Sex Bias in the Naming of Stimulus Persons

Joseph Kasof

Researchers often use sex-typed names (e.g., John vs. Joan) to identify stimulus persons' sex, assuming that such names communicate sex only. In fact, however, such names also create impressions that have little or nothing to do with sex. Study 1 analyzed the age connotations, intellectual-competence connotations, and attractiveness of sex-typed names used in 230 published studies on sexism and fear of success. On each of these variables, the literature was pervasively confounded in a manner favoring male stimulus persons. Study 2 found that the name biases reported in Study 1 were positively correlated with outcome measures in a sample of sexism studies, but only when names were presented with limited other information. Possible causes of the bias are discussed, and recommendations for naming stimulus persons are presented, including a list of male names and female names matched on several key variables.

In a classic experiment on sex discrimination, Goldberg (1968) asked subjects to evaluate essays attributed to either male or female authors, whose sex was indicated by sex-typed names such as John (or Joan) McKay, Stephen (or Stella) Hamilton, and Paul (or Pauline) Conger. In general, subjects evaluated essays more positively when attributed to male authors than to female authors—clear evidence, it would seem, of sex discrimination.

In another classic study, Horner (1968) measured sex differences in fear of success (FOS) by comparing stories written by male and female subjects in response to verbal cues depicting competitive success. Horner altered names and personal pronouns to give male subjects a version depicting male success and female subjects a version depicting female success: “After first-term finals, John (Anne) finds himself (herself) at the top of his (her) medical school class.” Female subjects wrote more stories expressing negative themes than did male subjects, a finding that Horner took as evidence that female subjects feared success more than did male subjects.

These experiments became famous partly because they seemed to provide an elegant and inexpensive way to operationalize stimulus-person sex. Following Goldberg (1968) and Horner (1968), the use of sex-typed names became, and remains, a standard procedure in research on sex stereotypes, sex discrimination, and FOS. Furthermore, names are often used to indicate stimulus-person sex in other fields of research, such as interpersonal attraction (e.g., Jamieson, Lydon, & Zanna, 1987; Lydon, Jamieson, & Zanna, 1988), altruism (e.g., Batson et al., 1988), psycholinguistics (e.g., Au, 1986; Crawley & Stevenson, 1990), self-monitoring (e.g., Snyder, Berscheid, & Matwyckuk, 1988), social stratification (e.g., Bose & Rossi, 1983), social cognition (e.g., Forgas, Burnham, & Trimboli, 1988), social control (e.g., Frank, Cullen, Travis, & Borntrager, 1989; Link & Cullen, 1983), social identity (Brewer, Ho, Lee, & Miller, 1987), and person perception (e.g., Rook, 1987).

In such experiments, it is implicitly assumed that sex-typed names vary dichotomously, being either male or female. In fact, however, sex-typed names communicate a great deal more than just the name bearer's sex. Such names differ in attractiveness and connote impressions of the name bearer's age, intellectual competence, race, ethnicity, social class, and other attributes. The names Michael and Edith, for example, differ on variables other than sex: Edith is considered less attractive, more old-fashioned, and connotative of lower intellectual ability than is Michael (from data in Buchanan & Brining, 1971; Dion, 1985; Dunkling, 1986; Mehrabian, 1988).

To use such mismatched names in tests of sexism or FOS is to confound sex with other variables and to bias subjects against female stimulus persons. Yet in the research to be reported herein, it was discovered that such mismatches constitute a large majority of male-female name pairings used in sexism and FOS studies.

In this article, several name-related variables are considered: forename attractiveness; age connotation; intellectual-competence connotation; title of address; surname; and racial, ethnic, and social class connotations. I present evidence that demonstrates a pervasive sex bias in the naming of stimulus persons. This bias is then related to outcome measures in a sample of published sexism studies, to determine what effect, if any, the biased naming of stimulus persons had in such studies. Finally, I discuss possible causes of this bias, and I present recommendations for the unbiased naming of stimulus persons, including a list of male names and female names matched on several key variables.

Forename Attractiveness

Forenames vary in attractiveness. This has been demonstrated in more than 20 studies, using various procedures including the methods of paired comparisons (Allen, Brown, Dickinson, & Pratt, 1941; Finch, Kilgren, & Pratt, 1944; Wal-
Sex Bias in the Naming of Stimulus Persons

Forenames also connote impressions of the name bearer's intelligence, creativity, knowledge, and academic aptitude. Furthermore, apparent nondiscrimination between male stimulus persons with old-fashioned forenames and female stimulus persons with new-fashioned forenames may result from age discrimination masking sex discrimination, rather than from a true absence of discrimination. Moreover, many sexism and FOS experiments—including several different name comparisons, often with varying patterns of age stereotype—produce experimental results inextricably confounded by potential age and sex effects.

Age Stereotype

Forenames also connote impressions of the name bearer's intelligence, creativity, knowledge, and academic aptitude. Furthermore, apparent nondiscrimination between male stimulus persons with old-fashioned forenames and female stimulus persons with new-fashioned forenames may result from age discrimination masking sex discrimination, rather than from a true absence of discrimination. Moreover, many sexism and FOS experiments—including several different name comparisons, often with varying patterns of age stereotype—produce experimental results inextricably confounded by potential age and sex effects.

Intellectual-Competence Connotation

Forenames also connote impressions of the name bearer's intelligence, creativity, knowledge, and academic aptitude. Furthermore, apparent nondiscrimination between male stimulus persons with old-fashioned forenames and female stimulus persons with new-fashioned forenames may result from age discrimination masking sex discrimination, rather than from a true absence of discrimination. Moreover, many sexism and FOS experiments—including several different name comparisons, often with varying patterns of age stereotype—produce experimental results inextricably confounded by potential age and sex effects.
trolled. Charles and Carl, for example, received virtually identical attractiveness scores in both the Buchanan and Bruning (1971) study and the Dion (1985) study, and each name fell out of fashion at approximately the same time and rate (Dunkling, 1986). In two recent studies, however, subjects expected significantly greater intelligence and creativity from persons named Charles than from persons named Carl (Mehrabian, 1988, 1990).

Dependent measures commonly used in sexism research are sensitive to intellectual-competence stereotypes. Expectations of intellectual achievement and competence influence evaluations of job applicant resumes (Dipboye, Fromkin, & Wiback, 1975), essays (Babadj, 1977; Chase, 1979; Dash, 1975; Friend, Kalin, & Giles, 1979; Gibb, 1983; Graham & Dyer, 1987; Guttmann & Bar-Tal, 1982; Hughes, Keeling, & Tuck, 1983; Peck, 1978; Sherif, 1935; Valasek, Avolio, & Forbinger, 1979; but see Panek, Deitchman, Burikholder, Speroff, & Haude, 1976), and paintings (Bernberg, 1953; Ellerman, Dowling, Hinschen, Kemp, & White, 1981; Ellerman & Smith, 1983; Etough & Sanders, 1974; Farnsworth & Beaumont, 1929; Farnsworth & Misumi, 1931; Gibbons & Kassin, 1987; but see Honig & Carterette, 1978; Pheterson, Kiesler, & Goldberg, 1971).

**Study 1**

Study 1 was an attempt to determine whether past research on sexism and FOS was confounded by forename attractiveness, age stereotype, and intellectual-competence connotation. If so, and if such confounding generally favored male stimulus persons, then such confounding may help explain the existing literature's general findings of apparent sexism and of greater FOS imagery among females than males (for reviews, see Baptow, 1986; Olian, Schwab, & Haberfeld, 1989; Tosi & Einbender, 1985; Wallston & O'Leary, 1981; for opposing conclusions, see Swim, Borgida, Maruyama, & Myers, 1989; Tresemmer, 1977). To test for such confounding, I examined the attractiveness, age stereotypes, and intellectual-competence connotations of male and female forenames used in such studies.

**Method**

**Literature search.** To find studies in which names were used to identify male and female stimulus persons, I checked the *Social Science Citation Index* for all references to Goldberg (1968), Horner (1968, 1972), and three other highly cited experiments in which stimulus-person sex was operationalized through the use of sex-typed names: Mischel's (1974) study of the evaluation of scholarly articles attributed to male or female authors in stereotypically masculine or feminine fields, Pheterson et al.'s (1971) experiment on women's evaluations of paintings attributed to male or female artists, and Rosen and Jerdee's (1973) vignette study of the evaluation of supervisory behavior in business situations as a function of the managers' sex. Studies indexed in *Psychological Abstracts* under sexism, fear of success, or sex discrimination were likewise checked for the use of sex-typed names to operationalize stimulus-person sex. All studies published through 1987 and conducted in the United States or Canada were considered. Then I consulted the *National Faculty Directory* (1986, 1987) and the *APA Membership Register* (Lazo, 1987, 1988) to locate authors of articles that did not specify all names used. I contacted these authors and received a complete list of names for 45 of 86 such articles. Only studies for which complete lists were available were included in the present review. The entire bibliography, represented in Appendix A, encompasses 198 articles reporting 230 studies. Of these studies, 160 (70%) were concerned with sexism, and 70 (30%) were concerned with FOS.

