Taxometric Analyses of Sexual Orientation and Gender Identity

Steven W. Gangestad
University of New Mexico

J. Michael Bailey
Northwestern University

Nicholas G. Martin
Queensland Institute of Medical Research

Taxa are nonarbitrary classes whose existence is an empirical question and not a matter of mere semantic convenience. Taxometric procedures detect whether numerical relations between purported indicators of conjectured taxa bear the hallmarks of true taxa. On the basis of theoretical considerations, the current study tested whether taxa underlie sexual orientation and related measures of gender identity. Two taxometric procedures, maximum covariance, making hits maximum (MAXCOV) and mean above minus below a cut (MAMBAC), were applied to Kinsey Scales and measures of childhood gender nonconformity and adult gender identity in a sample of nearly 5,000 members of the Australian Twin Registry. Results suggest that latent taxa underlie these measures. About 12–15% of men and 5–10% of women belong to latent taxa associated with homosexual preference. These percentages are greater than those of individuals who report homosexual preference, however, and hence it appears that an appreciable proportion of individuals in these taxa have heterosexual preference. An understanding of the origins of these latent taxa may be important to understanding the development of sexual orientation and gender identity.

Are homosexuality and heterosexuality categorically different, or do they differ only as a matter of degree? Pioneer researcher Alfred Kinsey, who introduced the 7-point Kinsey Scale to measure gradations and found that men exist at every Kinsey level (Kinsey, Pomeroy, & Martin, 1948), argued the latter:

> Males do not represent two discrete populations, heterosexual and homosexual. The world is not to be divided into sheep and goats. Not all things are black nor all things white. It is fundamental to taxonomy that nature rarely deals with discrete categories. Only the human mind invents categories and tries to force facts into separated pigeon-holes. (Kinsey et al., 1948, p. 639)

Though Kinsey's observations (also seen for women; Kinsey, Pomeroy, Martin, & Gebhard, 1953) suggest that dimensional assessment of sexual orientation conveys more information than dichotomizing people into homosexual or heterosexual, they do not settle the issue of whether any dichotomy in the domain of sexual orientation reflects an artificial cut on a continuous distribution or a more fundamental difference. This is an issue of taxonomy.

Taxon or Dimension?

Meehl (1992) defined a taxon as "a nonarbitrary class whose existence is conjectured as an empirical question, not a mere semantic convenience" (p. 117). A domain containing taxa is *taxonic.* Examples include biological sex, biological species, some disease entities (e.g., measles), and some ideological systems in politics or religion (Meehl, 1992). Many taxa are characterized by their causal simplicity. Taxonic domains are more likely than dimensional ones to have *specific etiologies,* including dichotomous necessary causal factors (Meehl, 1977; Meehl, 1992). For example, infectious diseases are taxa, and their causes consist of specific microbes. (More complex causal processes, such as thresholds and polarization effects, may also underlie taxa [Meehl, 1977].) The existence of taxa can be supported either by the demonstration of requisite causal processes or by formal mathematical taxometric methods, which decide whether latent taxa underlie a set of candidate indicators of a conjectured taxon based on numerical relations between them. If so, the formal-numerical taxa that are thereby defined are empirical. Their causal basis must

---

Steven W. Gangestad, Department of Psychology, University of New Mexico; J. Michael Bailey, Department of Psychology, Northwestern University; Nicholas G. Martin, Epidemiology Unit, Queensland Institute of Medical Research, Brisbane, Queensland, Australia.

This research was supported by Grant MH47227 from the National Institute of Mental Health and by a small Commonwealth AIDS Research Grant (Australia).

Correspondence concerning this article should be addressed to Steven W. Gangestad, Department of Psychology, Logan Hall, University of New Mexico, Albuquerque, New Mexico 87131-1161. Electronic mail may be sent to sgangest@unm.edu.

Journal of Personality and Social Psychology, 2000, Vol. 78, No. 6, 1109–1121
Copyright 2000 by the American Psychological Association, Inc. 0022-3514/00/$5.00 DOI: 10.1037//0022-3514.78.6.1109

be discovered through additional research, and, thereby, taxometric findings can guide future inquiry into the causes of variation in the domain.

Waller and Meehl (1998) clarified several common misconceptions about taxa. First, bimodality—sometimes thought a hallmark of a categorical variable—is neither necessary nor sufficient to show latent taxonicity. An indicator must separate two taxa very strongly to be bimodal and, hence, is too stringent a criterion. Although mean height of the sexes differs by about two standard deviations, height is not clearly bimodal. But because bimodality can arise from scaling or sampling artifacts, it is also insufficient to show taxa (Grayson, 1987). Second, latent taxa do not imply a lack of true dimensions. Taxonicity is an *additional,* not alternative, feature of a domain. Biological sex underlies variation in height, but that does not mean that height is not a dimension or doesn't vary within the sexes. "Taxonic versus dimensional" should thus be "taxonic–dimensional versus dimensional only." Third and relatedly, the structure of a domain (taxonic or nontaxonic) should not be confused with the nature of the variables (qualitative or quantitative) used to detect that structure. "Types" may mistakenly be thought to imply differences in kind, not degree. But taxa may be detected with quantitative variables only.

Taxometric analyses are most appropriate when theory suggests a specific causal factor such as a dichotomous necessary factor or threshold effect. An example is Meehl's (1962) theory of social anhedonia as a component of schizotypy. He proposed that a genetically based integrative defect (schizotaxia) is the basis of a personality organization (schizotypy) predisposing schizophrenia. Although he conjectured that all individuals with schizotaxia develop schizotypy, only a small portion of these develop schizophrenia. This theory thus conjectured that the domain of schizotypy is taxonic, the base rate of the taxon in the population thought to be about 10%. Though originally proposing anhedonia as a key indicator of the schizotypy taxon, Meehl (1989, 1990) later speculated, "hypohedonia is one of a *dozen normal-range (nontaxonic) individual difference factors (dimensions)* that raise or lower the probability of decompensation" (Meehl, 1990, p. 24, emphasis in original). J. J. Blanchard, Gangestad, Brown, and Horan (2000) asked whether indicators of social anhedonia (subscales of the Social Anhedonia Scale; Chapman, Chapman, & Raulin, 1976) are taxonic. Taxometric analyses revealed that social anhedonia is in fact taxonic, the base rate of a group predisposed to social anhedonia being about 5–8% in a college population. These findings do not imply that social anhedonia is not a quantitative dimension; as noted above, taxa are not inconsistent with dimensionality. They do suggest, however, that some dichotomous factor affects the domain of social anhedonia, justifying and potentially helping to direct attempts to identify a critical factor accounting for "schizotypy." Evidence for taxa also exists for self-monitoring (Gangestad & Snyder, 1985; see also von Davier & Rost, 1997), Type A (Strube, 1989), schizotypy (more broadly; Tyrka et al., 1995), perceptual aberation (Korfine & Lenzenweger, 1995; Lenzenweger & Korfine, 1992), psychopathy (Harris, Rice, & Quinsey, 1994), and dissociative disorder (Waller, Putnam, & Carlson, 1996).

At least one theory suggests that sexual orientation is taxonic–dimensional rather than dimensional only. The *neurohormonal theory* (Ellis & Ames, 1987; see also LeVay, 1996) specifies that sexual orientation results from prenatal hormonal processes that also affect some other morphological and behavioral sex differences, including sex-typical gender orientation and play patterns. Specifically, the theory suggests that sex-atypical patterns of prenatal hormonal processes are one major reason for gay or lesbian sexual orientation, atypical gender identity in adulthood, and atypical childhood play patterns, resulting in their covariation (see Bailey & Zucker, 1995). Because canalization processes may shunt individuals' development of the neural bases of sexual orientation down one of two sex-typical tracks, hormonal processes may be taxonic in nature, the majority shunted down a sex-typical track, a minority shunted down a sex-atypical track. This theory does not deny continuous variation on a dimension of sexual orientation. Moreover, it does not say that the *only* pathway to gay or lesbian sexual orientation is via a neurohormonal route. Finally, it does not claim that *all* individuals with atypical prenatal hormonal patterns possess gay or lesbian sexual orientation. Rather, it suggests that *some* variation in sexual orientation and gender identity is associated with latent taxa. Other notions (such as any single gene theory of sexual orientation) also entail taxa of the domain.

In the only formal taxometric analysis of sexual orientation, Haslam (1997) performed commingling and taxometric analyses of the Minnesota Multiphasic Personality Inventory Masculinity–Femininity scale (MMPI-2 *Mf*), a correlate of sexual orientation, on 1,138 men in the restandardization sample of the MMPI-2. Haslam found no evidence that male sexual orientation is taxonic.

The Present Study

We examined whether latent taxa underlie measures of sexual orientation and two related domains, childhood gender nonconformity (CGN) and continuous gender identity (CGI), using a large sample of Australian twins. (Because our study was not concerned with distinguishing genetic and environmental determinants, participants' status as twins was ignored.) Our study differs from Haslam's (1997) in several ways. First, our sample included men and women, analyzed separately. Second, our measures included sexual orientation itself, as assessed by the Kinsey Scale. Finally, we used the taxometric procedure maximum covariance, making hits maximum (MAXCOV) as well as mean above minus below a cut (MAMBAC). MAXCOV, not used by Haslam (1997), is the taxometric procedure that has enjoyed the most widespread application to date (see Meehl, 1995a). Given many readers' unfamiliarity with these methods, we briefly discuss their logic.

