Journal of Personality and Social Psychology
1990, Vol. 59, No. 5, 1051-1065

0022-3514/90/$00.75

Gender Diagnosticity: A New Bayesian Approach to Gender-Related Individual Differences

Richard Lippa and Sharon Connelly
California State University, Fullerton

Gender diagnosticity refers to the Bayesian probability that an individual is predicted to be male or female on the basis of some set of gender-related diagnostic indicators. We computed gender diagnostic probabilities from occupational preference ratings made by 117 male and 110 female subjects. Subjects also completed the Personal Attributes Questionnaire and the Bem Sex-Role Inventory and were assessed on a number of gender-related criterion variables. Gender diagnostic probabilities proved to be factorially distinct from PAQ and BSRI masculinity and femininity and generally displayed greater predictive utility than did masculinity and femininity. Unlike existing scales, gender diagnosticity measures are not based on gender stereotypes, and they do not reify gender-related individual differences or freeze them into specific constructs such as instrumental or expressive traits. Furthermore, they are well suited to developmental and cross-cultural research.

Since the 1936 publication of Terman and Miles's *Sex and Personality*, considerable research has focused on assessing and understanding gender-related individual differences. Until the 1970s such research was dominated by the assumption that masculinity–femininity is best conceptualized as a bipolar unidimensional personality trait (Constantinople, 1973). Since the 1970s, research has investigated another possibility, that masculinity and femininity are two separate and independent dimensions (Bem, 1974; Cook, 1985; Heilbrun, 1976; Spence, Helmreich, & Stapp, 1974). Despite the paradigm shift, the newer two-dimensional approach to gender-related individual differences has not proven to be dramatically more successful than older one-dimensional approaches in predicting gender-related behaviors. Indeed, a number of recent reviews suggest that gender-related behaviors are intrinsically inconsistent, situationally variable, and, at best, weakly predicted by self-report measures of masculinity–femininity, masculinity, or femininity (Deaux, 1985, 1987; Deaux & Major, 1987; Spence, Deaux, & Helmreich, 1985; Spence & Helmreich, 1978, 1980; Spence & Sawin, 1985).

In this article we present a new approach to assessing gender-related individual differences, an approach we term *gender diagnosticity*. In brief, gender diagnosticity refers to the Bayesian probability that an individual is predicted to be male or female based on some set of gender-related diagnostic indicators. The concept of gender diagnosticity is theoretically linked to the diagnostic ratio approach to measuring stereotypes (McCauley & Stitt, 1978; McCauley, Stitt, & Segal, 1980). We propose here that gender diagnosticity—the Bayesian probability that an individual is classified as male or female—can serve as a measure of individual differences in gender-related behaviors within as well as across the sexes.

To support this contention, we present data that demonstrate the utility of gender diagnosticity measures in predicting a varied sample of gender-related behaviors in a sample of 227 male and female subjects, and we compare the predictive utility of gender diagnosticity measures with that of commonly used masculinity and femininity scales. We argue that gender diagnosticity measures have a number of advantages over previous assessment strategies: They have a better theoretical and mathematical rationale than do previous approaches; they do not reify gender-related individual differences or freeze them into specific constructs; and, most important, they prove to be more successful at predicting gender-related behaviors.

This article is divided into three related sections. In the first, we briefly review research on gender-related individual differences and then discuss the nature of gender diagnosticity and its theoretical rationale. In the second section, we describe a study that demonstrates the ability of gender diagnosticity measures to predict a broad array of gender-related behaviors. In the final section, we explore some of the theoretical and practical advantages of gender diagnosticity measures, and we speculate how similar measures might be applied to other research domains.

Preliminary reports of this research were presented at the 1988 and 1989 Nags Head conferences on personality and social behavior, Nags Head, North Carolina.

We wish to thank the following research assistants who worked conscientiously on this research project and who devoted many hours to running subjects, rating and coding data, and entering data into the computer: Lori Campos, Michele Cojerean, Carole DiCenso, Treseen Fleming, Nancy Olson, and Wilma Tope.

Correspondence concerning this article should be addressed to Richard Lippa, Department of Psychology, California State University, Fullerton, California 92634.

The Concept of Gender Diagnosticity

Review of Past Research on Gender-Related Individual Differences

Terman and Miles (1936) noted early in their classic work on sex and personality that:

> the typical woman is believed to differ from the typical man in the greater richness and variety of her emotional life and in the extent to which her everyday behavior is emotionally determined. In particular, she is believed to experience in greater degree than the average man the tender emotions, including sympathy, pity, and parental love; to be more given to cherishing and protective behavior of all kinds. Compared with man she is more timid and more readily overcome by fear. She is more religious and at the same time more prone to jealousy, suspicion, and injured feelings. Sexually she is by nature less promiscuous than man, is coy rather than aggressive, and her sexual feelings are less specifically localized in her body. Submissiveness, docility, inferior steadfastness of purpose, and a general lack of aggressiveness reflect her weaker conative tendencies. Her moral life is shaped less by principles than by personal relationships, but thanks to her lack of adventurousness she is much less subject than man to most types of criminal behavior. Her sentiments are more complex than man's and dispose her personality to refinement, gentility, and preoccupation with the artistic and cultural. (p. 2)

Although this passage may seem to the modern reader simply a catalog of gender stereotypes (for recent reviews of research on gender stereotypes, see Basow, 1986; Eagly & Steffen, 1986; Lippa, 1990a; Martin, 1987; Spence et al., 1985), it does set the stage for Terman and Miles's core assumption that gender-related individual differences *within the sexes* are defined by differences *between the sexes.*

Terman and Miles's research on masculinity–femininity commenced in 1922 when Terman noted substantial sex differences in the play and activity preferences of intellectually gifted children. From this beginning he and Miles developed a multifaceted masculinity–femininity test that comprised, among other things, attitude scales, word association tests, a standardized projective test, and a measure of subjects' interest in various activities and hobbies. The items selected for these tests were those that best differentiated male and female respondents. Terman and Miles's approach laid the foundation for many subsequent masculinity–femininity scales, including those of the Strong Vocational Interest Blank (Campbell, 1966; Strong, 1943), the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943), and the California Psychological Inventory (Gough, 1964).

The early 1970s witnessed a number of important conceptual and empirical critiques of the tradition of masculinity–femininity scaling begun by Terman and Miles (Bem, 1974; Block, 1973; Constantinople, 1973; Spence, Helmreich, & Stapp, 1974). In a comprehensive review of masculinity–femininity research Constantinople (1973) argued that existing masculinity–femininity measures were actually multidimensional, too tied to demographic factors such as social class and age, and overly defined in terms of gender stereotypes.

Partly in response to criticism like Constantinople's, researchers developed new scales that assessed masculinity and femininity as separate and independent dimensions (Bem, 1974; Heilbrun, 1976; Spence et al., 1974). In these newer scales, masculinity was defined in terms of instrumental personality traits and femininity in terms of expressive traits.[1] Although increasing the dimensionality of gender-related traits, these scales, perhaps even more than the earlier one-dimensional masculinity–femininity scales, explicitly defined masculinity and femininity in terms of gender stereotypes—this time, stereotypes of personality.

Separate masculinity and femininity scales led to new ways of classifying research subjects—individuals could be high on both masculinity and femininity (androgynous), low on both dimensions (undifferentiated), high on masculinity and low on femininity (masculine), or low on masculinity and high on femininity (feminine). Despite controversy over the proper classification of subjects on the basis of their assessed masculinity and femininity, a consensus seems to have emerged that researchers must assess the independent contributions of masculinity and femininity to dependent measures as well as their possible interaction through analysis of variance or regression techniques (Taylor & Hall, 1982).

In her early research on masculinity and femininity, Sandra Bem (1974, 1975; Bem & Lenney, 1976; Bem, Martyna, & Watson, 1976) proposed that extreme sex typing might be maladaptive and limiting, and that androgyny (the possession of equal degrees of masculinity and femininity, and in later research, the possession of high degrees of masculinity and femininity) might serve to define a new standard of mental health. In general, Bem's hypothesis that androgyny is linked to mental health has not been supported by research studies. Meta-analytic reviews of the relevant literature suggest that although both masculinity and femininity sometimes correlate with measures of mental health and adjustment, the effects of masculinity tend to be stronger and more consistent than those of femininity (e.g., Bassoff & Glass, 1982). In other words, because of their higher levels of masculinity, androgynous and masculine people tend to be better adjusted than feminine and undifferentiated people.

Since the mid-1970s hundreds of studies have explored the implications of the new generation of masculinity and femininity tests (Cook, 1985). In a comprehensive meta-analysis, Taylor and Hall (1982) found that masculinity scales tend to correlate with measures of instrumental behavior and femininity scales with measures of expressive behaviors. Spence and Helmreich (1978, 1980) have argued that "masculinity" and "femininity" scales are, in fact, instrumentality and expressiveness scales, and thus are at best weakly and indirectly related to other kinds of gender-related behaviors.

In a recent meta-analytic review of the relationship among masculinity, femininity, and cognitive abilities, Signorella and Jamison (1986) found masculinity to be positively related and femininity negatively related to performance on tests of mathematics and visual–spatial ability. These findings are interesting

[1] In keeping with the labels used by instruments such as the Personal Attributes Questionnaire and the Bem Sex-Role Inventory, we shall use the terms *masculinity* and *femininity* to refer to interpersonally instrumental and expressive traits. However, as other researchers have noted (e.g., Deaux, 1985; Paulhus, 1987; Spence, 1984), it may be more appropriate to label these constructs as *dominance* and *nurturance.*

because they link masculinity and femininity to behavioral domains that have shown documented sex differences (Hyde & Linn, 1984). Perhaps because of Bem's early hypotheses that androgyny was related to psychological adjustment and greater behavioral flexibility and adaptability (Bem, 1975; Bem & Lenney, 1976), research has often ignored the most obvious potential behavioral correlates of masculinity and femininity—namely, behaviors that show clear sex differences.

