pairs of curves from a given sample were missorted, but pairs of curves involving

other variables were sorted correctly and in a direction consistent with base-rate estimates, kurtosis for individual variables, and MAXCOV graphs.

We can consider the base-rate estimates from the variables that generated the curves in conjunction with the missorted MAMBAC curves. If base-rate estimates are near .50, the MAMBAC curves should be high (humped) in the middle. The primary mistake of the sorters (who did not know the base-rate estimates when they looked at the curves) was to confuse nontaxonic curves (which sometimes are higher on the right end due to sampling fluctuations) with taxonic curves from samples with low base rates (which may rise to a cusp on the right end). A base-rate estimate near .50 and MAMBAC curves with a dish shape are conjointly evidence for nontaxonicity. A sorter considering the MAMBAC curves and the base rates together should not have been misled.

When we use another of the cut kinetics procedures as a consistency test, MAXCOV curves using the variables in the missorted MAMBAC curves are clearly either taxonic or nontaxonic for all 19 of the missorted cases. Although a given pair of these MAMBAC curves by themselves might have left a researcher in doubt about the latent structure, we can now be confident in every one of the cases.

This is as good a place as any to clarify an important point about the over-all approach in the formalism. Mathematical expressions have been treated throughout as if they were true values rather than estimates, except for a few places where the usual statistician's "hat" (circumflex) has been used. Nowhere have we asked whether a certain statistical estimator as obtained in a taxometric sample is a maximum likelihood estimator, or similar questions customary in inferential statistics, e.g., sufficient? efficient? minimum variance bound estimator? (cf. Barnett, 1973). There are several reasons for this, which we will simply state briefly without arguing for them. The reader may choose to disagree with us but at least will understand our frame of reference. First, the inferred values in taxometrics are quite complicated functions of the observations, and, not being mathematical statisticians or experts in distribution theory, we are simply incapable of answering some of those conventional Fisherian questions, e.g., "is this estimator an MLE?" Second, the whole idea of getting a best single estimator of a statistical parameter becomes fuzzed up when we are talking about latent quantities as numerical attributes of hypothetical constructs, a problem Fisher did not face in agronomy; there all of the measures are operationally defined observables (such as pounds of fertilizer or yield in bushels of wheat), so that the *source* of error is random sampling fluctuation. Third, Fisher explicitly stated (1947, pp. 435-436, 1951, p. 54; cf. discussion in Johnstone, 1987) that his methods presuppose the physical process of randomization, which means that one either employs a randomizing procedure in determining which individuals in a sample are subjected to such-and-such a treatment or, in nonexperimental

contexts, specifies an actual physical population from which a sample is then randomly drawn. Neither of these conditions is met in the overwhelming majority of research domains where taxometric analysis is appropriate. If the theoretical structure is idealized sufficiently to make the formalism of conventional inferential statistics tractable, one knows for certain that this idealization is literally false and consequently faces the problem of robustness in the model. The robustness can hardly be determined analytically, since the reason for the idealization was that the "real" physical situation is mathematically intractable, either literally (as in the three-body problem in physics) or because we just are not smart enough to do it. But if the mathematics describing the real situation is intractable (which is why we idealized in the first place, to get going with the formalism), then obviously the discrepancy between the two, which is what robustness is about, will be intractable *a fortiori*. If we cannot work the mathematics for the real world, then we surely cannot work the mathematics for the robustness, which would compare the intractable formalism of the real world with the tractable formalism for the idealization! That is why so many taxometric methods in numerical taxonomy, such as the cluster algorithms, are far more often investigated by Monte Carlo methods than with rigorous analytical derivations. More mathematically competent readers may wish to look into the conventional statistician's questions about some of the procedures in the coherent cut kinetics method, and we certainly do not discourage such efforts. Meanwhile, we rely on Monte Carlo runs to answer the important questions concerning bias, random fluctuation, and tolerances.

REFERENCES

BARNETT, V. (1973) *Comparative statistical inference*. New York: Wiley.

CAMPBELL, D. T., & FISKE, D. W. (1959) Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

CLEVELAND, W. S. (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74, 829-836.

CRONBACH, L. J., & MEEHL, P. E. (1955) Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302. [Reprinted in P. E. Meehl, *Psychodiagnosis: selected papers*. Minneapolis, MN: Univer. of Minnesota Press, 1973. Pp. 3-31.]

CURETON, E. E. (1957) Recipe for a cookbook. *Psychological Bulletin*, 54, 494-497.