Meehl's taxometric methods (referred to as "coherent cut kinetic" methods) are based in philosophy of science concerns that distinguish them from many other statistical procedures. Waller and Meehl (1998) argued that the *riskiness* of a theory's predictions has been heavily weighed in the history of science: Successful theories tend to predict observations that would, in the words of philosopher Wesley Salmon (1984), be "strange coincidences" were the theory false. Scientists wisely reject "strange coincidence" as an explanation. One way theories take risks is to make precise numerical predictions (a famous example is the prediction of light bending during the 1919 solar eclipse made by relativity theory; Salmon, 1984). Meehl has argued that one can render precise predictions from a claim that a latent taxon underlies a domain through *consistency tests.*

As Waller and Meehl (1998) explained, "in its broadest sense, a consistency test is an algorithmic procedure that operates on observed or inferred quantities to yield a derived numerical value that should, within tolerance, agree with a numerical value derived differently and not shown to be equal by virtue of mathematical identity" (p. 98). Consistency tests have been used within the history of science on numerous occasions, without being called such. For instance, in the early 1900s a variety of methods were used to estimate Avogadro's number, the number of molecules in a mole, based on atomic theory. The theory did not offer one a priori value. However, the theory states that, if atoms truly exist, the values estimated by mathematically independent methods should agree with one another. The fact that multiple estimates did agree led many physicists to accept atomic theory as more than just a "convenient fiction" (Salmon, 1984).

Within taxometrics, the two major questions presented by a data set are (a) Do latent taxa exist? and (b) If so, what are the relative frequencies (base rates) of the taxa? Meehl's methods offer multiple means of estimating the base rates of latent taxa. One form of consistency test, then, is simply *whether the latent base-rate estimates converge.* If latent taxa exist, they surely have base rates and independent estimates of the base rates should be consistent. If latent taxa do not exist, then the convergence of mathematically nonequivalent estimates of the base rates would be a "Salmonean coincidence" (Waller & Meehl, 1998); there is no reason to expect independent estimates to agree if no latent taxa even exist (moreover, Monte Carlo simulations show this to be true). As scientists generally reject coincidence as explanation, consistent base-rate estimates for latent taxa through independent methods constitutes *evidence that latent taxa do indeed exist.*

We applied two different taxometric methods. MAXCOV yields several different estimates of the base rate itself, and these estimates should be similar. MAMBAC also yields a base-rate estimate, which should agree with MAXCOV estimates. Monte Carlo simulations of data that possess taxonic–dimensional and dimensional-only latent structures provide evidence that base-rate consistency is suggestive of latent taxa.

Consistent base-rate estimates may not be sufficient to show taxonicity (Waller & Meehl, 1998). Other tests, however, can be examined. Within both MAXCOV and MAMBAC, specific function forms of derived relationships (see below on peaked covariance curves, U-shaped distribution of probability of class membership and humped MAMBAC curves) indicate taxonicity (as shown by Monte Carlo simulations). These function forms therefore can be used to test a taxonic model, particularly when examined across both procedures and coupled with base-rate estimation. Other consistency tests, when violated, suggest parameters such as base rates may not be reliably estimated. We used a variety of consistency tests in this work. In fact, we know of no realm in which both MAXCOV and MAMBAC procedures have been applied to the same data set. Typically, published taxometric analyses report just one method.

In Meehl's coherent cut kinetic methods, no significance test is used to reject a null hypothesis; nor is a goodness-of-fit statistic computed (see Waller & Meehl, 1998). Meehl (1992, 1995a) argued strongly that the methods are nonetheless more rigorous than many procedures that do involve significance testing or goodness-of-fit tests. First, he has argued that passing a series of consistency tests is a much greater hurdle than is rejection of a weak null hypothesis. Second, goodness-of-fit statistics used in structural modeling (and used in classification procedures such as latent class analysis and commingling analyses) require auxillary assumptions that may not be realistic (e.g., constrained distributions of variables or latent response tendencies within classes), leading to false negative results; some classification methods may yield results consistent with a taxonic model when there are no latent taxa. Latent class analysis can yield fit when purely dimensional models also fit (e.g., Reise & Gomel, 1995). Cluster analysis often does not effectively recover true taxa from real data (see Golden & Meehl, 1980). Again, requiring that consistency tests be passed may test a taxonic conjecture more rigorously than a nonsignificant goodness-of-fit test. (See Meehl, 1995a, for further discussion of philosophy of science issues.)

Method

Participants

Participants were drawn from the Australian National Health and Medical Research Council Twin Register (ATR), a volunteer register begun in 1978 with about 25,000 twin pairs of all zygosity types and all ages enrolled and in various stages of active contact. (For more details, see Bailey, Dunne, & Martin, 2000; Dunne et al., 1997.) In 1992 all ATR twins aged 18 to 50 years who had completed a postal "Health and Lifestyle" survey between 1988 and 1990 ($N = 9{,}112$) were asked whether they'd be willing to receive a questionnaire regarding sex. All who agreed were mailed it. They were asked to complete a consent form with their name, date of birth, and signature and to return this separately to indicate whether or not they had consented to complete the sex questionnaire. Because twins attached identification numbers to the questionnaires that only they and their twins knew, anonymity was assured. About two weeks after initial mailing of the questionnaire, all twins were sent a reminder letter.

Twenty-eight percent explicitly refused to participate, and 54% (4,901) completed questionnaires. The remaining 18% initially agreed but did not respond when contacted (following one letter or phone call). Our response rate was not substantially lower than that of other recent large-scale mail sex surveys (typically 55–65%; Biggar & Melbye, 1992; Johnson et al., 1989; Sundet, Magnus, Kvalem, Samuelsen, & Bakketeig, 1992).

Measures

Sexual orientation. There are several alternative measures of sexual orientation, including the Kinsey scale of sexual attraction and fantasy, the Kinsey scale of behavioral experience, and categorical sexual identity (i.e., "homosexual," "bisexual," "heterosexual"). We chose the former, because we believe that it is the variable least likely to be constrained by societal pressures and in this sense most fundamental. Furthermore, in contrast to categorical measures, the 7-point Kinsey Scale (Kinsey et al., 1948) allows variation in degree of homosexual versus heterosexual preference, ranging from 0 (*exclusively heterosexual*) to 6 (*exclusively homosexual*). Kinsey Fantasy and Attraction correlated .92 for men and .67 for women. To be consistent with the sexual orientation literature, we rounded the averaged Kinsey scores down to the nearest integer.

CGN. The male and female measures of CGN included items retrospectively assessing childhood sex-typed behavior (i.e., participation in sex-stereotypic games and activities) and gender identity (i.e., internal feelings of maleness or femaleness). *Childhood* was defined as before the age of 12. Our measures were adapted from several published scales by taking relevant items and, in some cases, rewriting them to be appropriate for Australian participants (e.g., "cricket" rather than "baseball"). For men, items were taken from the Gender Identity Scale for Males (Freund, Langevin, Satterberg, & Steiner, 1977), the Childhood Play Activities

Checklist (Grellert, Newcomb, & Bentler, 1982), the Recalled Childhood Gender Behaviors Questionnaire (Mitchell & Zucker, 1992), and the Physical Aggressiveness Scale (R. Blanchard, McConkey, Roper, & Steiner, 1983). For women, items were taken from the Childhood Play Activities Checklist, the Recalled Childhood Gender Behaviors Questionnaire, and the Masculine Gender Identity Scale (R. Blanchard & Freund, 1983). All of these scales differ reliably between homosexual and same-sex heterosexual individuals, as have all scales of similar content (Bailey & Zucker, 1995). Although retrospective measures raise concerns about recall biases and forgetting, a study found moderate-to-high agreement between gay men and their mothers regarding the men's CGN (Bailey, Nothnagel, & Wolfe, 1995).

Because different contents define male and female sex-atypicality, some items on the male and female versions differ. We know of no reason why this should affect results of taxometric analyses. Items varied in their response format and included both dichotomously rated items and rating scales. Scree tests of the principal components suggested that, for each sex, one general factor primarily accounted for the item intercorrelations. Items were standardized within sexes and summed to yield a total CGN score. Coefficient alpha was .79 for both male and female CGN.