Gender Diagnosticity

The concept of gender diagnosticity has its roots in the diagnostic ratio approach to measuring stereotypes (McCauley & Stitt, 1978; McCauley et al., 1980). On the basis of Bayes' theorem, the diagnostic ratio is a measure of the degree to which the knowledge that another person is a member of a social group (e.g., Germans) revises an individual's probability estimate that the person possesses a given characteristic (e.g., is "nationalistic"). At its simplest, the diagnostic ratio is an individual's probability estimate that a member of a group possesses a given trait divided by the individual's probability estimate that people in general possess the trait. McCauley and Stitt (1978), for example, found that, on average, a group of junior college women estimated that 56% of Germans were "extremely nationalistic," whereas they estimated that 35% of "all the world's people" possessed the same trait. These estimates (converted to probabilities) yield a diagnostic ratio of 1.59, indicating that Germans were perceived to possess this trait considerably more than people in general.

McCauley and Stitt, in general, applied their concept of a diagnostic ratio to subjective beliefs. In contrast, we apply our concept of gender diagnosticity to observable behaviors. Furthermore, we reverse the direction of Bayesian inference: Whereas McCauley and Stitt were concerned with estimating an individual's perceived traits given his or her group membership, we are concerned with the converse—estimating an individual's likelihood of being a member of a group given possession of a trait.

A simple example using actual data will serve to illustrate the concept of gender diagnosticity. Martin (1987) asked a group of male and female Canadians to estimate the percentage of North American men and women who possess various traits (e.g., who are "aggressive") and also to rate whether or not the same traits were self-descriptive. For example, Martin's subjects estimated that 66% of men and 40% of women are aggressive. In their self-descriptions, 55% of the men and 22% of the women surveyed described themselves as aggressive.

In this example, McCauley and Stitt's diagnostic ratio applies to subjects' estimates of the percentage of men and women who are aggressive compared with their estimates for people in general (which in this case can be taken to be the average of the estimates for men and women). Gender diagnosticity addresses a related, yet quite different, issue: Given that an individual rates *aggressive* to be self-descriptive, what is the probability that that individual is female (or male)? Applying Bayes' theorem to this example, we note that p(female|describes self as aggressive) = p(female) × p(describes self as aggressive|female)/ p(describes self as aggressive). Assuming for the sake of simplicity that the base-rate probability of being female—p(female)—is .5, we compute that the probability that an individual is female given that he or she describes self as aggressive is .5 × .22/.385 = .29. The probability that an individual is male given the same diagnostic information is simply 1 minus the computed probability for being female, or .71.

This simple example of gender diagnosticity suggests a number of important points. First, gender diagnosticity indicates in probabilistic terms how "male-like" or "female-like" an individual is given one or more pieces of gender-diagnostic information. To compute gender diagnosticity from Martin's (1987) data, it is irrelevant whether or not people are accurately reporting their aggressiveness. Indeed, one can take the gender diagnostic probability computed in our example at face value: Given that an individual reports on a self-report questionnaire that he or she is aggressive, the probability is .29 that the individual is female and .71 that the individual is male.

It should be clear from our example that the gender diagnosticity of a given behavioral indicator may vary over time and across different populations of men and women. For example, the behavior "wearing pants" was undoubtedly more gender diagnostic 100 years ago than it is today in the United States. The variability of gender diagnosticity over time and across populations is a potential advantage over previous approaches to assessing gender-related traits. We argue that gender-related individual differences can be defined only in terms of the behaviors that differentiate men and women in a particular population in a particular culture during a particular historical era. The method of gender diagnosticity makes no assumptions as to why specific behaviors serve to probabilistically predict an individual's gender in such populations.

Gender diagnosticity provides a convenient metric (a probability) that can be computed for any gender-related behavior. It seems likely that gender-related individual differences, like most others, express themselves differently over the individual's life span (Caspi, 1987; Maccoby, 1987). Thus, for example, gender-related individual differences may be displayed in young children through play preferences, crying, and physical aggressiveness, whereas they may be displayed in adults more through hobbies, career choices, behaviors in close relationships, styles of dress, and nonverbal behaviors. Gender diagnosticity provides a uniform metric for assessing gender-related dispositions over the life span.

In the simple example we used earlier to illustrate the concept of gender diagnosticity, the conditional probability of being male or female was computed on the basis of a single piece of diagnostic information—whether or not an individual rated aggressive to be self-descriptive. A more reliable measure of gender diagnosticity would necessarily be based on many gender-related indicators.

How might gender diagnostic probabilities be computed on the basis of sets of indicators? The statistical technique of discriminant analysis provides a useful means of computing such probabilities. Discriminant analysis identifies the linear combination of predictor variables—the discriminant function—that optimally discriminates membership in two (or more) categories (see Goldstein & Dillon, 1978; Hand, 1981; Lachenbruch, 1975). To compute gender diagnostic probabilities for individuals in a particular population of males and females, a discriminant analysis is applied to a set of gender-related variables assessed for all individuals in the population; such an analysis identifies the weighted combination of some subset of these variables that best classifies individuals as male or female. Bayes' theorem is then applied to individuals' discriminant function scores to compute the probability that an individual is male or female. (The computation of such probabilities is a standard option in computerized statistical packages that perform discriminant analyses.) In essence, the diagnostic probabilities produced by such discriminant analyses indicate how male-like or female-like an individual's set of diagnostic indicators is.

The usual purpose of discriminant analysis is, of course, to predict membership in classes—in our case, to predict whether an individual is male or female on the basis of some set of predictor variables. We are proposing a different use for discriminant analysis: to compute diagnostic probabilities of class membership that can then serve as individual difference measures. In other words, we propose defining gender-related individual differences explicitly in terms of the behavioral indicators that best serve to differentiate the two social groups that index the individual difference—in the case of gender-related individual differences, men and women. Unlike contemporary masculinity and femininity scales, we do not define gender-related individual differences in terms of gender stereotypes, but rather in terms of the behavioral differences between the sexes that occur in a specific population of individuals. We propose that gender diagnostic probabilities serve not only to classify individuals as male or female but also to measure meaningful gender-related individual differences within the sexes.

To demonstrate the utility of gender diagnostic probabilities in predicting gender-related behaviors, we conducted a study in which we computed such probabilities for 117 male and 110 female subjects on the basis of their ratings of occupational preferences. Subjects also completed the two masculinity and femininity scales used most frequently in recent research: those of the Personal Attributes Questionnaire (PAQ; Spence et al., 1974) and the Bem Sex-Role Inventory (BSRI; Bem, 1974, 1981a). The primary focus of our study was to compare gender diagnosticity measures with these scales in predicting a wide sample of gender-related behaviors.

In an excellent review of research on sex differences, Eagly (1987) noted that researchers have frequently failed to obtain reliable, aggregated measures of gender-related behaviors when studying gender-related behaviors and dispositions. Taking note of this fundamentally important methodological point, we assessed our subjects on a broad sample of behaviors, many of which show reliable sex differences according to recent meta-analytic reviews. These include measures of mathematical ability, visual–spatial ability, stereotypic masculinity–femininity of college majors, self-reported aggressiveness, smiling, rated masculinity–femininity of physical appearance, rated masculinity–femininity of written self-descriptions, and rated masculinity–femininity of handwriting styles. These gender-related behaviors, both individually and in aggregate, served as the criterion measures against which we validated gender diagnosticity.

Method

Subjects

Subjects were 227 (117 male and 110 female) undergraduate introductory psychology students at California State University, Fullerton.

Materials

Subjects were administered a questionnaire packet in class that included the PAQ (as presented in Spence & Helmreich, 1978) and the BSRI (Bem, 1974; short form scales were computed in the current research as described in Bem, 1981a). The packet's cover sheet asked subjects to report demographic information, including their age, sex, whether they were born in the United States, and their ethnic group (Hispanic, Asian, Black, Caucasian, American Indian, or other). Subjects were also asked to list their major. If they had not declared a major, they were asked to list the two majors they thought they were most interested in pursuing. Finally, the questionnaire packet included a blank sheet of paper titled "Personal Narrative," which included the following instructions: "In the space provided below, in your own handwriting, please describe yourself in terms of your personality and your interests."

Subjects were subsequently scheduled for a laboratory session that lasted approximately 1 hr. During this session, subjects took the Vandenberg Mental Rotations Test (Vandenberg & Kuse, 1978), and they completed the verbal aggressiveness and physical aggressiveness scales of the Interpersonal Behavior Survey (Mauger & Adkinson, 1980; sample items are "I usually tell people off when they disagree with me" and "There are times when I would like to pick fist fights"). Subjects also rated their preference (by three possible responses: *like, indifferent,* or *dislike*) for 70 occupations selected from Part I of the Strong–Campbell Interest Inventory, Form T325 (Campbell & Hansen, 1981). Finally, subjects were asked to report whether or not they participated in 22 hobbies (such as aerobics, working on cars, clothes shopping, weight-lifting, dancing, and cooking). Pretesting had indicated that each of the hobbies in our list was endorsed by large numbers of students in our subject population and, furthermore, that there were significant sex differences in the frequency with which subjects reported participating in many of these hobbies.

Procedure

During the laboratory sessions subjects were run in groups ranging in size from 3 to 12. Two or three research assistants were present at these sessions. Prior to the laboratory sessions, subjects had completed the questionnaire packet containing the PAQ and BSRI. At the start of the session a research assistant read a prepared statement to subjects describing the laboratory procedures.

