THE MEASUREMENT OF PSYCHOLOGICAL ANDROGYNY [1]

SANDRA L. BEM [2]

Stanford University

This article describes the development of a new sex-role inventory that treats masculinity and femininity as two independent dimensions, thereby making it possible to characterize a person as masculine, feminine, or "androgynous" as a function of the difference between his or her endorsement of masculine and feminine personality characteristics. Normative data are presented, as well as the results of various psychometric analyses. The major findings of conceptual interest are: (*a*) the dimensions of masculinity and femininity are empirically as well as logically independent; (*b*) the concept of psychological androgyny is a reliable one; and (*c*) highly sex-typed scores do not reflect a general tendency to respond in a socially desirable direction, but rather a specific tendency to describe oneself in accordance with sex-typed standards of desirable behavior for men and women.

Both in psychology and in society at large, masculinity and femininity have long been conceptualized as bipolar ends of a single continuum; accordingly, a person has had to be either masculine or feminine, but not both. This sex-role dichotomy has served to obscure two very plausible hypotheses: first, that many individuals might be "androgynous"; that is, they might be *both* masculine and feminine, *both* assertive and yielding, *both* instrumental and expressive—depending on the situational appropriateness of these various behaviors; and conversely, that strongly sex-typed individuals might be seriously limited in the range of behaviors available to them as they move from situation to situation. According to both Kagan (1964) and Kohlberg (1966), the highly sex-typed individual is motivated to keep his behavior consistent with an internalized sex-role standard, a goal that he presumably accomplishes by suppressing any behavior that might be considered undesirable or inappropriate for his sex. Thus, whereas a narrowly masculine self-concept might inhibit behaviors that are stereotyped as feminine, and a narrowly feminine self-concept might inhibit behaviors that are stereotyped as masculine, a mixed, or

androgynous, self-concept might allow an individual to freely engage in both "masculine" and "feminine" behaviors.

The current research program is seeking to explore these various hypotheses, as well as to provide construct validation for the concept of androgyny (Bem, 1974). Before the research could be initiated, however, it was first necessary to develop a new type of sex-role inventory, one that would not automatically build in an inverse relationship between masculinity and femininity. This article describes that inventory.

The Bem Sex-Role Inventory (BSRI) contains a number of features that distinguish it from other, commonly used, masculinity–femininity scales, for example, the Masculinity–Femininity scale of the California Psychological Inventory (Gough, 1957). First, it includes both a Masculinity scale and a Femininity scale, each of which contains 20 personality characteristics. These characteristics are listed in the first and second columns of Table 1, respectively. Second, because the BSRI was founded on a conception of the sex-typed person as someone who has internalized society's sex-typed standards of desirable behavior for men and women, these personality characteristics were selected as masculine or feminine on the basis of sex-typed social desirability and not on the basis of differential endorsement by males and females as most other inventories have done. That is, a characteristic qualified as masculine if it was judged

TABLE 1

Items on the Masculinity, Femininity, and Social Desirability Scales of the BSRI

Masculine items	Feminine items	Neutral items
49. Acts as a leader	11. Affectionate	51. Adaptable
46. Aggressive	5. Cheerful	36. Conceited
58. Ambitious	50. Childlike	9. Conscientious
22. Analytical	32. Compassionate	60. Conventional
13. Assertive	53. Does not use harsh language	45. Friendly
10. Athletic	35. Eager to soothe hurt feelings	15. Happy
55. Competitive	20. Feminine	3. Helpful
4. Defends own beliefs	14. Flatterable	48. Inefficient
37. Dominant	59. Gentle	24. Jealous
19. Forceful	47. Gullible	39. Likable
25. Has leadership abilities	56. Loves children	6. Moody
7. Independent	17. Loyal	21. Reliable
52. Individualistic	26. Sensitive to the needs of others	30. Secretive
31. Makes decisions easily	8. Shy	33. Sincere
40. Masculine	38. Soft spoken	42. Solemn
1. Self-reliant	23. Sympathetic	57. Tactful
34. Self-sufficient	44. Tender	12. Theatrical
16. Strong personality	29. Understanding	27. Truthful
43. Willing to take a stand	41. Warm	18. Unpredictable
28. Willing to take risks	2. Yielding	54. Unsystematic

