Journal of Abnormal Psychology
2003, Vol. 112, No. 3, 457–468

0021-843X/03/$12.00 DOI: 10.1037/0021-843X.112.3.457

Finding the "Genuine" Schizotype: A Model and Method for Resolving Heterogeneity in Performance on Laboratory Measures in Experimental Psychopathology Research

Mark F. Lenzenweger, Shane T. Jensen, and Donald B. Rubin
Harvard University

Heterogeneity in the performance of persons affected with schizophrenia or schizotypic psychopathology on various laboratory tasks has long been recognized, both for its consistency across tasks and studies and for the massive methodological and substantive challenges it poses for experimental psychopathology, genetic, and other investigations. Traditional multivariate techniques, such as factor analysis, discriminant function analysis, and cluster analysis, have all been deemed inadequate for resolving heterogeneity, because of one or another statistical limitation. Here, an objective statistical approach based on a formal statistical model that uses the ubiquitous and well-developed expectation-maximization (EM) algorithm (A. P. Dempster, N. M. Laird, & D. B. Rubin, 1977) is presented, which enables one effectively to partition a group of experimental subjects, in this case identified initially using the well-known Perceptual Aberration Scale (L. J. Chapman, J. P. Chapman, & M. L. Raulin, 1978), in a manner that reduces heterogeneity and allows for the separation of what are termed *genuine* and *false-positive* schizotypes. The validity of the parsing strategy was supported by reference to other laboratory indexes of relevance to schizophrenia and schizotypy that were not included in the initial EM-based analyses. The potential utility of this approach is discussed with reference to future schizophrenia and schizotypy research.

Schizophrenia has long been known to be characterized by considerable heterogeneity in its clinical presentation (Bleuler, 1911/1950; Kraepelin, 1919/1971); indeed, Bleuler (1911/1950) referred to the "group of schizophrenias," and this heterogeneity has been a major source of frustration for schizophrenia researchers. It is well known that heterogeneity exists in all aspects of the illness, including symptoms, cognitive and behavioral dysfunctions, age and type of onset of illness, longitudinal course, and long-term outcome. Attempts to resolve this heterogeneity have usually taken the form of clinical subtyping approaches or multidimensional conceptualizations that seek to identify either homogeneous subgroups of patients or homogeneous dimensions of phenomenology or other characteristics (cf. Andreasen, Arndt, Alliger, Miller, & Flaum, 1995; Neale & Oltmanns, 1980). For example, the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994), like other systems before it, allows for the clinical subtypes (i.e., disorganized, paranoid, catatonic, and undifferentiated) even though these subtypes are well known to be arbitrary and unstable over time within individuals (Andreasen et al., 1995). An empirical approach to the heterogeneity of schizophrenia phenomenology, based on confirmatory factor analytic results, organized the symptoms into four major dimensions: negative, disorganized, reality distortion, and premorbid social functioning (Lenzenweger & Dworkin, 1996). Nonetheless, echoing the trenchant view of Gottesman (1987), aspiring cartographers of the heterogeneous phenomenological terrain of schizophrenia have been frustrated by this aspect of the disease for nearly a century.

Recognition of the Challenge Posed by Heterogeneity of Performance

Experimental psychopathology laboratory investigations of schizophrenia have similarly encountered the effects of heterogeneity (Maher, 2003), and this reality has forced many researchers to focus on the development of statistical or methodological strategies for dealing with the heterogeneity (e.g., L. J. Chapman & Chapman, 1977, 1989). For example, although there are some well-established laboratory findings about the illness (e.g., eye-tracking dysfunction; Levy, Holzman, Matthysse, & Mendell, 1993; and deficits in sustained attention; Cornblatt & Keilp, 1994), not all schizophrenia patients show deviance on any single given index, and more important, most schizophrenia cases do not display more than one of these deficits.

Although heterogeneity has long been observed and even taken as an assumed reality characteristic of the symptoms of schizophrenia and the laboratory task performance patterns of those affected with the illness, there is also an underlying presumption (one might say, theoretical hope) that there must be some core illness pattern or set of performance features that would uniquely demarcate those persons who genuinely have the illness from those with other psychiatric illnesses, or demarcate those who carry the

Mark F. Lenzenweger, Department of Psychology, Harvard University; Shane T. Jensen and Donald B. Rubin, Department of Statistics, Harvard University.

We thank Michael J. Coleman, Philip S. Holzman, Deborah L. Levy, Gillian O'Driscoll, and Sohee Park for their collaborative efforts with Mark F. Lenzenweger on the original studies that provided portions of the data for this illustration.

Correspondence concerning this article should be addressed to Mark F. Lenzenweger, who is now at the Department of Psychology, State University of New York at Binghamton, Science IV (Room G-08), Binghamton, New York 13902-6000. E-mail: mlenzen@binghamton.edu

liability for schizophrenia from those who do not (assuming either a polygenic [or oligogenic] system with a threshold effect or a single major locus operating against a background of polygenic effects; Gottesman, 1987; Korfine & Lenzenweger, 1995). The effect of heterogeneity on any investigation of schizophrenia has long been known to increase noise and, thereby, obscure signal in nearly every study.

Our understanding of schizophrenia and its manifestations is further complicated by the fact that the illness picture is blurred by additional complicating *third-variable* factors such as medication, institutionalization, deterioration, and stigma of diagnosis effects. One way to deal with this issue has been to study those persons at risk for the illness who are currently unaffected with schizophrenia. The at-risk individuals do not present the complicating features noted above. There are three approaches to defining risk in schizophrenia research: (a) genetic risk (e.g., the study of biological offspring, the study of first-degree biological relatives), (b) clinical definition of schizotypes (e.g., *DSM–IV* schizotypal personality disorder), and (c) laboratory definition of schizotypes (e.g., defined in terms of deviance on a reliable and valid indicator). However, even though the study of at-risk populations avoids the complications of the third variables, heterogeneity still exists in the laboratory performance of groups known to be at risk for the illness (Lenzenweger, 1998).

Prior Approaches to the Problem of Heterogeneity and Their Inadequacy

Traditional statistical approaches to the reduction of heterogeneity have been unable to address this vexing issue in schizophrenia research. For example, factor analytic procedures help to reduce large numbers of variables to a smaller set of factors, and in doing so, variables are reorganized, *not* persons. Discriminant function analysis seeks linear combinations of variables that efficiently separate diagnostic groups, and profile analysis seeks to determine if groups differ in their configuration of performance on variables or measures of interest. However, both discriminant function and profile analysis presume that group membership of the subjects is already known, and these techniques seek merely to sharpen the distinction between known (i.e., preexisting) groups, using variables of interest (but, most important, the subjects themselves are not reassigned or sorted in any way to new groups with either technique when applied initially to a data set), or they may be used to predict group membership for future unassigned subjects. In short, none of these commonly known methods can adequately separate out individuals who might represent the schizophrenia signal from those who represent noise within an at-risk-for-schizophrenia group, nor do they take into account adequately the heterogeneity in performance of normal subjects. Although traditional cluster analysis seeks to classify subjects using quantitative (i.e., continuous) data into meaningful subclasses and can be thought of as an aid in the investigation of latent classes, the marked limitations of cluster analysis for even simple classification tasks (e.g., male vs. female) are known (see Golden & Meehl, 1980) and limit enthusiasm for the technique in studies of heterogeneity. Another severe limitation of cluster analysis it that the method lacks a formal statistical procedure for the determination of the proper (or correct) number of classes underlying a multivariate space.

Proposition of a New Model and Method for Resolving Heterogeneity

Prior statistical and data-analytic approaches to the problem of heterogeneity of performance in laboratory measures are beset with shortcomings, and they are inherently insufficient as methods for this problem. Therefore, one must ask, what sort of model and method would one want to more effectively address the problem of heterogeneity in laboratory performance measures? What should it be able to do? How should it work? What aspects of data should it use to accomplish its goals? Ideally, one would want an approach that (a) was statistically well principled, (b) was not inordinately dependent on excessively large sample sizes (as many psychopathology studies have small sample sizes), and (c) would allow one to use performance data from a known normal group to assist in the subdivision of a hypothetically pathological group into meaningful subgroups. We argue that such a method should take into account the conditional probability structure of selected laboratory measures known to exist within a normal group and use such information to sift through the observed data on the same measures in the pathological group. In doing so, the method should be able to identify those cases (i.e., individuals) within the pathological group who are most likely misclassified as putatively pathological. Such an identification process would be empirically enabled as each individual within the putatively pathological group would be assigned a *posterior* probability (meaning given observed data), derived using established Bayesian principles, as to their likely status as a genuine instance of the pathological group (i.e., a true-positive case) as opposed to an incorrect instance of classification to the pathological group (i.e., a false-positive case). Such a model and method should not merely identify more severely affected cases within the putative pathological group vis-à-vis the normal group, as existing strategies for identification of severity of impairment are generally adequate as a basis for parsing a sample. Moreover, the method should be automatic in the sense that from a specified model, it can be applied to data without adjustment or tuning by the investigator.