CGI. This scale consisted of seven items taken from Finn (1987) and assesses self-concepts as masculine or feminine (e.g., "In many ways I feel more similar to women/men than to men/women") using 7-point rating scales. A subscale including these items distinguished homosexual from same-sex heterosexual individuals (Finn, 1987). Moreover, a study of gay and lesbian couples found that CGI scores moderately covaried with partners' ratings of respondents' masculinity–femininity (Bailey, Finkel, Blackwelder, & Bailey, 1996). Separate scree tests for each gender were both consistent with a single factor underlying CGI item intercorrelations. Items were summed to yield total scores, and coefficient alpha was .52 for men and .57 for women.[1]

Results

Descriptive Statistics

Table 1 gives the means and standard deviations for all relevant variables, and Table 2 gives the frequencies for Kinsey scores. About 3.6% of the men and 1.0% of the women endorsed Kinsey 5 (*primarily homosexual*) or 6 (*exclusively homosexual*), close to percentages of gay men and lesbian women estimated by recent large national sex surveys (see Diamond, 1993). Kinsey scores moderately correlated with CGN (r = .35 and .23 for men and women, respectively) and CGI (.22 and .30). The latter variables also significantly covaried (r = .34 and .47; $p < .00001$ for all rs).

All taxometric analyses were conducted on men and women separately for several reasons. First, sexual orientation distributions differ between the sexes. Second, the most influential etiological theory of sexual orientation hypothesizes that opposite processes cause male versus female homosexuality (Ellis & Ames, 1987; atypically small and large amounts of androgen action for men and women, respectively); male homosexuality is associated with childhood femininity and female homosexuality with childhood masculinity (Bailey & Zucker, 1995). Third, recent debate points to a number of developmental differences between homosexual and heterosexual men and women (Peplau, Garnets, Spalding, Conley, & Veniegas, 1998), suggesting that male and female results could differ.

Table 1
Descriptive Statistics for Kinsey Scores, Childhood Gender Nonconformity, and Continuous Gender Identity

	Men			Women		
Variable	*M*	*SD*	*N*	*M*	*SD*	*N*
Kinsey scores	0.37	1.19	1,759	0.34	0.83	2,747
Childhood gender nonconformity	−7.33	6.51	1,803	−7.25	6.74	3,004
Continuous gender identity	9.07	2.39	1,749	7.44	2.17	2,966

Table 2
Frequencies for Kinsey Scores

	Men		Women	
Frequency	*N*	%	*N*	%
0	1,502	82.4	2,142	69.5
1	136	7.5	451	14.6
2	31	1.7	75	2.4
3	12	0.7	33	1.1
4	12	0.7	16	0.5
5	20	1.1	15	0.5
6	46	2.5	15	0.5

The Maximum Covariance, Making Hits Maximum (MAXCOV-HITMAX) Procedures

General rationale. The taxometric method MAXCOV-HITMAX (also known simply as MAXCOV) was applied to the data (see Meehl, 1965, 1973, 1995a, 1995b; Meehl & Golden, 1982; Meehl & Yonce, 1996; Waller & Meehl, 1998). This procedure requires at least three quantitative indicators of a conjectured latent taxon. Our indicators were Kinsey scores, CGN, and CGI. One variable serves as an "input variable." It is dissected into successive intervals, which define subsamples. For instance, Kinsey scores range from 0 to 6. Each integer value in this range can define an interval of the scale, which a subsample scores. Within each subsample, the covariance between the other two variables is computed. Meehl (1973) showed that, if (a) each variable is, as conjectured, a valid indicator of a latent taxon and (b) the indicators do not covary within the taxon or its complement, then the covariances between the two indicators plotted as a function of the intervals on the input indicator will be "peaked": near zero in the extreme intervals and positive in less extreme intervals. The rationale for this expectation is straightforward. The covariance between two indicators, x and y, in a mixed distribution of a taxon and a complement is

---

[1] The variation underlying the CGI is actually quite distinct from that underlying measures of masculine and feminine personality or sex roles, for example, the Bem Sex Role Inventory (BSRI; Bem, 1974) or Personal Attributes Questionnaire (Spence, Helmreich, & Stapp, 1974). Self-ratings on the traits "masculine" and "feminine" tap dimensions separate from those measured by the BSRI (argued to be instrumental and expressive characteristics; e.g., Spence & Helmreich, 1979; Pedhauzer & Tetenbaum, 1979). The low internal consistency of the CGI may introduce weakened power but should not be able to account for spuriously positive taxometric results.

$$cov_{x,y} = pq(\Delta x)(\Delta y) + p(cov_{x,y|t}) + q(cov_{x,y|c}),$$

where p, q = base rates of the taxon (t) and the complement (c), respectively; Δx, Δy = difference between the taxon and complement means of x and y; and $cov_{x,y|t}$, $cov_{x,y|c}$ = covariance between x and y in the taxon and the complement. When the last two terms are assumed to be zero, then

$$cov_{x,y} = pq(\Delta x)(\Delta y).$$

Near the extremes of the distribution on the input indicator, z, pq is expected to be near zero (because either p or q is expected to be near zero). As one moves toward less extreme intervals of z, pq increases and reaches a maximum at $pq = .5 \times .5$. Because z is also assumed to be unrelated to x and y within taxa, Δx and Δy within each interval are expected to be the Δx and Δy within the population. Hence, these terms are not expected to change across intervals. Covariance within intervals thus changes as a function of pq, which results in a peaked covariance curve. The one exception to this expectation is when the taxon has a very low base rate such that, even at the extreme high interval, pq is still rising or near .5. Then, a rising or cusped covariance curve is expected (Meehl & Yonce, 1996). Thus, a taxonic situation is indicated by a covariance curve that is *peaked* or *rising*.

With three indicators, this procedure can be applied three times, with each indicator serving as the input variable for the other two. A taxonic situation should of course consistently give rise to peaked or rising covariance curves.

Sampling variability (e.g., in Δx and Δy in intervals) and assumption departures can result in covariance curves that do not peak as expected even when a taxon exists. Through Monte Carlo modeling, however, Meehl (1973) showed that, in the face of moderate departures from assumptions, these procedures are robust in *not* indicating that a taxon exists when it does not (see also Meehl, 1995b; Meehl & Yonce, 1996).

Findings. Results are shown in Figures 1 and 2. As can be seen, within each sex, the procedures yielded peaked covariance curves. Results were clearest for men. For women, the peak of the curve occurred when CGN was the input variable at an extreme interval. As just noted, however, when pq is close to .5 in an interval at one extreme of the input variable (particularly when the input variable is not highly skewed, as is the case for CGN), the peak of the covariance curve may be near or at one extreme.

Base-Rate Estimation

Procedures. MAXCOV allows estimation of the base rates of the two taxa when peaked covariance curves emerge. Under the model it is assumed that in the HITMAX interval (in which the highest covariance is observed), the covariance is approximated by

$$cov_{x,y} = pq(\Delta x)(\Delta y) = (.5 \times .5)(\Delta x)(\Delta y) = (1/4)(\Delta x)(\Delta y).$$

$Cov_{x,y}$ in this interval is known, and $(\Delta x)(\Delta y)$ can be estimated and plugged into

$$pq = cov_{x,y}/[(\Delta x)(\Delta y)]$$

for all other intervals. As $q = (1 - p)$, this formula can be expanded and rearranged to

$$p^2 - p + cov_{x,y}/[(\Delta x)(\Delta y)] = 0.$$

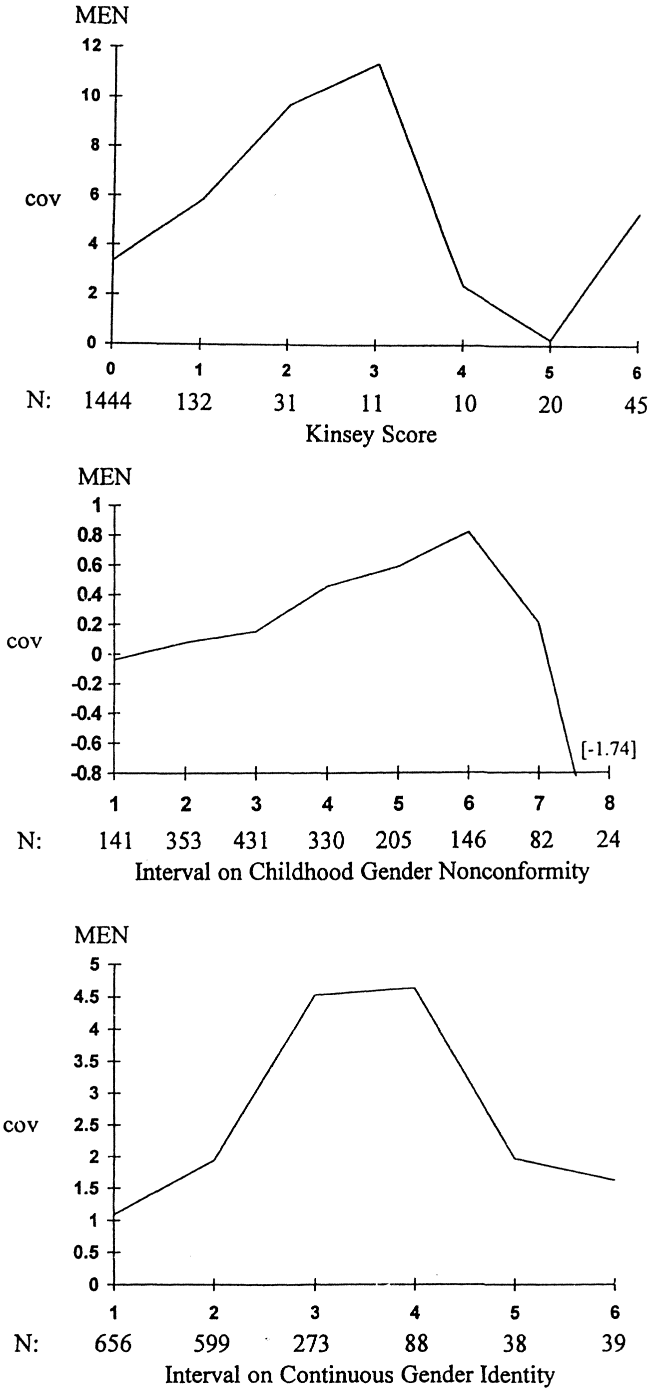


Figure 1. Results for men. Top: Covariance (cov) between childhood gender nonconformity (CGN) and continuous gender identity (CGI) as a function of Kinsey scores. Middle: Covariance between Kinsey scores and CGI as a function of CGN. Bottom: Covariance between Kinsey scores and CGN as a function of CGI.

