

The Display of Masculinity, Femininity, and Gender Diagnosticity in Self-Descriptive Photo Essays

Richard Lippa

California State University, Fullerton

ABSTRACT Thirty-seven college men and 57 college women assessed on Gender Diagnosticity (GD), Masculinity (M), and Femininity (F) created self-descriptive photo essays, which were then rated by six judges on 38 personality characteristics, including masculinity and femininity. Lay judges reliably rated men and women's masculinity and femininity from photo essay information. Men's GD strongly correlated with their judged masculinity and femininity, M with judged extraversion, and F with judged warmth and nurturance. However, women's GD correlated most strongly with their judged maladjustment and athleticism, M with dominance and extraversion, and F with adjustment and physical attractiveness. Naive judgments of men and women's masculinity-femininity were strongly linked to other judged personality characteristics, and physical attractiveness was correlated with judgments of women's but not men's masculinity and femininity. The results show that masculinity and femininity make sense to laypeople, are readily judged from multidimensional information, and that for men, GD predicts lay judgments of masculinity and femininity better than M and F do.

Feminine, female, effeminate, womanly, and womanish are used to describe women or their qualities.
Feminine characterizes the qualities that are regarded

I wish to thank the following research assistants who worked on various aspects of this study: Sara Arad, Rob Bailey, Lindsey Bernardin, Cris Chaparro, Tanjula Drumgole, Eric Jensen, Jesus Rodriguez, Edgar Sillas, Francisco Tan, Erin Ulrich, Richard Veira, and Craig Wagstaff. Correspondence concerning this article should be addressed to Richard Lippa, Department of Psychology, California State University, Fullerton, CA 92834-6846. E-mail may be sent via the Internet to rlippa@fullerton.edu.

Journal of Personality 65:1, March 1997.
Copyright © 1997 by Duke University Press.

as belonging particularly to women; *female* denotes sex strictly, without further implications: *feminine* modesty, a *female* voice. *Effeminate* is applied only to men, and describes attributes that are regarded as unseemly in a man, though appropriate to a woman: an *effeminate* walk. *Womanly* and *womanish* are both applied to women; *womanly* refers to things that are admirable and *womanish* to those that are not: *womanly* pity, *womanish* tears. —Ant. See synonyms for *MASCULINE*.

—*Funk & Wagnall's Standard College Dictionary*

As the above excerpt illustrates, lexicographers as well as psychologists have long grappled with the concepts of masculinity and femininity. Whatever the merit (or lack of merit) of *Funk & Wagnall's* definitions, they provide a rich entrée into two topics frequently studied by gender researchers: gender stereotypes and gender-related individual differences. Embedded in *Funk & Wagnall's* definitions are a number of interesting and undoubtedly controversial assumptions: Some characteristics “belong” particularly to women, and some to men; some of women’s supposed characteristics are admirable, others are not; and the characteristics of one sex may be “unseemly” when displayed by the other.

Dictionary definitions of “masculine” and “feminine” remind us that people *do* use these terms in everyday life and that lexicographers and laypeople—as well psychologists—make many and varied assumptions whenever they contemplate the nature of masculinity and femininity. Research psychologists’ assumptions about masculinity and femininity, perhaps more than lexicographers’ and laypeople’s, have evolved dramatically over the past 60 years as they have attempted to develop scientifically coherent conceptions and reliable operational definitions of these concepts.

How well do psychologists’ conceptions and measures capture what is commonsensically meant by the words masculine and feminine? The research to be described here tries to validate (or invalidate) psychologists’ formal attempts to measure masculinity and femininity by examining their relationship to people’s everyday judgments of others’ masculinity and femininity. At the same time, this research investigates the meaning and structure of lay judgments of masculinity and femininity by examining how they correlate with other judged charac-

teristics of people (for example, their judged dominance, nurturance, physical attractiveness, and so on).

Most recent empirical analyses of masculinity and femininity have emphasized that these constructs embrace a multidimensional collection of traits, behaviors, and predispositions (Lippa & Connelly, 1990; Myers & Gonda, 1982; Spence & Buckner, 1995; Spence & Sawin, 1985). For example, masculinity may comprise personality traits (e.g., aggressiveness, dominance), attitudes (a tough "law and order" stance toward crime, traditional sex-role ideology), nonverbal behaviors (expansive personal space, little smiling), physical traits (muscular build, heavy facial hair), hobbies and interests (interest in cars and spectator sports), and sexual behaviors (heterosexuality, strong interest in sex).

It seems likely that laypeople as well as psychologists hold multidimensional conceptions of masculinity and femininity (e.g., see Myers & Gonda, 1982; Spence & Sawin, 1985), and this implies that if laypeople are to reasonably judge an individual's masculinity and femininity, they must be given access to rich and varied information about that individual. In the current research, such rich and varied information was presented to lay judges in the form of self-descriptive photographic essays created by college men and women. In constructing their photo essays, these men and women were asked in essence to take 12 photographs that showed "who they are" and then to assemble these photographs in a booklet with captions and a summary self-descriptive essay.

Based on the content of their photo essays, participants were then rated by lay judges on a number of personality characteristics, including masculinity and femininity. These ratings provided a means to investigate the structure of lay conceptions of masculinity and femininity and to study the relationship between participants' masculinity and femininity (as judged by others) and their scores on more formal measures of gender-related individual differences (masculinity scales, femininity scales, and gender diagnosticity measures).

Psychological Approaches to the Assessment of Masculinity and Femininity

To understand the formal measures of gender-related individual differences used in the current research, it helps to briefly review the history of attempts to assess masculinity and femininity. Modern research on

this topic began with the 1936 publication of Terman and Miles's *Sex and Personality*, which presented a bipolar conception of masculinity-femininity (M-F). In essence, this approach held that M-F is a single dimension and that masculinity and femininity have an "either-or" quality—that is, the more masculine an individual is, the less feminine he or she is, and vice versa. Terman and Miles and their many successors selected items for inclusion in their M-F scales that showed reliable and strong sex differences in normative populations. The bipolar approach to M-F that began in the 1930s waned by the early 1970s in the face of conceptual and empirical critiques (e.g., Block, 1973; Constantinople, 1973), which argued in part that the supposedly unidimensional M-F scales were in fact multidimensional measures that didn't tap consistent and coherent domains of gender-related behavior.

The 1970s witnessed the general demise of the bipolar, unidimensional approach to M-F and the concurrent rise of two-dimensional conceptions of masculinity and femininity. The two-dimensional approach, which has been dominant for the past 20 years, holds that masculinity and femininity are separate dimensions, with Masculinity (M) defined in terms of instrumental personality traits (e.g., aggressive, dominant, independent) and Femininity (F) defined in terms of expressive traits (warm, sensitive, nurturant). During the 1970s a number of self-report inventories were developed to assess M and F as two separate dimensions. The best known of these are the Bem Sex-Role Inventory (BSRI; Bem, 1974, 1981a) and the Personal Attributes Questionnaire (PAQ; Spence & Helmreich, 1978; Spence, Helmreich, & Stapp, 1974). A huge empirical literature now exists on the psychometric properties and correlates of M and F as assessed by these scales (see Ashmore, 1990; Cook, 1985; Lenney, 1991).

The PAQ and BSRI continue to be widely used in research on gender-related individual differences. However, like earlier M-F scales, M and F scales have been subject to telling critiques. Indeed, the authors of the best-known M and F scales have themselves modified their original conceptions of their constructs, sometimes to the point of arguing that their scales do not really measure "masculinity" and "femininity" at all. For example, Spence and Helmreich (1980) have argued that M and F scales are in fact instrumentality and expressiveness scales, and as such, they show at best weak and inconsistent relationships to other kinds of gender-related behaviors and attitudes (see Spence & Buckner, 1995, for a recent theoretical discussion of M and F).

