# Male–Female Differences: A Computer Simulation

Richard F. Martell
*Department of Social–Organizational Psychology, Columbia University*

David M. Lane
*Department of Psychology, Rice University*

Cynthia Emrich
*Department of Management, University of Otago*

"The Science and Politics of Comparing Women and Men" (Eagly, March 1995) raised an important question: What constitutes a practically significant sex effect? The practice of relying on proportion of variance measures (e.g., $r^2$, $\omega^2$) to address this question has been judged inappropriate, for such measures are not intuitively accessible and can mislead researchers into ignoring the practical significance of small effects. In response, more easily understood metrics such as the binomial effects size display (Rosenthal & Rubin, 1982) and the common language effect (McGraw & Wong, 1992) have been introduced. However, Eagly concluded that even these approaches are not entirely sufficient:

> The evaluation of the . . . importance of sex-related differences should not end with the translation of them into metrics that are easily understood. In practical terms, the importance of a difference depends on the consequences of the behavior in natural settings. (p. 152)

We agree with Eagly and recommend the use of computer simulations as a tool for assessing the impact of sex differences.

Consider male–female differences in work performance ratings and how a computer simulation might help to resolve some of the interpretive difficulties raised by Eagly (1995). As discussed in a recent American Psychological Association Amicus Curie Brief (American Psychological Association, 1991) filed in the *Price Waterhouse v Hopkins* (1989) Supreme Court case of alleged workplace discrimination, women's work is often evaluated less favorably than men's. However, several rebuttals argued against the practical importance of the research findings cited in the brief noting, among other things, that sex bias effects are quite small, accounting for only approximately 1% to 5% of the variance in work performance ratings (e.g., Barrett & Morris, 1993). Overlooked in this debate, and what a computer simulation would force researchers to consider, are the "real world" parameters likely to influence the effects of male–female differences in performance ratings. Two organizational characteristics are especially relevant. First is the pyramid structure of most organizations, in which there are increasingly fewer positions as one attempts to climb to the top. Second, because most organizations rely on a tournament model, in which early career success is a necessary precondition for subsequent promotion, initial performance ratings strongly influence the likelihood of whether an individual reaches a top management position (Rosenbaum, 1979). To assess the extent to which these two facts of organizational life might limit the upward mobility of women, even when male–female differences in performance ratings are quite small, the following computer simulation was conducted..

The simulation, which depicted an organization comprised of eight levels—with 500 incumbents at the bottom and only 10 at the very top—began with an equal number of men and women awaiting promotion into one of the eight levels. Each person was assigned a performance evaluation score. The scores of men and women were distributed so as to be normal and identical ($\mu = 50$, $s = 10$). Incumbents with the highest scores became eligible for promotion once a position was available. The simulation began by removing 15% of the incumbents. These positions were filled from within the organization, with eligible individuals (those with the highest scores) being promoted into the position. The simulation continued to apply the 15% attrition rule until the organization was staffed entirely with "new" employees. That is, all individuals within the organization at the start of the simulation had been replaced with individuals from the initial pool. For each simulation, 20 computer runs were conducted to ensure an adequate degree of reliability.

To assess the impact of male–female differences, "bias points" were added to the performance score of each man. In Simulation 1, 4.58 bias points, accounting for 5% of the variance in scores, were added; in Simulation 2, 2.01 bias points, accounting for 1% of the variance, were added. Proportion of variance was calculated by converting the standardized mean differences in performance scores between men and women into $r^2$. Given the standard deviation in the scores, the distributions of male and female scores were overlapping, even after the introduction of bias points.

Detailed results are shown in Tables 1 and 2; the main findings are highlighted in Figure 1. It can be seen that a very high percentage of upper-level positions were filled by men, whereas women tended to cluster at the lower levels of the organization. With 5% of the variance in ratings attributed to sex, only 29% of the incumbents at the very top level of the organization were women, whereas 58% of the very bottom level positions were filled by women.

**Table 1**
*Results of Computer Simulation 1: Effect Size 5% of the Variance*

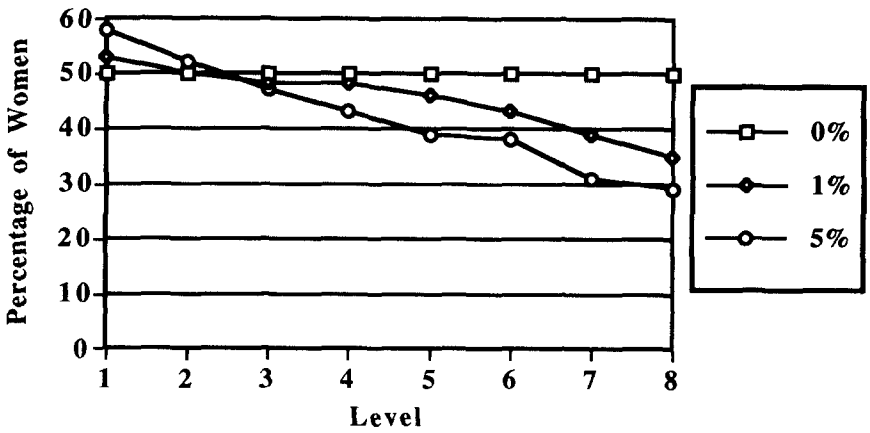
| Level | Incumbent's mean score | Number of positions | Percentage of women |
|---|---|---|---|
| 8 | 76.95 | 10 | 29 |
| 7 | 68.80 | 40 | 31 |
| 6 | 63.79 | 75 | 38 |
| 5 | 60.80 | 100 | 39 |
| 4 | 57.85 | 150 | 43 |
| 3 | 55.06 | 200 | 47 |
| 2 | 50.93 | 350 | 52 |
| 1 | 45.00 | 500 | 58 |