The Vandenberg Mental Rotations Test—a timed 8-min test—was then administered. During this time, the research assistants rated each subject, on the basis of his or her physical appearance, on a 7-point scale of femininity–masculinity ranging from *extremely feminine* (1) to *extremely masculine* (7). We used a unidimensional rating scale of femininity–masculinity because previous studies have consistently shown that lay conceptions hold femininity and masculinity to be opposite sides of a single continuum (Deaux, 1987; Deaux & Lewis, 1984; Major, Carnevale, & Deaux, 1981).

After completing the mental rotations test, subjects were given questionnaire packets that included the verbal and physical aggressiveness scales of the Interpersonal Behavior Survey and the occupation and hobby preference questionnaires. As subjects were completing these forms, they were briefly interrupted, one at a time, and escorted to a neighboring room where they were photographed with an instant camera. Subjects were asked to stand against a wall while being photographed, and photographs were framed to include subjects' entire bodies. After their photographs were taken, subjects returned to the original room and completed their questionnaire packets.

Judges' ratings of femininity–masculinity. Subsequent to the laboratory session, panels of research assistants rated the femininity–masculinity of subjects' handwriting, self-descriptive narratives, and photographed appearances using the same 7-point scale described above. (For a discussion of similar methods used to rate nonverbal masculinity–femininity, see Lippa, 1978, 1983; Lippa, Valdez, & Jolly, 1983.) In rating handwriting, judges observed the last line of photocopied handwriting in subjects' self-descriptive narratives. Judges were instructed to make their ratings solely on the basis of the style of the handwriting (see Lippa, 1977, for another research example of rating the perceived masculinity–femininity of handwriting). In rating the masculinity–femininity of self-descriptive narratives, judges rated type-written transcriptions of the narratives.

Six judges independently rated all subjects' handwriting samples. Seven judges independently rated subjects' self-descriptive narratives, and nine judges independently rated subjects' photographs. Judges' ratings were averaged to produce composite ratings of femininity–masculinity based on subjects' handwritings, self-descriptive narratives, and photographs.

Nine judges independently rated the degree to which subjects smiled in their photographs. For this rating, judges used a 3-point scale: 1 = *no smile*, 2 = *partial smile*, and 3 = *full smile*. Again, judges' ratings were averaged to form one composite rating of smiling.

To assess the stereotypic masculinity–femininity of subjects' college majors, a list of all the majors available at California State University, Fullerton, was presented to 35 social psychology students, who rated each major on a 7-point scale of how stereotypically masculine–feminine it was. This scale was identical to the other scales used by raters except that it was reversed in direction. The masculinity–femininity of a subject's declared major was defined as the mean rated stereotypic *m–f* of that major. If a subject did not declare a major, it was the rated m–f of his or her most preferred major. Some subjects (21 of 117 men and 22 of 110 women) listed neither a declared major nor preferred majors or listed a major that had not been rated on masculinity–femininity and thus had missing values for the m–f of their major.

SAT scores. The SAT verbal and math scores of most subjects (102 of 117 men and 97 of 110 women) were obtained from the Admissions and Records Office of the university. SAT scores were unavailable for some subjects because they had taken alternate college entrance examinations (such as the ACT).

Results

Reliability of Criterion Measures and Sex Differences

Table 1 presents the mean scores for men and women on 12 assessed criterion variables: SAT verbal scores; SAT math scores; mental rotation scores; rated m–f of self-descriptions; m–f of majors; self-reported verbal, physical, and total aggressiveness (sum of all aggression scale items); rated smiling from photos; rated in-person m–f; m–f rated from photos; and m–f rated from handwriting.[2] Reliabilities (Cronbach's coefficient alpha) are presented for mean judges' ratings, computed for all subjects and for men and women separately.

In general, criterion measures showed acceptable levels of reliability. The least reliable of the rated variables was in-person masculinity–femininity, and this relatively low reliability was likely due to the fact that these ratings were obtained from only two or three raters. (To compute reliabilities for these ratings, we averaged two of the ratings for those subjects rated by three judges, thus producing two ratings for all subjects.)

In general, our data showed sex differences that were consistent with the findings of recent meta-analytic reviews (Eagly, 1987; Hyde & Linn, 1986): Men, on average, scored higher than women on the math SAT and the Vandenberg test of mental rotation, and mental rotation scores showed the largest sex difference among the three cognitive variables. Self-report scales showed men to be more aggressive than women, and consistent with recent meta-analyses (e.g., Eagley & Steffen, 1986) this difference was greater for physical than for verbal aggressiveness. On average, men chose college majors that were rated to be more stereotypically masculine than those chosen by women. Consistent with meta-analytic reviews of sex differences in nonverbal behavior (Hall, 1984), women smiled significantly more in photographs than men. Not surprisingly, raters judged men, both in person and from photographs, to be more masculine and less feminine than women. Raters perceived the handwriting of men to be, on average, significantly more masculine in style than women's. Finally, men's self-descriptive paragraphs were rated as more masculine than those of women. The only variables listed in Table 1 that failed to show sex differences at conventional levels of statistical significance were SAT verbal scores (see Hyde & Linn, 1988, for a recent meta-analytic review that concludes that sex differences no longer exist in measures of verbal ability) and self-reported verbal aggressiveness.

A composite measure of gender-related criterion behaviors was computed by summing Z scores of eight variables that showed the strongest sex differences: SAT math, mental rotation, self-reported total aggressiveness, smiling, m–f of college major, m–f of self-descriptions, m–f rated from photos, and m–f rated from handwriting styles. All Z scores were keyed so that means for men were positive and means for women were negative. Masculinity–femininity rated in person was not included in the composite because of its low within-sex reliability. The composite measure was computed only for subjects possessing complete data on all of the eight component variables (82 men and 77 women). We hoped that forming the composite would serve to extract particularly that variance in variables accounted for by gender and gender-related individual differences. The aggregated measure of gender-related criteria showed reasonably good reliability over all subjects ($\alpha = .76$) and correlated .86 with sex of subject ($p < .001$).

Computing Gender Diagnosticity

Our primary measure of gender diagnosticity was computed from subjects' ratings of their preferences for the 70 occupations selected from the Strong–Campbell Interest Inventory. To compute gender diagnostic probabilities and to assess the reliability of these probabilities, we conducted seven discriminant analyses using the discriminant analysis procedure of SPSS/PC+ (*SPSS/PC+ V2.0 Base Manual* [Norusis, 1988] and *Advanced Statistics, SPSS/PC+* [Norusis, 1986]) with stepwise selection of variables. Each discriminant analysis was con-

[2] Foreign-born subjects (56 of 199 subjects, or 28% of the sample assessed on SAT scores) received, on average, lower SAT verbal scores than native-born subjects (*M*s = 336 and 414, respectively), $t(197) = 5.49$, $p < .001$. However, foreign-born subjects did not receive lower SAT math scores than native-born subjects (*M*s = 481 and 467, respectively), $t(197) = -.89$, *ns*. The percentage of male subjects who were foreign born (30%) was roughly the same as the percentage of women who were foreign born (26%), and thus this factor seems unlikely to be responsible for differences between male and female subjects. Both foreign-born and native-born subjects showed significant sex differences in SAT math and mental rotation scores, consistent with the results for the total sample of subjects.

Table 1
Gender-Related Criteria and Their Means for Men and Women

Criterion variable	Male *M*	*n*	Female *M*	*n*	Correlation with sex	2-tailed *p* level
SAT verbal	404	102	381	97	.12	.09
SAT math[a]	505	102	435	97	.33	.001
Mental rotation[a]	19.6	117	13.6	110	.38	.001
m–f of self-descriptions[a] (.92 .87 .88)	4.90	117	2.95	110	.63	.001
m–f of major[a,b]	3.27	96	3.82	88	.37	.001
Physical aggressiveness	4.40	117	2.48	110	.20	.001
Verbal aggressiveness	5.78	117	5.17	110	.10	.13
Total aggressiveness[a]	14.56	117	11.81	110	.25	.001
Smiling[a] (.98 .97 .98)	1.49	117	2.07	110	.39	.001
m–f in person (.84 .38 .17)	5.27	117	2.75	110	.89	.001
m–f from photo[a] (.98 .82 .85)	5.82	117	2.12	110	.95	.001
m–f handwriting[a] (.87 .65 .77)	5.18	117	3.13	110	.71	.001
Composite	.49	82	−.52	77	.86	.001

Note. Numbers in parentheses in left column are alpha coefficients computed for all subjects, for men, and for women.
[a] *Z* scores of these variables summed to form the composite. [b] Scale reversed from others.

ducted on 10 of the rated occupations, and each analysis yielded the Bayesian probability, computed from individuals' discriminant function scores, that a given subject was male (or, by subtracting this probability from one, female).

Thus, each subject was assigned seven gender diagnostic probabilities, each based on a distinct set of 10 occupational preference ratings. These probabilities were quite reliable when combined into a single average probability: Coefficient alpha was .87 over all subjects, .65 for men, and .66 for women. The last two reliabilities are noteworthy because they suggest that the mean gender diagnostic probabilities assessed relatively stable individual differences within the sexes.

Figure 1 presents histograms of the gender diagnostic probabilities for men and women. The histograms show clearly that these probabilities classified subjects as male or female quite accurately. (A subject with a probability greater than .5 is classified as male, and a subject with a probability less than .5 is classified as female.) One hundred eight of 117 men were correctly classified as male, and 97 of 110 women were correctly classified as female. The closer an individual's computed probability was to one or to zero, the "surer" the discriminant analyses were that the individual was male or female, respectively. The closer an individual's probability was to .5, the less "sure" the discriminant analyses were of an individual's gender.