Bem (1974, 1975, 1976) proposed in her early work on M and F that “sex-typed” individuals (men high on M and low on F, and women low on M and high on F) show traitlike consistencies in their gender-related behaviors, whereas androgynous individuals (men and women who are high on both M and F) display more variable and inconsistent gender-related behaviors. By the early 1980s, however, Bem had abandoned her earlier conceptions of M and F in favor of gender schema theory (Bem, 1981b, 1985), which holds that masculinity and femininity are cognitive constructs rather than psychological realities. In the context of gender schema theory and research, M and F scales serve to assess whether people are gender schematic or aschematic—that is, whether they strongly structure their cognitive and perceptual worlds in terms of gender categories—not necessarily the degree to which they possess traitlike masculinity or femininity.

Still other critiques have been offered of M and F scales. I have argued elsewhere (Lippa, 1991, 1995a, 1995b; Lippa & Connelly, 1990) that M and F scales and their associated constructs suffer from a number of problems. As Bem properly noted, M and F scales reify gender-related individual differences. Elaborating on this point, I further noted that this reification may restrict masculinity and femininity to overly limited domains of behavior. Because M and F scales define M and F in terms of gender-stereotypic instrumental and expressive personality traits, they fail to embrace a host of other characteristics that are highly relevant to everyday conceptions of masculinity and femininity—characteristics such as gender-related appearances, non-verbal behaviors, hobbies, interests, sexual behaviors, and ways of relating to friends, spouses, and lovers. Because of their fixed and limited item content, M and F scales fail to acknowledge that masculinity and femininity are fluid concepts that are culturally and historically relative.

To address some of these problems, I proposed a new approach to assessing within-sex gender-related individual differences—an approach termed “gender diagnosticity” (Lippa, 1991, 1995a, 1995b; Lippa & Connelly, 1990). In brief, gender diagnosticity (GD) refers to the Bayesian probability that an individual is predicted to be male or female based on some set of gender-related indicators (such as occupational preference ratings). According to the GD perspective, a masculine person is an individual who shows “male-like” behaviors in comparison to a normative group of males and females, and a feminine person is an individual who shows “female-like” behaviors. That

is, gender-related individual differences—within as well as across the sexes—are defined by behaviors that distinguish men and women in a given population at a given time. Indeed, such gender-related individual differences can be defined and measured *only* in relation to particular populations of men and women at particular points in time.

One virtue of the GD approach is that it acknowledges that a given indicator of masculinity or femininity may vary over time and over different populations of men and women. For example, the behavior of “wearing pants” was more gender diagnostic 100 years ago than it is today in the United States, and it is currently more gender diagnostic in some countries than in others. By implication, an American woman wearing pants 100 years ago would have been judged more masculine as a result of her behavior than an American woman wearing pants today, and a woman wearing pants in contemporary Saudi Arabia would very likely be judged by members of her culture to be more masculine as a result of her behavior than would a woman wearing pants in contemporary America be judged by members of her culture.

Similarly, American college students’ interest in becoming a lawyer was probably more gender diagnostic 50 years ago than it is today, whereas students’ interest in becoming a mechanical engineer was strongly gender diagnostic 50 years ago and continues to be so today. By implication, interest in becoming a lawyer is probably not viewed as being particularly “masculine” among today’s American college women, but it may have been so viewed 50 years ago.

GD is formally computed from sets of indicators (such as occupational preference ratings) through the application of discriminant analyses (see Lippa, 1990, 1995b; Lippa & Connelly, 1990; this process will be described more fully later in this article). Prior research on GD shows that it can be measured reliably within the sexes from self-report data such as occupational preference ratings and that GD measures are factorially distinct from M and F as assessed by the PAQ and BSRI (Lippa, 1991, 1995b; Lippa & Connelly, 1990). Furthermore, GD measures are largely independent of the Big Five personality factors, whereas M and F are not (Lippa, 1991, 1995b). Indeed, M and F correlate substantially with the Big Five dimensions, with M loading highly on Extraversion and Neuroticism and F on Agreeableness (Lippa, 1991, 1995b). Finally, and perhaps most important in the context of the current research, GD measures often predict varied gender-related behaviors and attitudes within the sexes (e.g., math SAT performance, visual-spatial ability, nonverbal masculinity-femininity, masculinity-

femininity of chosen college major, self-ascribed masculinity-femininity, attitudes toward women's roles, and attitudes toward gay people) better than M and F do (Lippa, 1991, 1995b; Lippa & Connelly, 1990).

The research to be described here assessed participants on three of the measures of gender-related individual differences just reviewed: M scales, F scales, and GD measures. The current study examined the links between participants' "real-life" masculinity and femininity, as judged from their self-descriptive photo essays, and their assessed levels of M, F, and GD. It thus provided new validity evidence on M, F, and GD by examining how well these measures predict lay judgments of participants' masculinity and femininity.

METHOD

Participants and Self-Report Measures

Participants were drawn from a larger population of introductory psychology students (103 men and 186 women) who completed a questionnaire packet that included the BSRI, a GD measure, and an Extraversion scale. The larger population of participants had been solicited from several introductory psychology classes over the course of two semesters.

Extraversion was assessed in the current research as a kind of comparison dimension. Because considerable recent research has shown that Extraversion is the most observable and readily judged of the Big Five personality factors (e.g., Funder & Dobroth, 1987; John & Robins, 1993; Kenny, 1994), Extraversion was compared with M, F, and GD in terms of how reliably and accurately it and the other three individual difference dimensions were judged from participants' photo essays.

The photo essay technique. Participants in two introductory psychology classes (one in the fall semester of 1994 and one in the spring semester of 1995) had the option of creating photo essays for extra credit. After completing and turning in their photo essays, students were informed by their instructor that a faculty member in the Psychology Department was interested in using their photo essays for research purposes, but that their photo essays would be used in this research only if they gave permission. A consent form was distributed to students who completed photo essays, and was signed by 37 men and 57 women (the differing numbers of men and women reflect their relative proportions in the introductory psychology classes, not a differing rate of consent). The photo essays of all students signing consent forms were those studied in the current research.

Photo essays served as a useful means to gather rich information about participants' social relationships, life settings, and personal characteristics

(see Dollinger & Clancy, 1993; Ziller, 1990). In the current study, participants were asked to describe “who they are” by taking 12 photographs and assembling them into a booklet. Following procedures similar to those described by Dollinger and Clancy (1993), participants were given the following written instructions:

[Your] project is to create and turn in a photo essay that describes who you are....

In your photo essay, please use your photographs to describe *how you see yourself*. To do this, take or have someone else take *12 NEW PHOTOGRAPHS THAT TELL WHO YOU ARE....* The photographs you take for your photo essay can be of anything, just as long as they tell something about who you are. You should not be interested in your skill as a photographer. Keep in mind that the photographs should describe who you are *as you see yourself*.

When you have a set of 12 photos you are satisfied with, put your photographs in a bound booklet. Use the kind of cardboard or plastic cover and binding that you would use for a class paper or report. Your report should include 12 pages, with one photograph attached to each page. Write a caption on each page that describes the attached photograph. Remember, each photograph should tell something about you, as you see yourself.

The final pages of the booklet should consist of a brief essay (either typed or handwritten) that discusses how the set of 12 photographs does or does not capture who you are. This essay is a kind of self-analysis in relation to your photographs.

As others have noted (Dollinger & Clancy, 1993; Ziller, 1990), people find the photo essay project quite engrossing. One indication of this is that many student participants in the current research desired to have their photo essays returned to them—an option that was offered to them. The information contained in students’ photo essays was quite varied (partial descriptions of four photo essays are presented later in this article).

