How does gender play a role in the earnings gap? an update

Although personal choices, occupational crowding,
and discrimination, contribute to the gender gap,
the higher share of women in an occupation
is still the largest contributor

Stephanie Boraas
and
William M. Rodgers III

Stephanie Boraas is
an economist in the
Office of
Employment and
Unemployment
Statistics, Bureau of
Labor Statistics.
William M. Rodgers
III is the Frances L.
and Edwin L.
Cummings
associate professor
of economics and
director of the
Center for the Study
of Equality at the
College of William
and Mary. The
views expressed in
this article are solely
the authors' and
do not represent
the policy positions
of the Bureau of
Labor Statistics or
the U. S.
Department of
Labor. E-mail:
boraas_s@bls.gov;
wmrodg@wm.edu

According to the Bureau of Labor Statistics, women earned approximately 77 percent as much as men did in 1999.[1] Although the existence of the gender pay gap is well documented, the factors that contribute to it are still debated. One such factor is the difference in the proportion of jobs held by women and men. However, understanding how occupational differences contribute to the gender pay gap is made more complicated by the fact that both men and women in predominately female occupations earn less than men and women in male dominated occupations.

An important question that has not been resolved is, why do predominately female occupations tend to pay less? A number of possibilities including worker characteristics, job characteristics, occupational crowding, devaluing by society of women, and discrimination have been posited.[2]

This article sheds some light on reasons for the gender earnings gap, focusing on the role that the share of women in an occupation plays. We utilize the methodology employed by George Johnson and Gary Solon to identify the sources of the relationship between wages and the share of women in an occupation.[3] Johnson and Solon used Current Population Survey (CPS) data to estimate the relationship between wages and the concentration of females within occupations. They found that the relationship was negative,

even after controlling for worker and job characteristics. Industry was found to have the largest effect on the relationship, primarily because predominately male industries, such as construction and manufacturing, pay higher wages.[4]

The data used in this article are from the 1989, 1992, and 1999 Outgoing Rotation Group Files of the CPS. The CPS has the advantage of including detailed occupational data on a large national sample of more than 50,000 households in addition to the more commonly found earnings, employment, human capital, and demographic information.[5]

One of the primary drawbacks of using the CPS for estimating the sources of the gender gap is that it does not include a measure for actual work experience. Because higher earnings are associated with more experience, and women tend to have less experience than men due to breaks in their labor force participation for childrearing and other reasons, this omission results in inaccurate estimates of the contribution of each characteristic to the wage gap.[6]

To approximate actual experience, researchers construct potential experience.[7] This method likely overestimates the experience of women, as they are more likely than men to take time out of the labor force.[8] However, it is well known that delayed marriages and lower fertility rates have contributed to a rise in women's labor force participation,

particularly among younger cohorts of women. Because of this pattern, the use of potential experience probably provides a better estimate of actual experience than it did in the past.

Prior to 1992, the Bureau of Labor Statistics asked CPS respondents to provide the number of years of school completed. However, in 1992, BLS changed its education question by asking respondents to provide the highest degree attained (for example, years of schooling for a bachelor's degree, and so forth), making it difficult to construct potential experience.[9] Instead of using a method for imputing years of schooling from an independent CPS file, which contains respondents' answers to both education questions, we chose to use age and age-squared to proxy for experience.

Although the negative relationship between wages and an occupation's percentage of women has diminished since 1989, occupations that are predominately made up of females are still found to have lower average wages for both men and women. In 1999, men employed in predominately female occupations earned 11 percent to 13 percent less than men did in predominately male occupations. Much of this negative impact on wages is due to the industry distribution, but the positive effect of education and experience offsets this effect somewhat. For women in 1999, working in a female dominated occupation lowers wages by approximately 26 percent. Accounting for education and age of women reduces this effect to 21 percent.

Our nonlinear specifications show that in 1999, women employed in occupations that are 90 percent female earn 29 percent less than women do in predominately male occupations. Accounting for characteristics such as education and age of women reduces the disadvantage to 19 percent. Women employed in occupations that are 50 percent female earn 21 percent less than comparable women in predominately male occupations. The disadvantage falls to 13 percent when personal characteristics (such as education, age, and region of residence) are added to the model. The disadvantage falls to 14 percent for women in occupations that are 30 percent female. Personal characteristics explain one-half of the lower earnings.