This literature search was more inclusive than that of any previous reviews of published research on sexism or FOS. Compared with the present study, for example, the largest previous review of sexism research (Swim et al., 1989) spanned fewer years (1968–1985), fewer articles (106) and fewer studies (123); fewer headings searched in *Psychological Abstracts* and fewer citations searched in the *Social Science Citation Index*, a narrower range of stimulus-person characteristics (e.g., age), and a narrower range of dependent variables (e.g., studies on attraction, attribution, and persuasion were excluded). Most important of all, Swim et al. excluded studies in which subjects rated stimulus persons with little or no individualizing information. Research on sexism, including that of Swim et al., has consistently shown that individualizing information significantly reduces or eliminates the impact of sex-typed names (Krueger & Rothbart, 1988; Locksley, Borgida, Brekke, & Hepburn, 1980; Locksley, Hepburn, & Ortiz, 1982; Swim et al., 1989; Tosi & Einbender, 1985). Therefore, by including only those studies that presented sex-typed names with individualizing information, Swim et al. inadvertently restricted their review to studies in which sexism effects would be least likely and weakest. The present review's inclusion of studies that used sex-typed names with little or no individualizing information thus provides a more representative sample of the sexism literature than did the earlier review. This is particularly important because research on names has found that individualizing information also mitigates the impact of forename and surname attractiveness (O'Sullivan, Chen, Mohapatra, Sigelman, & Lewis, 1988; Steele & Smithwick, 1989) and, as Study 2 demonstrates, even within the restricted range of studies included in Swim et al.'s review, individualizing information significantly reduced the impact of name-related confounding on "sexism" effect sizes.

**Measures of forename attractiveness.** The 199 forenames used in these 230 studies were coded for attractiveness using Dion's (1985) desirable-undesirable ratings of 786 forenames and Buchanan and Bruning's (1971) like-dislike ratings of 1,060 forenames. I chose these two data sets because they cover far more names than any other comparable set. Both studies used 7-point bipolar scales, with higher scores indicating greater attractiveness. For the 478 names common to both studies, Dion's ratings, gathered in Toronto in 1978, correlate positively with Buchanan and Bruning's ratings, gathered in Ohio in 1970 (r = .66, p < .0001). Moreover, in neither data set do male and female names differ generally in attractiveness, t(770) = 0.55 and t(101) = 0.45, ns (Dion, 1985, p. 290). Of the 230 studies under review, 207 (90%) used at least one forename comparison in which both names were rated in Buchanan and Bruning's (1971) study, and 210 studies (91%) used at least one forename comparison in which both names were rated in Dion's (1985) study. To calculate mean attractiveness ratings for each study and across all studies, I used only those forename comparisons for which both names were rated.

Because Dion (1985) and Buchanan and Bruning (1971) did not provide ratings for all 199 forenames used in the studies under review, I also gathered new attractiveness ratings. Thirty-two undergraduates (6 male and 26 female) at the University of Texas received course credit...
for judging the 199 forenames in terms of both attractiveness and intellectual-competence connotation. The forenames were arranged into three different random sequences, which were presented to subjects on a random basis. Subjects rated each name on a 7-point bipolar scale ranging from 'strongest disliking' (1) to 'strongest liking' (7). Before rating names for attractiveness, half the subjects first rated the same names, presented in a different random sequence, for connotations of intellectual competence (to be described). An obvious threat to the validity of these data is that they were collected more than 1 decade after the mean publication date of the studies under review (1978) and more than 2 decades after the original Goldberg (1968) and Horner (1968) studies. Name fashions have changed in the years since many of these studies were conducted (Dunkling, 1986). Nevertheless, despite this substantial interval, the 1990 ratings correlate positively with those of Buchanan and Bruning for the 125 names common to both studies (r = .66, p < .0001) and with those of Dion for the 156 names studied in common (r = .78, p < .0001).

Measures of age stereotype. Each forename used in the 230 studies under review was classified into one of three categories: "older adult," "younger adult," and "age unassociated." Dunkling's (1986) frequency rankings for the United States for the years 1925, 1940, 1950, 1960, and 1970 were converted into two scores for each name: the name's mean rank for 1925 and 1940 and its mean rank for 1950, 1960, and 1970. A name was categorized as "older adult" if its mean prewar rank was at least 10 positions higher than its mean postwar rank, as "younger adult" if its postwar mean rank was at least 10 positions higher than its prewar mean rank, and as "age unassociated" if its mean prewar and postwar ranks were less than 10 positions apart. Dunkling's published data were limited to the 50 most popular names in a given year. For 1950, 1960, and 1970, unpublished rankings beyond the top 50 were obtained. For 1925 and 1940, however, such data were unavailable; for these years, I coded a name's position as 51 when it was not among the top 50. To avoid bias caused by the availability of unpublished rankings only for postwar years, I eliminated names from analysis if no data for prewar years were available but unpublished data were available for postwar years. I categorized shortened names and variant spellings consistent with their original form; for example, Steve and Stephen were given the same age categorization as Steven.

This procedure resulted in 124 age-categorized forenames and 8 variant spellings (see Appendix B). Of these, 47% were categorized as "older adult," 29% as "younger adult," and 24% as "age unassociated." Consistent with negative stereotyping of the elderly, mean attractiveness ratings, averaged across all three scales, were 3.16 for "older adult" names, 4.61 for "age unassociated," and 4.54 for "younger adult" names. A one-way analysis of variance (ANOVA) indicated that these age differences in attractiveness were significant on all three measures (F = 27.72, 23.55, and 35.76, respectively, on the Buchanan and Bruning, 1971, Dion, 1985, and 1990 measures, all ps < .0001).

Measures of intellectual-competence connotation. Forenames used in the 230 studies under review were coded for intellectual-competence stereotype using data from Mehriban's (1988) study of name connotations. Subjects in Mehriban's study rated on a 9-point bipolar scale ranging approximately from 0 (lowest intelligence) to 100 (highest intelligence) the degree of intellectual competence they expected from people bearing each of the 199 forenames used in the studies under review. Subjects rated each name on a 7-point bipolar scale ranging from lowest intellectual competence (1) to highest intellectual competence (7). As described earlier, names were arranged in three different random sequences, which were presented to subjects on a random basis. Before rating names for intellectual competence, half of the subjects first judged the attractiveness of the same names presented in a different random order.

For an additional measure of sex bias, Mehriban's (1990) data on intellectual-competence connotations were used. On 9-point rating scales, subjects indicated how much intelligence, creativity, ambition, and success they associated with 795 female and 973 male forenames. These ratings, gathered at UCLA in 1989, were then averaged and linearly transformed into separate male and female scales ranging approximately from 0 (lowest competence) to 100 (highest competence). Because these scales represented separate within-sex transformations, they could not be used to estimate the absolute magnitude of difference between male and female names and hence could not be used to assess general cultural differences in the intellectual-competence connotations of male and female names. (The raw, untransformed ratings on which these scales were based were unavailable) These scales could be used, however, for determining whether researchers selected male and female names from corresponding locations within each sex's distribution of intellectual-competence connotation and hence for determining whether researchers were biased in naming stimulus persons. Of the 230 studies under review, 228 (99%) used at least one forename comparison in which both names were rated by Mehriban. In using these data to compare mean intellectual-competence ratings for each study and across all studies, only those forename comparisons for which both names were rated were considered.


2 Unpublished data were supplied through the courtesy of Cleveland Evans, upon whose data Dunkling's (1986) rankings were based.

3 Each measure of competence also correlated positively, although less highly, with attractiveness. Mehriban's (1988) ratings correlated .56 with the attractiveness ratings of Buchanan and Bruning (1971) for the 89 names common to both studies, .63 with those of Dion (1985) for the 91 names common to both studies, and .60 with the 1990 attractiveness ratings for the 104 names common to both studies (all ps < .0001). Similarly, my 1990 Texas competence ratings correlated .54 with the attractiveness ratings of Buchanan and Bruning (1971) for the 125 names common to both studies, .80 with those of Dion (1985) for the 156 names common to both studies, and .82 with my 1990 attractiveness ratings (all ps < .0001). For female names, Mehriban's (1990) ratings correlated .56 with Buchanan and Bruning's (1971) ratings of the 64 names common to both studies, .65 with Dion's (1985) ratings of the 77 names common to both studies, and .69 with the 1990 attractiveness ratings of the 90 names common to both studies (all ps < .0001). For male names, Mehriban's (1990) ratings correlated .35 with Buchanan and Bruning's (1971) ratings for the 60 names common to both studies, .67 with Dion's (1985) ratings for the 70 names common to both studies, and .55 with my 1990 attractiveness ratings for the 90 names common to both studies (all ps < .0001). The mean correlation between measures of attractiveness and measures of intellectual-competence connotation (r = .62) was lower than the mean correlation...
Results

Forename attractiveness. Results indicated that the sexism literature under review was confounded by forename attractiveness in a direction favoring males. On all three measures of forename attractiveness, the mean rating of male names exceeded that of the female names with which they were compared at below the .0001 level (see Table 1). In addition, using the Buchanan and Bruning (1971) data, the mean attractiveness rating of male forenames exceeded that of female forenames with which they were compared in 131 (94%) of 139 sexism studies. On the Dion (1985) ratings, the mean attractiveness score of male forenames exceeded that of female forenames with which they were contrasted in 126 (90%) of 140 sexism studies. Using the 1990 ratings, the mean attractiveness score of male forenames exceeded that of female forenames with which they were contrasted in 112 (70%) of 160 sexism studies. Each of these results was significant at below the .0001 level, using the binomial distribution with the null hypothesis that name pairings would favor female stimulus persons as frequently as they favored male stimulus persons.