Participants often included in their essays photographs of themselves in various life settings—for example, at home, at work, with friends, engaging in hobbies and activities. Because the majority of participants in the current research were commuters who did not live in university housing, their photo essays displayed greater variety than might be observed in similar essays created by resident students. For example, students’ photo essays frequently included photographs of their nonuniversity homes and apartments, family members, work settings, and past (precollege) friends. The photo essays also typically included information about a participant’s appearance (physical attractiveness, demeanor, nonverbal style, grooming, and dress); social relationships (friends, family members, coworkers); life settings (home, room, workplace, social venues such as outings, parties, and sororities/fraternities);

hobbies and interests (collections; prized possessions such as pets, stuffed animals, cars, compact disks; and activities such as back packing, sports, dancing, going to church or to museums). Photo essays also included richly descriptive verbal content in the form of photo captions and summary essays. Because they contained such varied information, these photo essays seemed a particularly promising route in assessing masculinity and femininity, which are richly multidimensional constructs.

Photo essay ratings. Each photo essay was independently read and rated by six (three male and three female) raters. All raters were undergraduates, except for one female rater who was a high-school junior (a participant in a program for exceptional high-school students who wished to work on university-level research projects). Raters rated each photo essay on 38 characteristics (e.g., introverted, extraverted, anxious, depressed, masculine, and feminine; see Table 1 for a complete list). These ratings were made on a 7-point scale that ranged from "not at all" (1) to "extremely" (7) and were based on raters' overall impressions of a participant from the information contained in his or her photo essay. Mean ratings were computed by averaging the ratings made by the six raters of each photo essay on each of the 38 characteristics.

RESULTS

Computation of Scales and Self-Report Measures

Short-form BSRI M and F scores were computed in the current research, as described in Bem (1981a). Extraversion was assessed by asking participants to rate themselves on the following trait adjectives, which all have been found to load highly on Big Five Extraversion (see John, 1990; McCrae & Costa, 1987): retiring, sociable, sober, fun loving, quiet, talkative, reserved, energetic, and outgoing. Participants rated themselves on these traits using a 7-point scale ranging from "never or almost never true" (1) to "always or almost always true" (7) of oneself, and participants' overall Extraversion scores were simply the sum of their self-ratings on these items, with the appropriate items reversed.

GD was computed from a self-report questionnaire that asked participants to rate their degree of preference for 131 occupations (see Lippa, 1991, for additional details). Participants rated each occupation on a 5-point scale: strongly dislike (1), slightly dislike (2), neutral or indifferent (3), slightly like (4), and strongly like (5). Gender diagnostic probabilities (GD scores) were computed from participants' occupational preference ratings. Specifically, 13 discriminant analyses were conducted

Table 1
Interjudge Reliabilities of Photo Essay Ratings (Coefficient Alpha)

| <i>r</i> | Photo essay rating |
|----------|--|
| .72 | Agreeable, nice, likable |
| .54 | Anxious |
| .92 | Athletic |
| .53 | Authentic, real |
| .48 | Calm, relaxed |
| .73 | Coarse, vulgar |
| .86 | Competitive |
| .71 | Conscientious |
| .66 | Conventional, ordinary |
| .76 | Creative |
| .64 | Disagreeable, unpleasant, unlikable |
| .77 | Dominant |
| .87 | Extraverted |
| .75 | Depressed |
| .91 | Family-oriented |
| .93 | Feminine (.78 for men only; .66 for women only) |
| .61 | Fraudulent, presents a false image of self |
| .81 | Happy |
| .74 | Has good self-esteem |
| .80 | Has sense of humor |
| .61 | Intelligent |
| .87 | Introverted |
| .73 | Irresponsible |
| .91 | Masculine (.81 for men only; .62 for women only) |
| .63 | Maladjusted |
| .86 | Nurturant, cares for others |
| .87 | Physically attractive |
| .88 | Physically unattractive |
| .72 | Possesses broad interests |
| .95 | Religious |
| .77 | Sad |
| .68 | Self-disclosing |
| .76 | Serious |
| .85 | Studious |
| .74 | Theatrical, self-dramatizing |
| .70 | Troubled |
| .63 | Unconventional, eccentric |
| .84 | Warm |

on discrete sets of occupations—12 analyses were on sets of 10 occupations each, and 1 analysis on a set of 11 occupations. Thus, the 13 discriminant analyses included all 131 occupational preference items.

Each discriminant analysis yielded the Bayesian probability, computed from each participant's discriminant function score, that a given participant was male (or, by subtracting this probability from 1, female). Thus, on the basis of their occupational preference ratings, each participant had 13 separate gender diagnostic probabilities, each computed from a distinct subset of rated occupations. A participant's overall GD score was simply the average of the 13 probabilities.

Multiple gender diagnostic probabilities were computed for each participant in order to provide a means to assess their reliability (see Lippa, 1991; Lippa & Connelly, 1990). The reliability of GD was acceptably high for all participants ($\alpha = .94$) as well as for men only ($\alpha = .74$) and women only ($\alpha = .74$). Reliabilities were also acceptably high for BSRI M ($\alpha = .83$), BSRI F ($\alpha = .89$), and Extraversion ($\alpha = .78$).

As noted before, photo essay participants were a subset of a much larger group of participants who filled out questionnaires. The mean scores and standard deviations of photo essay participants' scores on M, F, GD, and Extraversion were quite similar to those of the larger population of students from which they were drawn. For example, for male photo essay participants, mean M, F, GD, and Extraversion scores were respectively 5.01, 5.15, .70, and 4.54, whereas the corresponding means for the larger population of men were 4.80, 5.08, .71, and 4.32. Similarly, for female photo essay participants, mean M, F, GD, and Extraversion scores were 4.73, 5.80, .31, and 4.67, whereas corresponding means for the larger population of women were 4.76, 5.68, .31, and 4.63. In other words, photo essay participants seemed quite comparable to their peers on the personality dimensions assessed in this study.

Reliability of Photo Essay Ratings

Table 1 presents the reliabilities for the mean ratings of each characteristic judged from participants' photo essays. In general, these reliabilities were high, indicating that the six raters agreed substantially in most of their ratings.

Over all photo essays, reliabilities for the judged traits of masculine and feminine were undoubtedly inflated by the fact that raters judged

men and women to be quite different on these traits. Stated another way, a substantial part of raters' consensus, over all participants, resulted from the fact that they agreed that men are masculine and women feminine. Thus, interrater reliabilities for these traits were also computed separately for just men's photo essays and for just women's photo essays. Interrater reliabilities remained high for judgments of men's masculinity and femininity (alphas = .81 and .78, respectively), while reliabilities were somewhat lower for judgments of women's masculinity and femininity (alphas = .62 and .66). Thus, raters displayed more consensus when rating men's than women's masculinity and femininity, suggesting that individual differences in masculinity and femininity may have been more readily judged or observed in men's photo essays than in women's.

Interrater reliability was high for the following rated characteristics: introverted, extraverted, happy, has sense of humor, physically attractive, physically unattractive, nurturant, warm, competitive, religious, family-oriented, studious, and athletic. Agreement was lowest for the following characteristics: anxious, calm/relaxed, authentic/real, fraudulent, maladjusted, and unconventional—characteristics that seem to tap relatively “internal” traits that have a strong component of social desirability or undesirability (see Funder & Colvin, 1988; John & Robins, 1993).

Relationship between M, F, GD, and Participants' Rated Characteristics

The current data provide a portrait of M, F, and GD in terms of their correlations with participants' rated photo essay characteristics. Correlations between photo essay ratings and participants' M, F, and GD scores were computed separately for men and women. Tables 2 through 4 present correlations for men, and Tables 5 through 7 present correlations for women. In all tables correlations are listed in order of their strength.

Table 2 presents correlations between BSRI M and photo essay ratings for men; Table 3 presents correlations between BSRI F and photo essay ratings for men; and Table 4 presents correlations between GD scores and photo essay ratings for men.