The human capital and occupational crowding theories have been put forth as possible explanations for the relationship between wages and the gender concentration of occupations, as have job attributes, preferences, and discrimination.[10]

How do these theories attempt to explain the relationship? Human capital theory suggests that women tend to select occupations that require small investments in human capital and those in which the necessary skills do not atrophy with disuse because they better suit the daily schedules or long-term labor force participation intentions of women. The occupational-crowding hypothesis suggests that, from childhood, individuals are socialized to view some occupations as "women's work" and others as "men's work" and that they must choose one appropriate to their sex. This socialization influences human capital development and results in the crowding of women into the relatively small number of appropriate occupations. The ready supply of workers to fill these "female occupations" works to depress the wages of individuals employed in these occupations.

Data

This analysis is based on data from the Outgoing Rotation Group Files of the 1989, 1992, and 1999 Current Population Survey (CPS). This is a slightly different data set than the May 1978 CPS data used by Johnson and Solon, but it is similar enough to allow for a close replication and update of their results. The sample includes male and female nonagricultural wage and salary workers, age 16 and older, who responded to supplementary questions about "usual weekly hours" and "usual weekly earnings."[11]

Other explanatory variables included were educational attainment (measured as years of schooling through 1991 and as the highest grade attained thereafter), potential experience in 1989, age in 1992 and 1999, dummy variables for region of residence, size of the metropolitan area, Black or other minority status, voluntary and involuntary part- time work status, marital status, membership in or coverage by a union, public-sector employment, major industries (with private household service being the omitted category), and the percent of an individual's occupation made up of females. The percent of each three-digit occupation that is female was created from the Outgoing Rotation Group files of the Current Population Survey.

Multicollinearity between the individual level industry data and the aggregate occupation data is not a problem for our analysis. The percent of female variable is based on 500 occupation categories. As a formal check for multicollinearity, we constructed pair-wise correlations between the percent of female variable and the 22-industry dummy variables. In 1989, the absolute values of the correlations range from 0.009 to 0.18. In 1992, they range from 0.005 to 0.17. In 1999, they range from 0.01 to 0.19. The values are far from the level at which we should be concerned about multicollinearity.

Sources of the gender wage gap

Decomposing wage differentials has become a popular and useful way to identify the sources of the wage gap and their contributions. The Oaxaca-Blinder decomposition separates the portion of the gap resulting from differing characteristics of men and women from the portion that is not explained by these personal characteristics. The unexplained portion of

the gap may result from discrimination. Men receive higher returns to investments in their human capital and other characteristics than women, or men with the same characteristics as women are simply paid more.

A basic decomposition that uses the mean characteristics of women and the coefficients of men as weights can be written as:

$$Gap = (\boldsymbol{b}_M - \boldsymbol{b}_F)\overline{X}_F + \boldsymbol{b}_M(\overline{X}_M - \overline{X}_F),$$

where *Gap* denotes the difference in wages of men and women, the first term on the right hand side of the equation measures the unexplained differential, and the second term measures the gap due to differences in characteristics of men and women. The unexplained differential has several interpretations. It may be due to differences between the characteristics of men and women that might not have been included in the model. For example, some argue that the quality of education has not been included and can play a large role in the ability and productivity of an employee. The remainder of the unexplained gap also may be due to discrimination. We also construct the decomposition with the mean characteristics of men and the coefficients of women as the weights. Doing this provides a lower and upper bound on the contribution of each characteristic.

Table 1 presents the wage gap decompositions for 1989, 1992, and 1999. The gender gap has narrowed from 30 percent in 1989, to 26 percent in 1992, and 24 percent in 1999.[12] In 1989, using the male coefficients as weights, we find that industry and the percentage of women in an occupation account for 19.7 percent ((0.0600/0.304)*100) and 21.0 percent ((0.064/0.304)*100) of the wage gap, with percentage of women being slightly larger. Using the female coefficients as weights increases the role of percentage of women to 30.6 percent and reduces industry's role to 12.5 percent. In 1992, the role of industry ranges from 10.1 to 20.0 percent and the contribution of percentage of women ranges from 20.0 to 32.0 percent. In 1999, the role of percentage of women in an occupation ranges from 18.6 to 31.6 percent, with industry's role ranging from 7.6 to 14.3 percent. In all the years of this analysis, education and age work to shrink the wage gap, but are overshadowed by the percentage of women in an occupation and industry. The explained effect meets or exceeds the unexplained effect in every year except 1999.