As the histograms indicate, the gender diagnostic probabilities of men and women showed an on-average difference (mean probabilities were .69 for men and .32 for women). Indeed, the correlation between sex and gender diagnosticity based on occupation ratings was .79 ($p < .001$). Thus, sex accounted for approximately 64% of the variance in these gender diagnostic probabilities. The large correlation between sex and diagnostic probabilities is not surprising given that the goal of the discriminant analyses was to classify men and women as accurately as possible given the sets of diagnostic information. Despite the large sex difference in diagnostic probabilities, the histograms also make clear that there was considerable variation within each sex as well, and some overlap in the distributions for the two sexes. Our central research question was this: Did this within-sex variance reflect meaningful gender-related individual differences?

We computed a second measure of gender diagnosticity based on subjects' ratings of whether or not they participated in each of 22 hobbies. Discriminant analyses were carried out on two sets of 5 and two sets of 6 hobbies each. Again, each discriminant analysis yielded a gender diagnostic probability for each subject, and again we averaged these probabilities to form one mean diagnostic probability based on hobby ratings. Because these probabilities were based on fewer total items (22 hobbies versus 70 occupations) and fewer component probabilities (4 versus 7), and because hobby participation was rated dichotomously rather than with the 3-degree preference rating used for occupations, it is not surprising that they were less reliable than probabilities based on occupations—coefficient alpha was .69 for all subjects, .29 for men, and .47 for women.[3]

[3] We suggest that in future studies on gender diagnosticity researchers use relatively large numbers of gender diagnostic behavioral indicators to compute gender diagnostic probabilities. If these indicators are questionnaire items, we recommend using 5- or 7-point scales (rather than the dichotomous or 3-point ratings obtained in the current study). Lippa (1990b) has conducted a follow-up study further investigating the correlates of gender diagnosticity as assessed from preference ratings for occupations and other kinds of behaviors. In this follow-up study, subjects rated 131 occupations (the full Strong–Campbell set) on a 5-point scale of preference to permit the computation of more

The aggregated diagnostic probabilities based on hobby ratings were again highly correlated with sex of subject, $r = .75$ ($p < .001$), and these diagnostic probabilities correctly classified 101 of 117 men to be male and 94 of 110 women to be female.

Interrelation of Masculinity, Femininity, and Gender Diagnosticity

Were our gender diagnosticity measures related to assessed masculinity and femininity? Table 2 presents the intercorrelations of our two measures of gender diagnostic probabilities, PAQ scales, and BSRI scales. These correlations suggest that the two gender diagnosticity measures correlated highly with one another and less so with the PAQ and BSRI scales.

To investigate further the dimensionality of these individual difference measures, we carried out factor analyses (principal components, orthogonal varimax rotation) on the following measures: PAQ masculinity, PAQ femininity, PAQ masculinity–femininity, BSRI masculinity, BSRI femininity, gender diagnosticity based on occupations, and gender diagnosticity based on hobbies. Three analyses were conducted—one on data for all subjects, one on data for men, and one on data for women. Each of these factor analyses showed consistent three-factor solutions. The rotated factor matrices for each solution are presented in Table 3.

Three distinct and quite pure factors were identified: masculinity, femininity, and gender diagnosticity. The PAQ and BSRI masculinity scales loaded highly on the masculinity factor; the PAQ and BSRI femininity scales loaded highly on the femininity factor; and the two gender diagnosticity measures loaded highly on the gender diagnosticity factor. The PAQ masculinity–femininity scale proved to be a hybrid scale that loaded on both masculinity and femininity, but not on gender diagnosticity. All three factor analyses suggested that the two gender diagnosticity measures were assessing a construct distinct from masculinity and femininity. It is particularly noteworthy that the factor structure maintained itself in the separate analyses for men and women.[4]

Correlations of Gender Diagnostic Probabilities, Masculinity, and Femininity with Criterion Measures of Gender-Related Behaviors

Tables 4, 5, and 6 present the correlations between the seven individual difference measures assessed in our study and the criterion measures. These were computed for all subjects (Table 4), men (Table 5), and women (Table 6). The correlations computed for all subjects indicate that gender diagnostic probabilities, particularly those based on occupational preferences, predicted virtually all criterion gender-related behaviors better than PAQ or BSRI scales. The only exception to this pattern occurred for self-reported aggression, which correlated somewhat more strongly with PAQ and BSRI femininity scales than with gender diagnosticity measures. Gender diagnostic probabilities based on occupational preferences correlated very strongly with the composite measure of gender-related behavior, $r = .82$, $p < .001$, and accounted for more than twice as much variance in this measure as any PAQ or BSRI scale. The correlation of gender diagnostic probabilities with the composite was significantly larger than any of the correlations of PAQ or BSRI scales with the composite ($p < .001$), using t tests for differences in correlations when two variables are correlated with the same third variable (see McNemar, 1962, p. 140).

To investigate further the comparative predictive utility of gender diagnosticity and PAQ and BSRI scales, we conducted a stepwise regression analysis, with the composite criterion serving as the dependent variable and gender diagnosticity based on occupations, PAQ scales, and BSRI scales as predictor variables. This regression showed that gender diagnosticity accounted for almost all explainable variance. Gender diagnosticity was entered into the equation in the first step and PAQ femininity in the second and final step. Gender diagnosticity accounted for 67% of the variance, and PAQ femininity accounted for only an additional 2%.

In general, correlations computed for male subjects showed a pattern of results similar to those for all subjects (see Table 5)—that is, gender diagnosticity correlated more strongly with

---

reliable gender diagnostic probabilities. In this research, alpha coefficients for gender diagnosticity computed from occupational preferences increased to .92 over all subjects, .76 for men, and .78 for women.

Ironically, the computation of gender diagnostic probabilities may not produce meaningful individual differences within the sexes if the diagnostic indicators are too strongly correlated with sex, for then discriminant analyses will classify subjects with complete "certainty." That is, many men will be assigned a probability of being male near 1.0 and many women will be assigned a probability of being male near 0.0, and thus there will be little within-sex variance.

In such cases, it may be appropriate to use discriminant function scores as individual difference measures rather than diagnostic probabilities. In the research reported here, discriminant function scores produced results similar to those reported for diagnostic probabilities. We encourage researchers to examine the utility of both discriminant function scores and diagnostic probabilities as individual difference measures.

[4] Factor-analytic studies of the long form of the BSRI have noted that the items "masculine" and "feminine" tend not to correlate highly with other items of their respective scales and correlate more strongly than other scale items with sex of subject (Gaudreau, 1977; Pedhazur & Tetenbaum, 1979). These items are of interest because they represent the most direct assessment of subjects' self-ascribed masculinity and femininity. The correlations of gender diagnosticity based on occupations with the 2-item scale defined by BSRI items "masculine" minus "feminine" were .80 for all subjects, .26 for men, and .39 for women (all $ps < .01$).

When we included this 2-item scale in factor analyses of BSRI and PAQ scales and gender diagnosticity measures, the loadings of the 2-item scale on the gender diagnosticity, masculinity, and femininity factors were, respectively, .86, .23, and −.25 for all subjects, .30, .62, −.03 for men; and .67, −.11, and −.45 for women. Thus, for subjects as a whole and for women, self-ascribed masculinity and femininity proved more related to gender diagnosticity measures than to PAQ and BSRI scales, whereas for men, self-ascribed masculinity and femininity were most related to PAQ and BSRI masculinity (i.e., interpersonal dominance). Thus, our data suggest that self-ascribed masculinity and femininity may mean different things to men and women.

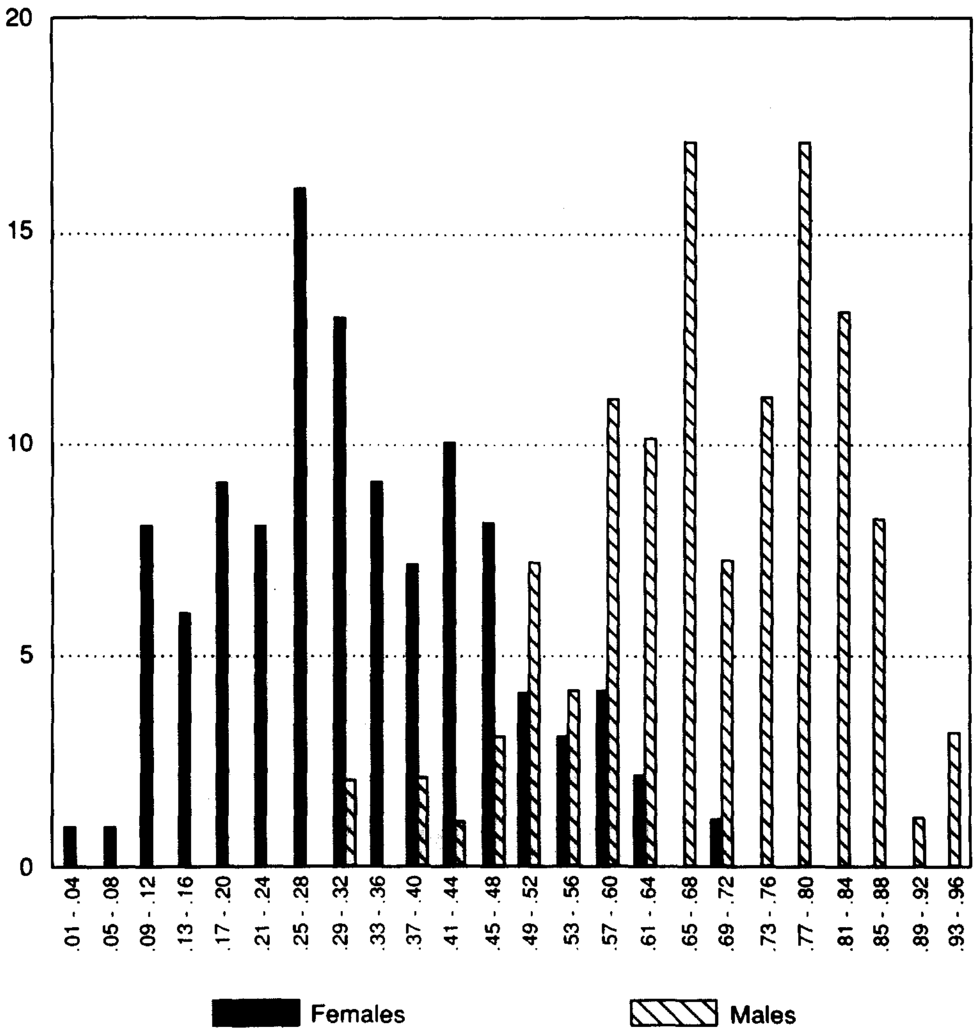

Figure 1. Histograms of gender diagnostic probabilities for men and women. (For women, $N = 110$, $M = .32$, and $SD = .14$; for men, $N = 117$, $M = .69$, and $SD = .13$. Coefficient alpha for all subjects = .87, for men = .65, and for women = .66.)