Therefore, we undertook the present study with a large, multivariate database gathered from individuals who were identified initially as putative schizotypes or nonschizotypes and used a statistical approach for mixture models, based on the expectation-maximization (EM) algorithm, designed to sift through the laboratory task performance patterns of these study subjects, guided by the probability structure of the performance of the normal subjects, to reduce heterogeneity among the schizotypes. In short, this study sought to segregate, in an almost automatic manner, those schizotypic subjects who are most likely to be genuine schizotypes from those who might best be thought of as false-positive cases.[1]

[1] We readily acknowledge that there could be a difference of opinion as to the use of terms such as *true* or *genuine* schizotypes. We have used *genuine* schizotype in a conceptual manner to convey the essence of what we are trying to accomplish with this exercise. Given that many investigators call those persons selected on the basis of high scores on a schizotypy indicator *putative* schizotypes, we wanted to avoid potential confusion (and awkwardness of expression) by describing those persons we term *genuine* schizotypes as *putative putative* schizotypes. We anticipate the reader will distill the meaning and intent of our use of the term *genuine* schizotype.

Method

Subjects

Subjects for the present study were drawn from a randomly ascertained sample of 1st-year undergraduates at a university in the Northeast, who voluntarily completed a 250-item psychological inventory titled Attitudes, Feelings, and Experiences Questionnaire, which included the Perceptual Aberration Scale (PAS; L. J. Chapman et al., 1978; see below for detail). This approach was used to maximize diversity within the pool of potential study subjects, as well as to minimize the effects of both subject self-selection factors and group-related test-taking attitudes often found in introductory-psychology-course-based sampling procedures.

Two thousand individuals were initially selected at random from a university roster of all 1st-year students who entered during a fall semester (approximately 3,000 students per year). Of the 2,000 potential subjects, 1,684 (51.3% women, 48.7% men) completed the inventory. The response rate for the screening was 84.2%.

To detect pseudorandom responding and invalid test-taking attitudes among those screened, a 14-item version of Jackson's (1984) Infrequency scale from his Personality Research Form was included in the 250-item screening inventory. Subjects scoring greater than 3 on the Infrequency scale were dropped from the sample; 35 (2.1%) were excluded from our sample on this basis. Three additional subjects were dropped because of extensive missing data on the inventory. The final sample consisted of 1,646 cases, from which two subject groups were composed for the experimental assessments described below. Separate group means and standard deviations for men and women on the PAS were computed and served as the basis for subject selection. Following L. J. Chapman and Chapman (1985), high-PAS subjects were required to have scored at least 2.0 standard deviations above the group mean on the PAS, whereas normal controls were required to have scored no higher than 0.5 standard deviations above the group mean. Study subjects for each of the two groups were selected at random from the two subsamples of subjects meeting the specified criteria. Subsequent testing was carried out while unaware of group membership. Twenty-six (14 female and 12 male) normal control subjects and 31 (16 female and 15 male) high-PAS subjects were tested. The proportions of male and female subjects across the two groups did not differ significantly, $\chi^2(1, N = 57) = 0.028$, $p = .87$. The mean ages of the high- and low-PAS subjects were 19.00 years ($SD = 0.52$) and 18.96 years ($SD = 0.53$), respectively. The mean PAS scores of the high- and low-PAS subjects were 19.00 ($SD = 6.35$) and 0.77 ($SD = 0.99$), respectively. There were no significant differences between the two groups in terms of agreement to participate in the study described below.

Although the individuals contained in the pool of 1,646 potential study subjects had been preselected initially for academic achievement (i.e., university admission), academic ability does not preclude risk for psychopathology (Lenzenweger, 1999; Lenzenweger, Loranger, Korfine, & Neff, 1997). The population from which the sample was drawn was most probably somewhat biased against particularly early-onset and clinically expressed variants of severe psychopathology. However, one would not necessarily anticipate any diminution in the prevalence of schizophrenia-spectrum-related liability in the undergraduate population studied.

Measures

Screening Measures

PAS. The PAS is a 35-item true–false self-report measure of disturbances and distortions in perceptions of body image as well as of other objects (L. J. Chapman et al., 1978), and its construction was inspired, in part, from Meehl's model and compendium of schizotypic signs (Meehl, 1962, 1964). Multiple converging lines of criterial evidence show that the PAS is a valid, although imperfect, psychometric indicator of some aspects of schizotypy (cf. Cronbach & Meehl, 1955), and extensive literature reviews bearing on the reliability and validity of the PAS as schizotypy (or, perhaps more broadly, psychosis-proneness) measures can be found elsewhere (J. P. Chapman, Chapman, & Kwapil, 1995; Lenzenweger, 1998). In short, the PAS is known to be significantly associated with (a) increased risk for schizophrenia in the first-degree biological relatives (but not risk for unipolar or bipolar illness), (b) poor Wisconsin Card Sorting Test (WCST) performance, (c) increased thought disorder, (d) Minnesota Multiphasic Personality Inventory schizophrenia-related deviance (Hathaway & McKinley, 1983), (e) increased schizotypal personality disorder symptomatology, (f) impaired smooth-pursuit eye movements, (g) poor antisaccade performance, (h) poor delayed response task (DRT) performance, and other schizophrenia-relevant processes (see Lenzenweger, 1998, for extensive review). At this time, no other psychometric measure of schizotypy is associated with such a multidimensional profile of validating evidence.

Psychosis screening. All subjects completed the self-administered computerized screening version of the Diagnostic Interview Schedule (DISSI; Robins, Helzer, Croughan, & Ratcliffe, 1981) to assess lifetime presence of a schizophrenia–schizophreniform psychosis. The DISSI screening assessment was done using a computerized, self-administered version of the DISSI. Subjects completed the DISSI alone at a computer workstation. No subject met DISSI screening criteria for a suspected prior schizophrenia–schizophreniform psychosis.

Laboratory Performance Measures

Each of the laboratory measures that generated the performance indexes for the present demonstration has been described in previously published articles, and therefore, extensive detail regarding the tasks' properties and the associated procedures for their administration have been omitted.

Measures Used in Expectation-Maximization-Based Classification Approach

Wisconsin Card Sorting Test. A computerized WCST was administered according to the standard guidelines specified in the WCST manual (Heaton, 1981) and scored using a computerized version of the test (Harris, 1988). The WCST is the well-known neuropsychological task that measures abstraction ability and cognitive flexibility and is commonly hypothesized to be associated with functioning of the dorsolateral prefrontal cortex (see Goldman-Rakic, 1991), although the specificity of this relationship is debated (see Wagman & Wagman, 1992). For this analysis, we used the WCST index known as the *failure to maintain set* (FMS), which assesses loss of the correct sorting principle needed to perform the WCST properly (Harris, 1988) and has been shown to discriminate schizotypes from nonschizotypes (Lenzenweger & Korfine, 1994; Park, Holzman, & Lenzenweger, 1995).

Eye-movement measurement. As described in detail in O'Driscoll, Lenzenweger, and Holzman (1998), oculomotor recordings for eight cycles were obtained in a darkened room, using an Optokinetograph (OKG) system. Eye-movement-pursuit-performance quality was evaluated independently by two expert raters (P. S. Holzman & G. A. O'Driscoll), unaware of subject group membership. The mean of the two raters, which was highly reliable (intraclass $r = .93$), served as the basic performance index for the eye-tracking performance (see O'Driscoll et al., 1998, for additional detail) and is denoted as our eye-tracking dysfunction (ETD) index below. ETD scores are recoded in the analyses to correspond to higher scores indicating poorer performance.