Note. The number preceding each item reflects the position of each adjective as it actually appears on the Inventory.

to be more desirable in American society for a man than for a woman, and it qualified as feminine if it was judged to be more desirable for a woman than for a man. Third, the BSRI characterizes a person as masculine, feminine, or androgynous as a function of the difference between his or her endorsement of masculine and feminine personality characteristics. A person is thus sex typed, whether masculine or feminine, to the extent that this difference score is high, and androgynous, to the extent that this difference score is low. Finally, the BSRI also includes a Social Desirability scale that is completely neutral with respect to sex. This scale now serves primarily to provide a neutral context for the Masculinity and Femininity scales, but it was utilized during the development of the BSRI to insure that the inventory would not simply be tapping a general tendency to endorse socially desirable traits. The 20 characteristics that make up this scale are listed in the third column of Table 1.

ITEM SELECTION

Both historically and cross-culturally, masculinity and femininity seem to have represented two complementary domains of *posi-*

tive traits and behaviors (Barry, Bacon, & Child, 1957; Erikson, 1964; Parsons & Bales, 1955). In general, masculinity has been associated with an instrumental orientation, a cognitive focus on "getting the job done"; and femininity has been associated with an expressive orientation, an affective concern for the welfare of others.

Accordingly, as a preliminary to item selection for the Masculinity and Femininity scales, a list was compiled of approximately 200 personality characteristics that seemed to the author and several students to be both positive in value and either masculine or feminine in tone. This list served as the pool from which the masculine and feminine characteristics were ultimately chosen. As a preliminary to item selection for the Social Desirability scale, an additional list was compiled of 200 characteristics that seemed to be neither masculine nor feminine in tone. Of these "neutral" characteristics, half were positive in value and half were negative.

Because the BSRI was designed to measure the extent to which a person divorces himself from those characteristics that might be considered more "appropriate" for the opposite sex, the final items were selected

TABLE 2

MEAN SOCIAL DESIRABILITY RATINGS OF THE MASCULINE, FEMININE, AND NEUTRAL ITEMS

Item	Male judges			Female judges		
	Masculine item	Feminine item	Neutral item	Masculine item	Feminine item	Neutral item
For a man	5.59	3.63	4.00	5.83	3.74	3.94
For a woman	2.90	5.61	4.08	3.46	5.55	3.98
Difference	2.69	1.98	.08	2.37	1.81	.04
t	14.41*	12.13*	.17	10.22*	8.28*	.09

* $p < .001$.

for the Masculinity and Femininity scales if they were judged to be more desirable in American society for one sex than for the other. Specifically, judges were asked to utilize a 7-point scale, ranging from 1 ("Not at all desirable") to 7 ("Extremely desirable"), in order to rate the desirability in American society of each of the approximately 400 personality characteristics mentioned above. (E.g., "In American society, how desirable is it for a man to be truthful?" "In American society, how desirable is it for a woman to be sincere?") Each individual judge was asked to rate the desirability of all 400 personality characteristics either "for a man" or "for a woman." No judge was asked to rate both. The judges consisted of 40 Stanford undergraduates who filled out the questionnaire during the winter of 1972 and an additional 60 who did so the following summer. In both samples, half of the judges were male and half were female.

A personality characteristic qualified as masculine if it was independently judged by both males and females in both samples to be significantly more desirable for a man than for a woman ($p < .05$).[3] Similarly, a personality characteristic qualified as feminine if it was independently judged by both males and females in both samples to be significantly more desirable for a woman than for a man ($p < .05$). Of those characteristics that satisfied these criteria, 20 were selected for the Masculinity scale and 20 were selected for the Femininity scale (see the first and second columns of Table 1, respectively).