There are two solutions for p, one above and one below .5. Intervals above the HITMAX interval are assumed to have p greater than .5; all intervals below the area are assumed to have p less than .5. Taking p times n within each interval, summing across

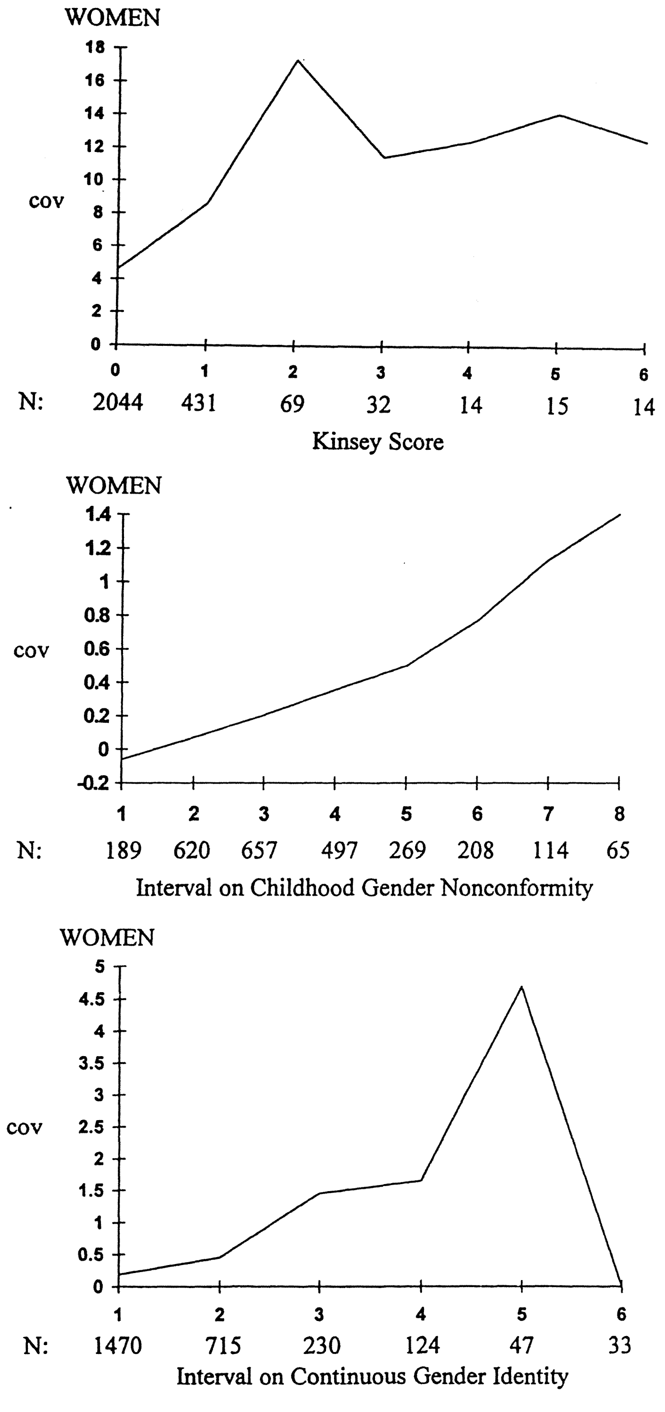


Figure 2. Results for women. Top: Covariance (cov) between childhood gender nonconformity (CGN) and continuous gender identity (CGI) as a function of Kinsey scores. Middle: Covariance between Kinsey scores and CGI as a function of CGN. Bottom: Covariance between Kinsey scores and CGN as a function of CGI.

intervals, and dividing by total sample size estimates the base rate (P) of the taxon within the full sample.

This procedure can be applied to all three peaked covariance curves. The consistency of these base-rate estimates is an additional test of the taxonic model. Meehl (1973) showed that, even when assumption departures are moderate, base rates tend to be fairly accurately estimated by these procedures.

Findings. The taxon base-rate estimates derived from the covariance curves with Kinsey scores, CGN, and CGI used as input variables were .13, .15, and .21 for men. The overall mean base-rate estimate was .16. The corresponding estimates for women were .11, .07, and .05. The overall mean was .08, about half that for men.

MAXCOV may overestimate base rates of small taxa when indicators covary within taxa, though the bias is generally not more than 3–4% except when nuisance correlation approaches .5 (Meehl, 1995b). In the present case, we estimated nuisance covariation liberally by assuming that all individuals in the most extreme categories toward the complement end of each dimension were pure complement cases and, using procedures described by Meehl (1995b), entered these values into the general covariance mixture formula and reestimated the base rates. For men, the estimated base rate was .12; for women, it was .05, no more than a few percent off the basic MAXCOV estimates.

Classification of Individuals

Rationale. Based on the three indicators, one can assign to individuals probabilities of belonging in each of the two taxa. Each indicator is split at the HITMAX interval and, using results from base-rate estimation, probabilities of an individual scoring (a) at or above this interval and (b) below this interval, contingent on belonging to the taxon, are computed. If each indicator independently taps taxon membership, Bayes' Theorem gives probabilities of belonging to the taxon associated with the eight combinations of scoring at least as high as or below the HITMAX interval on the three indicators:

$$p(t|x', y', z') = P[p(x'|t)p(y'|t)p(z'|t)]/[P \times p(x'|t) \times p(y'|t)p(z'|t) + (1 - p)p(x'|c)p(y'|c)p(z'|c)],$$

where x', y', z' is a particular combination of scoring at least as high as or, alternatively, below the HITMAX interval on indicators x, y, and z, and t and c refer to membership in the taxon and the complement, respectively.

Monte Carlo simulations have shown that most of these probabilities are near zero or one when a latent taxon underlies the three indicators. The distribution tends to be humped in the middle or flat when a continuous variable underlies the association between the indicators (Meehl, 1973). When the base rate of the taxon is low, most individuals naturally will be assigned a probability of belonging to the taxon near zero. When the assumption of local independence of no covariation between indicators is violated, the formula above gives biased probabilities of class membership. Monte Carlo simulations nonetheless indicate that a U-shaped distribution of class memberships will result from a taxonic situation even when considerable nuisance covariation within classes exists.

Findings. Distributions of individual probabilities of belonging to the taxon are given in Figure 3. As can be seen, most individuals are assigned a probability of belonging to the taxon near zero or one, consistent with the existence of a latent taxon.

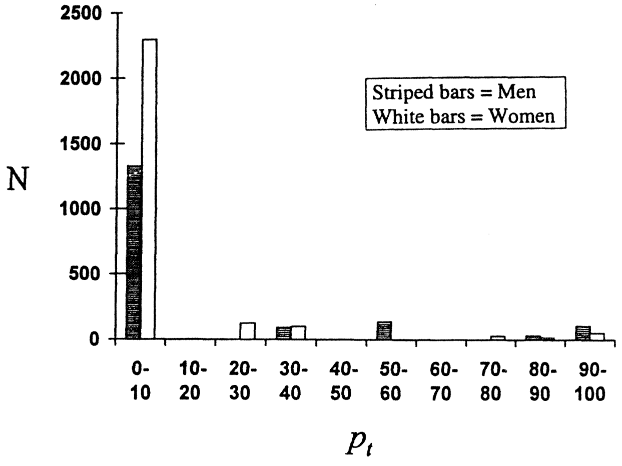


Figure 3. Distribution of probabilities of belonging to the latent taxon for men and women on the basis of maximum covariance, making hits maximum (MAXCOV) results.

Indicator Validities

Procedures. The MAXCOV procedures estimate the validities of the three indicators. For each covariance curve, an estimate of the product of the validities of the two indicators whose covariances were computed (e.g., $\Delta x \Delta y$) is obtained (see above). Solving the three simultaneous equations for the unknown parameters Δx, Δy, and Δz yields estimates of the three raw validities (differences between the taxon and complement means for x, y, and z), which are readily converted into estimated γs, differences between the taxon means expressed in within-taxon standard deviation units (Cohen, 1977). A second set of validities can be estimated on the basis of the indicator distributions derived for purposes of estimating the base rate. Each method may have error due to (a) the possibility that the maximum covariance is on average systematically overestimated (as it is chosen to be the maximum on empirical grounds), which may result in slightly inflated estimates overall, and (b) nuisance covariation between indicators within taxa, which may lead to overestimation of one set of estimates and underestimation of the other set.