FISHER, R. A. (1947) Development of the theory of experimental design. *Proceedings of the International Statistical Conferences*, 3, 434-439.

FISHER, R. A. (1951) Statistics. In A. E. Heath (Ed.), *Scientific thought in the twentieth century*. London: Watts. Pp. 31-55.

FISHMAN, G. S., & MOORE, L. R. (1982) A statistical evaluation of multiplicative congruential random number generators with modulus $2^{31} - 1$. *Journal of the American Statistical Association*, 77, 129-136.

GOLDEN, R., & MEEHL, P. E. (1973) *Detecting latent clinical taxa, V: A Monte Carlo study of the maximum covariance method and associated consistency tests*. (Report No. PR-73-3) Minneapolis, MN: Univer. of Minnesota, Research Laboratories of the Department of Psychiatry. [Pp. 15-16, called the "Maximum Mean Difference" consistency test, shape of expected distribution described]

GOLDEN, R., TYAN, S., & MEEHL, P. E. (1974) *Detecting latent clinical taxa, VI: Analytical development and empirical trials of the consistency hurdles theory*. (Report No. PR-74-4)

Minneapolis, MN: Univer. of Minnesota, Research Laboratories of the Department of Psychiatry. [Used as a consistency hurdles test to cull items, not on continuous variables. First examples of curve drawn. P. 4, expected curve shape described, maximum near the hitmax cut ". . . Analytical demonstration of this fact has unfortunately proved elusive so far," i.e., the quasi-proofs of Meehl (1968) and Golden and Meehl (1973) are invalid.]

JOHNSTONE, D. J. (1987) Tests of significance following R. A. Fisher. *British Journal for the Philosophy of Science*, 38, 481-499.

KENNEDY, W. J., JR., & GENTLE, J. E. (1980) *Statistical computing*. New York: Marcel Dekker.

LANYON, R. I. (1968) *A handbook of MMPI group profiles*. Minneapolis, MN: Univer. of Minnesota Press.

LOEVINGER, J. (1957) Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635-694. (Monogr. Suppl. V3)

MEEHL, P. E. (1956) Wanted—a good cookbook. *American Psychologist*, 11, 263-272. [Reprinted in P. E. Meehl, *Psychodiagnosis: selected papers*. Minneapolis, MN: Univer. of Minnesota Press, 1973. Pp. 63-80.]

MEEHL, P. E. (1962) Schizotaxia, schizotypy, schizophrenia. *American Psychologist*, 17, 827-838. [Reprinted in P. E. Meehl, *Psychodiagnosis: selected papers*. Minneapolis, MN: Univer. of Minnesota Press, 1973. Pp. 135-155.]

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

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, MN: Univer. of Minnesota, Research Laboratories of the Department of Psychiatry. [Pp. 11ff, the procedure was first described as an alternative procedure for locating the hitmax cut.]

MEEHL, P. E. (1973a) MAXCOV-HITMAX: a taxonomic search method for loose genetic syndromes. In P. E. Meehl, *Psychodiagnosis: selected papers*. Minneapolis, MN: Univer. of Minnesota Press. Pp. 200-224.

MEEHL, P. E. (1973b) *Psychodiagnosis: selected papers*. Minneapolis, MN: Univer. of Minnesota Press.

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

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

MEEHL, P. E. (1990b) *Corroboration and verisimilitude: against Lakatos' "sheer leap of faith."* (Working Paper, MCPS-90-01) Minneapolis, MN: Univer. of Minnesota, Center for Philosophy of Science.

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

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

MEEHL, P. E. (1990e) 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*. Hillsdale, NJ: Erlbaum, 1991. Pp. 13-59.]

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

MEEHL, P. E., & ROSEN, A. (1955) Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52, 194-216. [Reprinted in P. E. Meehl, *Psychodiagnosis: selected papers*. Minneapolis, MN: Univer. of Minnesota Press, 1973. Pp. 32-62.]

MORRISON, D. E., & HENKEL, R. E. (Eds.) (1970) *The significance test controversy*. Chicago, IL: Aldine.

NYE, M. J. (1972) *Molecular reality*. London: Macdonald.

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

STATISTICAL SCIENCES. (1992) *S-Plus for DOS* (Version 2.0). Seattle, WA: Statistical Sciences, Inc.

TUKEY, J. W. (1977) *Exploratory data analysis*. Reading, MA: Addison-Wesley.

WALLER, N. G., PUTNAM, F. W., & CARLSON, E. B. (submitted) Types of dissociation and dissociative types: a taxometric analysis of dissociative experiences.

Accepted April 6, 1994.