To find the source of the relationship between wages and the concentration of women within an occupation and its decline, we use the decomposition framework from Johnson and Solon.[13] That study found that the largest contribution to the relationship was from the link between wages and industry of employment and the correlation between an individual's industry of employment and the occupation's concentration of women. To decompose the effect of the various control variables on the relationship between the concentration of women within an occupation and the wages of individuals in that occupation, Johnson and Solon found the difference between simple and multiple regression estimates of the coefficient on the percentage of women in each occupation. This can be expressed as:

$$\hat{\boldsymbol{g}} - \tilde{\boldsymbol{g}} = -\sum_{j=1}^{k} \hat{\boldsymbol{b}}\,\hat{b}_{jF},$$

where $\hat{\boldsymbol{g}}$ is the wage-concentration of women within occupations relationship in the full specification, $\tilde{\boldsymbol{g}}$ is the wage-concentration of women within occupations relationship in the bivariate specification, the $\boldsymbol{b}_j$'s are the estimated coefficients on the k control variables, and the b_{jF}'s are the coefficients on the percent female from the regression of each of the control variables on the percent female in an occupation. This approach identifies the factors that explain why the earnings of men and women are lower in occupations that have larger shares of women.

Table 2 shows the coefficients on percent female from both the bivariate and full regressions.[14] For both men and women in the bivariate specification, earnings in 1989, 1992, and 1999 were found to be lower in occupations with more women, with the relationship between low earnings and "female occupations" being stronger amongst women. After taking account of worker characteristics, the coefficient for men was reduced by about 30 percent in 1989 and 1992, and by between 26 percent and 28 percent for women (the difference divided by the bivariate). In 1999, taking account of worker characteristics leads to no decline in the men's coefficient, largely due to the weakening in the simple bivariate relationship during the 1990s. The decline in the coefficient for women in 1999 is smaller than earlier years.

To illustrate the impact of accounting for worker characteristics on the relationship between log wages and percent of the occupation that is female, chart 1 (page 34) illustrates the relationship at various points of the percent of occupation distribution. These figures were created by using predicted wages that were standardized to the earnings of a woman in an all-male occupation. The top panel shows that in 1989, the relationship for women was slightly nonlinear in the bivariate regression and less so in the full regression. As the concentration of females within an occupation increases, the negative impact on wages grows more slowly. By 1999, the bivariate regression is less nonlinear and the adverse impact ranges from 10 percentage points to 12 percentage points less than in 1989, as shown in the bottom panel. The negative impact of the concentration of women within occupations fell at all points of the percent female distribution, but the declines are largest in the upper one-half of the distribution. As the concentration of women increases

Table 1. Decomposition of the gender wage gap, 1989, 1992, and 1999

Variable	Weighting schemes					
	1989		1992		1999	
	Male coefficient	Female coefficient	Male coefficient	Female coefficient	Male coefficient	Female coefficient
Gender wage gap	0.304	0.304	0.256	0.256	0.237	0.237
Explained effect (in percent)	58	57	53	55	39	42
Percent female	.064	.093	.051	.082	.044	.075
Schooling and experience	.011	.009	.006	.005	−.001	−.001
Region	0005	.0005	.0003	.0002	.0002	.0002
SMSA size	.001	.002	.001	.001	.0001	.0001
Minority status	.003	.001	.004	.002	.004	.001
Part-time status	.019	.019	.012	.013	.000	.004
Marital status	.003	.000	.004	.001	.004	.002
Union membership	011	.012	.008	.008	.005	.006
Government employment	001	−.001	.001	.0003	.003	.001
Industry	.060	.038	.050	.026	.034	.018
Unexplained	.129	.130	.119	.116	.147	.133

Table 2. Coefficients on the percent female in an occupation, for men and women from the bivariate and full regressions, 1989, 1992, and 1999

Year	Men			Women		
	Bivariate	Full	Difference	Bivariate	Full	Difference
1989	−0.231	−0.155	−0.076	−0.315	−0.227	−0.088
1992	−.184	−.129	−.055	−.284	−.208	−.076
1999	−.125	−.123	−.002	−.259	−.209	−.050

NOTE: The only control variable in the bivariate specification is the percent of an occupation that is female. The full specification contains all of the control variables (educational attainment, potential experience in 1989 and age in 1992 and 1999, dummy variables for region of residence, size of metropolitan area, black or other minority status, voluntary or involuntary part-time work status, marital status, membership in or coverage by a union contract, public sector employment, major industries (with private household service being the omitted category), and the percent of an individual's occupation that is made up of women.

beyond this point, the impact on wages becomes less negative.