Measures Used in Validation Phase of This Demonstration

Continuous Performance Test. Sustained attention was measured using the well-known Continuous Performance Test—Identical Pairs Version (CPT–IP; Cornblatt, Lenzenweger, & Erlenmeyer-Kimling, 1989; Cornblatt, Risch, Faris, Friedman, & Erlenmeyer-Kimling, 1988). The CPT–IP is a high momentary processing load, low a priori signal probability

attention task that taps effortful or controlled information processing. The CPT–IP has been described in great technical detail elsewhere (Cornblatt, Lenzenweger, & Erlenmeyer-Kimling, 1989; Cornblatt, Risch, et al., 1988; Lenzenweger, Cornblatt, & Putnick, 1991). The performance measure derived from the CPT–IP for use in this particular investigation is the average reaction time (in milliseconds) of subjects recorded in connection with their correct target detections (i.e., hits; see Lenzenweger, 2001, for additional detail).

Thought Disorder Index. Thought disorder ratings were made from speech samples using the thought disorder scoring approach developed by Holzman and colleagues (Holzman et al., 1986; Johnston & Holzman, 1979), which is considered reliable (Coleman et al., 1993). The TDI manual contains complete descriptions of the thought disorder categories that are rated as well as the psychometric characteristics of the validation studies (Holzman et al., 1986; Johnston & Holzman, 1979). This thought disorder rating system produces a total Thought Disorder Index (TDI) score, among other scores; the higher the TDI score, the higher the level of formal thought disorder in the subject's performance. (See Coleman, Levy, Lenzenweger, & Holzman, 1996, for additional detail.)

Oculomotor delayed response task. The DRT used in this study, which involved an oculomotor delayed response that assessed spatial working memory, was that originally described by Park et al. (1995). Extensive detail regarding the properties of the task and its administration can be found in Park et al. The primary performance index for this DRT was the percentage of correct performance on those trials that required working memory. The percentage correct was therefore an accuracy score in which higher scores indicated better working memory performance.

Recruitment and Testing Procedures

Potential study participants were contacted by telephone and were invited to participate voluntarily in a study of young adult development for which they would receive a $50 honorarium. A complex coding scheme was used to disguise the group status of the subjects, and all study staff were unaware of a subject's group membership throughout the study. Subjects were tested in a quiet darkened room; all gave informed consent. The order of the administration of the study measures (i.e., laboratory measures, questionnaires, and other measures) was randomized across subjects.

Statistical Analysis

In this section, we describe our statistical approach to the problem at hand. This section is necessarily technical in some respects; however, we have sought to convey the logic and concepts behind our method to highlight the utility of this statistical approach.

We begin by noting that each subject can be represented by membership in a particular cell of a two-way contingency table with dimensions defined by their possible outcomes on the FMS index (Categories 1, 2, . . . , 6) and ETD (Categories 1, 2, . . . , 8, which correspond to possible mean ETD values) tasks. We have separate two-dimensional tables of cell counts (i.e., number of subjects in each cell) for classified normals and classified schizotypes, and the location of each individual is dictated by their performance on FMS and ETD tasks. The sample size for this analysis was 46, rather than the original 57 subjects, because complete data were required for both the FMS and ETD indices.

Several assumptions were made about these observed counts (i.e., distribution of subjects in the two-way table):

1. The classification was correct for all individuals initially classified as normal.

2. The classification was potentially incorrect for individuals initially classified as schizotypes, that is, some of those subjects classified as schizotypes were false positives.

3. Normal individuals misclassified as schizotypes would have the same model on their observed FMS and ETD measures as normal individuals who were correctly classified, that is, the false positives mentioned above would perform similarly to individuals classified as true normals on the FMS and ETD tasks.

4. The FMS and ETD measures were assumed to be independent in true normal individuals, whereas no independence between FMS and ETD measures was assumed for genuine schizotypic subjects (the actual correlation between the FMS and ETD scores in the classified normals was only .13 and reasonably consistent with this assumption). We are unaware of any published empirical data that would suggest that this assumption is unfounded or unreasonable. In fact, the FMS and ETD variables were selected a priori for the reasonableness of this assumption.

In schizotypy research, the ideal situation would be one where we knew that our selection measure (i.e., the PAS in this instance) was functioning perfectly and only would select genuine schizotypes. However, simply put, it was not known which individuals within the group originally classified as schizotypic on the PAS were genuine schizotypes and which were actually normal (but plainly misclassified by the fallible PAS). Thus, our assumptions implied that the classified schizotype group was actually a mixture of genuine schizotypes and true normals, whereas all individuals in the classified normal group were true normals. We note that the notion that a group of individuals selected on the basis of a fallible measure of schizotypy consists of a mixture of true-positive and false-positive schizotypes is well known (cf. Lenzenweger & Moldin, 1990), and the presence of this admixture is a psychometric or classification reality, not merely an assumption of our statistical approach. However, this mixture complicates the estimation of the unknown parameters of interest, namely, the separate cell probabilities for the distinct true normal and genuine schizotype tables, and the proportion of genuine schizotypes within the classified schizotype group. In this analysis, we focus on maximum likelihood estimation of these parameters of interest. (For a detailed technical analysis of this problem from the Bayesian perspective, see Jensen, Lenzenweger, & Rubin, 2002).

If we were somehow omniscient and knew which schizotypes were false positives, we could just subtract the counts corresponding to these false positives from the classified schizotype table and add them into the classified normal table, thereby creating both a true normal table and a genuine schizotype table. The proportion of genuine schizotypes within the classified schizotype group would then be a known quantity, and maximum likelihood estimation of the true cell probabilities would be trivial (and technically uninteresting). Then, the maximum likelihood estimates of the true normal probabilities for each cell would be simply the products of the row and column proportions (because of the assumed independence), whereas maximum likelihood estimates of the genuine schizotype probabilities for each cell would be the cell counts divided by the number of genuine schizotypes (or simple proportions in the cells).

Of course, we do not know which schizotypes are false positives. How can one proceed if one does not know which of the putative PAS-selected schizotypes is a false positive? We propose that this situation can be viewed as a missing data problem, with the missing data being an indicator variable for each subject in the classified schizotype group that indicates whether that subject is a genuine schizotype. By the first assumption, we do not need an unknown indicator variable for any subject classified as normal, because they are all assumed to be true normals. This assumption is reasonable given the location in the distribution of PAS scores from which the normals were selected initially.

How does one then go about estimating the missing indicator variable given what is known about the subjects? This is a challenging problem that cannot be addressed satisfactorily using conventional procedures (e.g., factor analysis, cluster analysis). One approach to this problem is to use the well-known and powerful iterative EM algorithm (Dempster et al., 1977), which can be used to calculate maximum likelihood estimates in the

presence of missing data. Although perhaps not well known among psychologists, the EM algorithm has had a major impact in statistical theory and has transformed many approaches to statistical and computational problems. The EM algorithm is, in fact, used somewhat invisibly by various statistical software programs that are often used by psychologists (e.g., hierarchical linear modeling, missing data procedures). In this study, we have used the fundamental EM algorithm and implemented it in the service of the problem at hand, that is, trying to determine which subjects in the schizotype pool are genuine schizotypes via the estimation of the so-called indicator variable.

Briefly, the EM algorithm is iterative, where each iteration consists of two statistical steps: the expectation (E) step and the maximization (M) step. In our problem, in the E step, the expectation of each missing value is calculated, conditional on the observed data and initial values of the parameters. In the M step, the expectations calculated in the E step are substituted for the missing values, so that the data no longer has any missing components. With this complete data, new maximum likelihood estimates of the parameters are calculated. This completes the first iteration. These new values of the parameters can then be used in the E step of the next iteration, which results in new expected values of the missing data that are then used in the next M step, and so on. Each step of EM strictly increases the observed-data likelihood, and if EM converges, it is to a maximal value. Most important, Dempster et al. (1977) demonstrated that the EM algorithm is guaranteed (essentially) to converge to the maximum likelihood estimate (i.e., conditional only on the observed data). This is a highly attractive feature of the EM method. Further details regarding the EM algorithm are rather technical in nature and beyond the scope of this report (for extensive technical detail, see Dempster et al., 1977; Little & Rubin, 2002; McLachlan & Krishnan, 1997).