Findings. Indicator validities for the three variables are given in Table 3. As can be seen, for both men and women the γ for CGN was greatest, followed by CGI and Kinsey scores. Estimated point-biserial correlations between taxon membership and the indicators are .77, .46, and .44 for CGN, CGI, and Kinsey scores for men; these estimates are .64, .43, and .56 for women. For men, these estimates are close to the loadings on a single factor estimated from a principal axis factor analysis of the three variables (.73, .49, and .45), as should be expected if the separation between taxon and complement means accounts for most of the covariation between variables. For women, these loadings were .77, .39, and .61, which suggests that greater nuisance covariance within taxa may exist for women, perhaps particularly between CGN and CGI. This result is consistent with the fact that, even in groups with extreme Kinsey scores, these variables covaried.[2]

Estimated distributions for Kinsey scores are provided in Figure 4. One striking outcome is that for both genders these scores vary considerably *within* the taxon. Indeed, for both men and women, approximately 50% of individuals in the taxon appear to score Kinsey 0 (*exclusively heterosexual;* 53% and 49%, respectively). Of men in the taxon, about a third (31% in this sample) are estimated to have Kinsey scores 4–6 (on the homosexual side of bisexual). Of women in the taxon, only about a tenth (11%) are estimated to have Kinsey scores 4–6. Hence, the taxon is not highly specific to homosexual preference. At the same time, relatively few men with homosexual preference (Kinsey 4–6) are not members of this taxon: 9%, a low value considering that 83% of the population belongs to the complement taxon. A somewhat greater proportion of women who score Kinsey 4–6 are estimated to be members of the complement taxon: 25%. (As these estimated distributions are more error prone than the base-rate estimates, however, these values can be taken as only crude estimates; see e.g., Meehl & Yonce, 1994).

MAXCOV Findings: Summary

In sum, MAXCOV procedures yielded a number of findings consistent with latent taxa underlying social orientation and gender identity: (a) consistently peaked or rising covariance curves; (b) consistent base-rate estimates for a taxon associated with sex-atypical sexual orientation and gender identity for both sexes, averaging about 16% for men and 8% for women; (c) a U-shaped distribution of taxon membership probabilities for both men and women. These results occurring in the absence of true latent taxa would be a "damn strange coincidence" and, therefore, strongly suggest that latent taxa exist. To further test the taxonic conjecture, however, we applied a second method.

MAMBAC: A Second Taxometric Method

Rationale. We also applied the MAMBAC (Meehl & Yonce, 1994). These procedures require at least two indicators. One variable is treated as an "input variable." Successive cuts on this input variable are made, and for each cut the means on the second variable (the "output variable") above and below the cut are computed. Meehl and Yonce (1994) showed that if nontaxonic dimensional variation underlies covariation between indicators, the difference between output means above and below the cut will be a dish-shaped function of the cut point on the input variable (i.e., higher for extreme cuts on the input variable than middle cuts). If taxonic variation exists and the base rate of the taxon is low, the difference between output means either will be *humped* or will *increase* as a function of the cut point on the input variable.

MAMBAC estimates the base rate as well. If nearly all individuals below the extreme low cut are complement taxon members, if nearly all individuals above the extreme high cut

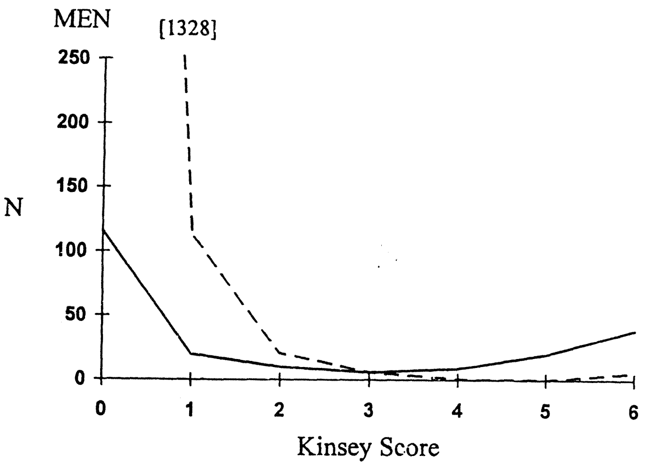
---

[2] It may seem odd to apply common factor analysis, sometimes regarded as a method applicable when continuous factors underlie covariation between observed variables, to measures conjectured to be taxonic-dimensional. The latent factors in factor analysis, however, need not be continuous. As Waller and Meehl (1998) discussed, Thurstone (1947, in Waller & Meehl, 1998) saw that factors could be discretely and dichotomously valued.

Table 3
Estimated Validities of Indicators of Latent Taxa

Variable	MAXCOV γ	MAXCOV r_{pb}	MAMBAC γ	MAMBAC r_{pb}	Mean γ	Mean r_{pb}
Men						
Kinsey scores	1.39	.46	1.70	.52	1.54	.48
Childhood gender nonconformity	3.24	.77	1.68	.51	2.27	.68
Continuous gender identity	1.33	.44	1.41	.45	1.37	.44
Women						
Kinsey scores	1.79	.43	1.60	.48	1.70	.46
Childhood gender nonconformity	3.14	.64	2.11	.58	2.88	.66
Continuous gender identity	2.56	.56	3.34	.75	2.66	.63

Note. Estimates based on the mean are not equal to the means of the individual estimates, because the estimates based on the mean are based on components that go into the individual estimates (taxon means and standard deviations) rather than the final estimates themselves. MAXCOV = maximum covariance; MAMBAC = mean above minus below a cut. γ = estimated difference between taxon means divided by within-taxon standard deviation; r_{pb} = estimated point-biserial correlation between indicator and latent taxon variable.



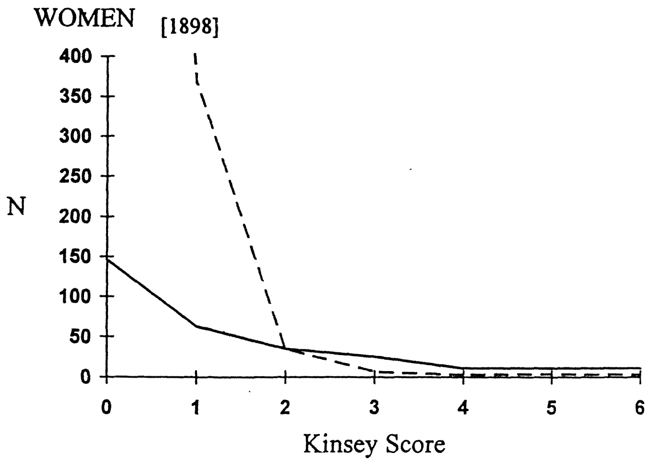


Figure 4. Estimated distributions of Kinsey scores for the taxon (solid line) and the complement (dotted line) for men (top) and women (bottom).

are taxon members, and if little nuisance covariation between the input and output variables exists within taxa, then the base rate is estimated by

$$P = 1/(1 + R),$$

where $R = ({}_{Hi}N_b/{}_{Lo}N_a) \times {}_{Hi}[d_y(x)]/{}_{Lo}[d_y(x)]$ and ${}_{Hi}N_b$ and ${}_{Lo}N_a$ are the number of individuals below the high cut and above the low cut; ${}_{Hi}[d_y(x)]$ and ${}_{Lo}[d_y(x)]$ are the differences between the means on the output variable (x) above and below the high cut and low cut (see Meehl & Yonce, 1994).

MAMBAC estimates output variable indicator validities with

$$\Delta y_{est1} = {}_{Hi}[d_y(x)]({}_{Hi}N_b)/(NQ), \text{ and}$$

$$\Delta y_{est2} = {}_{Lo}[d_y(x)]({}_{Lo}N_a)/(NP).$$

With three variables, each indicator serves as the output variable twice and, hence, MAMBAC gives four estimates of each indicator's validity, averaged to give the best.

Findings. MAMBAC may work best to estimate latent parameters, such as the base rate, when 10 or more cuts can be made on the input variable (Meehl & Yonce, 1994). Two indicators, Kinsey scores and CGI, permitted only six cuts. Nonetheless, we applied MAMBAC to all possible pairs of input and output variables.