Results indicated that the FOS literature under review was also confounded by forename attractiveness in a direction favoring males. As Table 1 shows, the mean rating of male names exceeded that of the female names with which they were compared on all three measures of forename attractiveness, although this difference was not significant for the 1990 ratings. The mean attractiveness rating of male forenames exceeded that of female forenames with which they were compared in 63 (93%) of 68 FOS studies (p < .0001). On the Dion (1985) ratings, the mean attractiveness score of male forenames exceeded that of female forenames with which they were contrasted in 67 (96%) of 70 FOS studies (p < .0001). On the 1990 data, however, this pattern was reversed: The mean attractiveness score of male forenames exceeded that of female forenames with which they were contrasted in only 17 (24%) of 70 FOS studies. The discrepancy with the 1990 data was restricted to only those name pairings for which Buchanan and Bruning had complete ratings, male and female forenames did not significantly differ in attractiveness (t = .91), and the mean attractiveness score of male forenames exceeded that of female names with which they were contrasted in only 15 (21%) of 70 FOS studies. Likewise, when the analysis of 1990 data was restricted to only those name pairings for which Buchanan and Bruning had complete ratings, male and female forenames did not significantly differ in attractiveness (t = .56), and the mean attractiveness score of male forenames exceeded that of female names with which they were contrasted in only 13 (19%) of 68 FOS studies.

Year of publication had little effect on the degree of naming bias. Within the sexism literature, year of publication and the attractiveness differential between male and female names varied significantly on the Dion (1985) scale, r(271) = .12, p < .05, but insignificantly on the Buchanan and Bruning (1971) scale, r(259) = .05, and on the 1990 attractiveness scale, r(330) = .06. Within the FOS literature, year of publication varied insignificantly with the male–female attractiveness differential on all three scales: r(90) = .20 on Dion, r(96) = −.09 on Buchanan and Bruning, and r(103) = .07 on the 1990 attractiveness ratings.

Age stereotype. Results indicated that the literature under review was confounded by age stereotype in a pattern favoring male stimulus persons. In the studies under review, there were 340 name comparisons in which both names could be classified as “older adult,” “younger adult,” or “age unassociated.” Of these, 265 (78%) were mismatched in terms of age connotation. Age-unassociated male names were contrasted with new-fashioned female names in 29 cases, and age-unassociated female names were contrasted with new-fashioned male names in 17 cases. The most important mismatches, however, were the 219 that contrasted an old-fashioned name of one sex with a new-fashioned or age-unassociated name of the other sex, because these were the cases in which ageism might occur. Of these 219 comparisons, 204 (93%) contrasted older female names against younger or age-unassociated male names. This result was significant at below the .0001 level, using the binomial distribution with the null hypothesis that name pairings would favor female stimulus persons as frequently as they favored male stimulus persons.

This bias appeared in both the sexism and FOS literatures. In sexism studies, of the 146 name pairings in which an old-fashioned name of one sex was contrasted with a new-fashioned or age-unassociated name of the other sex, 131 (90%) favored male stimulus persons (p < .0001). In FOS studies, of the 73 name pairings that contrasted an old-fashioned name of one sex with a new-fashioned or age-unassociated name of the other sex, all 73 (100%) favored male stimulus persons (p < .0001).

Year of publication had no effect on age-stereotype confounding. Mean dates of publication were 1977.67 for name pairings that favored females, 1978.83 for pairings that favored males, and 1978.24 for pairings that favored neither, F(2) = 1.24, ns. There was no significant difference in year of publication in either the sexism or FOS studies.

4 This discrepancy resulted largely from the ratings of a single name pair, John versus Anne, which following Horner (1968) was used very commonly in FOS studies, including in 73% of those under review and 37 (53%) that used no other name pairing. In 1990, subjects rated Anne as more attractive than John. However, in earlier studies of forename attractiveness in which both names were rated, John without exception was rated higher than Anne (Buchanan & Bruning, 1971; Crisp, Apostal, & Luessenheide, 1984; Dion, 1985; Garwood, Baer, Levine, Carroll, & O'Neal, 1981; Lawson, 1980).
Table 1
Mean Attractiveness Ratings for Male Names and Females Names Used in Sexism and Fear-of-Success (FOS) Research

<table>
<thead>
<tr>
<th>Research</th>
<th>Male names</th>
<th>Female names</th>
<th>t</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Sexism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1970 data</td>
<td>259</td>
<td>4.91</td>
<td>0.68</td>
<td>4.29</td>
</tr>
<tr>
<td>1978 data</td>
<td>271</td>
<td>4.69</td>
<td>0.68</td>
<td>3.95</td>
</tr>
<tr>
<td>1990 data</td>
<td>330</td>
<td>4.06</td>
<td>0.99</td>
<td>3.57</td>
</tr>
<tr>
<td>FOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1970 data</td>
<td>96</td>
<td>5.17</td>
<td>0.37</td>
<td>4.16</td>
</tr>
<tr>
<td>1978 data</td>
<td>90</td>
<td>4.83</td>
<td>0.41</td>
<td>4.61</td>
</tr>
<tr>
<td>1990 data</td>
<td>103</td>
<td>4.15</td>
<td>0.57</td>
<td>4.08</td>
</tr>
</tbody>
</table>

Note. 1970 data are from Buchanan and Bruning (1971), 1978 data are from Dion (1985), and 1990 data are from the present study.
* p < .0001.

Intellectual-competence connotation. Results indicated that the sexism literature under review was confounded by intellectual-competence connotation in a direction favoring males. On all three measures, the intellectual-competence connotation of male names exceeded that of the female names with which they were compared at below the .0001 level (see Table 2). In addition, on Mehrabian's (1988) scale, the mean intellectual-competence rating of male forenames exceeded that of female forenames with which they were compared in 104 (85%) of 123 sexism studies. On the 1990 ratings, the mean intellectual-competence score of male forenames exceeded that of female forenames with which they were contrasted in 127 (79%) of 160 sexism studies. On Mehrabian's (1990) male scale and female scale, the mean intellectual-competence rating of male forenames exceeded that of female forenames with which they were compared in 136 of 158 (86%) sexism studies. Each of these results was significant at below the .0001 level, using the binomial distribution with the null hypothesis that name pairings would favor female stimulus persons as frequently as they favored male stimulus persons.

Results indicated that the FOS literature under review was also confounded by intellectual-competence connotation in a direction favoring males. On all three measures, the intellectual-competence connotation of male names exceeded that of the female names with which they were compared at below the .0001 level (see Table 2). In addition, on Mehrabian (1988) scale, the mean intellectual-competence rating of male forenames exceeded that of female forenames with which they were compared in 62 (97%) of 64 FOS studies. Using the 1990 ratings, the mean intellectual-competence rating of male forenames exceeded that of female forenames with which they were contrasted in 63 (90%) of 70 FOS studies. On Mehrabian's (1990) male scale and female scale, the mean intellectual-competence rating of male forenames exceeded that of female forenames with which they were contrasted in 68 of 70 (97%) FOS studies. Each of these results was significant at below the .0001 level, using the binomial distribution with the null hypothesis that name pairings would favor female stimulus persons as frequently as they favored male stimulus persons.

Year of publication had no effect on intellectual-competence confounding. On each measure, the annual mean difference between male and female names correlated insignificantly with

Table 2
Connotations of Intellectual-Competence for Male Names and Female Names Used in Sexism and Fear-of-Success (FOS) Research

<table>
<thead>
<tr>
<th>Research</th>
<th>Male names</th>
<th>Female names</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sexism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1987 data</td>
<td>209</td>
<td>5.72</td>
</tr>
<tr>
<td>1989 data</td>
<td>324</td>
<td>74.44</td>
</tr>
<tr>
<td>1990 data</td>
<td>330</td>
<td>4.65</td>
</tr>
<tr>
<td>FOS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1987 data</td>
<td>84</td>
<td>5.88</td>
</tr>
<tr>
<td>1989 data</td>
<td>103</td>
<td>78.96</td>
</tr>
<tr>
<td>1990 data</td>
<td>103</td>
<td>4.74</td>
</tr>
</tbody>
</table>

Note. 1987 data are from Mehrabian (1988), 1989 data are from Mehrabian (1990), and 1990 data are from the present study.
* p < .0001.
Discussion

Taken together, these findings indicate that research on sexism and FOS is pervasively confounded by forename attractiveness, age stereotype, and intellectual-competence connotation in a manner favoring male stimulus persons.