In the present situation, implementation of the E step is really quite straightforward.[2] The expectation of an indicator variable of genuine schizotypy for a classified schizotype is simply the probability of that subject being a genuine schizotype, conditional on (a) his or her cell in the classified schizotype table and (b) current maximum likelihood estimates of the parameters (from the previous M step). (For those readers interested in the technical aspects of these probability calculations, those details are given in the appendix.)

The data are then completed by filling in the missing indicator variables with their expectations from the E step. For example, say that Subject A has a probability of .7 of being a genuine schizotype, given the cell of Subject A in the estimated genuine schizotype table and the current parameter values. This means that .7 of Subject A's count is placed in the genuine schizotype table, whereas .3 of Subject A's count is placed in the true normal table. Similarly, all other subjects in the putative schizotype table are split between the genuine schizotype and true normal table on the basis of their probability of being a genuine schizotype, as calculated in the E step.

For our problem, the M step is straightforward because maximum likelihood estimates are easy to calculate when the true normal and genuine schizotype counts are known. As mentioned previously, the maximum likelihood estimates of the true normal probabilities for each cell are simply the products of the marginal probabilities in the true normal table, whereas maximum likelihood estimates of the genuine schizotype probabilities for each cell are the true schizotype cell counts divided by the number of genuine schizotypes. The maximum likelihood estimate of the proportion of genuine schizotypes within the classified schizotype group is simply the total of all counts in the genuine schizotype table divided by all counts in the classified schizotype table. The EM procedure then cycles through the E and M steps many times to refine the estimates.

Finally, a word about the assumptions of this statistical method. All of the assumptions that are required for this approach have been presented above, and no additional assumptions are required. The approach we have taken, and implemented using the EM algorithm, *does not* assume multivariate normality. It is not a technique that is restricted to use with extremely large sample sizes, as the stability of the estimations produced by the method can be assessed using a validation simulation (as we have done in this study as described below). Finally, note also that this particular method shares essentially nothing with the taxometric analysis procedures developed by Meehl (e.g., Waller & Meehl, 1998).

Validation of the Subgroup Membership: Contrast Analyses

The EM analysis generated an estimated posterior probability (i.e., the probability of the indicator variable being positive) for each of those initially classified as schizotypes, with a high probability suggesting that a subject was likely to be a genuine schizotype rather than a false-positive schizotype. We used these estimated posterior probabilities to partition those individuals initially identified as schizotypes, by virtue of their initial PAS scores, into either the genuine or false-positive schizotype groups. We then sought to evaluate the validity of this parsing of the schizotypic subjects on the basis of the EM analysis by using laboratory indices that were not used in the original EM computations.

We conducted two sets of contrast analyses with analysis of variance (ANOVA; Rosenthal, Rosnow, & Rubin, 2000), each of which was guided by theoretical considerations, using other laboratory task performance data that were not included in the classification analysis (which used only FMS and ETD data). These other laboratory task data were derived from assessments of sustained attention, spatial working memory, and thought disorder (measures described above).

The first contrast, which represented the strongest expression of our theoretical model, held the genuine schizotypes apart from the other two subject groups, which were constrained not to differ (contrast weights: genuine schizotypes = 2, false-positive schizotypes = –1, normal controls = –1). The second contrast, which was orthogonal to the first contrast, held aside the genuine schizotypes and contrasted the performance of the false-positive schizotypes and the normal control subjects (contrast weights: genuine schizotypes = 0, false-positive schizotypes = –1, normal controls = 1).

For the first contrast analysis, our prediction was that the genuine schizotypes would show poorer performance than the combined false-positive schizotype and normal subject groups, thereby providing support for the parsing produced by the EM-based procedure. The second contrast tested whether the performance of the false-positive schizotypes differed from that of the normal controls. If the performance of the false-positive schizotypes did not differ significantly from that of the normal controls, this pattern of results would provide additional support for the parsing produced by the EM-based procedure (i.e., false-positive schizotypes are similar to normals in their performance).

The specific dependent variables for the contrast analyses were (a) mean reaction time for hits (i.e., correct detections) on the CPT–IP, (b) total number of thought disorder responses scored according to the TDI procedure, and (c) percentage correct performance on a spatial delayed response task. Given the clear-cut directionality of our first contrast, the *p* values associated with the contrast *t* tests (equal-variance *t* tests) for that set of variables are one-tailed; whereas the *p* values for the second contrast *t* tests (equal-variance *t* tests) are two-tailed, given the absence of a directional hypothesis. Effect-size estimates for the contrast are reported as r_{contrast} following Rosenthal et al. (2000).

Results

The true normal and true schizotype cell probabilities calculated with the EM algorithm as described above are given in Table 1, along with the empirical probabilities from the tables for those

[2] The EM-algorithm-based approach used in this study was implemented using code written by Shane T. Jensen. Inquiries regarding the code may be sent to jensen@fas.harvard.edu.

Table 1
Values From the Expectation-Maximization Algorithm (FMS × ETD)

	Eye-tracking dysfunction															
	Normal classification								Schizotype classification							
FMS	−3	−2.75	−2.5	−2.25	−2	−1.75	−1.5	−1	−3	−2.75	−2.5	−2.25	−2	−1.75	−1.5	−1
							Initial proportion									
0	.381	.048			.048			.095	.080	.080		.040	.040	.040		.080
1	.143	.048			.048		.048	.048	.080		.080		.080	.040		.120
2		.048			.048								.040			.040
3														.040		
4															.040	.040
6											.040					
							Final proportion									
0	.214	.071			.095		.014	.126				.105		.105		
1	.158	.053			.070		.011	.093			.210			.105		
2	.039	.013			.017		.003	.023					.039			.017
3														.105		
4															.105	.105
6											.105					

Note. ETD (eye-tracking dysfunction) values are reversed from the original scoring so that higher scores (less negative) are indicative of poorer smooth pursuit performance. FMS (failure to maintain set on the WCST) values are coded so that higher scores are indicative of poorer WCST performance. WCST = Wisconsin Card Sorting Test.

initially classified as normal and as schizotypes. The remaining parameter, the proportion of true schizotypes (designated as λ in our statistical notation in the appendix and in Figure 1) within the classified schizotype group, was estimated to be .38.[3]

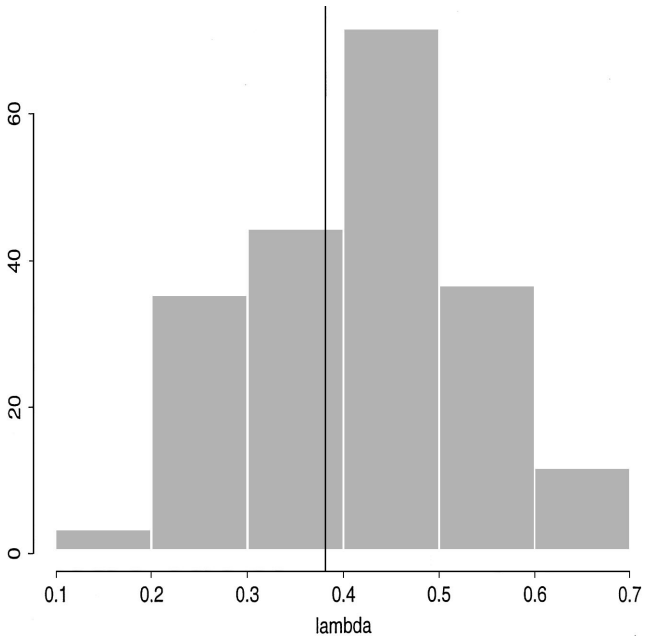

Figure 1. Distribution of lambda values generated by the validation simulation procedure using several hundred generated datasets, as detailed in the text. The *x*-axis covers the range of the estimated values of lambda, and the *y*-axis is the observed counts in each bin of the lambda values. The black vertical line represents the true value of lambda derived from the analysis of the actual data for the study subjects.