MAMBAC curves are shown in Figures 5 (for men) and 6 (for women). In most cases the function is not dished but rather is sloped upward or humped, consistent with a taxon with low base. As evidence that visual discrimination of the form of MAMBAC curves validly indicates taxonicity, Meehl and Yonce (1994; Appendix D) had 17 individuals sort curves generated from simulated taxonic or nontaxonic latent situations into two groups based on relative similarity to two sets of curves, one generated by taxonic situations, the other by pure dimensions. Most judges had no more than a B.A. degree. Judges' discriminations were "correct" nearly 98% of the time. No single data set was misclassified by more than 2 judges (12%). We asked 5 judges (3 graduate students and 2 faculty at the University of New Mexico) to perform the sorting

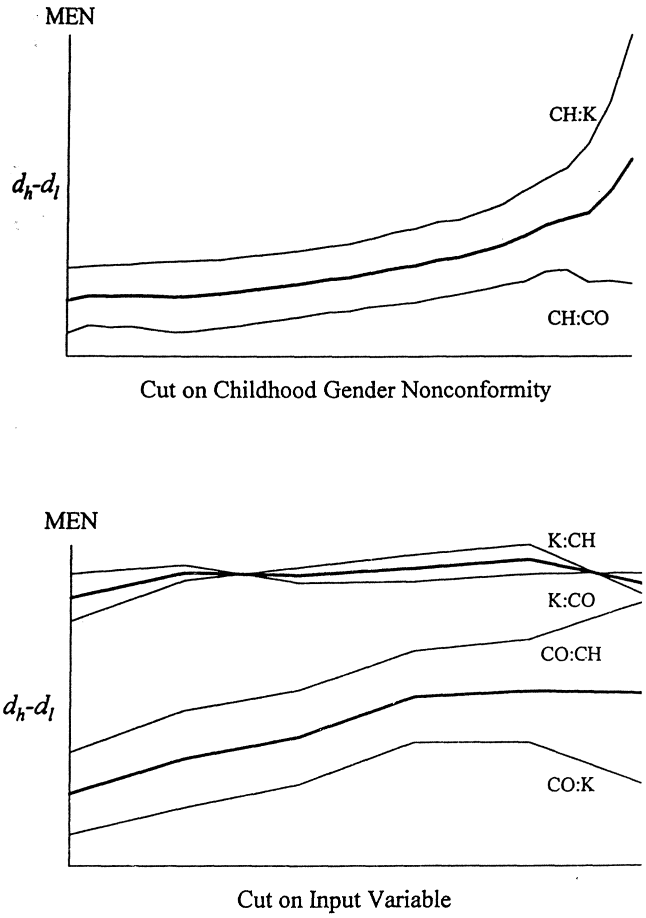


Figure 5. Mean above minus below a cut (MAMBAC) results for men. Top: Differences between means above and below cuts on childhood gender nonconformity (CH; 27 cuts). Bottom: Differences between means above and below cuts on Kinsey scores (K) and continuous gender identity (CO; 6 cuts). The first variable listed in the ratio (CH, K, or CO) is the input variable; the second is the output variable. Thick lines = averaged MAMBAC functions. Results are displayed in terms of z score differences. Lines are displaced for graphical convenience.

task with the MAMBAC curves in Figures 5 and 6. None were familiar with taxometric methods or knew how data were generated. All 5 judged the curves to be more similar to those generated by taxonic situations. Based on Meehl and Yonce's Monte Carlo data, the probability that all 5 judges would *incorrectly* classify a purely dimensional situation as taxonic is close to zero.[3]

For base-rate estimation, extreme cuts included about 50 individuals above and below the cut to ensure reasonable sample size. The base rate of the taxon was estimated to be .15 for men, almost identical to the MAXCOV estimate (.16). For women, the estimate was .13, slightly higher than the MAXCOV estimate (.08). MAMBAC taxon base-rate estimates tend to be positively biased when the taxon base rate is very low (in Meehl and Yonce's, 1994, simulations, a 10% base rate was estimated on average to be 12%). The base rate for women may thus be somewhat overestimated by MAMBAC.

Meehl and Yonce (1994) argued that a low MAMBAC base-rate estimate itself indicates taxonicity. In their Monte Carlo runs, nontaxonic situations yielded base-rate estimates close to .5 and never more extreme than .4. If these simulations are taken as a guide, the base-rate estimates for both men and women clearly indicate latent taxa.

Estimated indicator validities are given in Table 3. MAMBAC's estimates of indicator validities are not as robust as its estimates of base rates (Meehl & Yonce, 1994); as can be seen, the estimates are similar, though not identical, to MAXCOV estimates. The mean estimated γ is 2.07 (range = 1.37–2.88). Estimated point-biserial correlations range from .44 to .66. The estimated validity differs across gender for one variable, CGI. For women, this variable is estimated to have the highest validity, for men the lowest.[4]

Discussion

Theoretical and empirical reasons led us to examine the domains of sexual orientation, CGN, and CGI for taxonicity. Two taxometric procedures, MAXCOV and MAMBAC, suggest that taxa do indeed underlie these domains. Peaked or rising MAXCOV curves, humped or rising MAMBAC curves, consistent base-rate estimates, U-shaped distributions of taxon membership probabilities, and low MAMBAC base-rate estimates should all occur when true taxa exist. They should not and, according to Monte Carlo simulations, do not occur when no true taxa exist. That all would occur were there no true taxa would be a "strange coincidence."

Haslam's Findings

Our findings contrast with those of Haslam (1997), who found no evidence that taxa underlie men's MMPI *Mf* Scale using two

---

[3] Were errors across judges independent, this probability would be about $.02^5$ or .000000003. Although the probability is larger than this because of nonindependent errors across judges, it is clearly much smaller than .05, that used in traditional significance testing. While judges in Meehl and Yonce's study had access to 12 MAMBAC curves whereas ours had 6, in a follow-up validation study of MAMBAC, judges sorting pairs of MAMBAC curves performed as well as those with 12 curves (Meehl & Yonce, 1994).

[4] As we have emphasized, Meehl (1965, 1973, 1995a; Meehl & Golden, 1982) has stressed the importance of consistency tests in the application of taxometric methods: Tests that tell whether the taxonomic "story" revealed by the formal–numerical relationships examined by the procedures is a coherent one. One key consistency test is the consistency of base-rate estimates across different taxonomic procedures. Given that the formal–numerical relationships involved in these estimates cannot be derived from one another but rather are linked only through a common taxonomic model, it would be coincidental were these estimates to converge on the same value and yet not refer to a "real" base rate of a true taxon (i.e., were the taxonomic model false). Hence, their consistency (which we observed in this case) is strong evidence for the taxonomic model. Other consistency tests derived by Meehl (1965, 1973; Meehl & Golden, 1982) reveal whether specific estimates of the model (e.g., estimates of the latent validities) can be trusted. MAXCOV-HITMAX has four consistency tests of this kind (see Meehl & Golden, 1982). In the current data, three of these consistency tests were passed. The fourth, the Sum of the Hit Rates test, was not. It was not surprising that this test failed, as it fails when the base rates are more extreme than .2 and .8—as in the present case. These results suggest that MAXCOV estimates of the indicator validities should not be trusted without corroboration, and we have not relied on them alone. The failure of a single consistency test that is designed to fail expressly when base rates are in the range we estimate is no basis for rejecting the taxonomic model.

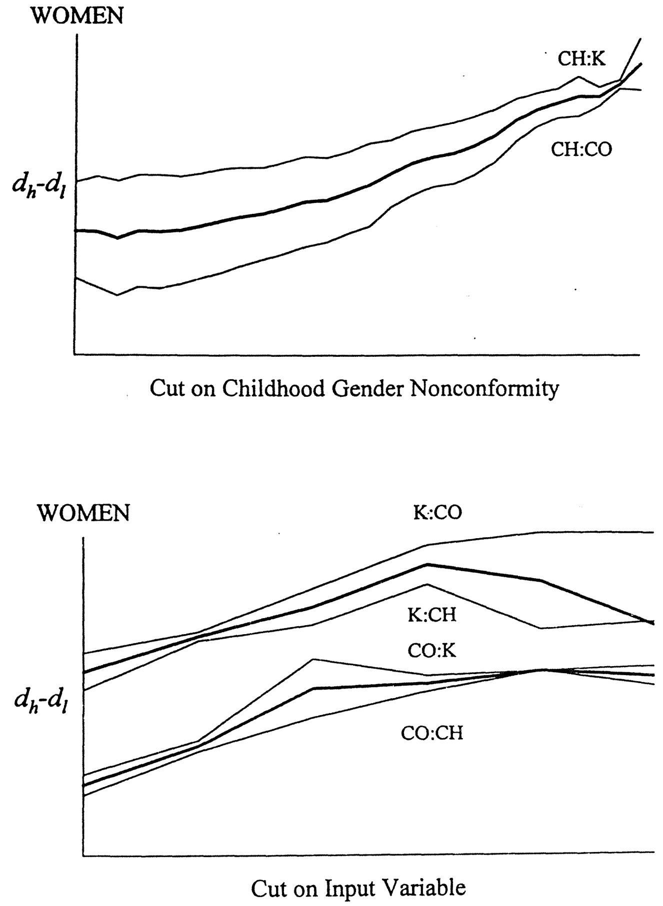


Figure 6. Mean above minus below a cut (MAMBAC) results for women. Top: Differences between means above and below cuts on childhood gender nonconformity (CH; 28 cuts). Bottom: Differences between means above and below cuts on Kinsey scores (K) and continuous gender identity (CO; 6 cuts). The first variable listed in the ratio (CH, K, or CO) is the input variable; the second is the output variable. Thick lines = averaged MAMBAC functions. Results are displayed in terms of z score differences. Lines are displaced for graphical convenience.