[3] All significance levels in this article are based on two-tailed t tests.

A personality characteristic qualified as neutral with respect to sex and hence eligible for the Social Desirability scale (a) if it was independently judged by both males and females to be no more desirable for one sex than for the other ($t < 1.2$, $p > .2$) and (b) if male and female judges did not differ significantly in their overall desirability judgments of that trait ($t < 1.2$, $p > .2$). Of those items that satisfied these several criteria, 10 positive and 10 negative personality characteristics were selected for the BSRI Social Desirability scale in accordance with Edwards' (1964) finding that an item must be quite positive or quite negative in tone if it is to evoke a social desirability response set. (The 20 neutral characteristics are shown in the third column of Table 1.)

After all of the individual items had been selected, mean desirability scores were computed for the masculine, feminine, and neutral items for each of the 100 judges. As shown in Table 2, for both males and females, the mean desirability of the masculine and feminine items was significantly higher for the "appropriate" sex than for the "inappropriate" sex, whereas the mean desirability of the neutral items was no higher for one sex than for the other. These results are, of course, a direct consequence of the criteria used for item selection.

Table 3 separates out the desirability ratings of the masculine and feminine items for male and female judges rating their *own* sex. These own-sex ratings seem to best represent the desirability of these various items as perceived by men and women when they are asked to describe *themselves* on the inven-

TABLE 3

MEAN SOCIAL DESIRABILITY RATINGS OF THE
MASCULINE AND FEMININE ITEMS
FOR ONE'S OWN SEX

Item	Male judges for a man	Female judges for a woman
Masculine	5.59	3.46
Feminine	3.63	5.55
Difference	1.96	2.09
t	11.94*	8.88*

* $p < 001$.

tory. That is, the left-hand column of Table 3 represents the phenomenology of male subjects taking the test and the right-hand column represents the phenomenology of female subjects taking the test. As can be seen in Table 3, not only are "sex-appropriate" characteristics more desirable for both males and females than "sex-inappropriate" characteristics, but the phenomenologies of male and female subjects are almost perfectly symmetric: that is, men and women are nearly equal in their perceptions of the desirability of sex-appropriate characteristics, sex-inappropriate characteristics, and the difference between them ($t < 1$ in all three comparisons).

SCORING

The BSRI asks a person to indicate on a 7-point scale how well each of the 60 masculine, feminine, and neutral personality characteristics describes himself. The scale ranges from 1 ("Never or almost never true") to 7 ("Always or almost always true") and is labeled at each point. On the basis of his responses, each person receives three major scores: a Masculinity score, a Femininity score and, most important, an Androgyny score. In addition, a Social Desirability score can also be computed.

The Masculinity and Femininity scores indicate the extent to which a person endorses masculine and feminine personality characteristics as self-descriptive. Masculinity equals the mean self-rating for all endorsed masculine items, and Femininity equals the mean self-rating for all endorsed feminine items. Both can range from 1 to 7. It will be recalled that these two scores are logically independent. That is, the structure of the

test does not constrain them in any way, and they are free to vary independently.

The Androgyny score reflects the relative amounts of masculinity and femininity that the person includes in his or her self-description, and, as such, it best characterizes the nature of the person's total sex role. Specifically, the Androgyny score is defined as Student's t ratio for the difference between a person's masculine and feminine self-endorsement; that is, the Androgyny score is the difference between an individual's masculinity and femininity normalized with respect to the standard deviations of his or her masculinity and femininity scores. The use of a t ratio as the index of androgyny—rather than a simple difference score—has two conceptual advantages: first, it allows us to ask whether a person's endorsement of masculine attributes differs significantly from his or her endorsement of feminine attributes and, if it does ($|t| \geq 2.025$, $df = 38$, $p < .05$), to classify that person as significantly sex typed; and second, it allows us to compare different populations in terms of the percentage of significantly sex-typed individuals present within each.[4]