Men who were high on M tended to be judged as being extraverted, not serious, not creative, not intelligent, athletic, humorous, competitive, and not studious. Men who were high on F were perceived to be

Table 2
Correlations of Photo Essay Ratings with the Bem Sex Role Inventory
(BSRI) Masculinity Scale for Men

| <i>r</i> | Photo essay rating |
|----------|--|
| -.51*** | Introverted |
| .49*** | Extraverted |
| -.39** | Serious |
| -.38** | Creative |
| -.38** | Intelligent |
| .37** | Athletic |
| .36** | Has sense of humor |
| .31* | Competitive |
| -.30* | Studious |
| -.28* | Conscientious |
| .24 | Warm |
| .24 | Irresponsible |
| .23 | Masculine |
| .22 | Coarse, vulgar |
| .22 | Dominant |
| .19 | Theatrical, self-dramatizing |
| .18 | Happy |
| -.18 | Calm, relaxed |
| .18 | Nurturant, cares for others |
| -.15 | Self-disclosing |
| .13 | Fraudulent, presents a false image of self |
| .13 | Has good self-esteem |
| .12 | Agreeable, nice, likable |
| .11 | Family-oriented |
| -.10 | Maladjusted |
| .10 | Anxious |
| -.09 | Feminine |
| .09 | Unconventional, eccentric |
| .08 | Physically unattractive |
| -.08 | Authentic, real |
| -.08 | Sad |
| -.07 | Religious |
| .06 | Troubled |
| .05 | Disagreeable, unpleasant, unlikable |
| .02 | Physically attractive |
| .01 | Conventional, ordinary |
| -.01 | Possesses broad interests |
| -.01 | Depressed |

* $p < .10$ (two-tailed)

** $p < .05$ (two-tailed)

*** $p < .01$ (two-tailed).

Table 3
Correlations of Photo Essay Ratings with the Bem Sex Role Inventory
(BSRI) Femininity Scale for Men

| <i>r</i> | Photo essay rating |
|----------|--|
| .48*** | Warm |
| .47*** | Nurturant, cares for others |
| -.30* | Coarse, vulgar |
| -.30* | Maladjusted |
| .29* | Agreeable, nice, likable |
| .28* | Feminine |
| .28 | Family-oriented |
| .26 | Studious |
| -.26 | Depressed |
| .25 | Self-disclosing |
| .24 | Religious |
| -.24 | Fraudulent, presents a false image of self |
| .23 | Happy |
| -.23 | Masculine |
| .18 | Calm, relaxed |
| -.18 | Sad |
| .18 | Intelligent |
| .18 | Has sense of humor |
| -.18 | Irresponsible |
| .18 | Conscientious |
| -.16 | Competitive |
| .15 | Authentic, real |
| -.15 | Physically unattractive |
| -.14 | Troubled |
| .14 | Has good self-esteem |
| -.13 | Anxious |
| .12 | Conventional, ordinary |
| -.12 | Dominant |
| -.11 | Introverted |
| -.10 | Disagreeable, unpleasant, unlikable |
| .10 | Creative |
| .10 | Extraverted |
| .10 | Theatrical, self-dramatizing |
| -.08 | Unconventional, eccentric |
| .07 | Serious |
| .02 | Physically attractive |
| .02 | Possesses broad interests |
| -.01 | Athletic |

* $p < .10$ (two-tailed)

*** $p < .01$ (two-tailed).

Table 4
Correlations of Photo Essay Ratings with Gender
Diagnosticity for Men

| <i>r</i> | Photo essay rating |
|----------|--|
| -.60**** | Feminine |
| .58**** | Masculine |
| -.42** | Self-disclosing |
| -.38** | Warm |
| -.35** | Creative |
| -.32* | Family-oriented |
| .31* | Coarse, vulgar |
| -.30* | Religious |
| .29* | Irresponsible |
| .29* | Conventional, ordinary |
| -.29* | Sad |
| .28 | Competitive |
| -.27 | Nurturant, cares for others |
| .27 | Has sense of humor |
| .26 | Athletic |
| -.25 | Serious |
| -.25 | Authentic, real |
| .24 | Has good self-esteem |
| -.24 | Intelligent |
| -.24 | Troubled |
| .23 | Physically attractive |
| .23 | Dominant |
| .22 | Fraudulent, presents a false image of self |
| -.22 | Unconventional, eccentric |
| -.22 | Studious |
| -.20 | Depressed |
| -.19 | Conscientious |
| .17 | Physically unattractive |
| .15 | Happy |
| -.15 | Agreeable, nice, likable |
| -.13 | Possesses broad interests |
| .11 | Disagreeable, unpleasant, unlikable |
| .09 | Calm, relaxed |
| -.09 | Extraverted |
| .06 | Introverted |
| .03 | Theatrical, self-dramatizing |
| -.03 | Anxious |
| -.03 | Maladjusted |

* $p < .10$ (two-tailed)

** $p < .05$ (two-tailed)

*** $p < .001$ (two-tailed).

warm, nurturant, not coarse or vulgar, not maladjusted, agreeable and nice, feminine, and family-oriented. And finally, men who were high on GD (that is, who showed "male-like" occupational preferences) were judged to be masculine and not feminine, not self-disclosing, not warm, not creative, not family-oriented, not religious, conventional, irresponsible, and not sad.

The corresponding correlations for women yielded a somewhat different pattern. Table 5 presents correlations between BSRI M and photo essay ratings for women; Table 6 presents correlations between BSRI F and photo essay ratings for women; and Table 7 presents correlations between GD scores and photo essay ratings for women.

Women who were high on M tended to be judged to be dominant, extraverted, theatrical, fraudulent, and irresponsible. Women who were high on F were judged to be happy, adjusted, physically attractive, have high self-esteem, and be agreeable. Finally, women who were high on GD (that is, who showed "male-like" occupational preferences) were judged to be disagreeable, athletic, anxious, sad, and (marginally) not feminine.

As noted earlier, participants' Extraversion was assessed to provide a comparison dimension in assessing the validity of photo essay judgments. Table 8 presents correlations (computed over all participants) of assessed Extraversion with rated photo essay characteristics. Because of the larger population (men and women combined), statistical significance is achieved by smaller correlations in Table 8 than in previous tables.

In general, assessed Extraversion showed an expected pattern of correlations with judged photo essay characteristics. Reasonably enough, participants who were high on Extraversion tended to be judged as extraverted, not introverted, having a sense of humor, and not serious. Stated another way, judges displayed a significant degree of accuracy in judging participants' self-reported extraversion from their photo essays.

A question may be raised regarding how much the photo essay judgments were based on photographic information versus textual information. Stated a bit differently, did judges reliably (and sometimes validly) perceive personality from photo essays in part or even largely because photo essay participants verbally described their personalities in their photo captions and self-descriptive essays?

To empirically answer these questions, I selected 9 photo essays by men and 13 by women that matched the total sample in terms of the mean and standard deviations of judges' M-F ratings within the sexes.