criterion behaviors than did PAQ and BSRI scales. Indeed, gender diagnostic probabilities based on occupation ratings correlated with the composite criterion measure of gender-related behaviors .54 ($p < .001$). The largest correlation between PAQ or BSRI scales and the composite was that with BSRI femininity ($r = -.24$, $p < .05$), and this was significantly smaller than the correlation between the composite and gender diagnosticity, $t(79) = 2.61$, $p < .01$. Gender diagnosticity accounted

Table 2
Intercorrelations of Gender Diagnosticity Measures, PAQ Scales, and BSRI Scales

Measure	Gender diagnosticity: occupation	Gender diagnosticity: hobby	PAQ masculinity	PAQ femininity	PAQ M–F	BSRI masculinity
Gender diagnosticity: hobby	***.71***					
PAQ masculinity	**.31**	**.22**				
PAQ femininity	**−.34**	**−.24**	.09			
PAQ m–f	**.42**	**.31**	**.46**	**−.34**		
BSRI masculinity	**.35**	**.29**	***.76***	.01	**.46**	
BSRI femininity	**−.53**	**−.38**	−.04	***.71***	**−.45**	.00

Note. $N = 227$. PAQ = Personal Attributes Questionnaire. BSRI = Bem Sex-Role Inventory. Boldface correlations are significant at $p < .001$. Italicized correlations are between two different measures of the same construct.

Table 3
Factor Analyses of PAQ Scales, BSRI Scales, and Two Measures of Gender Diagnosticity (Varimax Orthogonally Rotated Factor Matrices)

	All subjects		
Measure	Masculinity factor	Femininity factor	Gender diagnosticity factor
All subjects (N = 227)			
PAQ masculinity	**.91**	.07	.13
PAQ femininity	.07	**.90**	−.11
PAQ m–f	**.65**	**−.54**	.12
BSRI masculinity	**.89**	.06	.19
BSRI femininity	.00	**.86**	−.32
Gender diagnosticity: occupation	.25	−.32	**.83**
Gender diagnosticity: hobby	.13	−.13	**.92**
Men (n = 117)			
PAQ masculinity	**.87**	.16	.07
PAQ femininity	.18	**.89**	−.06
PAQ m–f	**.64**	**−.50**	.17
BSRI masculinity	**.87**	.22	.05
BSRI femininity	.09	**.92**	−.10
Gender diagnosticity: occupation	.27	−.32	**.71**
Gender diagnosticity: hobby	.05	−.04	**.92**
Women (n = 110)			
PAQ masculinity	**.92**	.10	.09
PAQ femininity	.02	**.82**	.11
PAQ m–f	**.65**	**−.47**	−.05
BSRI masculinity	**.90**	.07	.04
BSRI femininity	.03	**.86**	−.22
Gender diagnosticity: occupation	.13	−.27	**.72**
Gender diagnosticity: hobby	−.04	.15	**.77**

Note. PAQ = Personal Attributes Questionnaire. BSRI = Bem Sex-Role Inventory. Boldface factor loadings are greater than .35.

for more than five times as much variance in the composite criterion as any PAQ or BSRI scale.

A stepwise regression analysis was conducted for male subjects like that reported earlier for all subjects. Only gender diagnosticity was entered into the regression equation, indicating that no PAQ or BSRI scales explained a significant amount of additional variance for male subjects.

Female subjects showed a somewhat different pattern of re-

Table 4
Correlations of Gender Diagnosticity Measures, PAQ Scales, and BSRI Scales with Criterion Measures for All Subjects

Measure	Composite	SAT math	SAT verbal	Mental rotation	m–f major	m–f self	m–f photo	m–f live	m–f handwriting	Smile	Total aggressiveness
Gender diagnosticity: occupation	**.82**	**.40**	**.24**	**.39**	**−.40**	**.67**	**.79**	**.76**	**.65**	**−.38**	**.28**
Gender diagnosticity: hobby	**.67**	**.35**	**.13**	**.36**	**−.29**	**.54**	**.75**	**.74**	**.58**	**−.26**	**.14**
PAQ masculinity	**.19**	.09	**.14**	**.12**	**−.13**	**.29**	**.26**	**.24**	**.13**	−.05	.03
PAQ femininity	**−.43**	**−.20**	−.08	**−.13**	**.25**	**−.36**	**−.31**	**−.35**	**−.21**	**.18**	**−.31**
PAQ m–f	**.36**	.05	−.01	**.17**	**−.23**	**.36**	**.37**	**.39**	**.29**	**−.13**	**.13**
BSRI masculinity	**.25**	.06	**.15**	**.11**	−.06	**.32**	**.35**	**.35**	**.22**	−.07	.11
BSRI femininity	**−.52**	**−.28**	.00	**−.24**	**.33**	**−.46**	**−.45**	**−.46**	**−.35**	**.25**	**−.37**

Note. PAQ = Personal Attributes Questionnaire. BSRI = Bem Sex-Role Inventory. Boldface correlations are significant at the .05 level or less.

Table 5
Correlations of Gender Diagnosticity Measures, PAQ Scales, and BSRI Scales with Criterion Measures for Men

Measure	Composite	SAT math	SAT verbal	Mental rotation	m–f major	m–f self	m–f photo	m–f live	m–f handwriting	Smile	Total aggressiveness
Gender diagnosticity: occupation	**.54**	**.22**	**.28**	**.22**	**−.17**	**.38**	.05	**.16**	.10	−.04	**.20**
Gender diagnosticity: hobby	**.37**	**.26**	**.18**	**.18**	−.02	**.21**	.14	**.23**	**.16**	−.03	.09
PAQ masculinity	.01	−.05	.09	.08	.05	.09	.12	.11	.05	.02	−.02
PAQ femininity	−.18	**−.17**	−.15	−.02	.12	**−.24**	.10	−.02	.14	.01	**−.35**
PAQ m–f	**.20**	−.09	.05	**.17**	−.12	.16	.10	**.27**	.10	.02	.08
BSRI masculinity	.03	−.08	.13	.06	.15	.03	**.20**	**.22**	.13	.06	.10
BSRI femininity	**−.24**	**−.19**	−.06	−.08	**.20**	**−.28**	−.06	−.12	**.19**	.02	**−.32**

Note. PAQ = Personal Attributes Questionnaire. BSRI = Bem Sex-Role Inventory. Boldface correlations are significant at the .05 level or less.

sults from male subjects (see Table 6). Gender diagnosticity based on occupational preferences and the PAQ and BSRI femininity scales correlated about equally with the composite criterion, *r*s = .43, −.35, and −.44, respectively. A stepwise regression analysis was conducted to determine which individual difference variables best predicted the composite criterion for female subjects. This regression analysis identified two significant predictors: BSRI femininity and gender diagnosticity, both of which accounted for significant amounts of variance (19% and 12%, respectively). Thus, although gender diagnosticity and BSRI femininity correlated almost equally with the composite criterion of gender-related behaviors, each accounted for largely independent variance. In the regression equation, BSRI femininity and gender diagnosticity were weighted almost identically (βs = −.34 and .33, respectively). The multiple correlation was .54 ($p < .001$), accounting for 29% of the variance in the composite criterion.

In general, both regression analyses and the simple correlations reported in Tables 4, 5, and 6 suggested that PAQ and BSRI femininity scales correlated more strongly with gender-related criteria than did the PAQ or BSRI masculinity scales and that femininity correlated more strongly with criteria for female than for male subjects.

Tables 4, 5, and 6 also present correlations between gender diagnostic probabilities and component measures of gender-related criteria. The correlations in Table 5 suggest that for men, gender diagnostic probabilities correlated most strongly with cognitive ability measures (SAT math, SAT verbal, and mental rotation), m–f rated from self-descriptive narratives, m–f of college major, and self-reported aggressiveness. Gender diagnosticity correlated more poorly with measures of men's gender-related nonverbal behaviors—rated m–f of photos, rated m–f of handwriting, rated in-person m—f, and smiling.

The gender diagnostic probabilities of female subjects correlated more consistently than men's with measures of nonverbal behaviors. For women, BSRI and PAQ femininity scores as well as gender diagnostic probabilities frequently showed significant correlations with criterion gender-related behaviors.

Factor Analyses of Individual Difference and Criterion Variables

The correlations presented in Tables 5 and 6 suggest a different patterning of variables for men and women. To examine further the structure of our data, we conducted factor analyses on all variables—both individual difference measures and criterion variables. Table 7 presents the rotated factor matrix for male subjects, and Table 8 presents the corresponding matrix for female subjects.