It should be noted that the greater the absolute value of the Androgyny score, the more the person is sex typed or sex reversed, with high positive scores indicating femininity and high negative scores indicating masculinity. A "masculine" sex role thus represents not only the endorsement of masculine attributes but the simultaneous rejection of feminine attributes. Similarly, a "feminine" sex role represents not only the endorsement of feminine attributes but the simultaneous rejection of masculine attributes. In contrast, the closer the Androgyny score is to zero, the more the person is androgynous. An "androg-

---

[4] A Statistical Package for the Social Sciences (SPSS) computer program for calculating individual t ratios is available on request from the author. In the absence of computer facilities, one can utilize the simple Androgyny difference score, Femininity − Masculinity, as the index of androgyny. Empirically, the two indices are virtually identical ($r = .98$), and one can approximate the t-ratio value by multiplying the Androgyny difference score by 2.322. This conversion factor was derived empirically from our combined normative sample of 917 students at two different colleges.

ynous" sex role thus represents the equal endorsement of both masculine and feminine attributes.

The Social Desirability score indicates the extent to which a person describes himself in a socially desirable direction on items that are neutral with respect to sex. It is scored by reversing the self-endorsement ratings for the 10 undesirable items and then calculating the subject's mean endorsement score across all 20 neutral personality characteristics. The Social Desirability score can thus range from 1 to 7, with 1 indicating a strong tendency to describe oneself in a socially undesirable direction and 7 indicating a strong tendency to describe oneself in a socially desirable direction.

PSYCHOMETRIC ANALYSES

Subjects

During the winter and spring of 1973, the BSRI was administered to 444 male and 279 female students in introductory psychology at Stanford University. It was also administered to an additional 117 male and 77 female paid volunteers at Foothill Junior College. The data that these students provided represent the normative data for the BSRI, and, unless explicitly noted, they serve as the basis for all of the analyses that follow.

Internal Consistency

In order to estimate the internal consistency of the BSRI, coefficient alpha was computed separately for the Masculinity, Femininity, and Social Desirability scores of the subjects in each of the two normative samples (Nunnally, 1967). The results showed all three scores to be highly reliable, both in the Stanford sample (Masculinity $\alpha = .86$; Femininity $\alpha = .80$; Social Desirability $\alpha = .75$) and in the Foothill sample (Masculinity $\alpha = .86$; Femininity $\alpha = .82$; Social Desirability $\alpha = .70$). Because the reliability of the Androgyny t ratio could not be calculated directly, coefficient alpha was computed for the highly correlated Androgyny difference score, Femininity $-$ Masculinity, using the formula provided by Nunnally (1967) for linear combinations. The reliability of the Androgyny difference score was .85

for the Stanford sample and .86 for the Foothill sample.

Relationship between Masculinity and Femininity

As indicated earlier, the Masculinity and Femininity scores of the BSRI are logically independent. That is, the structure of the test does not constrain them in any way, and they are free to vary independently. The results from the two normative samples reveal them to be empirically independent as well (Stanford male $r = .11$, female $r = -.14$; Foothill male $r = -.02$, female $r = -.07$). This finding vindicates the decision to design an inventory that would not artifactually force a negative correlation between masculinity and femininity.

Social Desirability Response Set

It will be recalled that a person is sex typed on the BSRI to the extent that his or her Androgyny score reflects the greater endorsement of "sex-appropriate" characteristics than of "sex-inappropriate" characteristics. However, because of the fact that the masculine and feminine items are all relatively desirable, even for the "inappropriate" sex, it is important to verify that the Androgyny score is not simply tapping a social desirability response set.