**Table 2**
*Results of Computer Simulation 2: Effect Size 1% of the Variance*

| Level | Incumbent's mean score | Number of positions | Percentage of women |
|---|---|---|---|
| 8 | 74.08 | 10 | 35 |
| 7 | 67.14 | 40 | 39 |
| 6 | 62.16 | 75 | 43 |
| 5 | 59.15 | 100 | 46 |
| 4 | 56.03 | 150 | 48 |
| 3 | 53.64 | 200 | 48 |
| 2 | 49.77 | 350 | 50 |
| 1 | 44.02 | 500 | 53 |

**Figure 1**

*Percentage of Women at Each Position Level, With 0%, 1%, and 5% of the Effect Size Variance Attributed to Sex*




Even more dramatic is the finding that when sex differences explained but 1% of the variance, an estimate that might be dismissed as trivial, only 35% of the highest-level positions were filled by women. Thus, relatively small sex bias effects in performance ratings led to substantially lower promotion rates for women, resulting in proportionately fewer women than men at the top levels of the organization.

These results confirm Eagly's (1995) point that the effects of male–female differences are best determined not by the magnitude of the effect but its consequences in natural settings. In this case, by taking into account the relative scarcity of very senior-level positions in organizations as well as the weight accorded early career performance ratings, a little bias hurt women a lot. We suggest, then, a salutary approach to assessing practical significance is not to reject traditional effect size measures but to translate them into estimates of real world impact. Computer simulations are ideal for this.

#### REFERENCES


American Psychological Association. (1991). In the Supreme Court of the United States: *Price Waterhouse v. Ann B. Hopkins: Amicus* curiae brief for The American Psychological Association. *American Psychologist, 46,* 1061–1070.

Barrett, G. V., & Morris, S. B. (1993). The American Psychological Association's amicus curiae brief in *Price Waterhouse v. Hopkins:* The values of science versus the values of the law. *Law and Human Behavior, 17,* 201–215.

Eagly, A. (1995). The science and politics of comparing women and men. *American Psychologist, 50,* 145–158.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111,* 361–365.

Price Waterhouse v. Hopkins, 109 Supreme Court 1775 (1989).

Rosenbaum, J. E. (1979). Tournament mobility: Career patterns in a corporation. *Administrative Science Quarterly, 24,* 220–241.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166–169.


## Differences Between Women and Men: Their Magnitude, Practical Importance, and Political Meaning


Alice H. Eagly
*Department of Psychology, Northwestern University*


The issue of the magnitude of differences between women and men continues to elicit diverse opinions, as the comments by Lott (1996) and Martell, Lane, and Emrich (1996) illustrate. Lott's (1996) view that "there is simply no getting around the fact that the differences so painstakingly identified are small indeed!" (p. 156) can be juxtaposed against Martell, Lane, and Emrich's insightful demonstration of the practical importance of differences that Lott, along with most psychologists, would surely label as extremely small. Martell, Lane, and Emrich thus showed that a difference in performance evaluations favoring men but accounting for only 1% of the variability in scores can produce an organizational structure in which only 35% of the highest level positions are filled by women. Also dramatic was Abelson's (1985) earlier demonstration that baseball players' batting skills have a substantial impact on their teams' success, despite the fact that the percentage of variance in any single batting performance that is explained by batting skill is approximately 0.3%. These simple illustrations of the practical importance of seemingly small effects thus underscore my point that psychologists are generally misled when they address magnitude issues in terms of percentage of variance.

In addition to translating research findings to more intuitively understandable metrics (e.g., the binomial effect-size display and the common-language effect size), psychologists should follow the example of Martell, Lane, and Emrich by examining consequences of group differences in natural settings. Lott (1996) and I are in agreement about the importance of explaining differences theoretically. As I wrote, "Empirical findings take on meaning and importance within theories that explain the antecedents of the findings" (Eagly, 1995, p. 148). Puzzlingly, Lott advocates the development of theories of difference but simultaneously opposes the identification of differences. In science, the identification of a phenomenon precedes explanation of it. As I argued (Eagly, 1995, p. 148), the 1970s consensus among research psychologists that sex differences are null or very small discouraged theoretical attention to differences because weak, unreliable effects seemed undeserving of theoretical explanation. Therefore, as a first phase of scientific activity, cataloging differences and similarities is extremely useful.

Research findings can be cataloged a manner that is more or less interesting, depending on whether reviewers attend to the context of the differences and their theoretical meaning. As I noted (Eagly, 1995, pp. 152–153), quantitative syntheses offer three excellent methods of investigating whether findings are context dependent: (a) the calculation of a statistical index that expresses the degree of homogeneity versus heterogeneity of findings in a sample of studies, (b) the identification of outliers among a set of findings, and (c) the identification of moderator variables that account for variability in findings. Using these methods, many contemporary meta-analysts scrutinize variability among effect sizes and the contextual variables that produce this variability. They also test theories using the data produced by their syntheses of research (see Miller & Pollock, 1994).

As Archer (1996) cautioned, the context of research findings cannot be investigated in meta-analyses unless the relevant contextual feature has varied across the avail-