The next step in this procedure was to address the central question of this study: Who is a genuine schizotype versus a false-positive schizotype (i.e., someone misclassified by the PAS)? In general, the posterior probabilities calculated for each subject in the classified schizotype group were near either 0 or 1, so that it was an easy and unambiguous operation to reclassify the 25 putative schizotypes (i.e., on the basis of the initial PAS-based classification) into 10 genuine schizotypes and 15 false-positive schizotypes.[4] The means for the three subject groups for the FMS and ETD were as follows: For normals, FMS = 0.52, *SD* = 0.68; ETD = 2.46, *SD* = 0.76; for false-positive schizotypes, FMS = 0.60, *SD* = 0.63; ETD = 1.97, *SD* = 0.89; for genuine schizotypes, FMS = 2.20, *SD* = 1.99; ETD = 1.95, *SD* = 0.50.

Given that our sample size was somewhat small, we undertook a rigorous but straightforward approach to evaluating the

[3] Clearly, this analysis required a check for the possible existence of multiple modes in the likelihood. To check for the possibility that the likelihood had multiple modes, we started the EM algorithm with several other well-dispersed initial parameter values. The EM algorithm consistently converged to the same final values, suggesting that the likelihood was indeed unimodal.

[4] We estimated the probability of each putative schizotype representing a genuine schizotype, conditional on their performance measures and our best estimates of the unknown parameters. This probability is known as a posterior probability, which is relatively easy to estimate for each of the putative schizotypic subjects by the probability calculation in the final E step of the EM algorithm given above (and detailed in the appendix), on the basis of maximum likelihood estimates of the unknown parameters (i.e., their values at the last M step). If this probability is reasonably unambiguous for a subject (i.e., near either 0 or 1), then the person can be reclassified as either a genuine schizotype (probability near 1) or a false-positive (probability near 0).

validity of our EM-based estimations. We performed a simple validation simulation to ensure that the EM algorithm used above could recover the truth. Three-hundred new data sets of 46 individuals each were drawn from multinomial distributions with parameters values given by the maximum likelihood probability estimates in Table 1, as well as the maximum likelihood estimate of the proportion of true schizotypes ($\lambda = .38$). The EM algorithm was then applied to each generated data set to see if the algorithm would converge back appropriately to the true parameter values. The distributions of the final values from each EM analysis, as obtained from each of the 300 generated data sets, were found. We have presented in Figure 1 the results of the validation simulation for the parameter lambda. In Figure 1, one sees the distribution of the 300 lambda values.

The *x*-axis in Figure 1 covers the range of the estimated values of lambda, and the *y*-axis is the observed counts in each bin of the lambda values. The black vertical line represents the true value of lambda derived from the analysis of the actual data for the study subjects. Thus, what one sees in this figure is the actual true parameter value for lambda obtained in our analysis of the actual data surrounded by a distribution of 300 estimates of lambda based on the 300 generated data sets. In general, the EM algorithm seems to recover effectively the truth (i.e., the final parameter estimates are close to the underlying true parameter), and these results support the validity of the solution we obtained. Had we not been able to recover the truth in this instance, we would have had considerably less confidence in the solution for our actual data set. We conducted analogous analyses for all other parameters. The results for these other parameters, which are not reported here but are available on request, were similar in pattern to those reported for lambda, the proportion of genuine schizotypes, in Figure 1. In short, the results of the validation simulation provide important support for the ability of this approach to work well in the context of a small sample.

Validation Analyses Using External Criteria

One might imagine that this reclassification of the schizotypic subjects did nothing more than partition schizotypic subjects into two new groups that had no differential relations with external criteria of validity for schizophrenia-related deviance. If that were the case, then this EM-based exercise would be less than illuminating. However, if the new classes containing the newly assigned schizotypes did, in fact, bear some meaningful relation to external validity criteria, then this reclassification approach would represent an increment in our ability to parse heterogeneity in laboratory index performance. To assess the relations of the new group memberships with external criteria of validity, we conducted the two sets of contrast analyses described above. The means and standard deviations for the CPT–IP reaction time, thought disorder, and delayed response task performance variables can be found in Table 2. The two sets of contrast analyses were performed for each of these dependent variables using the classification results from the initial EM estimations using FMS and ETD. The first contrast reflected the most conservative (i.e., strongest theoretical statement) hypothesis that genuine schizotypes performed more poorly than either of the other two groups, which were hypothesized not to differ (contrast weights for genuine schizotypes, false-positive schizotypes, and normal controls were 2, –1, and –1, respectively). The results were as follows: CPT–IP reaction time (in milliseconds), $t(43) = 1.83$, $p < .035$, $r_{contrast} = .27$; sum of thought disordered responses, $t(42) = 1.72$, $p < .045$, $r_{contrast} = .26$; delayed response task performance $t(43) = 2.13$, $p < .02$, $r_{contrast} = .31$. Our hypothesis reflected in this contrast was clearly directional, and therefore, we report one-tailed *p* values for the contrast *t* tests. We note in this context a supplementary series of *t*-test comparisons of the genuine schizotypes and the normal controls, holding aside the false-positive schizotypes, reveal substantial (i.e., large effect sizes) and

Table 2
Reaction Time Thought Disorder and Delayed Response Task Performance in Three Participant Groups Using the Two-Variable Expectation-Maximization Model Classifications

	Subject groups					
	Normal controls ($n = 21$)		False szt ($n = 15$)		Genuine szt ($n = 10$)	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
CPT–IP reaction time (hits)	543.53	53.69	572.48	60.18	595.44	59.19
Delayed response task	93.26	5.40	90.85	7.95	86.65	8.70
Total thought disorder	2.57	3.17	5.79	8.27	9.50	15.01

Note. The reaction time index is in milliseconds. Delayed response task values represent percentage of correct performance. Total thought disorder is the total number of thought-disordered responses for an individual as scored by the Thought Disorder Index (Holzman, Shenton, & Solovay, 1986). Contrast analyses for these data can be found in the text. The sample size for the false schizotypes for total thought disorder is 14 owing to one missing case. The means and standard deviations for the combined schizotype group (false and genuine schizotypic participants) were as follows: for CPT–IP reaction time, $M = 581.66$, $SD = 59.65$; for the delayed response task, $M = 89.17$, $SD = 8.34$; and for Total thought disorder, $M = 7.33$, $SD = 11.41$. Szt = schizotype; CPT–IP = continuous performance test—Identical Pairs Version (Lenzenweger, 2001).

statistically significant differences among the groups in the predicted direction.[5]

The second set of contrasts, which was orthogonal to the first set of contrasts, held aside the genuine schizotypes and tested whether the performance of the false-positive schizotypes differed significantly from the normal controls (contrast weights for genuine schizotypes, false-positive schizotypes, and normal controls were 0, –1, and 1, respectively). The hypothesis reflected in this contrast was nondirectional, and therefore, we report two-tailed *p* values for the contrast *t* tests. The results were as follows: CPT–IP reaction time (in milliseconds), $t(43) = 1.50$, $p = .14$, $r_{contrast} = .23$; sum of thought-disordered responses, $t(42) = 1.08$, $p = .29$, $r_{contrast} = .16$; and delayed response task performance, $t(43) = 1.01$, $p = .32$, $r_{contrast} = .15$ (see Footnote 5).

Comparison of Expectation-Maximization Subdivision of Schizotypes With Other Commonly Used Approaches