Accordingly, product-moment correlations were computed between the Social Desirability score and the Masculinity, Femininity, and Androgyny scores for the Stanford and Foothill samples separately. They were also computed between the Social Desirability score and the absolute value of the Androgyny score. These correlations are displayed in Table 4. As expected, both Masculinity and Femininity were correlated with Social Desirability. In contrast, the near-zero correlations between Androgyny and Social Desirability confirm that the Androgyny score is not measuring a general tendency to respond in a socially desirable direction. Rather, it is measuring a very specific tendency to describe oneself in accordance with sex-typed standards of desirable behavior for men and women.

TABLE 4

CORRELATION OF MASCULINITY, FEMININITY, AND ANDROGYNY WITH SOCIAL DESIRABILITY

Sample	Masculinity with social desirability		Femininity with social desirability		Androgyny with social desirability		|Androgyny| with social desirability	
	Males	Females	Males	Females	Males	Females	Males	Females
Stanford	.42	.19	.28	.26	.12	.03	.08	−.10
Foothill	.23	.19	.15	.15	−.07	.06	−.12	−.09
Stanford and Foothill combined	.38	.19	.28	.22	.08	.04	.03	−.10

Test–Retest Reliability

The BSRI was administered for a second time to 28 males and 28 females from the Stanford normative sample. The second administration took place approximately four weeks after the first. During this second administration, subjects were told that we were interested in how their responses on the test might vary over time, and they were explicitly instructed not to try to remember how they had responded previously. Product-moment correlations were computed between the first and second administrations for the Masculinity, Femininity, Androgyny, and Social Desirability scores. All four scores proved to be highly reliable over the four-week interval (Masculinity $r = .90$; Femininity $r = .90$; Androgyny $r = .93$; Social Desirability $r = .89$).

TABLE 5

CORRELATION OF THE MASCULINITY–FEMININITY SCALES OF THE CALIFORNIA PSYCHOLOGICAL INVENTORY (CPI) AND GUILFORD-ZIMMERMAN SCALE WITH THE MASCULINITY, FEMININITY, AND ANDROGYNY SCALES OF THE BSRI

Scale	CPI		Guilford-Zimmerman	
	Males	Females	Males	Females
BSRI Masculinity	−.42	−.25	.11	.15
BSRI Femininity	.27	.25	.04	−.06
BSRI Androgyny	.50	.30	−.04	−.06

Note. The CPI scale is keyed in the feminine direction, whereas the Guilford-Zimmerman scale is keyed in the masculine direction.

Correlations with Other Measures of Masculinity–Femininity

During the second administration of the BSRI, subjects were also asked to fill out the Masculinity–Femininity scales of the California Psychological Inventory and the Guilford-Zimmerman Temperament Survey, both of which have been utilized rather frequently in previous research on sex roles. Table 5 presents the correlations between these two scales and the Masculinity, Femininity, and Androgyny scales of the BSRI. As can be seen in the table, the Guilford-Zimmerman scale is not at all correlated with any of the three scales of the BSRI, whereas the California Psychological Inventory is moderately correlated with all three. It is not clear why the BSRI should be more highly correlated with the CPI than with the Guilford-Zimmerman scale, but the fact that none of the correlations is particularly high indicates that the BSRI is measuring an aspect of sex roles which is not directly tapped by either of these two scales.

NORMS

Table 6 presents the mean Masculinity, Femininity, and Social Desirability scores separately by sex for both the Stanford and the Foothill normative samples. It also presents means for both the Androgyny t ratio and the Androgyny difference score. As can be seen in the table, males scored significantly higher than females on the Masculinity scale, and females scored significantly higher than males on the Femininity scale in both samples. On the two measures of androgyny, males scored on the masculine side of zero