These factor analyses indicated that gender-related behaviors were factorially complex, both for men and women. Consistent with the factor analyses reported earlier, these analyses indicated that PAQ and BSRI femininity and masculinity scales were factorially distinct from gender diagnosticity measures. Furthermore, consistent with our correlational analyses,

Table 6
Correlations of Gender Diagnosticity Measures, PAQ Scales, and BSRI Scales with Criterion Measures for Women

Measure	Composite	SAT math	SAT verbal	Mental rotation	m–f major	m–f self	m–f photo	m–f live	m–f handwriting	Smile	Total aggressiveness
Gender diagnosticity: occupation	**.43**	**.28**	**.21**	.08	**−.20**	**.29**	**.21**	.14	**.24**	**−.20**	.09
Gender diagnosticity: hobby	−.11	.06	−.07	.07	−.01	.05	**.25**	**.26**	.00	.10	**−.25**
PAQ masculinity	−.04	.05	.14	−.03	−.14	**.24**	.00	−.08	**−.18**	.09	−.06
PAQ femininity	**−.35**	.02	.15	.01	**.19**	**−.19**	**−.21**	**−.32**	−.06	.13	−.12
PAQ m–f	.01	−.10	**−.22**	−.12	−.12	**.21**	.06	.03	.01	.00	−.01
BSRI masculinity	−.15	−.07	.10	−.10	−.01	**.24**	−.02	−.01	−.15	.08	−.06
BSRI femininity	**−.44**	−.10	**.26**	−.06	**.21**	**−.19**	**−.17**	−.14	**−.25**	.17	**−.26**

Note. PAQ = Personal Attributes Questionnaire. BSRI = Bem Sex-Role Inventory. Boldface correlations are significant at the .05 level or less.

Table 7
Factor Analysis of PAQ Scales, BSRI Scales, Gender Diagnosticity Measures, and Criterion Measures for Men (Varimax Orthogonally Rotated Factor Matrix)

Measure	Factor 1 (Cognitive Abilities)	Factor 2 (Macho)	Factor 3 (Work Self-Concept)	Factor 4 (Nonverbal Behavior)	Factor 5 (Femininity)	Factor 6 (Masculinity)
Gender diagnosticity: occupation	**.44**	.30	**.44**	.13	−.25	.20
Gender diagnosticity: hobby	**.40**	**.60**	.16	.00	−.05	−.10
PAQ masculinity	.03	−.01	.02	.04	.15	**.81**
PAQ femininity	−.08	.13	−.04	.03	**.91**	.11
PAQ m–f	.12	.14	.20	−.15	**−.39**	**.65**
BSRI masculinity	−.01	.19	−.15	.04	.11	**.82**
BSRI femininity	−.06	.01	−.17	.09	**.89**	−.01
Criterion measures						
Composite	**.53**	.32	**.51**	**.44**	−.21	−.02
SAT math	**.82**	−.09	.07	.08	−.04	−.17
SAT verbal	**.82**	−.03	−.21	.16	−.11	.18
Mental rotation	**.61**	.14	.24	−.32	.21	.20
Major	.04	.07	**−.89**	.09	.05	.07
Self-description	.10	.06	**.55**	**.50**	−.30	.09
Photo	−.14	**.80**	−.16	.18	.10	.09
In-person m–f	.02	**.65**	.09	.18	−.05	.34
Handwriting	.02	.09	.13	**.67**	.17	.30
Smile	−.09	−.13	.14	**−.69**	−.03	.31
Total aggressiveness	−.07	**.43**	.03	.10	**−.57**	−.04

Note. PAQ = Personal Attributes Questionnaire. BSRI = Bem Sex-Role Inventory. Boldface factor loadings are greater than .35.

gender diagnostic probabilities, more than masculinity or femininity, loaded on factors that tapped substantive gender-related behaviors. For men, gender diagnosticity measures loaded on factors tapping cognitive abilities (Factor 1), masculinity–femininity of appearance and aggressiveness (Factor 2, labeled the "macho" factor), and masculinity–femininity of major and work-related self-concept (Factor 3), whereas PAQ and BSRI femininity loaded on a factor defined by only one other variable (self-reported aggressiveness) and PAQ and BSRI masculinity defined a factor that was linked to no criterion variables.

This pattern was partially replicated for female subjects (see Table 8): The factor defined by PAQ and BSRI femininity was linked to aggressiveness and SAT verbal ability, whereas the factor defined by PAQ and BSRI masculinity was not strongly linked to any criterion variable. The factor analysis for women showed that gender diagnosticity measures loaded on factors tapping cognitive abilities (Factor 1), feminine interests and aggressiveness (Factor 2), work-related self-concept (Factor 3), and feminine appearance (Factor 4).

In both factor analyses conducted for male and female subjects, the composite gender-related criterion was included as a marker variable. The composite tended to load on factors on which gender diagnosticity also loaded highly but not on factors tapping femininity or masculinity. Interestingly, both factor analyses indicated that there were clusters of gender-related nonverbal behaviors (a factor marked by smiling, m–f of self-descriptions, and m–f of handwriting for men and a factor marked by m–f of photographs, in-person m–f, m–f of self-descriptions, and smiling for women) that were not strongly correlated with either gender diagnosticity measures or PAQ or BSRI scales.

Undoubtedly, the inclusion of the PAQ and BSRI scales contributed to the factorial complexity of our rotated solutions, for these scales tended not to be related to other variables. Thus, we also conducted factor analyses on just our two gender diagnosticity measures and the criterion variables. These factor analyses were carried out for all subjects, for men, and for women.

The factor analysis for all subjects yielded a two-factor unrotated solution, with the first factor accounting for 48% and the second factor 13% of the variance. The composite measure of gender-related criteria and gender diagnostic probabilities based on occupations and on hobbies loaded highly on the first general factor (loadings = .97, .89, and .78, respectively). Thus, this factor can reasonably be labeled the Gender factor. The second factor was defined by the three cognitive variables—SAT verbal, SAT math, and mental rotation (loadings = .81, .61, and .55, respectively). Both SAT math and mental rotation loaded on the general gender factor as well as on the cognitive ability factors (loadings = .47 and .49). The rotated solution was quite similar to the unrotated solution, except for the fact that cognitive variables loaded more purely on the second factor and gender diagnosticity loaded on both factors rather than only on the first.

The factor analysis conducted on men also showed evidence for a general factor. The first factor extracted in the unrotated solution accounted for 27% of the variance, followed by four additional factors accounting, respectively, for 14%, 11%, 10%, and 8% of the variance. The composite criterion and gender diagnosticity based on occupations and hobbies loaded highly on the first, general factor (.91, .75, .58), as did a number of the component criterion variables—SAT verbal (.53), SAT math (.54), mental rotation (.42), m–f of handwriting (.36), in-person m–f (.40), and m–f of self-description (.61). The varimax rotated solution for men showed factors similar to those reported in the

Table 8
Factor Analysis of PAQ Scales, BSRI Scales, Gender Diagnosticity Measures, and Criterion Measures for Women (Varimax Orthogonally Rotated Factor Matrix)

Measure	Factor 1 (Cognitive Abilities)	Factor 2 (Feminine Interests)	Factor 3 (Work Self-Concept)	Factor 4 (Feminine Appearance)	Factor 5 (Femininity)	Factor 6 (Masculinity)
Gender diagnosticity: occupation	**.35**	.19	**.62**	.15	.02	.14
Gender diagnosticity: hobby	.00	**.84**	.01	.20	−.06	−.09
PAQ masculinity	.01	.00	.06	−.04	.05	**.90**
PAQ femininity	−.10	.16	−.01	−.32	**.74**	−.16
PAQ m–f	−.29	.10	.06	−.04	**−.49**	**.56**
BSRI masculinity	−.01	.03	−.12	.04	.00	**.86**
BSRI femininity	−.04	.00	**−.35**	−.06	**.81**	.05
Criterion measures						
Composite	.28	−.34	**.72**	**.38**	−.26	−.08
SAT math	**.73**	−.08	**.42**	−.15	−.03	−.03
SAT verbal	**.62**	−.02	−.01	.00	**.54**	.17
Mental rotation	**.80**	−.10	−.24	.04	−.12	−.19
Major	.21	.15	**−.55**	.08	.19	−.26
Self-description	−.01	−.17	.28	**.57**	−.10	.34
Photo	−.01	.19	.17	**.85**	−.06	−.04
In-person m–f	.02	.25	−.16	**.74**	−.22	−.08
Handwriting	−.07	.05	**.77**	.06	−.13	−.19
Smile	.24	.21	**−.48**	**−.51**	−.12	.18
Total aggressiveness	.04	**−.59**	.08	.00	**−.40**	−.26

Note. PAQ = Personal Attributes Questionnaire. BSRI = Bem Sex-Role Inventory. Boldface factor loadings are greater than of .35.

previous factor analysis (excluding, of course, the masculinity and femininity factors): a cognitive ability factor, a Macho (masculine appearance and aggressiveness) factor, an m–f self-concept factor, and two nonverbal behavior factors. Gender diagnosticity based on occupations loaded .49 on the cognitive factor, .30 on the Macho factor, and .68 on the m–f self-concept factor. Gender diagnosticity based on hobbies loaded .41 on the cognitive factor and .52 on the Macho factor.

The factor analysis conducted on female subjects showed weaker evidence for a general factor: The unrotated solution extracted a first factor that accounted for 24% of the variance followed by three additional factors accounting for 15%, 14%, and 10%. The composite criterion loaded .91 on the first factor. The varimax rotated solution for women showed factors similar to those reported in the previously reported factor analysis (excluding the masculinity and femininity factors): a cognitive abilities factor, a gender-related self-concept and nonverbal behavior factor, an m–f of appearance factor, and a feminine interests and aggressiveness factor. Gender diagnosticity based on occupations loaded on the first two of these factors (factor loadings were .44 and .56, respectively), and gender diagnosticity based on hobbies loaded on the last two (loadings were .47 and −.52).

Discussion

In general, our data provided strong evidence that gender diagnosticity measures were empirically distinct from masculinity and femininity as assessed by the PAQ and BSRI.[5] Furthermore, gender diagnosticity measures often correlated more strongly with gender-related criterion behaviors than did the masculinity, femininity, and masculinity–femininity scales of the PAQ and BSRI.