Given that our approach involves a complex statistical method, one might wonder if our EM-based reclassifications of the schizotypic subjects into false-positive and genuine schizotypes might have been achieved with more commonly available statistical techniques or simpler approaches. Therefore, we used both cluster-analytic and median-split approaches to address this issue using the same variables that were used in the EM-based computations, namely FMS and ETD. A central issue for this set of analyses concerned whether these more commonly used methods would essentially subdivide the study subjects in the same manner as that achieved with the EM-based approach. A *K*-means cluster solution that retained three clusters (resulting cluster *n*s = 1, 12, and 33, respectively) did not remotely recover the same classification of the study subjects as was achieved with the EM-based approach; rather, it misclassified a relatively large number of normal subjects as well as schizotypic subjects when compared with the EM-based memberships. For the three-cluster solution, the breakdown of the subjects in terms of their initial PAS-based classification was as follows: for Group 1 ($n = 1$), 1 putative schizotype; for Group 2 ($n = 33$), 17 normals, 16 putative schizotypes; for Group 3 ($n = 12$) 4 normals, 8 putative schizotypes. A *K*-means cluster analysis of ETD and FMS that was constrained to two clusters failed to recover (or approximate) even the basic high- versus low-PAS status of the subjects, combining 42 of the 46 subjects (the 42 subjects included 21 normals and 21 putative schizotypes; the second group of 4 subjects included only putative schizotypes). One might consider how the *K*-means-clustering-assigned group memberships were related to the validation indexes, namely working memory (DRT), sustained attention (CPT–IP hit rate reaction time; CPT–IP RT), and thought disorder (TDI). The heterogeneous grouping of normal control and putative schizotype subjects produced by the three-cluster *K*-means solution, as well as the fact that one of the clusters contained but a single subject, to our minds, precluded a serious consideration of this solution. However, performance of the two clusters (Group 2 [$n = 33$] vs. Group 3 [$n = 12$]) on the validation measures was compared using *t* tests, and the two groups did not differ significantly on any of the three measures (DRT, $p < .17$, CPT–IP RT, $p < .47$, and TDI, $p < .60$). For the two-cluster *K*-means solution, which yielded two groups (Group 1 [$n = 42$], Group 2 [$n = 4$]), the two groups differed on only one of the validation variables (DRT), $t(44) = 2.03$, $p < .05$, but not on the other measures (CPT–IP RT, thought disorder). Note that we are not advocating the use of cluster analysis as an analytic approach to reducing heterogeneity. We have used cluster analysis here only because it is one of the techniques that some psychopathologists might elect to use to classify subjects. We note in this context that cluster analysis is typically used with sample sizes considerably larger than those in this study.

The median-split approach is also a commonly used approach for the classification of subjects based on performance measures. It is one in which the median values for the ETD and FMS variables among the normal subjects were used to subdivide the two subject groups (controls and putative schizotypes), and it generated four new subject classifications: good smooth-pursuit eye tracking/low FMS, poor smooth-pursuit eye tracking/low FMS, good smooth-pursuit eye tracking/high FMS, and poor smooth-pursuit eye tracking/high FMS. This new redistribution of subjects, based on the median-split classification, when crossed with the EM-based classification yielded the following results: (a) good smooth-pursuit eye tracking/low FMS (2 false-positive schizotypes, 0 genuine schizotypes, 8 controls); (b) poor smooth-pursuit eye tracking/low FMS (5 false-positive schizotypes, 2 genuine schizotypes, 4 controls); (c) good smooth-pursuit eye tracking/high FMS (2 false-positive schizotypes, 0 genuine schizotypes, 3 controls); and (d) poor smooth-pursuit eye tracking/high FMS (6 false-positive schizotypes, 8 genuine schizotypes, 6 controls). In short, the distribution of subjects achieved in this simpler median-split approach did not resemble (or correspond to) the EM-based analysis results in any compelling fashion. Analysis of the performance of the subjects on the validation measures could also be examined from the standpoint of the median-split approach to subdividing the subjects. This would take the form of an ANOVA with two between-subjects factors, namely, median-split-based grouping (four levels) and initial PAS-based grouping (two levels), which would yield a 2 × 4 ANOVA design. If the median-split strategy provided enhanced the resolving power with respect to efficient parsing of the subjects, then a substantial and significant Median-Split Group × Initial PAS Classification interaction would reveal itself. We did not observe significant interactions for any of the three validation measures: For CPT–IP RT, $F(3, 46) = 2.63$, *ns*; for DRT, $F(3, 46) = 2.41$, *ns*; for TDI, $F(3, 45) = 0.18$, *ns*. Moreover, inspection of the cell means for each of the three validation measures in each of the ANOVAs did not

[5] Although we have a preference for focused contrasts in ANOVAs (see Rosenthal et al., 2000), at the request of a reviewer, we also performed *t* tests on the reaction time (CPT), working memory (DRT), and thought disorder (TDI) indexes for two important comparisons: (a) genuine schizotypes versus controls and (b) genuine schizotypes versus false-positive schizotypes. In these additional *t* test analyses, we focused on the direction of predicted effects and effect sizes rather than *p* values (given the small sample sizes in this exercise). The genuine schizotypes differed substantially (average Cohen's $d = .91$) from the controls on the three validity indicators of CPT, DRT, and TDI (these differences were all significant at $p < .025$). The genuine schizotypes differed from the false-positive schizotypes in the predicted direction on all three validity measures with appreciable effect sizes (average Cohen's $d = .40$), although the associated *p* values were nonsignificant owing to the small sample sizes. Note that our first contrast analysis reported in the text tested the genuine schizotypes versus the combined pool of false-positive schizotypes and normal controls. The complete details of this supplementary set of *t* test analyses may be requested from Mark F. Lenzenweger.

reveal a pattern that could be readily interpreted with confidence (e.g., the slowest RTs on the CPT–IP were found among schizotypes in the good smooth-pursuit eye tracking/low-FMS cell). Finally, cell sizes began to shrink dramatically with the application of the median-split strategy owing to the necessary resultant increase of cells, a feature of the approach that directly affects power.[6] We are not taking a formal position on the relative merit of the median-split approach to data analysis; however, it clearly did not provide a clearer picture than was provided by our EM-based analysis. Note also in this context that the median values obtained from the normals that are used to parse subjects into subgroups will necessarily be influenced by the sampling approach (and any related artifacts, e.g., sample size) used to collect data from the normal subjects, which will in turn influence (perhaps not optimally) the redistribution of subjects in any application of this approach.

One might also conjecture that the EM-based solution merely organized the schizotypes in terms of their severity of impairment on the FMS and ETD (smooth-pursuit eye tracking) indexes and that the most deviant performers on these two measures were also the most deviant performers on the TDI, DRT, and CPT–IP. We tested this conjecture directly by constructing a *z* score-based deviance index for the FMS and ETD variables, as well as one for the TDI, CPT, and DRT variables. In short, those subjects who were most deviant on the FMS–ETD index (i.e., 2 standard deviations above the mean) were not the same subjects who were deviant on the TDI–CPT–DRT *z* score index. In fact, only 2 of the 5 subjects designated as deviant on the FMS–ETD index were found among the 6 subjects designated as deviant on the TDI–CPT–DRT index. Clearly, the EM approach we describe is not merely detecting severity of impairment on the laboratory measures.

When considered together, these results led us to think that the EM-based method we propose here generates results that are not easily obtained with other more common procedures, and these alternative approaches (particularly the cluster-analysis approach) did not appear to generate more valid classifications. Note, however, that we do not view the EM-based classifications as a gold standard against which the other approaches must be evaluated, but rather the EM-based approach yields a classification of subjects that could not be resolved with the other commonly available procedures.[7]

Did the Expectation-Maximization Approach Merely Subdivide Subjects in Terms of Severity of Schizotypic Features?

It is important to raise an additional question here regarding our EM-based parsing, namely, did we merely ended up parsing the schizotypic group in terms of severity on the initial selection index (i.e., the PAS) or correlated schizotypic phenotypic features? If we just ended up subdividing the schizotypic subjects on the basis of their initial PAS severity (deviance), then one might question the utility of our approach. To test this possibility, we contrasted the false-positive and genuine schizotypic groups on their initial PAS values and found them to not differ significantly (genuine schizotypes: $M = 18.10$, $SD = 3.57$ vs. false-positive schizotypes: $M = 20.07$, $SD = 7.69$), $t(23) = 0.75$, $p = .46$. Thus, the genuine schizotypes were not more severely affected in terms of the initial selection index, the PAS. Further, we contrasted the two schizotype groups for the number of schizotypal personality disorder criteria met on the self-report International Personality Disorder Examination—Screen (Lenzenweger et al., 1997) and found them not to differ significantly (genuine schizotypes: $M = 5.10$, $SD = 1.45$ vs. false-positive schizotypes: $M = 5.33$, $SD = 1.50$), $t(23) = 0.39$, $p = .70$. Thus, the two schizotype groups would not appear different in terms of their clinical symptom–feature presentation. These supplementary data clearly suggest that our EM-based approach did not merely represent a sophisticated statistical technique used to divide the schizotypic group that indirectly reflected simply a difference in the severity of phenotypic features, as measured on the PAS, or the number of schizotypal features. Rather, our approach enabled us to go into the sample of schizotypes and objectively select a subgroup of subjects who possessed significantly poorer correlated performance across a number of laboratory measures independent among normals, even though their phenotypic (symptomatic) presentation did not demarcate them from the others.