Table 5
Correlations of Photo Essay Ratings with Bem Sex Role Inventory
(BSRI) Masculinity Scale for Women

| <i>r</i> | Photo essay rating |
|----------|--|
| .47**** | Dominant |
| .46**** | Theatrical, self-dramatizing |
| .44*** | Extraverted |
| .42*** | Irresponsible |
| -.41*** | Introverted |
| .39*** | Fraudulent, presents a false image of self |
| -.38*** | Studious |
| .37*** | Has sense of humor |
| -.32** | Serious |
| .33** | Competitive |
| .31** | Has good self-esteem |
| .30** | Unconventional, eccentric |
| .30** | Masculine |
| -.28** | Depressed |
| -.27** | Maladjusted |
| .24* | Happy |
| -.23* | Intelligent |
| -.21 | Feminine |
| .21 | Athletic |
| -.20 | Conscientious |
| -.20 | Conventional, ordinary |
| -.19 | Sad |
| -.19 | Authentic, real |
| -.19 | Calm, relaxed |
| .19 | Coarse, vulgar |
| -.18 | Troubled |
| -.17 | Warm |
| .16 | Disagreeable, unpleasant, unlikable |
| -.15 | Self-disclosing |
| -.10 | Religious |
| -.10 | Nurturant, cares for others |
| -.08 | Agreeable, nice, likable |
| -.07 | Physically unattractive |
| .05 | Physically attractive |
| -.04 | Creative |
| -.02 | Possesses broad interests |
| -.02 | Anxious |
| .01 | Family-oriented |

* $p < .10$ (two-tailed)

** $p < .05$ (two-tailed)

*** $p < .01$ (two-tailed)

**** $p < .001$ (two-tailed).

Table 6
Correlations of Photo Essay Ratings with the Bem Sex Role Inventory
(BSRI) Femininity Scale for Women

| <i>r</i> | Photo essay rating |
|----------|--|
| -.33** | Troubled |
| .31** | Physically attractive |
| -.30** | Maladjusted |
| .28** | Happy |
| -.28** | Depressed |
| .27** | Has good self-esteem |
| .26** | Agreeable, nice, likable |
| -.26* | Physically unattractive |
| -.22 | Self-disclosing |
| .21 | Extraverted |
| -.21 | Introverted |
| -.21 | Anxious |
| -.20 | Sad |
| .20 | Calm, relaxed |
| -.20 | Disagreeable, unpleasant, unlikable |
| -.18 | Authentic, real |
| .16 | Feminine |
| .15 | Fraudulent, presents a false image of self |
| .15 | Nurturant, cares for others |
| -.14 | Serious |
| .12 | Has sense of humor |
| .12 | Competitive |
| .10 | Conventional, ordinary |
| .09 | Creative |
| -.08 | Studious |
| .08 | Conscientious |
| -.08 | Coarse, vulgar |
| .07 | Warm |
| -.06 | Masculine |
| .05 | Athletic |
| .04 | Possesses broad interests |
| -.04 | Dominant |
| -.03 | Irresponsible |
| .02 | Intelligent |
| -.01 | Religious |
| -.01 | Unconventional, eccentric |
| -.01 | Family-oriented |
| .01 | Theatrical, self-dramatizing |

* $p < .10$ (two-tailed)

** $p < .05$ (two-tailed).

Table 7
Correlations of Photo Essay Ratings with Gender
Diagnosticity for Women

| <i>r</i> | Photo essay rating |
|----------|--|
| -.38*** | Agreeable, nice, likable |
| .36** | Disagreeable, unpleasant, unlikable |
| .31** | Athletic |
| -.29** | Calm, relaxed |
| .28** | Sad |
| -.26* | Feminine |
| .25* | Anxious |
| .25* | Competitive |
| -.24* | Warm |
| .22 | Troubled |
| .21 | Masculine |
| .21 | Dominant |
| -.21 | Nurturant, cares for others |
| -.20 | Happy |
| -.20 | Creative |
| -.19 | Conscientious |
| .19 | Irresponsible |
| .18 | Depressed |
| .17 | Fraudulent, presents a false image of self |
| .15 | Unconventional, eccentric |
| .13 | Maladjusted |
| -.11 | Possesses broad interests |
| -.11 | Conventional, ordinary |
| .09 | Self-disclosing |
| -.09 | Physically attractive |
| -.08 | Introverted |
| -.08 | Intelligent |
| -.08 | Has good self-esteem |
| .07 | Coarse, vulgar |
| .06 | Theatrical, self-dramatizing |
| .05 | Serious |
| -.04 | Has sense of humor |
| -.03 | Extraverted |
| -.01 | Religious |
| -.02 | Studious |
| .00 | Family-oriented |
| .00 | Physically unattractive |
| .00 | Authentic, real |

* $p < .10$ (two-tailed)

** $p < .05$ (two-tailed)

*** $p < .01$ (two-tailed).

Table 8
Correlations of Photo Essay Ratings with Extraversion
for All Participants

| <i>r</i> | Photo essay rating |
|----------|--|
| .43**** | Extraverted |
| -.43**** | Introverted |
| .34*** | Has sense of humor |
| -.31*** | Serious |
| .26** | Theatrical, self-dramatizing |
| .25** | Happy |
| -.24** | Studious |
| .24** | Has good self-esteem |
| -.22** | Depressed |
| -.22** | Physically unattractive |
| .21** | Family-oriented |
| -.21* | Calm, relaxed |
| -.19* | Intelligent |
| .18* | Dominant |
| -.18* | Maladjusted |
| -.18* | Sad |
| .18* | Competitive |
| -.17 | Creative |
| .17 | Irresponsible |
| .16 | Religious |
| .14 | Athletic |
| -.12 | Conscientious |
| .12 | Fraudulent, presents a false image of self |
| .11 | Warm |
| .11 | Unconventional, eccentric |
| .10 | Agreeable, nice, likable |
| .10 | Physically attractive |
| -.09 | Authentic, real |
| .08 | Feminine |
| -.08 | Self-disclosing |
| -.08 | Conventional, ordinary |
| -.07 | Troubled |
| .06 | Nurturant, cares for others |
| -.04 | Possesses broad interests |
| -.03 | Anxious |
| -.03 | Disagreeable, unpleasant, unlikable |
| .02 | Coarse, vulgar |
| .01 | Masculine |

* $p < .10$ (two-tailed)

** $p < .05$ (two-tailed)

*** $p < .01$ (two-tailed)

**** $p < .001$ (two-tailed).

(As noted earlier in the Method section, some photo essays had been returned to participants after their initial rating, so it was impossible to re-rate the entire original sample of photo essays.) The mean M-F rating for the initial full set of men's photo essays was 15.4 with a standard deviation of 9.49, and the corresponding statistics for the subsample of 9 men's photo essays were 18.4 and 10.31. The mean M-F rating for the initial full set of women's photo essays was -14.8 with a standard deviation of 8.58, and the corresponding statistics for the subsample were -11.15 and 11.22. Thus, the subsamples comprised the full range of M-F ratings that were present in the total initial samples.

The selected 9 men's and 13 women's photo essays were rated by four new judges on the 38 characteristics listed in Table 1. However, this time two judges (a man and a woman) rated photo essays based only on photographs (with the text blocked out), and two additional judges (again, a man and a woman) rated the photo essays based only on text (with photographs blocked out). The ratings of the two "photo judges" were averaged together, as were the ratings of the two "text judges."

In general, these new data showed that both photo and text impressions were related to impressions formed from complete photo essays. For example, judges' ratings of men's M-F based on complete information correlated significantly with M-F ratings based only on photos ($r = .74$, one-tailed $p = .01$) and with M-F ratings based only on text ($r = .75$, one-tailed $p = .01$). Judges' ratings of women's M-F based on total information correlated significantly with M-F ratings based only on text ($r = .75$, one-tailed $p = .002$) but not with M-F ratings based only on photos ($r = .29$, one-tailed $p = .17$). Finally, judges' ratings of all participants' Extraversion based on total information correlated significantly with Extraversion ratings based only on photographs ($r = .77$, one-tailed $p < .001$) and with Extraversion ratings based only on text ($r = .59$, one-tailed $p = .003$).