Our data demonstrated the multidimensionality of gender-related behaviors and suggested somewhat different patternings of gender-related behaviors in men and women, but they also indicated that there was a reasonable degree of coherence in gender-related behaviors, particularly for men. Gender diagnostic probabilities based on occupations correlated .54 with a composite gender-related criterion for men and .43 for women. Our data demonstrate the value of aggregating gender-related behaviors to demonstrate their dispositional consistencies. In our study, we "aggregated" at one level—we combined a number of different gender-related behaviors into a composite. Clearly, each of the composite variables could have been based on an aggregate as well. For example, the ratings of the masculinity–femininity of subjects' appearance could have been based on a number of photographs taken at different times and in different settings, and the measures of aggressiveness could have been based on aggregated ratings made by a number of friends of each subject. We suspect that with such hierarchical aggregation, it would be possible to demonstrate even more coherence and consistency in gender-related behaviors than were apparent in our study.

Future Directions for Research on Gender Diagnosticity

Does gender diagnosticity represent a return to unidimensional masculinity–femininity and to the tradition begun by

[5] Gender diagnosticity also proved to be unrelated to conventional assessments of androgyny. When subjects were classified as high or low on masculinity and femininity on the basis of median splits of the PAQ and BSRI scales, ANOVAS on subjects' gender diagnostic probabilities based on these median splits showed no interactions between masculinity and femininity. Regression analyses (which used continuous variables rather than dichotomized masculinity and femininity) yielded the same results: The interaction between masculinity and femininity accounted for no significant variance in gender diagnostic probabilities.

Terman and Miles? We think not. We reiterate that our data indicated that gender-related behaviors are multidimensional, and furthermore that their dimensionality may differ for men and women. The notion that gender-related variables may be patterned differently in men and women is not original with us (see Maccoby, 1966, for a classic discussion, and Ozer, 1987, for a recent discussion of this issue in relation to intellectual abilities). The method of gender diagnosticity leads to a single continuum of individual differences by virtue of its fundamental methodological assumption: If an individual difference is defined by the behavior of two indexing groups, A and B, then the discriminant function computed from a given set of diagnostic indicators will yield a Bayesian probability that describes how "A-like" or "B-like" an individual is.

Clearly, the method of gender diagnosticity bears some resemblance to empirical methods of scale construction. Like empirical masculinity–femininity tests, the method of gender diagnosticity ultimately rests on observed differences between males and females. However, despite the obvious similarities between the two approaches, there are some noteworthy differences as well: The method of gender diagnosticity is not tied to normative samples, as are empirically derived scales. Gender diagnostic probabilities are computed anew for each sample of subjects. The statistical process of computing gender diagnostic probabilities differs from the statistical processes conventionally used in constructing and scoring empirically developed tests. Most fundamentally, the metric of individual differences yielded by the method of gender diagnosticity is quite different from those operationally defined by traditional scales. Diagnostic probabilities are true ratio scales of individual differences (see Buss & Craik, 1984, for a discussion of this issue in relation to the act-frequency approach to personality assessment). They have rich interpretations as probabilities and may possess psychometric properties different from conventional personality measures. Clearly, studying the psychometric characteristics of gender diagnosticity measures is a fruitful direction for further investigation.

Although gender diagnostic probabilities seem to be weakly related to existing measures of masculinity and femininity and unrelated to androgyny, they may be related to other personality constructs. Future research can help calibrate gender diagnosticity against commonly found dimensions of personality such as the "Big Five" (John, Angleitner, & Ostendorf, 1988; McCrae & Costa, 1987).[6]

Our data showed that gender diagnostic probabilities computed from occupations and hobbies were related, but that these two measures did not equally correlate with all gender-related criteria. For example, our analyses indicated that for female subjects, gender diagnosticity based on occupations was correlated with cognitive abilities and work-related self-concept, whereas gender diagnosticity based on hobbies was more strongly correlated with feminine appearance and aggressiveness. We believe it would be fruitful to compute and compare highly reliable measures of gender diagnostic probabilities from a number of different sets of gender diagnostic behaviors (e.g., occupation ratings, hobby ratings, attitude measures, and assessments of family roles and activities). In this sense, our research represents a return to Terman and Miles's strategy of assessing many domains of gender-related behaviors. Research on multiple measures of gender diagnosticity would help assess the consistency of such measures and show how well different gender diagnosticity measures predict various criteria.

In the current study we computed our most reliable measure of gender diagnostic probabilities from occupational preference ratings. These probabilities correlated significantly with cognitive ability measures and proved to predict gender-related criteria somewhat better for men than for women. It is important to determine how much these findings are dependent on the particular set of gender diagnostic behaviors used to compute the probabilities. Gender diagnostic probabilities computed from occupational preference data may have particularly fruitful applications in research on the career choices of women and men (e.g., see Betz & Fitzgerald, 1987; Helson, Elliott, & Leigh, 1989).

Clearly, to establish the construct validity of gender diagnosticity, it will be necessary to establish a network of relationships between this new class of measures and established psychological constructs. Much research has examined the relationships among masculinity, femininity, and such constructs as self-esteem, psychological adjustment, and behavioral flexibility (see Cook, 1985; Heilbrun, 1981; Paulhus & Martin, 1988). Similar research is required to determine the links between gender diagnosticity and these same constructs.

Potential Advantages of Gender Diagnosticity Measures

Gender diagnosticity represents a new approach to assessing individual differences. It is important to note that gender diagnosticity is not a "test" per se, but rather a method for assessing the degree to which individuals are prototypic of indexing groups given a set of diagnostic indicators of group membership. Stated this way, it is clear that gender diagnosticity is but one example of a broader concept—group diagnosticity. Just as an individual's behavior can serve as the basis to assess the degree to which the individual is "male-like" or "female-like," so could a properly chosen set of diagnostic behaviors serve to assess how "middle-class-like" versus "working-class-like," or "fundamentalist" versus "non-fundamentalist," an individual is. The concept of group diagnosticity proposes that one define certain kinds of individual differences specifically in terms of indexing social groups, and it provides a statistical methodology for doing so.

The method of gender diagnosticity may have certain advantages over more traditional methods of assessing individual differences. We argue that "masculinity" and "femininity" are not necessarily universal across cultures, but rather are, to some degree, defined by behaviors that vary across cultures and subcultures. Thus, if psychologists want to assess how masculine or feminine an individual's behavior is, this can be done only within the context of a particular social group and cultural environment. Masculinity and femininity are not necessarily equivalent to instrumental or expressive personality traits, or to wearing pants or skirts. Rather, they are defined by behaviors that discriminate men from women in a particular population in a particular society during a particular historical era.

[6] Lippa (1990b) has collected preliminary data showing that whereas masculinity and femininity as assessed by the PAQ and BSRI load highly on "Big Five" dimensions, gender diagnosticity is independent of the "Big Five" both across and within the sexes.

Gender diagnosticity tailors its definition of gender-related individual differences to specific populations of subjects. In our study, gender-related individual differences in a group of California college students were defined by behavioral indicators (e.g., occupational preference ratings) that discriminated men from women in that specific population.

The concept of gender diagnosticity encourages researchers to acknowledge that gender-related individual differences are not to be defined or assessed solely in terms of self-ascribed personality traits. Recent research on gender stereotypes (e.g., Deaux & Lewis, 1983, 1984) suggests that laypeople perceive greater differences in the social and family roles and physical characteristics of men and women than in their personality traits. The method of gender diagnosticity draws upon this insight and permits researchers to assess gender-related individual differences from a host of variables such as occupational preferences, participation in hobbies, attitudes, activities performed in the home, family roles, and so on.

Because gender diagnosticity is a method and not a test, it can be computed in a wide array of data sets. Typically, when a new individual difference measure is proposed to supplant older measures, it creates an immediate discontinuity in the research literature, for researchers are not sure how the new measure calibrates against older measures. Gender diagnosticity, however, can be computed for any data set, past or present, that comprises a set of assessed gender-related behaviors in a population of males and females. In this sense, it augments, not replaces, previous measures and permits researchers to reanalyze old data. Our data for female subjects illustrate that gender diagnosticity measures in combination with other personality measures may improve prediction of gender-related behaviors. For women, both gender diagnosticity and assessed femininity explained significant amounts of independent variance in gender-related criterion behaviors.

As noted at the beginning of this article, the method of gender diagnosticity may be particularly well suited to developmental research on gender-related behaviors. Existing masculinity and femininity tests, for example, cannot be directly applied to preschool populations. However, gender diagnostic probabilities can readily be computed for all age groups. The method of gender diagnosticity creates a metric of gender-related individual differences that is directly comparable over the entire life span, and it seems likely to be particularly useful if the same group of individuals is repeatedly assessed. Whether the method of gender diagnosticity ultimately documents temporal consistency or inconsistency in gender-related individual differences, it provides a useful methodology for addressing such questions.

The gender diagnosticity approach suggests that rather than labeling people as masculine or feminine, one should more neutrally construe them, given the context of their particular culture and subculture, as male-like or female-like to varying degrees. Such terminology does not reify masculinity and femininity into eternal verities, yet at the same time acknowledges that some people adhere more and some less to their society's gender prescriptions (see Bem, 1981b). Stated differently, some people are more and some less conventionally sex typed.

Unlike existing measures of masculinity–femininity, masculinity, and femininity, the method of gender diagnosticity in a sense self-destructs in a society without gender differences or dichotomies, for in such a society behavioral indices would not diagnose gender. In the sample of subjects assessed in our study, gender diagnosticity proved to constitute a meaningful measure of gender-related individual differences, and it correlated with a number of psychologically significant criterion variables. We believe that gender diagnostic probabilities will serve as meaningful measures of gender-related individual differences in other groups in contemporary society as well. If our society evolves into one in which people display fewer or even no sex differences in many behavioral domains, then gender diagnosticity can provide both a method and a metric to document that change.