Discussion

The primary purpose of this study was to apply a statistical method, which uses the EM algorithm, to empirical data derived from laboratory measures completed by the same subjects, to resolve observed heterogeneity in performance on those measures among the schizotypic subjects. The overall results of the study indicated that the group originally designated as schizotypic on the basis of initial deviance on the PAS consisted of an admixture of genuine and false-positive schizotypes and, importantly, the schizotypic sample could be meaningfully divided to reveal this situation. Note, however, under our independence and dependence assumptions, that our statistical approach did not invent the admixture of genuine and false-positive schizotypes within the original schizotypic group. Rather, the admixture resided in the original data, and we were able to discover it with our approach. Specifically, for the two-variable estimations (using the FMS variable from the WCST as well as a measure of smooth-pursuit eye-tracking performance), the EM procedure led to substantial differences between starting values for individual schizotypic probabilities and the final convergent values for schizotypic probabilities, with the proportions in several cells of the schizotype table disappearing toward 0 as a result of some putative schizotypic individuals moving to be classified in the normal table. This is not a surprising result, because an anticipated product of this analysis was that the schizotype count table might change dramatically as a result of our fundamental assumption that some of the classified schizotype counts really belonged in the normal table.

When we extended our model to include a third variable, the TDI, its addition did not seem to substantially change the majority

[6] The means and standard deviations for the cells in the ANOVAs based on the median-split approach were omitted to conserve space, but they are available on written request to Mark F. Lenzenweger. Additionally, the unfocused ANOVAs reported here for the median-split strategy were requested by a reviewer. We did not conceptualize a set of a priori planned contrasts for this supplementary analysis.

[7] Extended detail regarding these alternative classification approaches and the results of the analyses obtained is available on written request to Mark F. Lenzenweger. This material was omitted from the present article to conserve space.

of the point estimates of the cell probabilities. In the context of our model, we note that there is a cost to adding additional variables because, although supplying additional information about the subjects, extra variables also place additional stress on the important assumption that the performance on each task is independent in normal individuals. The trade-off between adding information via additional variables and the potential violation of the independence assumption is an unstudied question that merits additional statistical research.

As noted above, each of the putative schizotypic subjects in the original schizotypic sample was eventually assigned a probability of being either a genuine schizotype or a false schizotype. To test the validity of this parsing, on the basis of the EM results, these two groups as well as the normal group were evaluated using three separate and independent biobehavioral laboratory measures known to be relevant to schizophrenia and schizotypic psychopathology, namely, CPT–IP RT, TDI, and DRT performance (i.e., spatial working memory). As predicted, our first set of contrast analyses using the three dependent variables (i.e., CPT–IP RT, TDI, and DRT performance) revealed that the genuine schizotypes performed significantly more poorly than either of the other two subject groups; that is, the genuine schizotypes displayed slower RTs on the CPT–IP, higher levels of thought disorder on the TDI, and worse performance on the spatial working memory task (DRT). The second set of contrast analyses revealed that the performance of the false-positive schizotypes did not differ significantly from the normal controls (holding aside the genuine schizotypes) on these dependent measures, although the effect directions indicated that the misclassified normals had somewhat worse performance than the pure normals. The overall pattern of the contrast analysis results (as well as the supplementary *t* test analyses reported in Footnote 3) suggested to us that our EM-based procedure performed quite well in helping to resolve heterogeneity of performance on the laboratory measures within the schizotypic group. The EM-based approach that we used did not constitute subject groups that could be easily recovered or discerned using the modal approaches used in psychopathology research for the reduction of heterogeneity in performance measures and the classification of subjects, namely, cluster-analytic, median-split, and standard score composite approaches. Moreover, note also that our approach did not merely reclassify the schizotypic subjects as a function of their initial level either on the PAS or on a measure of schizotypic features; thus, it was not just tapping into a severity dimension.

Overall, we offer this EM-based approach as an objective method for approaching the heterogeneity problem in schizophrenia and schizotypy research. In very practical terms, what this approach offers is a useful and statistically principled method of reducing heterogeneity in samples of subjects that have been preselected as at risk on a psychometric (or other) measure of liability. This method adds a refining step to the selection process and extends the ability of the investigator to sift through putatively at-risk cases to detect signal cases rather than noise cases. As such, we suggest that our approach can be viewed as a useful adjunct to the biobehavioral–psychometric high-risk paradigm, wherein risk for schizophrenia (or putative schizophrenia liability) is thought to be reflected merely in deviance on a laboratory index (e.g., the PAS). As noted previously, it has long been recognized that any sample selected on the basis of deviance on a putative schizophrenia liability marker (or any psychopathology risk measure) is likely to consist of an admixture of genuine schizotypes (i.e., truly at risk for schizophrenia and schizophrenia-related pathologies) and false-positive schizotypes (e.g., Holzman et al., 1995; Lenzenweger & Moldin, 1990).

In sum, we used an EM-based statistical procedure to laboratory data drawn from a study of putative schizotypes and normal subjects and were able to divide the schizotypic group into genuine and false-positive schizotypes. An approach such as the one we describe could serve to help diminish or resolve heterogeneity in laboratory or other indexes commonly used in schizotypy and schizophrenia research and thereby better separate signal from noise in the detection of genuine schizotypy. Such an advance would be of considerable use to other schizophrenia-related or other psychopathology studies, such as in neuropharmacologic and genetic research, that demand a well-defined phenotype that is as valid and refined as possible. More generally, for example, this approach might represent a potentially valuable advanced quantitative method for use in disssecting neurobehavioral profiles in personality disorders research (Depue & Lenzenweger, 2001), identifying treatment responders from large groups of individuals who have received some intervention (Pilkonis, 2001), or refining subject partitioning in cognitive neuroscience efforts (Barch et al., 2002).

References

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

Andreasen, N. C., Arndt, S., Alliger, R. M., Miller, D., & Flaum, M. (1995). Symptoms of schizophrenia: Methods, meanings, and mechanisms. *Archives of General Psychiatry, 52,* 341–351.

Barch, D. M., Carter, C. S., Braver, T. S., Sabb, F. W., MacDonald, A., Noll, D. C., & Cohen, J. D. (2002). Selective deficits in prefrontal cortex function in medication-naive patients with schizophrenia. *Archives of General Psychiatry, 58,* 280–288.

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

Chapman, J. P., Chapman, L. J., & Kwapil, T. R. (1995). Scales for the measurement of schizotypy. In A. Raine, T. Lencz, & S. Mednick (Eds.), *Schizotypal personality* (pp. 79–106). New York: Cambridge University Press.

Chapman, L. J., & Chapman, J. P. (1977). Selection of subjects in studies of schizophrenic cognition. *Journal of Abnormal Psychology, 86,* 10–15.

Chapman, L. J., & Chapman, J. P. (1985). Psychosis proneness. In M. Alpert (Ed.), *Controversies in schizophrenia: Changes and constancies* (pp. 157–172). New York: Guilford Press.

Chapman, L. J., & Chapman, J. P. (1989). Strategies for resolving the heterogeneity of schizophrenics and their relatives using cognitive measures. *Journal of Abnormal Psychology, 98,* 357–366.

Chapman, L. J., Chapman, J. P., & Raulin, M. L. (1978). Body-image aberration in schizophrenia. *Journal of Abnormal Psychology, 87,* 399–407.

Coleman, M. J., Carpenter, J. T., Waternaux, C., Levy, D. L., Shenton, M. E., Perry, J., et al. (1993). The thought disorder index: A reliability study. *Psychological Assessment, 5,* 336–342.

Coleman, M. J., Levy, D. L., Lenzenweger, M. F., & Holzman, P. S. (1996). Thought disorder, perceptual aberrations, and schizotypy. *Journal of Abnormal Psychology, 105,* 469–473.

Cornblatt, B. A., & Keilp, J. G. (1994). Impaired attention, genetics, and the pathophysiology of schizophrenia. *Schizophrenia Bulletin, 20,* 31–46.

Cornblatt, B. A., Lenzenweger, M. F., & Erlenmeyer-Kimling, L. (1989). The Continuous Performance Test, Identical Pairs Version: II. Contrasting attentional profiles in schizophrenic and depressed patients. *Psychiatry Research, 29,* 65–85.