Not only did both photo and text ratings generally correlate significantly with ratings based on total information, but they often showed substantial validity as well. For example, men's GD scores correlated .47 with M-F ratings based on photos (one-tailed $p = .10$) and .85 with M-F ratings based on text (one-tailed $p = .002$). Women's GD scores correlated .47 with M-F ratings based on photos (one-tailed $p = .06$) and .17 with M-F ratings based on text (*ns*). And participants' assessed Extraversion correlated significantly with Extraversion ratings based on photos ($r = .51$, one-tailed $p = .01$) and marginally with Extraversion ratings based on text ($r = .37$, one-tailed $p = .06$). These data indicate

that there was valid information communicated through photographs as well as through text, and indeed sometimes validity was greater for photo ratings than for text ratings. The correlations just reported are particularly impressive given the small sample size and the fact that photo and text ratings were each based on just two raters, and therefore possessed lower reliabilities than the photo essay ratings reported in the main study (which were based on six raters).

The Relationship between Judged M-F and Other Judged Photo Essay Characteristics

How did photo essay judges assess a participant's masculinity and femininity? One way to answer this question is to observe how judgments of masculinity and femininity correlated with other judged characteristics.

Consistent with previous research (e.g., Pedhazur & Tetenbaum, 1979; Spence, 1993), judges in the current research tended to view masculinity and femininity as bipolar opposites. That is, for ratings of both men's and women's photo essays, judged masculinity and femininity were strongly negatively correlated ($r = -.86$ for ratings of men and $-.77$ for ratings of women, both significant at $p < .001$). Therefore, these two ratings were combined into a single rating of judged M-F by subtracting judged femininity from judged masculinity.

Tables 9 and 10 present the correlations of judged M-F with other judged characteristics; Table 9 presents correlations for men and Table 10 for women. In both tables correlations are listed in order of strength. These correlations provide a rich portrait of the characteristics that were linked to judged M-F in raters' minds.

Men who were judged to be high on masculinity tended to be seen as athletic, competitive, coarse and vulgar, not warm, not studious, not conscientious, not nurturant, dominant, irresponsible, not intelligent, and not self-disclosing. This list provides a multidimensional (if not very flattering) portrait of lay conceptions of men's masculinity.¹

1. The current characterizations of lay conceptions of masculinity and femininity are largely in terms of personality traits. It is important to note that this is so because most of the characteristics judged from photo essays in the current research were personality traits, with a few possible exceptions (e.g., physically attractive, religious, family-oriented). Lay conceptions of masculinity and femininity undoubtedly comprise other components as well (such as aspects of appearance, sexuality, interests), which were not well captured in the current photo essay ratings. This does not mean that these other components are unimportant in lay conceptions of masculinity and femininity.

Table 9

**Correlations of Photo Essay Ratings of Men's Masculinity-Femininity
(Masculinity Minus Femininity) with Other Rated Characteristics**

| <i>r</i> | Photo essay rating |
|----------|--|
| .71**** | Athletic |
| .66**** | Competitive |
| .58**** | Coarse, vulgar |
| -.55**** | Warm |
| -.53*** | Studious |
| -.52*** | Conscientious |
| -.50*** | Nurturant, cares for others |
| .49*** | Dominant |
| .49*** | Irresponsible |
| -.47*** | Intelligent |
| -.43*** | Self-disclosing |
| .38** | Disagreeable, unpleasant, unlikable |
| -.37** | Agreeable, nice, likable |
| -.35** | Authentic, real |
| -.35** | Religious |
| .35** | Fraudulent, presents a false image of self |
| -.33** | Creative |
| -.33** | Serious |
| -.27 | Family-oriented |
| .25 | Has sense of humor |
| .22 | Maladjusted |
| .21 | Physically attractive |
| .21 | Physically unattractive |
| .18 | Theatrical, self-dramatizing |
| .13 | Has good self-esteem |
| -.13 | Sad |
| -.10 | Troubled |
| -.10 | Introverted |
| .09 | Extraverted |
| .06 | Anxious |
| .05 | Unconventional, eccentric |
| -.05 | Depressed |
| -.06 | Calm, relaxed |
| -.02 | Conventional, ordinary |
| .01 | Happy |
| .00 | Possesses broad interests |

** $p < .05$ (two-tailed)

*** $p < .01$ (two-tailed)

**** $p < .001$ (two-tailed).

Table 10
**Correlations of Photo Essay Ratings of Women's Masculinity-
 Femininity (Masculinity Minus Femininity)**
with Other Rated Characteristics

| <i>r</i> | Photo essay rating |
|----------|--|
| .70**** | Coarse, vulgar |
| .70**** | Dominant |
| -.58**** | Conventional, ordinary |
| .53**** | Has sense of humor |
| .52**** | Unconventional, eccentric |
| -.52**** | Warm |
| -.50**** | Physically attractive |
| -.46**** | Conscientious |
| .46**** | Physically unattractive |
| -.44*** | Family-oriented |
| .43*** | Theatrical, self-dramatizing |
| .42*** | Disagreeable, unpleasant, unlikable |
| .41*** | Irresponsible |
| -.37*** | Introverted |
| .37*** | Competitive |
| .37*** | Extraverted |
| -.35*** | Serious |
| -.35*** | Calm, relaxed |
| -.32** | Nurturant, cares for others |
| -.25* | Studious |
| .25* | Anxious |
| .24* | Athletic |
| -.24* | Agreeable, nice, likable |
| -.20 | Intelligent |
| .14 | Has good self-esteem |
| -.14 | Religious |
| .14 | Fraudulent, presents a false image of self |
| -.12 | Self-disclosing |
| -.10 | Sad |
| .09 | Happy |
| -.08 | Depressed |
| .04 | Maladjusted |
| -.03 | Possesses broad interests |
| .02 | Creative |
| .02 | Authentic, real |
| -.02 | Troubled |

* $p < .10$ (two-tailed)

** $p < .05$ (two-tailed)

*** $p < .01$ (two-tailed).

**** $p < .001$ (two-tailed).

Women who were judged to be high on femininity tended to be seen as not coarse or vulgar, not dominant, conventional and ordinary, not having a sense of humor, warm, physically attractive, conscientious, family-oriented, not self-dramatizing, agreeable, and responsible. This list includes both positive and negative characteristics. For example, feminine women were seen to possess stereotypic feminine virtues such as propriety, warmth, attractiveness, conscientiousness, agreeableness, and responsibility, but more negatively, they were also judged to be passive and nondominant, conventional, and lacking in humor.

It's worth highlighting here one important difference between the correlations for men and women that will be emphasized later: Women's judged M-F was significantly correlated with their judged physical attractiveness, whereas men's was not.

A Brief Description of Four Photo Essays

As noted earlier, judges were able to rate photo essay participants' masculinity and femininity reliably. This is noteworthy because almost none of the photo essay participants showed dramatic instances of gender nonconformity. While the previous tables of correlations give a sense of some of the characteristics that may have led photo essay participants to be judged to be masculine or feminine, brief descriptions of several actual photo essays will enrich the reader's sense of how such judgments were made. This section provides sketches of the partial content of four representative photo essays—two by men and two by women. One of the two men was rated to be quite masculine, the other not, and one of the two women was rated to be quite feminine, the other not.

Participant #1. This man was at the 16th percentile of rated masculinity among male participants.

One photograph shows the man in an apartment living room with his girlfriend. They are swinging a toddler between them. The caption reads: "...Here I'm playing with my roommate's baby and my girlfriend. This photo shows the playful side of me."

Another photograph shows the man taking a tray of cookies out of the oven. The caption reads, "...Every now and then, I like to bake fresh chocolate chip cookies. They're always better than those pre-baked at the store. I guess it is supposed to represent my 'superb' culinary [sic] skills."

A third photo shows the man holding his roommate's toddler. Caption: "I LOVE LITTLE KIDS!!! (at least ... cute ones that don't always cry!) This is my roommate's little one-year-old. His name is _____ and he is very smart. I love this kid very much...."

In his final summary essay, the man writes, "I'm really an easy-going guy. I usually get along with everyone. I'm definitely an extrovert. At parties, I'm usually the guy whos [sic] making everyone laugh, either with me or at me...."