Many researchers using forenames in sexism research also used photographs to convey the stimulus persons' sex (e.g., Cann, Siegfried, & Pearce, 1981; Cash & Kilcullen, 1985; Di- boye et al., 1975). To avoid confounding sex and physical attractiveness, most researchers pretested photographs and selected those of males and females that were equally attractive. Research on names, however, suggests that these methodological precautions may have failed to standardize physical attractiveness halo effects (Eagly, Ashmore, Makhijani, & Longo, 1991; Webster & Driskell, 1983). Several studies indicated that subjects of both sexes considered female stimulus persons more beautiful when identified by attractive forenames than by unattractive forenames (Garwood, Cox, Kaplan, Wasserman, & Sulzer, 1980; Hensley & Spencer, 1985; Infante et al., 1980, Table 1; but see Hassebrauck, 1988). Because forename attractiveness affects impressions of physical attractiveness, sexism experiments confounded by forename attractiveness and using photographs of males and females rated as equally physically attractive in pretesting (when unidentified by name) were confounded not only by forename attractiveness but also by physical attractiveness.

Age-typed forenames convey impressions not only of the name bearer's age but also of his or her generation. Indeed, some apparent ageism may in fact be discrimination and stereotyping on the basis of cohort membership. Virtually all research in these fields was conducted during the 1960s, 1970s, and 1980s, using subjects born after World War II. Some age stereotypes found in such research probably derive less from impressions of senescence than from impressions of historically and culturally specific differences between prewar and postwar birth cohorts. One such difference is female participation in professional education and employment: The proportion of female applicants and admissions to medical schools, for example, increased drastically from prewar to postwar cohorts (Cole, 1986). Most FOS studies under review (57%) used "older adult" female names in verbal cues depicting success in medical school, thereby activating stereotypes not only of women succeeding in the traditionally masculine field of medicine (Cherry & Deaux, 1978; Feather & Raphaelson, 1974; Monahan, Kuhn, & Shaver, 1974; Shapiro, 1979) but also, perhaps, of women doing so before the 1970s, when such success became much more gender appropriate (Cole, 1986). If so, the often-replicated Horner procedure not only measures stereotypes rather than or in addition to motivation, but specifically it measures stereotypes of a time when female success in medical school was more inappropriate than at the time of testing for FOS, casting further doubt on the test's validity.

Connotation of intellectual competence is a particularly important confounding variable because the major dependent variables in most sexism and FOS studies include impressions of intellectual worth. Essay evaluation, resume evaluation, and person perception tap impressions of intellectual competence. In addition, FOS cues typically depicted stimulus persons succeeding not just in competitive tasks per se, but specifically in competitive intellectual tasks, such as medical school exams. Thus, such projective instruments may measure stereotypes or fantasies of intelligent versus unintelligent people succeeding at intellectual tasks. Given the present findings, it may be stated that sexism and FOS experiments generally compared impressions of intellectually inferior males with impressions of intellectually superior females on tasks requiring intellectual competence.

Study 2

Study 1 documented a pervasive bias in the naming of stimulus persons in past studies of sexism and fear of success. In such studies, researchers generally selected male names that were more attractive, more youthful, and more intellectually competent in their connotations than were the female names with which researchers contrasted them.

Did this naming bias actually make a difference in the results of such studies? A recent meta-analysis of sexism studies by Swim et al. (1989) found only a weak bias against female stimulus persons, thus challenging earlier reviews of research using the Goldberg paradigm (Basow, 1986; Olian et al., 1988; Tosi & Einbender, 1985; Wallston & O'Leary, 1981) and casting doubt on the importance of the naming bias documented in Study 1. Swim et al.'s review is difficult to interpret, however, because its literature search was restricted to studies in which sex-typed names were presented with individuating information. As noted earlier, research on sexism, including that of Swim et al., has consistently shown that individuating information mitigates the impact of sex-typed forenames (Krueger & Rothbart, 1988; Locksley et al., 1980; Locksley et al., 1982; Swim et al., 1989; Tosi & Einbender, 1985). Therefore, the unintended effect of this selection criterion was to limit the meta-analysis to studies in which sexism findings should be least likely and weakest. In 71% of the studies meta-analyzed by Swim et al., stimulus persons were identified not only with sex-typed names but also with tape-recordings, photographs, films or videos, a paragraph or more of verbal description, resumes, job applications, or live presentation of an experimental confederate. (In the remaining 29% of studies meta-analyzed by Swim et al., stimulus persons were presented with less than one paragraph of written description.) Swim et al.'s exclusion of studies in which stimulus persons were identified without individuating information, that is, by sex-typed name only, is particularly important because research on names has found that individuating information also mitigates the impact of forename and surname attractiveness (O'Sullivan et al., 1988; Steele & Smithwick, 1989). Moreover, as Study 2 demonstrates, even within the restricted range of studies included in Swim et al.'s review, individuating
information significantly reduced the impact of name-related confounding on “sexism” effect sizes.

Study 2 was undertaken to determine whether naming bias influenced experimental outcomes in a portion of the sexism literature reviewed in Study 1. Specifically, I predicted that (a) the difference in attractiveness and in intellectual-competence connotation between male and female names would vary directly with the degree of reported bias against female stimulus persons and (b) the correlation between name confounding and observed bias against female stimulus persons would be higher in studies that presented sex-typed names with little individuating information than in studies that presented sex-typed names with much individuating information. These predictions were tested using naming-bias data from Study 1 and effect-size and individuating-information data from Swim et al. (1989).3

**Method**

Swim et al. (1989) meta-analyzed 123 studies, of which 57 were included among the 230 reviewed in Study 1. Effect-size data were unavailable for 4 of these studies. Of the 53 remaining studies, 24 were excluded from the present analysis because the results were reported insufficiently for calculation of effect sizes, regardless of the results' significance (Swim et al. had assigned these studies effect sizes of zero and had analyzed them separately from studies with known effect sizes). Additionally, 1 study that was removed from Swim et al.'s meta-analysis because it was an outlier was similarly removed from the present analysis (its removal did not significantly affect results). The present analysis was conducted with the 28 remaining studies, which were reported in the 25 articles denoted by asterisk in Appendix A. These 28 studies encompassed 3,337 male and 3,241 female subjects; the number of subjects per study ranged from 36 to 1,587.

The sexism effect sizes used in the present analysis were those calculated by Swim et al. (1989) as mean effect sizes across all findings within each study. Correlations between naming bias and sexism effect sizes were obtained by weighted regression analysis, with effect sizes weighted by the reciprocal of their variances (Hedges & Olkin, 1985). Negative values indicated more favorable ratings of males, and positive values indicated more favorable ratings of females.

Studies were also coded for amount of individuating information. Although Swim et al. (1989) coded four levels of individuating information, the relatively small number of studies in the present analysis required that these four levels be collapsed into just two: high and low. Studies high in individuating information were those coded by Swim et al. as Levels 3 or 4 and identified stimulus persons not only with sex-typed names but also with photographs, tape recordings, films or videos, a paragraph or more of verbal description, resumes, job applications, or live presentation of an experimental confederate. Examples of such studies are Dipboye et al. (1975), Fidell (1970), and Pheterson et al. (1971). Studies low in individuating information were those coded by Swim et al. as Levels 1 or 2 and presented sex-typed names with less than one paragraph of written description of the stimulus person. Examples of such studies are Baruch (1972), Chobot, Goldberg, Abramson, and Abramson (1974), and Goldberg (1968). As noted earlier, studies in which stimulus persons were identified without individuating information, that is, by sex-typed name only, were not included in the Swim et al. review. Of the 28 studies in the present analysis, 12 were coded as high and 16 as low in individuating information.

Naming bias was calculated by subtracting mean ratings of each study's male names from mean ratings of female names used in the study. For example, if three name pairings were used in an experiment, mean ratings for the three male names were averaged into a single male mean for the study and then subtracted from the averaged mean ratings for the three female names. Hence, if naming bias affected the results of sexism studies as predicted, then naming bias and sexism effect sizes would be positively correlated.

Naming-bias measures were restricted to ratings of forename attractiveness and intellectual-competence connotation. Age-stereotype data are nominal and cannot be averaged in studies that used multiple name comparisons with varying patterns of age stereotype. In only two of the remaining studies did age stereotype favor female stimulus persons, thus providing too few cases for analysis. In addition, data from Mehrabian (1990) were not used in Study 2 because, as described earlier, these scales represented separate within-sex linear transformations and could not be used to estimate the absolute magnitude of difference between male names and female names.

Of the 28 studies analyzed in Study 2, complete name data were available for 16 studies using Buchanan and Bruning’s (1971) attractiveness ratings, 19 studies using Dion’s (1985) attractiveness ratings, 12 studies using Mehrabian’s (1988) intellectual-competence ratings, and all 28 studies using the 1990 ratings of attractiveness and intellectual-competence connotation.

**Results**

Results indicated that naming bias was highly correlated with sexism effect sizes but only when stimulus persons were presented with limited individuating information. For studies low in individuating information, name bias correlated positively with effect sizes on all five measures of forename attractiveness and intellectual-competence connotation (see Table 3). The mean correlation was .65, with a range from .44 to .79. Only for the two 1990 scales, however, were the sample sizes large enough (n = 12) to reach significance. By contrast, for studies high in individuating information, the mean correlation was -.11, with a range from -.64 to .19. Only one of the five measures of forename attractiveness and intellectual-competence connotation reached significance (see Table 3).