Cornblatt, B. A., Risch, N. J., Faris, G., Friedman, D., & Erlenmeyer-Kimling, L. (1988). The Continuous Performance Test, Identical Pairs Version (CPT–IP): I. New findings about sustained attention in normal families. *Psychiatry Research, 26,* 223–238.

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

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39,* 1–38.

Depue, R. A., & Lenzenweger, M. F. (2001). A neurobehavioral dimensional model of personality disorders. In W. J. Livesley (Ed.), *The handbook of personality disorders* (pp. 136–176). New York: Guilford Press.

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

Goldman-Rakic, P. S. (1991). Prefrontal cortical dysfunction in schizophrenia: The relevance of working memory. In B. J. Carroll & J. E. Barrett (Eds.), *Psychopathology and the brain* (pp. 1–23). New York: Raven Press.

Gottesman, I. I. (1987). The psychotic hinterlands or the fringes of lunacy. *British Medical Bulletin, 43,* 557–569.

Harris, M. E. (1988). *Wisconsin Card Sorting Test* (Research ed.) [Computer software]. Odessa, FL: Psychological Assessment Resources.

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

Heaton, R. K. (1981). *Wisconsin Card Sorting Test manual.* Odessa, FL: Psychological Assessment Resources.

Holzman, P. S., Coleman, M., Lenzenweger, M. F., Levy, D. L., Matthysse, S., O'Driscoll, G., & Park, S. (1995). Working memory deficits, antisaccades, and thought disorder in relation to perceptual aberrations. In A. Raine, T. Lencz, & S. A. Mednick (Eds.), *Schizotypal personality* (pp. 353–381). New York: Cambridge University Press.

Holzman, P. S., Shenton, M. E., & Solovay, M. R. (1986). Quality of thought disorder in differential diagnosis. *Schizophrenia Bulletin, 12,* 360–372.

Jackson, D. N. (1984). *Manual for the Personality Research Form* (3rd ed.). Port Huron, MI: Research Psychologists Press.

Jensen, S. T., Lenzenweger, M. F., & Rubin, D. B. (2002). A Bayesian approach to reducing heterogeneity in laboratory performance measures: An illustration from schizophrenia research. In C. Gatsonis (Ed.), *Case studies in Bayesian statistics* (Vol. 6). New York: Springer-Verlag.

Johnston, M. H., & Holzman, P. S. (1979). *Assessing schizophrenic thinking.* San Francisco: Jossey-Bass.

Korfine, L., & Lenzenweger, M. F. (1995). The taxonicity of schizotypy: A replication. *Journal of Abnormal Psychology, 104,* 26–31.

Kraepelin, E. (1971). *Dementia praecox and paraphrenia* (R. M. Barclay, Trans. & G. M. Robertson, Ed.). Huntington, NY: Krieger. (Original work published 1909–1913)

Lenzenweger, M. F. (1998). Schizotypy and schizotypic psychopathology: Mapping an alternative expression of schizophrenia liability. In M. F. Lenzenweger & R. H. Dworkin (Eds.), *Origins and development of schizophrenia: Advances in experimental psychopathology* (pp. 93–121). Washington, DC: American Psychological Association.

Lenzenweger, M. F. (1999). Stability and change in personality disorder features: The longitudinal study of personality disorders. *Archives of General Psychiatry, 56,* 1009–1015.

Lenzenweger, M. F. (2001). Reaction time slowing during high load vigilance task performance in psychometric schizotypy. *Journal of Abnormal Psychology, 110,* 290–296.

Lenzenweger, M. F., Cornblatt, B. A., & Putnick, M. E. (1991). Schizotypy and sustained attention. *Journal of Abnormal Psychology, 100,* 84–89.

Lenzenweger, M. F., & Dworkin, R. H. (1996). The dimensions of schizophrenia phenomenology? Not one or not two, at least three, perhaps four. *British Journal of Psychiatry, 168,* 432–440.

Lenzenweger, M. F., & Korfine, L. (1994). Perceptual aberrations, schizotypy and the Wisconsin Card Sorting Test. *Schizophrenia Bulletin, 20,* 345–357.

Lenzenweger, M. F., Loranger, A. W., Korfine, L., & Neff, C. (1997). Detecting personality disorders in a nonclinical population: Application of a two-stage procedure for case identification. *Archives of General Psychiatry, 54,* 345–351.

Lenzenweger, M. F., & Moldin, S. O. (1990). Discerning the latent structure of hypothetical psychosis proneness through admixture analysis. *Psychiatry Research, 33,* 243–257.

Levy, D. L., Holzman, P. S., Matthysse, S., & Mendell, R. (1993). Eye tracking dysfunction and schizophrenia: A critical perspective. *Schizophrenia Bulletin, 19,* 461–536.

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed). New York: Wiley-Interscience.

Maher, B. A. (2003). Psychopathology and delusions: Reflections on methods and models. In M. F. Lenzenweger & J. M. Hooley (Eds.), *Principles of experimental psychopathology: Essays in honor of Brendan A. Maher* (pp. 9–28). Washington, DC: American Psychological Association.

McLachlan, G. J., & Krishnan, T. (1997). *The EM algorithm and extensions.* New York: Wiley-Interscience.

Meehl, P. E. (1962). Schizotaxia, schizotypy, schizophrenia. *American Psychologist, 17,* 827–838.

Meehl, P. E. (1964). *Manual for use with Checklist of Schizotypic Signs.* Minneapolis: University of Minnesota.

Neale, J. M., & Oltmanns, T. F. (1980). *Schizophrenia.* New York: Wiley.

O'Driscoll, G., Lenzenweger, M. F., & Holzman, P. S. (1998). Antisaccades and smooth pursuit eye tracking and schizotypy. *Archives of General Psychiatry, 55,* 837–843.

Park, S., Holzman, P. S., & Lenzenweger, M. F. (1995). Individual differences in working memory in relation to schizotypy. *Journal of Abnormal Psychology, 104,* 355–363.

Pilkonis, P. A. (2001). Treatment of personality disorders in association with symptom disorders. In W. J. Livesley (Ed.), *Handbook of personality disorders: Theory, research, and treatment* (pp. 541–554). New York: Guilford Press.

Robins, L. N., Helzer, J. E., Croughan, J., & Ratcliffe, K. S. (1981). National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. *Archives of General Psychiatry, 38,* 381–389.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach.* New York: Cambridge University Press.

Wagman, A., & Wagman, W. (1992). On the Wisconsin. In E. F. Walker, R. H. Dworkin, & B. A. Cornblatt (Eds.), *Progress in experimental personality and psychopathology research* (Vol. 15, pp. 162–182). New York: Springer.

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

Appendix

In the following, $i = 1, \ldots, 6$ indexes the failure to maintain set (FMS) values on the Wisconsin Cord Sorting Test and $j = 1, \ldots, 8$ indexes the eye-tracking dysfunction (ETD) categories. The given data included 21 individuals classified as normal and 25 individuals classified as schizotypic, so ϕ, the proportion of individuals classified as normal, is 21/46.

$$\text{Normal cell counts: } N_{ij} \sim \text{multinomial}(\pi_{ij}^{N}) \quad \text{Independence model on } \pi_{ij}^{N}$$

$$\text{Schizotypic cell counts: } S_{ij} \sim \text{multinomial}(\pi_{ij}^{S}) \quad \text{Saturated model on } \pi_{ij}^{S}$$

Now, if λ, the proportion of classified schizotypes that were true schizotypes, was known, then we would be able to augment our data with I_k, a Bernoulli variable for each classified schizotype that would indicate their true status ($I_k = 1$ if true schizotype, $I_k = 0$ if true normal).

$$I_k \sim \text{Bernoulli}(p),$$

where p is the probability the kth classified schizotype is truly schizotypic, given the location of that position in the schizotypic table. This p can be calculated simply by applying Bayes rule (FMS = A, ETD = B):

$$P(I_k = 1 | A = i, B = j) = \frac{P(A = i, B = j | I_k = 1) \cdot P(I_k = 1)}{P(A = i, B = j)}$$

$$= \frac{\pi_{ij}^{S} \cdot (1 - \phi)\lambda}{\pi_{ij}^{S} \cdot (1 - \phi)\lambda + \pi_{ij}^{N} \cdot [\phi + (1 - \phi)(1 - \lambda)]}.$$

Received May 16, 2001
Revision received January 24, 2003
Accepted January 27, 2003 ■