taxometric procedures. First, he found no evidence that the full *Mf* distribution is a mixture of two latent distributions. One drawback of mixture analyses is that within-taxon distributions are constrained, while multi-item psychological scales can have irregular distributions, owing to their dependence on the distribution of item difficulties and discrimination (e.g., Grayson, 1987). Second, MAMBAC analyses yielded no evidence consistent with taxa underlying five obliquely rotated factors extracted from the *Mf* Scale. One problem with these analyses resides with the nature of the measures. For MAMBAC to detect underlying taxa, the measures it is applied to should covary, even if only moderately (Meehl, 1995a). The five factors Haslam extracted correlated from −.07 to .12, the mean near zero. Although the full *Mf* Scale has been shown to discriminate heterosexual and homosexual men moderately well (γ close to 1.5 standard deviations; see Haslam, 1997, for a review), the specific factors derived from the full measure cannot all perform as well. MAMBAC and MAXCOV work well when taxon separations on indicators are 1.25 standard deviations or more, but Meehl (1995b) warned that they may not reliably do so when separations are lower (especially when the taxon base rate is small). Two indicators that covary solely because two taxa are separated by 1.25 standard deviations on each, with taxon base rates of .10 and .90, will correlate .14, greater than the *largest* correlation in Haslam's data. Our indicators correlated .32 on average, with latent taxa separations estimated to be over 2 standard deviations. We suggest, then, that we detected taxonicity while Haslam did not because of power: Taxometric procedures applied to our measures could more powerfully detect true taxa.[5]

This difference in power may well be due to differences in the content of our measures and the *Mf* subscales. One of our measures directly taps sexual orientation, whereas the others tap variables known to covary moderately with sexual orientation. *Mf* items do not explicitly concern sexual orientation, childhood interests, or gender identity. That *Mf* subscales show no evidence of being taxonic, then, does not contradict the current evidence; the studies are not truly comparable. Although replication using two different procedures suggests that our results are not due to a sampling fluke (Meehl, 1995a), we encourage attempts to replicate on other samples using alternative measures.

Alternative Methods of Testing for Latent Classes

In addition to commingling models, statisticians have relied on a number of other procedures to test for latent classes. Perhaps most notable of these are latent class analyses (e.g., Heinen, 1996; Kendler, Karkowski, & Walsh, 1998; Lazarsfeld & Henry, 1968; Rost & Langeheime, 1997), which determine the minimum number of classes required to account for covariation between indicators (though recent models also allow covariation because of quantitative variables). The analyses provide goodness-of-fit tests of whether a model can account for indicator covariation. In his recent *American Psychologist* article on his taxometric procedures, Meehl (1995a) did not cite work on latent class models (though see Waller & Meehl, 1998).

Meehl (e.g., 1995a) has reasons for not recommending latent class analysis to test taxonic conjectures. For latent class analyses to yield goodness-of-fit tests, assumptions about the nature of latent distributions within classes, covariation between indicators within classes (and their homo- or heterogeneity across classes), and so forth, must be made. Though few in number, these assumptions constitute auxillary hypotheses of the test of a taxonic conjecture. Rejection of a particular taxonic conjecture, then, may depend on these assumptions rather than the taxonic conjectures themselves. Although Meehl's methods require that provisional auxillary assumptions be made, the critical feature is not that they be met but that results of the analyses are *robust* to assumption departures (Meehl, 1995a). Goodness-of-fit tests in latent variable analysis are designed to be sensitive, not robust, to assumption departures. Lack of constraining assumptions means that no goodness-of-fit statistic can be computed from Meehl's proce-

---

[5] Haslam (1997) performed Monte Carlo simulations showing that, under reasonable assumptions of taxonicity, his commingling analyses should have been powerful enough to delete a difference between a mixed model and a single normal distribution model. However, these simulations also assumed particular taxon distributions. If these assumed distributions are not valid, the Monte Carlo simulations could be misleading.

dures. But, as emphasized above, the methods' rigor tests consistency, coherence, and "Salmonean coincidence," not goodness-of-fit. As Meehl (1995a) pointed out, cluster analysis, another method popular with applied statisticians, cannot perform well the simple task of recovering biological sex as a taxon in real data, which his methods do (Golden & Meehl, 1980). Comparisons between latent class analysis and Meehl's methods in recovering latent classes from simulated data generated from a variety of assumptions would be worthwhile.

Unlike Meehl's methods, latent class analysis does allow one to ask whether latent classes greater than two might fit data better than a two-class model (though, again, under restrictive assumptions). The results we present indicate that two latent classes do account for variance underlying social orientation and gender identity. If three or four ordered classes existed, the results obtained would not be expected. Nonetheless, the current results do not rule out the possibility that subclasses within the two latent classes indicated by our analyses exist, perhaps discriminated by measures other than those we observed.

One additional issue is whether homosexual or heterosexual interests differentially discriminate the latent classes. Although these interests independently measured covary negatively (at least in men), one or another may better discriminate the latent taxa our analyses reveal. This issue remains for future research to explore.

The Significance of Taxonicity

As noted earlier, the primary implication of taxonicity is that it indicates that a dichotomous factor plays an important role in the processes that create variation. The neurohormonal theory of sexual orientation states that one cause of gay or lesbian sexual orientation is a sex-atypical hormonal profile affecting neurodevelopment. Canalization tends to lead to sex-typical outcomes within the genders and, therefore, while the development of most individuals' structures affecting sexual orientation may be shunted down a sex-typical track, a minority of individuals' development may be canalized down a sex-atypical track. This implies that a threshold effect influences sexual orientation (Meehl, 1977).

While this theory provides one account of the taxa, others are possible. A single gene affecting male sexual orientation may account for taxa in men (Bailey, 1995; Hamer, Hu, Magnuson, Hu, & Pattatucci, 1993; Hu et al., 1995; see also Bailey & Pillard, 1991). Perhaps an as-yet-unidentified gene similarly affects women's sexual orientation (see Bailey, Pillard, Neale, & Agyei, 1993).

The issue of taxonicity should not be conflated with the issue of genetic influence (Garber, 1995; Weinberg, Williams, & Pryor, 1994; Wertz, 1996); they are independent. Height is highly heritable but dimensional. Belief in Trotskyist political ideology is taxonic (Meehl, 1992) but clearly requires much social influence and is less heritable than height. Our finding does not imply that sexual orientation is highly heritable. Although we cannot presently offer a plausible social psychological account of our results, socialization factors may be responsible for taxa. The key finding here is that some dichotomous factor appears to play a very important role in the development of sexual orientation, which should inform future attempts to explain the development of sexual orientation.

Base Rates of the Latent Classes

Perhaps the most remarkable aspect of our findings is not that they indicate taxonicity but the base rates of the latent classes. According to our estimates, 12–15% of men and 5–10% of women belong to latent taxa associated with greater degrees of homosexual desires and fantasies—considerably higher than the percentages of individuals who endorse homosexual desires and fantasies on the Kinsey scales (5% of men and 3% of women score at least 3). Hence, the "joints of nature" that the underlying taxa reflect do *not* correspond to the bimodal distributions of men's and women's self-reported sexual orientation. Indeed, as many as half of the men and women in the latent taxa associated with greater degrees of homosexual desires claim to be *exclusively heterosexual.*

The latent taxa, then, appear *not* to be classes of, on the one hand, heterosexual and, on the other, homosexual individuals, unless one of the base-rate estimates is greatly biased. Although homosexual desires may be underreported in national samples, this probably cannot fully account for the discrepancy. And, whereas MAXCOV and MAMBAC may overestimate base rates of small latent taxa when indicators covary substantially within classes (Meehl, 1995b), our adjustments for this covariation were only a few percent (consistent with Meehl's, 1995b, Monte Carlo data). The most reasonable view is that the proportions of heterosexual men and women in the latent taxa associated with homosexual preference is probably appreciable.

Whereas many heterosexual individuals may belong to latent taxa in which most nonheterosexual individuals belong, it seems that few nonheterosexual individuals belong to the complement classes within each gender. The latent taxa may hence reflect a causal factor that is *necessary* (or near necessary) to the development of nonheterosexual orientation (particularly for men, for whom a very small proportion of homosexual men belonged to the complement taxon) but not *sufficient* to account for it. For instance, the sex-atypical patterns of early hormonal processes conjectured by the neurohormonal theory of sexual orientation may be (nearly) necessary but not sufficient for the development of homosexual orientation. Why this might be so is a matter for future research. Possibly, gay sexual orientation does not reflect development shunted down the sex-atypical track so much as it reflects a lack of canalization down the sex-typical track, with the result that about 50% of such individuals will possess a sex-typical and 50% a sex-atypical sexual orientation. Any alternative account of the latent taxa (e.g., in terms of social processes) must also account for why one taxon consists primarily of heterosexual individuals while the other consists of a mixture of both sexual orientations.

Whatever the cause of latent taxa, these findings may be critical to understanding the development of sexual orientation and gender identity. The fact that taxa do underlie sexual orientation and gender identity argues against simple additive models of causality (i.e., many independent events affecting these outcomes, each with a small effect).