As predicted, results also indicated that the correlation between naming bias and effect sizes was higher in studies low in individuating information than in studies high in individuating information. On each measure of attractiveness and competence connotation, correlations in studies low in individuating information were higher than in studies high in individuating information, with a mean difference of .76 across all five measures. Using Fisher’s r-to-z transformation for unequal sample sizes, such differences reached one-tailed significance or marginal significance on four of the five measures, despite the small number of studies analyzed (z = 1.79, p = .04, for 1970 attractiveness ratings; z = 1.13, p = .13, for 1978 attractiveness ratings; z = 2.39, p = .008, for 1990 attractiveness ratings; z = .90, ns, for 1987 competence ratings; and z = 2.43, p = .008, for 1990 competence ratings).

**Discussion**

These findings are consistent with the hypothesis that the biased naming of stimulus persons documented in Study 1 influenced the outcomes of many sexism studies. Nevertheless, several limitations should be emphasized. First, the number of

---

3 Effect-size and individuating-information data as well as valuable technical advice were generously provided by Janet Swim.
studies included in the present analysis was small, particularly for the 1970, 1978, and 1987 ratings of attractiveness and competence. Although a larger sample size would increase the significance of the correlations found in Study 2, it is possible that a larger sample would lower the correlations between effect size and naming bias. Second, the present analysis examined only studies that included at least some individuating information. On the basis of past research and theory (e.g., Anderson, 1974), it seems likely that biased naming of stimulus persons had an even stronger impact in studies that lacked individuating information, but this has yet to be demonstrated. Third, the present analysis excluded studies with unknown effect sizes. Such studies, had their effect sizes been calculable, might have reduced the observed association between naming bias and outcome measures. Fourth, because Study 2 examined sexism studies only, the results may not be generalized to FOS studies. Although FOS researchers typically presented sex-typed names with little individuating information and would be presently classified as low in individuating information, such studies used projective measures of unknown sensitivity to forename effects.

General Discussion

Study 1 documented a pervasive sex bias in the naming of stimulus persons in research on sexism and fear of success. In such research, investigators generally gave subjects more favorable impressions of male stimulus persons than of female stimulus persons by selecting male names that were more attractive, more youthful, and more intellectually competent in their connotations than the female names with which researchers contrasted them. In Study 2, the naming bias reported in Study 1 was found to be correlated with outcome measures in a sample of sexism studies, but only when sex-typed names were presented with limited individuating information. The general discussion below focuses on three issues: possible causes of the biased naming of stimulus persons, other confounding vari-
SEX BIAS IN THE NAMING OF STIMULUS PERSONS

randomly selected sample of forenames, it is unknown whether male forenames and female forenames differ generally in their intellectual-competence connotations in this culture. As reported in Study 1, however, the analysis of Mehrabian's (1990) male and female scales clearly indicates that researchers did not select male names and female names from corresponding locations within each sex's distribution of intellectual-competence connotation. That researchers selected forenames from higher positions within the distribution of male names than from within the distribution of female names suggests an experimenter bias beyond whatever general cultural differences may exist between male forenames and female forenames' connotations of intellectual competence.

Although the cultural stock of available names does not appear to be biased, there remains the possibility of sexually biased usage of forenames in our culture. From an unbiased stock, parents may select more positive names for their boys than for their girls. If such a bias were usual, positive male forenames would be more common and familiar than their female equivalents. Hence, researchers generating names for stimulus persons might find that positive male forenames come to mind more readily than equally positive female forenames. The favorable naming of male stimulus persons would then be representative of usage, though not of the stock of available forenames.

The possibility of sex bias in the naming of children has not been studied in the United States, but Dion (1985) reported evidence suggesting a sexist naming bias among Canadian parents. In Canadian newspaper birth announcements, male forenames were found to be more attractive than female forenames as measured by both the Dion (1985) and Buchanan and Bruning (1971) scales. Furthermore, in a sample of graduation rosters from the University of Toronto, male graduates had more attractive forenames than did female graduates as measured by the Dion scale. Several serious methodological flaws, however, cast doubt on the validity of Dion's (1985) findings. Both the birth-announcement and graduation data greatly oversampled children high in socioeconomic status. This is particularly true for Dion's sample of university graduates, of whom more than 50% received honors and remained at the university for 1 year beyond the graduation requirement. Research on names indicates that naming practices vary by socioeconomic status (Alford, 1988; Taylor, 1974; Willis, Willis, & Gier, 1982; Zweigenhaft, 1977) and that forename characteristics are related to academic achievement (Busse & Seraydarian, 1978b; Garwood, 1976; Nelson, 1977; Savage & Wells, 1948; Zweigenhaft, 1977; but see Crisp et al., 1984). Furthermore, several studies suggest interactive effects of sex and class or academic achievement on forename characteristics (Busse & Seraydarian, 1978b; Nelson, 1977; Willis et al., 1982; Zweigenhaft, 1977, Study 2). Thus, Dion's samples of newspaper birth announcements and university graduation records may not reflect general naming practices in Canada, much less in the United States. Additionally, several key omissions in reporting results make Dion's study difficult to interpret. For example, although the Buchanan and Bruning scale was used to analyze newspaper birth announcements, no such analysis was reported for the graduation lists; nor was any indication provided of the number of names for which attractiveness ratings were unavailable or of the number of male subjects and female subjects in the graduation sample, so that t tests might be performed (they were not performed by Dion). Finally, the possibility that the biased naming of stimulus persons is attributable to a general bias in the naming of children seems a particularly unlikely explanation of favoritism in age stereotypes, which were coded in Study 1 on the basis of actual usage rather than subject ratings. According to usage rankings, American parents are no more likely to assign names that have fallen out of fashion to their girls than to their boys; indeed, there is a weak tendency in the opposite direction (Dunkling, 1986). For example, as measured by the classification method used in Study 1, more "older adult" names appeared among the 50 most commonly used male names than among the 50 most commonly used female names in both 1960 and 1970 (8 male vs. 7 female for 1960; 7 male vs. 2 female for 1970; Dunkling, 1986, pp. 43, 47). Clearly, more research on this issue is needed.

If the biased naming of stimulus persons is attributable to the researchers, how might it be explained? One possibility is that the naming bias represents an "experimenter expectancy effect" (Rosenthal, 1966). Researchers expecting or hoping to find evidence of sexism or FOS may have unwittingly chosen name pairs that would make these outcomes likelier.

If so, the researchers' sex may have influenced their naming of stimulus persons, just as researchers' sex was found to influence experimental outcomes on sex differences in conformity (Eagly & Carli, 1981). Analysis of the present data reveals weak evidence for a sex-of-experimenter effect. Female principal investigators (Pis) named male stimulus persons more favorably than did male Pis on five of the six measures of attractiveness and intellectual-competence connotation. On only three of these measures, however, did such differences attain significance. With 1990 competence ratings, the mean differential between male and female stimulus persons was significantly greater among female Pis (t = 286, M = .65) than among male Pis (t = 147, M = .42), t(431) = 2.60, p < .005. Similarly, with Mehrabian's (1990) male and female scales, the mean differential between male and female stimulus persons was significantly greater among female Pis (t = 283, M = 13.82) than among male Pis (t = 144, M = 10.18), t(425) = 2.08, p < .02. With 1990 attractiveness ratings, the mean differential between male and female stimulus persons was marginally greater among female Pis (t = 286, M = .44) than among male Pis (t = 147, M = .27), t(431) = 1.51, p = .07. A comparison of results for articles with mostly male versus mostly female authors yielded similarly weak evidence for a sex-of-experimenter effect. On all six measures, name pairings from articles with mostly female authors evidenced greater favoritism toward male stimulus persons than did those from articles with mostly male authors. As with PI data, however, these differences attained or approached significance on only four measures. Using the 1990 competence

6 Mehrabian's (1988) ratings of intellectual-competence connotation, like the 1990 competence ratings, were available only for names used in the studies under review. Hence, such ratings cannot be used to determine general cultural norms. In addition, as noted earlier, the untransformed ratings on which Mehrabian's (1990) within-sex transformed data were based were unavailable.
scale, the mean difference between male and female stimulus persons was significantly greater in articles with mostly female authors ($n = 222, M = .63$) than in articles with mostly male authors ($n = 133, M = .41$), $t(353) = 2.37, p < .01$. Similarly, on Mehrabian’s (1990) ratings of competence connotation, the mean difference between male and female stimulus persons was significantly greater in articles with mostly female authors ($n = 220, M = 13.75$) than in articles with mostly male authors ($n = 130, M = 9.95$), $t(348) = 2.03, p < .05$. On Mehrabian’s (1988) competence scale, the mean difference between male and female stimulus persons was marginally greater in articles with mostly female authors ($n = 145, M = .63$) than in articles with mostly male authors ($n = 86, M = .48$), $t(229) = 1.42, p = .08$. Using the 1990 attractiveness ratings, the mean difference between male and female stimulus persons was marginally greater in articles with mostly female authors ($n = 222, M = .43$) than in articles with mostly male authors ($n = 133, M = .25$), $t(353) = 1.49, p = .07$.