Participant #2. This man scored at the 84th percentile on rated masculinity among male participants.

One photograph shows the man standing in front of a blank wall and a door. His hands are hooked in his pockets, and he's not smiling. The captions reads: "I should be a model!"

A second photograph is a picture of a plaque hanging on the wall that reads: "Freshman Football 1989 Coaches Award." The caption reads, "I deserved it!!"

A third photo shows the man in jeans, a baggy white T-shirt, and a white baseball cap, leaning over an engine under the open hood of his car. Caption: "There's nothing like a little fine tuning."

The summary essay reads in part: "My name is _____ and I am 18 yrs. old. I am a freshman in college as a full-time student. I am undecided on a major but am thinking about business. I played sports in high school (football, baseball) and earned a few awards. My main goal in life is to be rich. I am not working right now, but I do make a little money selling car stereo equipment. I spend a lot of time working on my car. It is my most prized possession [sic]...."

Participant #3. This woman was at the 5th percentile of rated femininity among female participants.

One photograph shows the woman sitting on a large cement turtle statue in an outdoor park sandlot. Caption: "Animals and the outdoors, two great combos."

A second photo shows the woman on Hollywood Boulevard in Los Angeles, crouched on the pavement behind Alfred Hitchcock's star. Caption: "Love of reading, especially mystery and horror."

A third photo looks through a jungle gym in a park and shows the woman hanging upside down, holding onto a bar with her knees. She's smiling and her hair is hanging straight down. Caption: "Lost, confused, and turned upside down."

The woman's summary essay reads in part: "Who am I? That's a question we've all asked ourselves at one time or another. My friends are a big part of who I am. They're very close to me and when one of them is in trouble and needs help I'll be right there by their side. In the first picture a group of friends and I celebrate Christmas with love. I was pleasantly surprised when a couple of my closest friends bought me an iguana....I've wanted one for as long as I can remember. My iguana, picture three, describes me. Not quite exotic, but not common. [He] can be calm and quiet and the next minute be a nutcase and run around his cage. I love animals, but wild animals always intrigued me...."

Participant #4. This woman was at the 93rd percentile on rated femininity among female participants.

One photograph shows the woman smiling, standing in front of a dresser and wall decorated with flowered wallpaper. There is a stuffed animal hanging on the wall, and stuffed bears sitting on the top of the dresser. The decor is traditional, with an early American flavor. Caption: "I am striking a pose by the fireplace in our country 'cottage.'"

Another photo shows the woman and her boyfriend, arms around each other's backs and holding hands with their circled outside arms, standing in a flower-filled backyard. The woman is wearing a flowered summer dress. Caption: "_____ and I at my aunt's house. Springtime flowers are lovely, aren't they!?"

A third photo shows the woman standing next to her boyfriend, ready to leave for a party. She is wearing a long, form-fitting black evening gown that exposes her shoulders. The skirt of the gown is slit up one side. She is wearing makeup and red nail polish. Her boyfriend is wearing a tuxedo shirt and tie and a black cowboy hat. Caption: "We're ready for a night on the town! Well, the company Christmas party will have to do."

The woman's photo essay reads in part: "...The photographs which I have included show people, places, and activities which are a part of my life. I am a very active person in various aspects. I love sports like water and snow skiing...."

"Overall, I presume that I am an average twenty-year-old. I love my family and friends dearly and am happy with my accomplishments so far."

DISCUSSION

The current data show that laypeople can reliably judge others' masculinity and femininity when they are provided with sufficiently rich information about the people whom they are asked to judge. The concepts of masculinity and femininity clearly are meaningful and judgable to laypeople. At the same time, these concepts are multifaceted and complex (see Spence & Buckner, 1995, for a cogent discussion of this issue).

The current photo essay data offer new insights into what is measured by M scales, F scales, and GD measures. The findings for men are particularly striking and clear: Assessed M correlated primarily with men's judged Extraversion, assessed F with men's judged warmth and nurturance, and assessed GD with men's judged masculinity and femininity. This pattern of correlations is consistent with the fact that M scales are in essence measures of instrumentality, and F scales are measures of expressiveness (Spence & Helmreich, 1980). Stated in Big Five terms, M scales load most strongly on Extraversion and F on Agreeableness (Lippa, 1991, 1995b). Thus for men, M and F scales show face validity in the current study—they correlate with precisely those judged photo essay characteristics relevant to their item content.

The pattern of correlations between men's GD scores and photo essay ratings provides strong evidence that, for men at least, GD measures tap into what laypeople mean intuitively by the words masculine and feminine more than M and F scales do. Indeed, the correlations between men's GD scores and their judged masculinity and femininity ($r = .58$ and $r = -.60$, respectively) are large by the standards of personality research in general, and larger than any other validity coefficients reported in this article, including those for Extraversion (which has often been considered to be the most judgable and observable of personality traits). These correlations are even more impressive when corrected for attenuation due to unreliability (corrected r s are respectively $.75$ and $-.79$), and the strength of these correlations is still more remarkable when one notes that GD measures, unlike M, F, and Extraversion scales, did not display obvious face validity. Occupational preference items that showed strong sex differences in this population of participants served to define a GD measure, which then proved to correlate strongly with within-sex judgments of men's masculinity and femininity based on photo essay information.

If masculinity and femininity are cognitive constructs, not palpable realities (as Bem, 1981b, 1993, has argued), then by implication, any

putative measure of masculinity and femininity must be validated against lay judgments of these constructs. The current data show clearly that GD measures correlate much more strongly with lay judgments of men's masculinity and femininity than M and F scores do.

The corresponding results for women are more complex and ambiguous, however. M correlated primarily with women's judged dominance and Extraversion, but unlike the corresponding findings for men, women's F correlated primarily with perceived adjustment and physical attractiveness, and women's GD correlated with judged disagreeableness, athleticism, maladjustment, and lack of femininity. Stated another way, M showed its expected face validity for women, but F did not—it correlated with judged adjustment and attractiveness rather than with judged warmth and nurturance. And while GD correlated weakly with a lack of judged femininity in women, it correlated most strongly with perceived disagreeableness, maladjustment, and athleticism.

Why were the results different for women and men? It is worth recalling that the interrater reliabilities of judged masculinity and femininity were lower for women than for men. Because of these lower reliabilities, judged masculinity and femininity simply could not correlate as strongly with assessed GD, M, or F for women as for men.

Perhaps a more fundamental explanation for the differing results for men and women is linked to a finding noted earlier: Judgments of women's masculinity and femininity seemed to be based more on physical attractiveness than were men's. This suggests that men may have been judged masculine or feminine based on their actual gender-related behaviors and individual differences, whereas women may have been judged masculine or feminine based on their physical appearance. Stated yet another way, women may have been judged to be feminine relatively less through "bottom-up" processing of the multidimensional information contained in their photo essays (e.g., their gender-related interests, activities, social relationships, nonverbal styles) and more through "top-down" application of a physical attractiveness stereotype.

To provide some empirical evidence for this point, I conducted stepwise regression analyses that used four predictor variables—M, F, GD, and rated physical attractiveness—to predict men's and women's rated M-F. For men, only one significant predictor variable entered into the regression equation in the first and final step: GD (multiple $r = .63$, $p < .001$). For women, however, physical attractiveness entered into the regression equation on the first step (multiple $r = .47$, $p < .001$),

and M entered the equation in the second and final step (multiple r increasing to .55, $p < .001$). These analyses show that GD was the primary predictor of men's judged M-F, whereas physical attractiveness was the primary predictor of women's judged M-F.

The influence of physical attractiveness on judgments of women's M-F may help explain, at least in part, the pattern of correlations between M, F, GD, and judged masculinity and femininity for women. If women were judged to be feminine or masculine more on the basis of physical attractiveness than men were, this would have the net effect of attenuating all correlations between judged masculinity, femininity, and assessed personality (i.e., M, F, and GD).