References

Bailey, J. M. (1995). Sexual orientation revolution. *Nature Genetics, 11,* 353–354.

Bailey, J. M., Dunne, M. P., & Martin, N. G. (2000). Genetic and environmental influences on sexual orientation and its correlates in an

Australian twin sample. *Journal of Personality and Social Psychology, 78,* 524–536.

Bailey, J. M., Finkel, E., Blackwelder, K., & Bailey, T. (1996). *Masculinity, femininity, and sexual orientation.* Unpublished manuscript.

Bailey, J. M., Nothnagel, J., & Wolfe, M. (1995). Retrospective measured individual differences in childhood sex-typed behavior among gay men: Correspondence between self- and maternal reports. *Archives of Sexual Behavior, 24,* 613–622.

Bailey, J. M., & Pillard, R. C. (1991). A genetic study of male sexual orientation. *Archives of General Psychiatry, 48,* 1089–1096.

Bailey, J. M., Pillard, R. C., Neale, M. C., & Agyei, Y. (1993). Heritable factors influence sexual orientation in women. *Archives of General Psychiatry, 50,* 217–223.

Bailey, J. M., & Zucker, K. J. (1995). Childhood sex-typed behavior and sexual orientation: A conceptual analysis and quantitative review. *Developmental Psychology, 31,* 43–55.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42,* 155–162.

Biggar, R. J., & Melbye, M. (1992). Responses to anonymous questionnaires concerning sexual behavior: A method to examine potential biases. *American Journal of Public Health, 82,* 1506–1512.

Blanchard, J. J., Gangestad, S. W., Brown, S. A., & Horan, W. P. (2000). Hedonic capacity and schizotypy revisited: A taxometric analysis of social anhedonia. *Journal of Abnormal Psychology, 109,* 87–95.

Blanchard, R., & Freund, K. (1983). Measuring masculine gender identity in females. *Journal of Consulting and Clinical Psychology, 51,* 205–214.

Blanchard, R., McConkey, J. G., Roper, V., & Steiner, B. W. (1983). Measuring physical aggressiveness in heterosexual, homosexual, and transsexual males. *Archives of Sexual Behavior, 12,* 511–524.

Chapman, L. J., Chapman, J. P., & Raulin, M. L. (1976). Scales for physical and social anhedonia. *Journal of Abnormal Psychology, 85,* 374–382.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). San Diego, CA: Academic Press.

Diamond, M. (1993). Homosexuality and bisexuality in different populations. *Archives of Sexual Behavior, 22,* 291–310.

Dunne, M. P., Martin, N. G., Bailey, J. M., Heath, A. C., Bucholz, K. K., Madden, P. A. F., & Statham, D. J. (1997). Participation bias in a sexuality survey: Psychological and behavioural characteristics of responders and non-responders. *International Journal of Epidemiology, 26,* 844–854.

Ellis, L., & Ames, M. A. (1987). Neurohormonal functioning and sexual orientation: A theory of homosexuality–heterosexuality. *Psychological Bulletin, 10,* 233–258.

Finn, S. E. (1987). *The structure of maculinity and femininity self ratings.* Unpublished manuscript.

Freund, K., Langevin, R., Satterberg, J., & Steiner, B. (1977). Extension of the Gender Identity Scale for Males. *Archives of Sexual Behavior, 6,* 507–519.

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

Garber, M. (1995). *Vice versa: Bisexuality and the eroticism of everyday life.* New York: Simon & Schuster.

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

Grayson, D. A. (1987). Can categorical and dimensional views of psychiatric illness be distinguished? *British Journal of Psychiaty, 151,* 355–361.

Grellert, E. A., Newcomb, M. D., & Bentler, P. M. (1982). Childhood play activities of male and female homosexuals and heterosexuals. *Archives of Sexual Behavior, 11,* 451–478.

Hamer, D. H., Hu, S., Magnuson, V. L., Hu, N., & Pattatucci, A. M. L. (1993, July 16). A linkage between DNA markers on the X chromosome and male sexual orientation. *Science, 261,* 321–327.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1994). Psychopathy as a taxon: Evidence that psychopaths are a discrete class. *Journal of Consulting and Clinical Psychology, 62,* 387–397.

Haslam, N. (1997). Evidence that male sexual orientation is a matter of degree. *Journal of Personality and Social Psychology, 73,* 862–870.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences.* Thousand Oaks, CA: Sage.

Hu, S., Pattatucci, A. M. L., Patterson, C., Li, L., Fulker, D. W., Cherny, S. S., Kruglyak, L., & Hamer, D. H. (1995). Linkage between sexual orientation and chromosome Xq28 in males but not in females. *Nature Genetics, 11,* 248–256.

Johnson, A. M., Wadsworth, J., Elliott, P., Prior, L., Wallace, P., Blower, S., Webb, N. L., Heald, G. I., Miller, D. L., Adler, M. W., & Anderson, R. M. (1989). A pilot study of sexual lifestyle in a random sample of the population of Great Britain. *AIDS, 3,* 135–141.

Kendler, K. S., Karkowski, L. M., & Walsh, D. (1998). The structure of psychosis: Latent class analysis of probands from the Roscommon family study. *Archives of General Psychiatry, 55,* 492–499.

Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male.* Philadelphia: W. B. Saunders.

Kinsey, A. C., Pomeroy, W. B., Martin, C. E., & Gebhard, P. H. (1953). *Sexual behavior in the human female.* Philadelphia: W. B. Saunders.

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

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin.

Lenzenweger, M. F., & Korfine, L. (1992). Confirming the latent structure and base rate of schizotypy: A taxometric analysis. *Journal of Abnormal Psychology, 101,* 567–571.

LeVay, S. (1996). *Queer science: The use and abuse of research into homosexuality.* Cambridge, MA: MIT Press.

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

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

Meehl, P. E. (1973). MAXCOV-HITMAX: A taxonomic search procedure for loose genetic syndromes. In P. E. Meehl (Ed.), *Psychodiagnosis: Selected papers* (pp. 200–224). Minneapolis, MN: University of Minnesota Press.

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

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

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

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

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

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

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

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

Meehl, P. E., & Yonce, L. J. (1996). Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure) [Monograph]. *Psychological Reports, 78* (Suppl. 1-V78), 1091–1227.

Mitchell, J., & Zucker, K. J. (1992, June). *The Recalled Childhood Gender Behaviors Questionnaire: Psychometric properties.* Paper presented at the meeting of the International Academy of Sex Research, Barrie, Ontario, Canada.

Pedhauzer, E. J., & Tetenbaum, T. J. (1979). Bem Sex Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology, 37,* 996–1016.

Peplau, L. A., Garnets, L. D., Spalding, L. R., Conley, T. D., & Veniegas, R. C. (1998). A critique of Bem's "Exotic becomes erotic" theory of sexual orientation. *Psychological Review, 105,* 387–394.

Reise, S. P., & Gomel, J. N. (1995). Modeling qualitative variation within latent trait dimensions: Application of mixed-measurement to personality assessment. *Multivariate Behavioral Research, 30,* 341–358.

Rost, J., & Langeheime, R. (Eds.). (1997). *Applications of latent trait and latent class models in the social sciences.* Muenster, Germany: Waxmann.

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

Spence, J. T., & Helmreich, R. L. (1979). The many faces of androgyny: A reply to Locksley and Colten. *Journal of Personality and Social Psychology, 37,* 1032–1046.

Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of sex-role stereotypes and masculinity–femininity. *JSAS Catalog of Selected Documents in Psychology, 4,* 43 (Ms. No. 617).

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

Sundet, J. M., Magnus, P., Kvalem, I. L., Samuelsen, S. O., & Bakketeig, L. S. (1992). Secular trends and sociodemographic regularities in coital debut age in Norway. *Archives of Sexual Behavior, 21,* 241–252.

Tyrka, A. R., Cannon, T. D., Haslam, N., Mednick, S. A., Schulsinger, F., Schulsinger, H., & Parnas, J. (1995). The latent structure of schizotypy: I. Premorbid indicators of a taxon of individuals at risk for schizophrenia-spectrum disorders. *Journal of Abnormal Psychology, 104,* 173–183.

von Davier, M., & Rost, R. (1997). Self-monitoring—A class variable? In J. Rost & R. Langeheime (Eds.), *Applications of latent trait and latent class models in the social sciences.* Muenster, Germany: Waxmann.

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

Waller, N. G., Putnam, F. W., & Carlson, E. (1996). Types of dissociation and dissociative types: A taxometric analysis of the Dissociative Experiences Scale. *Psychological Methods, 3,* 300–321.

Weinberg, M. S., Williams, C. J., & Pryor, D. W. (1994). *Dual attraction: Understanding bisexuality.* New York: Oxford University Press.

Wertz, D. C. (1996). Genetics and homosexuality. *The Gene Letter, 1.* Retrieved from the World Wide Web: http://www.geneletter.org/1196/homosexuality.htm

Received September 8, 1998
Revision received July 15, 1999
Accepted November 22, 1999 ■