Another way to test for experimenter effects would be to search for evidence of sexually biased naming of stimulus persons in fields other than sexism and FOS. Such a bias, in experiments lacking sexually differentiated results, would point to a more general cause than the experimenter effect. This is an area ripe for future research.

One such general cause was proposed and investigated in unpublished work by Brown and Khosroshahi (1989). In the English language, male and female terms are typically conjoined in a “frozen order” in which the male term is primary (Cooper & Ross, 1975). For example, most people speak of he and she, Mr. and Mrs., male and female, and John and Joan rather than vice versa. The primacy of male terms in English construction may have caused researchers to think of male names first. By Brown and Khosroshahi’s count, about half of the name pairings in the studies under review consist of phonological derivatives, such as John versus Joan, Paul versus Pauline, and Robert versus Roberta. This suggests that many researchers attempted to match names phonologically, perhaps in an effort to control forename impressions. If such researchers first generated male names and later derived phonologically similar female names, then male names and female names would have been produced by two different processes. Male names would have been generated on the basis of familiarity and would have been highly attractive, because forename familiarity varies directly with forename attractiveness (Allen et al., 1941; Busse & Seraydarian, 1978a; Colman, Hargreaves, & Sluckin, 1981; Crisp et al., 1984; Harrison, 1969; West & Shults, 1976). Female names, however, would have been generated on the basis of phonological resemblance, which would not necessarily have resulted in matched attractiveness. Furthermore, because forename attractiveness varies directly with intellectual-competence connotations in both sexes, the familiar male names initially generated by researchers would connote higher intellectual competence than would the phonologically similar female names with which researchers paired them.

Brown and Khosroshahi’s (1989) preliminary evidence provides equivocal support for this explanation. In an experiment requiring subjects to produce phonologically derived opposite-sex name pairings, most undergraduates of both sexes initially generated familiar male names rather than female or unfamiliar male names. Furthermore, among highly mismatched name pairings used in the studies under review, the more attractive names, whether male or female, contained fewer syllables than did their less attractive counterparts. Among less mismatched name pairings, more attractive names did not significantly differ from less attractive names in number of syllables. Because primary elements in frozen orders usually contain fewer syllables than do secondary elements (Cooper & Ross, 1975), this finding supports the notions that initially generated names are generally more attractive than are derived names and that male names were usually generated first in the strongly mismatched pairings. Contrary to the Brown-Khosroshahi hypothesis, however, phonologically derived pairings did not significantly differ from nonderived pairings in the attractiveness differential between male names and female names.

Surnames may provide an additional test of the Brown-Khosroshahi hypothesis. As with forenames, researchers who used surnames may have initially generated a surname for the male rather than the female stimulus person. Initially generated surnames would have been generated on the basis of familiarity and would be unattractive, because highly familiar surnames generally are unattractive (Colman, Sluckin, & Hargreaves, 1981; see also Arthaud, Hohneck, Ramsey, & Pratt, 1948). Thus, for male stimulus persons, forename attractiveness should vary inversely with surname attractiveness, male full names generally consisting of highly familiar and attractive forenames paired with highly familiar and unattractive surnames (e.g., John Smith, Marcia Jones, Robert Williams). Correlations between forename and surname attractiveness should be less negative for female stimulus persons given the same surname as male stimulus persons from which their forenames were derived, with female full names generally consisting of relatively unfamiliar and unattractive forenames paired with highly familiar and unattractive surnames (e.g., Joan Smith, Marcia Jones, Roberta Williams). This is another area for further research.

### Other Name-Related Confounding Variables

**Title of address.** Many sexism researchers identified stimulus-person sex not only by sex-typed titles of address (Baruch, 1972; Burstin, Doughtie, & Raphaeli, 1980; Calder & Ross, 1977; Chobot et al., 1974; Garland, 1977; Goldberg, 1968, 1974; M. B. Isaacs, 1981; Panek et al., 1976; Pheterson et al., 1971; Rosen & Jerdee, 1973; Rosen, Jerdee, & Prestwich, 1975; Swenson, Nash, & Roos, 1984; Valasek et al., 1979). For example, the alleged authors in the Goldberg (1968) study were titled Mr. or Miss.

Such titles emphasize the stimulus person’s sex, but they also give impressions having little or nothing to do with sex. Miss and Mrs. denote marital status, which has been found to affect person perception. Married stimulus persons of either sex are judged more positively than unmarried people on many variables, including job success, reliability, security, career dedication, and happiness (Bryan, Coleman, Ganong, & Bryan, 1986; Etaugh & Kasley, 1981; Etaugh & Malstrom, 1981; Etaugh & Petroski, 1985; Etaugh & Riley, 1983; Etaugh & Stern, 1984; Oliphant & Alexander, 1982). Experiments that contrasted a female titled Miss to a male titled Mr. did not merely contrast female to male. Such experiments pitted an unmarried female...
agreement, and 78% called him a man (the remainder were unsure). The latter group made a decision in favor of the man's claims 67% of the time, in favor of the woman's in 32%, and were undecided 1% of the time. The majority of subjects also showed a preference for the man, and his gender was the most influential variable. 

The following analyses examine the extent to which the study's findings are consistent with empirical evidence and theoretical explanations. 

**Results:** The data indicate that the participants were influenced by the given gender and the submissive or dominant role of the speaker. The analysis of variance (ANOVA) revealed a significant main effect of gender, *F*(1, 99) = 15.54, *p* < .001, with the group of male speakers receiving higher ratings than the group of female speakers. Furthermore, the interaction between gender and role was significant, *F*(1, 99) = 4.90, *p* = .03, indicating that the effect of gender is moderated by the role. Finally, the main effect of role was significant, *F*(1, 99) = 24.30, *p* < .001, with the group of submissive speakers receiving lower ratings than the group of dominant speakers. 

**Discussion:** The findings are consistent with previous research on gender and power in communication. The results suggest that social norms and stereotypes about gender roles continue to influence perceptions and evaluations of speakers. The study also highlights the importance of considering the role of the speaker in understanding the impact of gender on communication. 

**Conclusion:** The study demonstrates the power of gender and role in shaping perceptions of effectiveness in communication. The findings have implications for understanding the dynamics of gender in professional and social interactions. Further research is needed to explore the mechanisms underlying these effects and to develop interventions to address gender biases in communication. 

SEX BIAS IN THE NAMING OF STIMULUS PERSONS

against a male whose marital status was unspecified and who might be presumed to be married. (Indeed, in some experiments comparing Miss and Mr., subjects might have reasonably inferred that the Mr. stimulus was married, because he was presented as possessing advanced degrees [e.g., Panek et al., 1976] or as being of an age at which most men are married [e.g., Pheterson et al., 1971]). Thus, studies in which the title Miss was used to operationalize sex are confounded, making sexism conclusions untenable. Moreover, because some studies used titles denoting marital status but other studies did not, it is impossible to know whether the differing results of such experiments stem from differences in sexism or from differing impressions of marital status.

Because marital status varies with age, titles of address denoting marital status may also convey impressions of age. Miss connotes youth. Experiments that contrasted a female titled Miss against a male titled Mr. may have created impressions of a young female but not of a young male. If so, age stereotypes further confound such studies.

When titles denoting marital status are paired with age-typed names, interactive effects on impression formation may result. The title Miss, when combined with names connoting old age, summons the image of a woman who has remained single long past the culturally approved age for marriage—the so-called “spinster” or “old maid” stereotype, with its various unfavorable connotations. All nine of the experiments under review that used the title Miss paired it with “older adult” forenames (Baruch, 1972; Chobot et al., 1974; Garland, 1977; Goldberg, 1968, 1974; M. B. Isaacs, 1981; Panek et al., 1976; Pheterson et al., 1971; Valasek et al., 1979).

There exists also some evidence that titles of address vary in affective value. Kleinke (1974) asked students in fourth, fifth, and sixth grades to evaluate various terms for males and females. Both boys and girls judged the title Miss more favorably than Mrs. Girls judged Mr. less positively than Mrs., whereas boys rated Mr. more positively than Miss. Thus far, there exists no evidence on whether such preferences are shared by older students, such as college students serving as subjects in sexism experiments, or on whether these preferences would affect ratings of stimulus persons so titled. Nevertheless, Kleinke’s findings raise the possibility that sexism that used titles of address are confounded by the titles’ differential affective value.