References

Bassoff, S. A. (1986). *Sex-role stereotypes: Traditions and alternatives* (2nd ed.). Monterey, CA: Brooks/Cole.

Bassoff, E. S., & Glass, G. V. (1982). The relationship between sex roles and mental health: A meta-analysis of twenty-six studies. *Counseling Psychologist, 10,* 105–112.

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

Bem, S. L. (1975). Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology, 31,* 634–643.

Bem, S. L. (1981a). *Bem Sex-Role Inventory: Professional manual.* Palo Alto, CA: Consulting Psychologists Press.

Bem, S. L. (1981b). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88,* 354–364.

Bem, S. L., & Lenney, E. (1976). Sex-typing and the avoidance of cross-sex behavior. *Journal of Personality and Social Psychology, 33,* 48–54.

Bem, S. L., Martyna, W., & Watson, C. (1976). Sex typing and androgyny: Further explorations of the expressive domain. *Journal of Personality and Social Psychology, 34,* 1016–1022.

Betz, N. E., & Fitzgerald, L. F. (1987). *The career psychology of women.* Orlando, FL: Academic Press.

Block, J. H. (1973). Conceptions of sex roles: Some cross-cultural and longitudinal perspectives. *American Psychologist, 28,* 512–526.

Buss, D. M., & Craik, K. H. (1984). Acts, dispositions, and personality. In B. A. Maher & W. B. Maher (Eds.), *Progress in experimental personality research* (pp. 241–301). New York: Academic Press.

Campbell, D. F. (1966). *Revised manual for Strong Vocational Interest Blank.* Stanford, CA: Stanford University Press.

Campbell, D. P., & Hansen, J. C. (1981). *Manual for the SVIB-SCII, Strong-Campbell Interest Inventory, Form T325 of the Strong Vocational Interest Blank* (3rd ed.). Stanford, CA: Stanford University Press.

Caspie, A. (1987). Personality in the life course. *Journal of Personality and Social Psychology, 53,* 1203–1213.

Constantinople, A. (1973). Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin, 80,* 389–407.

Cook, E. P. (1985). *Psychological androgyny.* New York: Pergamon Press.

Deaux, K. (1985). Sex and gender. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 36, pp. 49–81). Palo Alto, CA: Annual Reviews, Inc.

Deaux, K. (1987). Psychological constructions of masculinity and femininity. In J. M. Reinisch, L. A. Rosenblum, & S. A. Sanders (Eds.), *Masculinity/femininity: Basic perspectives* (pp. 289–303). New York: Oxford University Press.

Deaux, K., & Lewis, L. L. (1983). Assessment of gender stereotypes: Methodology and components. *Psychological Documents, 13,* 25.

Deaux, K., & Lewis, L. L. (1984). The structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology, 46,* 991–1004.

Deaux, K., & Major, B. (1987). Putting gender into context: An interac-

tive model of gender-related behavior. *Journal of Personality and Social Psychology, 94,* 369–389.
Eagly, A. H. (1987). *Sex differences in social behavior: A social role interpretation.* Hillsdale, NJ: Erlbaum.
Eagly, A. H., & Steffen, V. J. (1986). Gender stereotypes, occupational roles, and beliefs about part-time employees. *Psychology of Women Quarterly, 10,* 252–262.
Gaudreau, P. (1977). Factor analysis of the Bem Sex-Role Inventory. *Journal of Consulting and Clinical Psychology, 45,* 299–302.
Goldstein, M., & Dillon, W. R. (1978). *Discrete discriminant analysis.* New York: Wiley.
Gough, H. G. (1964). *California Psychological Inventory: Manual.* Palo Alto, CA: Consulting Psychologists Press.
Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style.* Hillsdale, NJ: Erlbaum.
Hand, D. J. (1981). *Discrimination and classification.* New York: Wiley.
Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory.* New York: Psychological Corporation.
Heilbrun, A. B. (1976). Measurement of masculine and feminine sex role identities as independent dimensions. *Journal of Clinical and Consulting Psychology, 44,* 183–190.
Heilbrun, A. B., Jr. (1981). *Human sex-role behavior.* New York: Pergamon Press.
Helson, R., Elliott, T., & Leigh, J. (1989). Adolescent personality and women's work patterns. In D. Stern & D. Eichorn (Eds.), *Adolescence and work: Influences of social structure, labor, markets, and culture* (pp. 259–289). Hillsdale, NJ: Erlbaum.
Hyde, J. S., & Linn, M. C. (1986). *The psychology of gender: Advances through meta-analysis.* Baltimore: Johns Hopkins University Press.
Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104,* 53–69.
John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality, 2,* 171–203.
Lachenbruch, P. A. (1975). *Discriminant analysis.* New York: Hafner Press.
Lippa, R. (1977). Androgyny, sex-typing, and the perception of masculinity-femininity in handwritings. *Journal of Research in Personality, 11,* 1–14.
Lippa, R. (1978). The naive perception of masculinity-femininity on the basis of expressive cues. *Journal of Research in Personality, 12,* 1–14.
Lippa, R. (1983). Expressive behavior. In L. Wheeler & P. Shaver (Eds.), *Review of personality and social psychology* (Vol. 4, 181–205). Beverly Hills, CA: Sage Publications.
Lippa, R. (1990a). Gender and social behavior. In R. Lippa, *Introduction to social psychology* (pp. 342–385). Belmont, CA: Wadsworth Publishing Co.
Lippa, R. (1990b). *Some psychometric characteristics of gender diagnosticity measures: Reliability, consistency across domains, and relationship to the "Big Five."* Unpublished manuscript.
Lippa, R., Valdez, E., & Jolly, A. (1983). The effects of self-monitoring on the expressive display of masculinity-femininity. *Journal of Research in Personality, 17,* 324–338.
Maccoby, E. E. (1966). Sex differences in intellectual functioning. In E E. Maccoby (Ed.), *The development of sex differences* (pp. 25–55). Stanford, CA: Stanford University Press.
Maccoby, E. E. (1987). The varied meanings of "masculine" and "feminine." In J. M. Reinisch, L. A. Rosenblum, & S. A. Sanders (Eds.), *Masculinity/femininity: Basic perspectives* (pp. 227–239). New York: Oxford University Press.
Major, B., Carnevale, P. J. D., & Deaux, K. (1981). A different perspective on androgyny: Evaluations of masculine and feminine personality characteristics. *Journal of Personality and Social Psychology, 41,* 988–1001.
Martin, C. L. (1987). A ratio measure of sex stereotyping. *Journal of Personality and Social Psychology, 52,* 489–499.
Mauger, P. A., & Adkinson, D. R. (1980). *Interpersonal Behavior Survey (IBS): Manual.* Los Angeles: Western Psychological Services.
McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology, 36,* 929–940.
McCauley, C., Stitt, C. L., & Segal, M. (1980). Stereotyping: From prejudice to prediction. *Psychological Bulletin, 87,* 195–208.
McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality. *Journal of Personality and Social Psychology, 52,* 81–90.
McNemar, Q. (1962). *Psychological statistics* (3rd ed.). New York: Wiley.
Norusis, M. J. (1986). *Advanced statistics: SPSS/PC+ for the IBM PC/XT/AT.* Chicago: SPSS, Inc.
Norusis, M. J. (1988). *SPSS/PC+ V2.0 base manual for the IBM PC/XT/AT and PS/2.* Chicago: SPSS, Inc.
Ozer, J. (1987). Personality, intelligence, and spacial visualization: Correlates of mental rotations test performance. *Journal of Personality and Social Psychology, 53,* 129–134.
Paulhus, D. L. (1987). Effects of group selection on correlations and factor patterns in sex role research. *Journal of Personality and Social Psychology, 53,* 314–327.
Paulus, D. L., & Martin, C. L. (1988). Functional flexibility: A new conception of interpersonal flexibility. *Journal of Personality and Social Psychology, 55,* 88–101.
Pedhazur, E. J., & Tetenbaum, T. J. (1979). Bem Sex Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology, 37,* 996–1016.
Signorella, M. L., & Jamison, W. (1986). Masculinity, femininity, androgyny, and cognitive performance: A meta-analysis. *Psychological Bulletin, 100,* 207–228.
Spence, J. T. (1984). Masculinity, femininity, and gender-related traits: A conceptual analysis and critique of current research. *Progress in Experimental Personality Research, 13,* 1–97.
Spence, J. T., Deaux, K., & Helmreich, R. L. (1985). Sex roles in contemporary American society. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 2, pp. 149–178). New York: Random House.
Spence, J. T., & Helmreich, R. L. (1978). *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents.* Austin: University of Texas Press.
Spence, J. T., & Helmreich, R. L. (1980). Masculine instrumentality and feminine expressiveness: The relationships with sex role attitudes and behaviors. *Psychology of Women Quarterly, 5,* 147–163.
Spence, J. T., Helmreich, R., & 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)
Spence, J. T., & Sawin, L. L. (1985). Images of masculinity and femininity: A reconceptualization. In V. E. O'Leary, R. K. Unger, & B. S. Wallston (Eds.), *Women, gender, and social psychology* (pp. 35–66). Hillsdale, NJ: Erlbaum.
Strong, E. K. (1943). *Vocational interests of men and women.* Stanford, CA: Stanford University Press.
Taylor, M. C., & Hall, J. A. (1982). Psychological androgyny: Theories, methods, and conclusions. *Psychological Bulletin, 92,* 347–366.
Terman, L. M., & Miles, C. C. (1936). *Sex and personality: Studies in masculinity and femininity.* New York: Russell & Russell.
Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47,* 599–604.

Received October 19, 1989
Revision received February 16, 1990
Accepted February 16, 1990 ■