TABLE 6

SEX DIFFERENCES ON THE BSRI

Scale score	Stanford University			Foothill Junior College		
	Males $(n = 444)$	Females $(n = 279)$	t	Males $(n = 117)$	Females $(n = 77)$	t
Masculinity						
M	4.97	4.57		4.96	4.55	
SD	.67	.69	7.62*	.71	.75	3.86*
Femininity						
M	4.44	5.01		4.62	5.08	
SD	.55	.52	13.88*	.64	.58	5.02*
Social Desirability						
M	4.91	5.08		4.88	4.89	
SD	.50	.50	4.40*	.50	.53	ns
Androgyny t ratio						
M	−1.28	1.10		−.80	1.23	
SD	1.99	2.29	14.33*	2.23	2.42	5.98*
Androgyny Difference score						
M	−0.53	.43		−.34	.53	
SD	.82	.93	14.28*	.97	.97	6.08*

* $p < .001$.

and females scored on the feminine side of zero. This difference is significant in both samples and for both measures. On the Social Desirability scale, females scored significantly higher than males at Stanford but not at Foothill. It should be noted that the size of this sex difference is quite small, however, even in the Stanford sample.

Table 7 presents the percentage of subjects within each of the two normative samples who qualified as masculine, feminine, or androgynous as a function of the Androgyny t ratio. Subjects are classified as sex typed, whether masculine or feminine, if the Androgyny t ratio reaches statistical significance

($|t| \geq 2.025$, $df = 38$, $p < .05$), and they are classified as androgynous if the absolute value of the t ratio is less than or equal to one. Table 7 also indicates the percentage of subjects who fall between these various cutoff points. It should be noted that these cutoff points are somewhat arbitrary and that other investigators should feel free to adjust them in accordance with the characteristics of their particular subject populations.

CONCLUDING COMMENT

It is hoped that the development of the BSRI will encourage investigators in the areas of sex differences and sex roles to ques-

TABLE 7

PERCENTAGE OF SUBJECTS IN THE NORMATIVE SAMPLES CLASSIFIED AS
MASCULINE, FEMININE, OR ANDROGYNOUS

Item	Stanford University		Foothill Junior College	
	Males $(n = 444)$	Females $(n = 279)$	Males $(n = 117)$	Females $(n = 77)$
% feminine ($t \geq 2.025$)	6	34	9	40
% near feminine ($1 < t < 2.025$)	5	20	9	8
% androgynous ($-1 \leq t \leq +1$)	34	27	44	38
% near-masculine ($-2.025 < t < -1$)	19	12	17	7
% masculine ($t \leq -2.025$)	36	8	22	8

tion the traditional assumption that it is the sex-typed individual who typifies mental health and to begin focusing on the behavioral and societal consequences of more flexible sex-role self-concepts. In a society where rigid sex-role differentiation has already outlived its utility, perhaps the androgynous person will come to define a more human standard of psychological health.

REFERENCES

BARRY, H., BACON, M. K., & CHILD, I. L. A cross-cultural survey of some sex differences in socialization. *Journal of Abnormal and Social Psychology,* 1957, **55,** 327–332.

BEM, S. L. Sex-role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology,* 1974, in press.

EDWARDS, A. L. The measurement of human motives by means of personality scales. In D. Levine (Ed.), *Nebraska symposium on motivation: 1964.* Lincoln: University of Nebraska Press, 1964.

ERIKSON, E. H. Inner and outer space: Reflections on womanhood. In R. J. Lifton (Ed.), *The woman in America.* Boston: Houghton Mifflin, 1964.

GOUGH, H. G. *Manual for the California Psychological Inventory.* Palo Alto, Calif.: Consulting Psychologists Press, 1957.

KAGAN, J. Acquisition and significance of sex-typing and sex-role identity. In M. L. Hoffman & L. W. Hoffman (Eds.), *Review of child development research.* Vol. 1. New York: Russell Sage Foundation, 1964.

KOHLBERG, L. A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences.* Stanford, Calif.: Stanford University Press, 1966.

NUNNALLY, J. C. *Psychometric theory.* New York: McGraw-Hill, 1967.

PARSONS, T., & BALES, R. F. *Family, socialization, and interaction process.* New York: Free Press of Glencoe, 1955.