To avoid female titles denoting marital status, some researchers substituted Ms. for Miss or Mrs. (Bernard, Keehauer, Elsworth, & Naylor, 1981; Dobbins, 1986; Etaugh & Sanders, 1974; Gutke, Mirasch, & Cohen, 1983; Honig & Carterette, 1978; London & Stumpf, 1983; Remland, Jacobson, & Jones, 1983; Rush, Thomas, & Lord, 1977; Stumpf & London, 1981; Swenson et al., 1984). However, although Ms. does not denote marital status, it connotes distinctive stereotypical impressions not directly related to sex (L. R. Anderson, Finn, & Leider, 1981; Connor, Byrne, Mindell, Cohen, & Nixon, 1986; Dion, 1987; Dion & Schuller, 1990; Heilman, 1975). Heilman conducted two experiments that found that college-age men expected more enjoyment and more intellectual stimulation from nontechnical college courses taught by instructors titled Ms. than by instructors titled Mrs. or Miss. L. R. Anderson et al. (1981) found that high school and college students of both sexes were more willing to subordinate themselves to authoritarian leaders titled Ms. than to equally authoritarian leaders titled Miss, Mrs., or Mr. Connor et al. (1986) asked adult subjects to read a paragraph about an achieving young woman titled Miss, Mrs., or Ms. or presented without any title. Subjects of both sexes rated the Ms. stimulus person as significantly less honest. Dion (1987), in two experiments using semantic differential ratings, found that a woman preferring the Ms. title “gives the impression of being more achievement oriented, socially assertive and dynamic, but less interpersonally warm, relative to her counterparts with traditional titles of address” (p. 21). Finally, Dion and Schuller (1990) found that adults of both sexes expected young women who preferred the Ms. title to be more competent, masculine, and achievement oriented and less warm and likable than women who preferred the titles Mrs. or Miss. Given the Ms. stereotype suggested in these studies, research comparing males titled Mr. with females titled Ms. did not merely contrast impressions of males against impressions of females. Rather, such research contrasted impressions of males with impressions of unmarried females—unusual in their ambition, dominance, and competence in traditionally masculine pursuits.


Surnames, like forenames and titles of address, affect person perception in ways wholly unrelated to sex. Surnames vary in attractiveness (Arthaud et al., 1948; Colman, Sluckin, & Hargreaves, 1981) and may convey impressions of ethnicity, intelligence, physical attractiveness, and personality (Dulin & Duran, 1977; Hartmann, 1985; Razzan, 1950; Waters, 1989; see also Arthaud et al., 1948). Moreover, a German study found that undesirable surname impressions affected subjects’ impressions of full names significantly more than did forename impressions (Hartmann, 1985).

Surname attractiveness varies curvilinearly with surname familiarity: Subjects have preferred moderately familiar surnames over unfamiliar or extremely familiar surnames (Colman, Sluckin, & Hargreaves, 1981; see also Arthaud et al., 1948). Experiments that have contrasted males given moderately familiar surnames with females given unfamiliar or extremely familiar surnames did not merely contrast impressions of males and females. Rather, such experiments contrasted impressions of attractively surnamed males and unattractively surnamed females.

Surnames also convey impressions of ethnicity, personality, intelligence, and physical attractiveness. To the extent that subjects hold similar ethnic stereotypes, experiments in which male and female stimulus persons’ surnames suggested a single ethnic identity may have yielded results specific to that ethnic group (cf. Corenblum, 1977; Noel & Allen, 1976). Conversely, experiments in which male and female stimulus persons’ surnames suggested differing ethnic identities or differing impres-
sions of intelligence, personality, or physical attractiveness were confounded by surname stereotype.

Race, ethnicity, and social class. Name usage varies by race, ethnicity, and social class (Alford, 1988; Busse & Seraydarian, 1977; Dunkling, 1977, 1986; Eagleson & Clifford, 1945; Fayer, 1988; Puckett, 1975; Taylor, 1974; Willis et al., 1982; Zweigenhaft, 1977). American Blacks, for example, use Marcus, LaToya, Darryl, and Latasha more commonly and Frank, Katherine, Adam, and Megan less commonly than do American Whites (Dunkling, 1986; Willis et al., 1982). American Blacks also use the surname Williams, for example, more commonly and Anderson less commonly than do American Whites (Eagleson & Clifford, 1945; Puckett, 1975, pp. 304–305). Within both races, Albert is more common among the poor; and Richard, George, and Thomas are more common among higher income and professional people (Willis et al., 1982).

Because of these differences in name usage, names may convey impressions of race, ethnicity, or social class. Zweigenhaft (1977, Experiment 3) found that Anglo-Saxon surnames that were used as forenames were rated as unusually “upper-class” and aristocratic (see also Gladding & Farrar, 1982). In another study, forenames received widely varying scores on a common-noble bipolar scale (Darden & Robinson, 1976). In still another study, names common among different races were used as part of an effective experimental race manipulation: Howard & Pike, 1986.

Sexism researchers commonly use dependent variables that may be sensitive to racial, ethnic, and social class stereotypes. Research suggests, for example, that stimulus persons’ race, ethnicity, and social class affect impressions of their competence (Abramson, Goldberg, Greenberg, & Abramson, 1977), assessments of their performance (Kraiger & Ford, 1985), semantic differential ratings (Rosenwasser, Gonzales, & Adams, 1985), attraction scores (Lott & Lott, 1986; Razran, 1950), and attributions of their success and failure (Howard & Pike, 1986; Whitehead, Smith, & Eichhorn, 1982; Yarkin, Town, & Wallston, 1982).

Sexism experiments that used dependent measures sensitive to racial, ethnic, or social class stereotypes and that used names varying in racial, ethnic, or class connotations were thereby confounded. Because no good data exist on such connotations for the names used in the studies presently under review, it is impossible to determine how many studies were so confounded. Several experiments, however, clearly did suffer such confounding. Fidell (1970), for example, compared an aristocratic-sounding male name (Thornton La Salle) to a less genteel female name (Thelma La Salle).

Implications for Naming Stimulus Persons in Future Research

Future investigators in various fields will want to prevent the name-related confounding that has characterized past research on sexism and fear of success. The research reviewed in this article implies several methodological safeguards to minimize such confounding. I offer the following suggestions as guidelines for naming stimulus persons in future research: (a) Sex-typed titles of address should not be used. (b) Surnames should not be used unless necessary. Usually forenames will suffice. (c) If a surname is used, the male and female versions of a stimulus person should be given the same surname. (d) Young stimulus persons should not be given old-fashioned forenames, and elderly stimulus persons should not be given new-fashioned forenames. (e) Age-typed forenames should be assigned only to stimulus persons behaving consistently with the age stereotype connoted by the name. A stimulus person studying medicine, for example, should not be given an old-fashioned forename. (f) Male and female forenames should be matched in terms of attractiveness and connotation.

Unfortunately, a list of forenames matched on attractiveness and connotation exists nowhere in the literature. Researchers wishing to use such names would first need to conduct extensive name research or piece together data from past name research—both tasks requiring inordinate amounts of time and money. For the benefit of such investigators, I undertook the task of searching for male and female names rated similarly across various data sets.

Additionally, this selection procedure was used to find pairs of same-sex names matched on attractiveness and connotation. Such a list would be valuable to FOS investigators using multiple cues with subjects of only one sex (e.g., Hyland, Curtis, & Mason, 1985) and to researchers, in various fields, who contrast stimulus persons of one sex with others of the same sex (e.g., Frank et al., 1989; Ginossar & Trope, 1987; Harkins & Petty, 1987; McClure, Lalljee, Jaspars, & Abelson, 1989).

To generate a list of forenames matched on key variables, several statistical operations were performed on all forenames for which Dunkling (1986) reported frequency rankings. Names coded as “older adult,” “younger adult,” or “age unassociated” by the method described earlier were matched for attractiveness and connotation with other names within each age category (see Appendix B). Names were considered similar if they fell within a ±0.5 range on all the rating scales used herein—Dion (1985), Buchanan and Bruning (1971), Mehrabian (1988), and the 1990 ratings of attractiveness and intellectual-competence connotation—as well as within a ±10 range on Mehrabian’s (1990) male and female scales of intellectual-competence connotation. Names were then matched for racial stereotype. Given the racial composition of the United States and Canada, subjects will probably assume a stimulus person to be Caucasian unless provided with contrary information, such as a noticeably Asian, Hispanic, or Black name. None of the names being studied were noticeably Asian or Hispanic. To assess Black racial connotations, Dunkling’s (1986) data on racial differences in name fashions were used. A name was considered Black-typed if its rank in 1984 (the only year for which such data were available) was at least 10 positions higher among Blacks than among Whites. None of the names under consideration are Black-typed according to this criterion.

This procedure netted 10 pairs of male and female forenames matched on attractiveness, intellectual-competence connotation, age stereotype, and racial connotation (see Table 4). These pairings are, for “younger adult” stimulus persons, Brian versus Karen, Gary versus Lisa, and Gary versus Lynn; for “older adult” stimulus persons, Albert versus Ruth, Audrey versus Frank, Dorothy versus Harry, Jack versus Jane, Jack versus Marilyn, and Louise versus Sam; and for stimulus persons of any age, Patricia versus Tom. Across these 10 pairings, the...
mean scores were 4.43 for females and 4.42 for males on Buchanan and Bruning's (1971) attractiveness scale; 4.03 for females and 3.91 for males on Dion's (1985) attractiveness scale; 3.58 for females and 3.48 for males on my 1990 attractiveness ratings; 4.91 for females and 4.95 for males on Mehrabian's (1988) intellectual-competence ratings; 4.04 for females and 4.07 for males on the 1990 competence ratings; and 59.5 for males and 61.3 for females on Mehrabian's (1990) male scale and female scale of intellectual competence. These names are recommended for use in future studies in which stimulus persons of one sex are compared with others of the same sex.