The importance of physical attractiveness in determining judgments of women's masculinity and femininity still leaves unexplained two interesting and potentially puzzling findings for women: (a) Women who were high on GD (that is, who expressed "male-like" occupational preferences) tended to be perceived as more maladjusted than women who were low on GD, and (b) women who were low on F tended to be perceived as less happy and more maladjusted than women who were high on F.

Because photo essay participants also took part in a larger research project on gender-related individual differences and psychological adjustment (Lippa, 1995b), they had completed various measures of psychological adjustment, including the Rosenberg Self-Esteem Scale (Rosenberg, 1965), the Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), and a trait anxiety scale. In the larger study, GD and F tended not to correlate with these self-report adjustment measures for women, suggesting that GD and F were in fact not much related to women's actual depression, anxiety, or self-esteem.

Why then were high GD and low F women judged to be more maladjusted than low GD and high F women? One hypothesis is the following: Insofar as GD and F do tap individual differences in women's culturally defined femininity, then women who enact feminine roles (low GD and high F women) may display the *appearance* of adjustment more than women who do not enact such roles. For example, women who enact feminine roles may attempt to appear cheerful, upbeat, and accommodating in their photo essays (and in social life in general), and this may be perceived, superficially at least, to reflect good adjustment. Women who do not enact stereotypic feminine roles, on the other hand, may express more (and more varied) negative affect (e.g., sadness, anger, alienation) than their more "feminine" counterparts.

The sketches presented earlier of two women's photo essays (and my impressions of other photo essays as well) provide at least anecdotal evidence supporting this hypothesis. The woman who was rated to be quite feminine (#4) displayed considerable surface cheer and perkiness in her photo essay, but one senses a degree of feminine pose here. On the other hand, the woman who was rated to be low on femininity (#3) displayed more mixed emotions, including confusion and alienation. However, this woman's photo essay also projected a humor, richness, and complexity not apparent in the other. Clearly, one interesting direction for future research on masculinity and femininity—as displayed both in photo essays and in real life—is to attempt to disentangle appearance from reality in gender-related self-presentations. The current data suggest that for women, F and GD relate to *apparent*, if not actual adjustment.

Whatever the proper interpretation of the differing results for men and women, the current data provide clear evidence that gender-related individual differences and their correlates are sometimes patterned differently for men and women (see Lippa, 1991, 1995b; Lippa & Connelly, 1990, for additional evidence in support of this point). Despite these sex differences and the problems of interpretation they present, the current results nonetheless support several strong conclusions: (a) Masculinity and femininity, while undoubtedly complex and multidimensional, are concepts that make sense to laypeople. (b) Others' masculinity and femininity can be reliably judged when judges are given sufficiently complex information on which to base their judgments. (c) GD measures predict lay judgments of men's masculinity and femininity better than M and F scales do; the findings for women, however, are more complex. (d) Physical attractiveness is linked to lay judgments of women's masculinity and femininity more strongly than it is to judgments of men's masculinity and femininity, and this finding has implications for the degree to which naive judgments of men's and women's masculinity and femininity will (or can) correlate with their assessed M, F, and GD.

REFERENCES

- Ashmore, R. D. (1990). Sex, gender, and the individual. In L. A. Pervin (Ed.) *Handbook of personality: Theory and research* (pp. 486–526). New York: Guilford.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 53–63.

- Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, **42**, 165–172.
- Bem, S. L. (1975). Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology*, **31**, 634–643.
- Bem, S. L. (1976). Probing the promise of androgyny. In A. G. Kaplan & J. P. Bean (Eds.), *Beyond sex-role stereotypes: Readings toward a psychology of androgyny* (pp. 47–62). Boston: Little, Brown.
- 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. (1985). Androgyny and gender schema theory: A conceptual and empirical integration. In T. B. Sonderegger (Ed.), *Nebraska Symposium on Motivation, 1984: Psychology and gender* (pp. 179–226). Lincoln: University of Nebraska Press.
- Bem, S. L. (1993). *The lenses of gender: Transforming the debate on sexual inequality*. New Haven: Yale University Press.
- Block, J. H. (1973). Conceptions of sex roles: Some cross-cultural and longitudinal perspectives. *American Psychologist*, **28**, 512–526.
- Cook, E. P. (1985). *Psychological androgyny*. New York: Pergamon.
- Constantinople, A. (1973). Masculinity-femininity: An exception to a famous dictum? *Psychological Bulletin*, **80**, 389–407.
- Dollinger, S. J., & Clancy, S. M. (1993). Identity, self, and personality: II. Glimpses through the autophotographic eye. *Journal of Personality and Social Psychology*, **64**, 1064–1071.
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, **55**, 149–158.
- Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, **52**, 409–418.
- John, O. P. (1990). The “Big Five” factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York: Guilford.
- John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, **61**, 521–551.
- Kenny, D. A. (1994). *Interpersonal perception*. New York: Guilford.
- Lenney, E. (1991). Sex roles and the measurement of masculinity, femininity, and androgyny. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of psychological attitudes: Vol. 1. Measures of personality and social psychological attitudes* (pp. 573–660). San Diego: Academic Press.
- Lippa, R. (1991). Some psychometric characteristics of gender diagnosticity measures: Reliability, validity, and consistency across domains and relationship to the Big Five. *Journal of Personality and Social Psychology*, **61**, 1000–1011.
- Lippa, R. (1995a). Do sex differences define gender-related individual differences within the sexes? Evidence from three studies. *Personality and Social Psychology Bulletin*, **21**, 349–355.

- Lippa, R. (1995b). Gender-related individual differences and psychological adjustment in terms of the Big Five and circumplex models. *Journal of Personality and Social Psychology*, **69**, 1184–1202.
- Lippa, R., & Connelly, S. C. (1990). Gender diagnosticity: A new Bayesian approach to gender-related individual differences. *Journal of Personality and Social Psychology*, **59**, 1051–1065.
- McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality. *Journal of Personality and Social Psychology*, **52**, 81–90.
- Myers, A. M., & Gonda, G. (1982). Utility of the masculinity-femininity construct: Comparison of traditional and androgyny approaches. *Journal of Personality and Social Psychology*, **43**, 514–522.
- Pedhazur, E. J., & Tetenbaum, T. J. (1979). Bem's Sex Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology*, **37**, 996–1016.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton: Princeton University Press.
- Spence, J. T. (1993). Gender-related traits and gender ideology: Evidence for a multifactorial theory. *Journal of Personality and Social Psychology*, **64**, 624–635.
- Spence, J. T., & Buckner, C. (1995). Masculinity and femininity: Defining the undefinable. In P. J. Kalbfleisch & M. J. Cody (Eds.), *Gender, power, and communication in human relationships* (pp. 105–138). Hillsdale, NJ: Erlbaum.
- 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: Their relationships with sex role attitudes and behaviors. *Psychology of Women Quarterly*, **5**, 147–163.
- Spence, J. T., Helmreich, R. L., & Stapp, J. (1974). The Personal Attributes Questionnaire: A measure of sex role stereotypes and masculinity-femininity. *JSAS*, **4**, 43–44.
- Spence, J. T., & Sawin, L. L. (1985). Images of masculinity and femininity: A reconceptualization. In V. E. O'Leary, R. K. Unger, & B. X. Wallston (Eds.), *Women, gender, and social psychology* (pp. 35–66). Hillsdale, NJ: Erlbaum.
- Terman, L. M., & Miles, C. C. (1936). *Sex and personality: Studies in masculinity and femininity*. New York: Russell & Russell.
- Ziller, R. C. (1990). *Photographing the self*. Newbury Park, CA: Sage.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.