Although these name pairings should prove useful to researchers in various fields, several limitations should be noted. First, such name pairings, matched on ratings by undergraduates, may be mismatched when used with younger or older subjects, because forename attractiveness varies by age of rater (Busse & Helfrich, 1975; see also Finch et al., 1944). Second, such names may differ in attractiveness or connotation when used with subjects in countries other than the United States and Canada. Third, because no attempt was made to match names on variables other than attractiveness, age, intellectual competence, and race, the possibility remains that names in recommended pairings differ in other important respects (e.g., social class connotation). Fourth, several recommended names (e.g., Dorothy) are uncommon in minority groups and hence may be inappropriate for use with Asian, Hispanic, or Black stimulus persons. Fifth, because most names in the recommended pairings differ in length, initial sound, and number of syllables, such pairings may be inappropriate for experiments sensitive to such mismatched variables (e.g., reaction time experiments). Sixth, the name Lynn is sexually ambiguous and may be inappropriate unless used with other indicators of sex, such as sex-typed pronouns. Finally, because of historical variation in forename attractiveness and connotation, the name pairs that survived this statistical selection will surely become mismatched at some future time.

As noted earlier, the names of stimulus persons have been commonly used as an experimental manipulation in many studies, often with the assumption that they are...
fields other than sexism and FOS. Future researchers in such fields may benefit from the use of matched names such as those presented in Tables 4 and 5. In addition, it seems prudent that past research in such fields be reassessed for possible name confounding.

As an example, consider the field of intergroup relations. Ju-Young Lee, at the University of Southern California, is presently investigating the possibility that the minimal group paradigm, a standard experimental procedure for studying intergroup discrimination, has been confounded by the use of names. Following a classic experiment by Tajfel, Billig, Bundy, and Flament (1971), dozens of researchers have operationally defined group identity by giving subjects false feedback on their performance on an aesthetic-preference task: Half the subjects are told that they consistently prefer paintings by Wassily Kandinsky over those by Paul Klee, whereas the other subjects are told that they consistently show the opposite preference. To make group membership conspicuous, experimenters highlight the Kandinsky-Klee distinction by writing the names prominently on a chalkboard, referring to subjects as Kandinsky and Klees, and blatantly juxtaposing the two surnames on each page of a reward-allocation booklet that typically constitutes the experiments' dependent measure. This procedure is called the minimal group paradigm because it is thought to isolate categorization per se from extraneous variables such as intergroup competition and prior history. However, because the names Wassily Kandinsky and Paul Klee differ greatly in prominence on a chalkboard, referring to subjects as Kandinsky and Klees, and blatantly juxtaposing the two surnames on each page of a reward-allocation booklet that typically constitutes the experiments' dependent measure. This procedure is called the minimal group paradigm because it is thought to isolate categorization per se from extraneous variables such as intergroup competition and prior history. However, because the names Wassily Kandinsky and Paul Klee differ greatly in

Table 5
Matched Names for Same-Sex Stimulus Persons

<table>
<thead>
<tr>
<th>Name</th>
<th>Attractiveness ratings</th>
<th>Competence ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger adult</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Karen vs.</td>
<td>5.296</td>
<td>4.45</td>
</tr>
<tr>
<td>Lisa</td>
<td>5.060</td>
<td>4.40</td>
</tr>
<tr>
<td>Sandra vs.</td>
<td>3.378</td>
<td>4.35</td>
</tr>
<tr>
<td>Theresa</td>
<td>3.653</td>
<td>4.05</td>
</tr>
<tr>
<td>Older adult</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barbara vs.</td>
<td>4.978</td>
<td>4.45</td>
</tr>
<tr>
<td>Carol</td>
<td>4.997</td>
<td>4.20</td>
</tr>
<tr>
<td>Beverly vs.</td>
<td>4.372</td>
<td>3.95</td>
</tr>
<tr>
<td>Joan</td>
<td>4.521</td>
<td>3.60</td>
</tr>
<tr>
<td>Frank vs.</td>
<td>4.064</td>
<td>3.90</td>
</tr>
<tr>
<td>Walter</td>
<td>3.732</td>
<td>3.40</td>
</tr>
<tr>
<td>Jane vs.</td>
<td>4.510</td>
<td>3.70</td>
</tr>
<tr>
<td>Marilyn</td>
<td>4.925</td>
<td>3.80</td>
</tr>
<tr>
<td>Age unassociated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>David vs.</td>
<td>5.290</td>
<td>5.50</td>
</tr>
<tr>
<td>Thomas</td>
<td>6.030</td>
<td>6.05</td>
</tr>
<tr>
<td>John vs.</td>
<td>5.334</td>
<td>5.00</td>
</tr>
<tr>
<td>Robert</td>
<td>5.223</td>
<td>4.95</td>
</tr>
<tr>
<td>Joseph vs.</td>
<td>4.720</td>
<td>5.00</td>
</tr>
<tr>
<td>Richard</td>
<td>4.767</td>
<td>4.90</td>
</tr>
</tbody>
</table>

Note. 1970 ratings are from Buchanan and Bruning (1971), 1978 ratings are from Dion (1985), 1990 ratings are from the present study, 1987 ratings are from Mehrabian (1988), and 1989 ratings are from Mehrabian (1990).
(Taylor & Fiske, 1978), such mismatched forename pairings may affect how subjects allocate causality between agent and patient. (Forename salience discrepancy is also a problem for research on anaphoric reference and discourse focus, e.g., Crawley & Stevenson, 1990; Crawley, Stevenson, & Kleinman, 1990; Sanford, Moar, & Garrod, 1988). Past research on implicit causality often used random assignment of forenames to semantic or grammatical roles, and so it is extremely unlikely that the consistent and robust findings on implicit causality were substantially affected by names. Nevertheless, to reduce “noise” and prevent possible experimenter expectancy effects in future research, investigators studying implicit causality and other psycholinguistic phenomena may wish to use carefully matched names such as those presented in Tables 4 and 5.

A more general area for future investigation is that of the experimenter expectancy effect. The results reported herein suggest that researchers’ hopes or expectations may influence the naming of stimulus persons in a manner that helps confirm the researchers’ hypotheses. Has this been true of other fields as well? In experiments on physical-attractiveness stereotypes, for example, did researchers favorably name physically attractive stimulus persons? Experimenter effects in naming stimulus persons in various fields is an area ripe for future research.

Foremost among the numerous topics for future research, of course, are sex discrimination, sex stereotypes, and fear of success. Although the research reported in this article raises serious questions about the results of past research, it provides few conclusive answers. In particular, the present research does not demonstrate a complete absence of true sexism or FOS effects and does not attempt to explain or describe these phenomena. In light of the present results, many experiments of the past may need to be redone in a more carefully controlled manner. Hence, this article will hopefully stimulate future investigators to correct unseen flaws of the past and so will make possible a better understanding of sexism and fear of success.

References


National Faculty Directory (1986). Detroit: Gale Research.


Shapiro, J. P. (1979). "Fear of success" imagery as a reaction to sex-role
SEX BIAS IN THE NAMING OF STIMULUS PERSONS


(Appendix follows on next page)
Bibliography of Studies Under Review


Miller, J. R. (1980). Relationship of fear of success to perceived paren-
tal attitudes toward success and autonomy in men and women. Psychological Reports, 47, 79–86.


SEX BIAS IN THE NAMING OF STIMULUS PERSONS

Role behavior as defined by the task, the mode, and the actor. Journal of Personality and Social Psychology, 32, 381–390.


Appendix B

Age-Categorized Forenames

Older adult names: Al, Albert, Alfred, Alice, Anne, Arthur, Audrey, Barbara, Betty, Beverly, Carl, Carol, Charles, Don, Donald, Dorothy, Edith, Edward, Elaine, Eleanor, Elizabeth, Eugene, Frank, Fred, Frederick, Gail, George, Gerald, Gloria, Harold, Harry, Helen, Jack, Janet, Janice, Jean, Joan, Joanne, Judy, Lawrence, Louis, Louise, Margaret, Marilyn, Nancy, Norma, Norm, Phyllis, Ralph, Ruth, Sally, Sam, Samuel, Sarah, Shirley, Ted, Walter

Younger adult names: Brenda, Brian, Chris (male), Christine, Christopher, Dan, Debbie, Deborah, Dennis, Donna, Eric, Gary, Greg, Jennifer, Karen, Kevin, Laura, Linda, Lisa, Lynn (female), Mark, Matthew, Michael, Michelle, Mike, Pam, Pamela, Pat (male), Patrick, Ronald, Sandra, Steve, Steven, Sue, Susan, Theresa

Age-unassociated names: Alan, Andrew, Andy, Bob, Dave, David, Diane, James, Jim, Joe, John, Joseph, Katherine, Ken, Kenneth, Larry, Mary, Pat (female), Patricia, Paul, Peter, Phil, Richard, Rick, Robert, Suzanne, Sharon, Thomas, Tom, William

Received June 4, 1991
Revision received January 31, 1992
Accepted January 10, 1992