Table 3 shows the results of the decomposition of the relationship between wages and percentage of women. Although diminishing, industry of employment always played the largest role for men (except in 1992 when marital status was stronger). Multicollinearity is not a problem. The coefficients on the percent female from the regression of each industry dummy variable on the percent female in an occupation, the $b_{jF}'s$ range from -0.182 to 0.249. The R^2's have a maximum of 0.045.

Years of schooling and experience help to offset industry's negative impact. Since 1989, the share of men, in female dominated jobs, who have more education and are older than the men in less female-dominated jobs has increased. For all women, industry of employment plays very little role in explaining the relationship between percentage of women and wages. Years of schooling and experience are the other key factors that explain the relationship. However, in 1999, schooling and age are the only factors that have a significant ability to explain the relationship between percentage of women and wages. Industry of employment explains slightly more than 10 percent of the relationship.

Table 3. Decomposition on the influence of control variables on the estimation of the relationship between wages and percentage of female (y), 1989, 1992, and 1999

Variable	Men		
	1989	1992	1999
Total effect	−0.076	−0.055	−0.002
Schooling and age	.140	.116	.153
Region	.030	.028	.005
Minority status	−.006	−.007	−.007
Part-time status	−.017	−.009	−.001
Marital status	−.042	−.348	−.028
Union status	−.020	−.016	−.014
Government employment	−.005	−.004	−.010
Industry	−.156	−.129	−.101
Variable	**Women**		
Total effect	−.088	−.076	−.050
Schooling and age	−.045	−.057	−.057
Region	−.005	.000	.001
Minority status	.000	−.001	−.001
Part-time status	−.022	−.015	−.007
Marital status	.002	.001	.002
Union status	−.003	−.001	.003
Government employment	.001	.000	−.002
Industry	−.016	−.003	.010

NOTE: Potential experience is used in 1989 and age is used in 1992 and 1999.

Chart 1. Relationship between wages and the percent of an occupation that is predominated by women, 1989,1992, and 1999
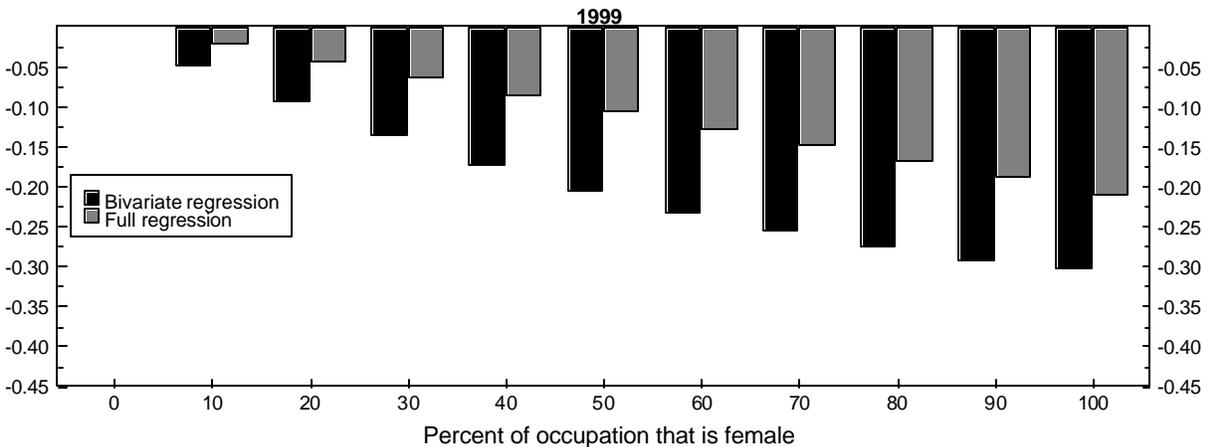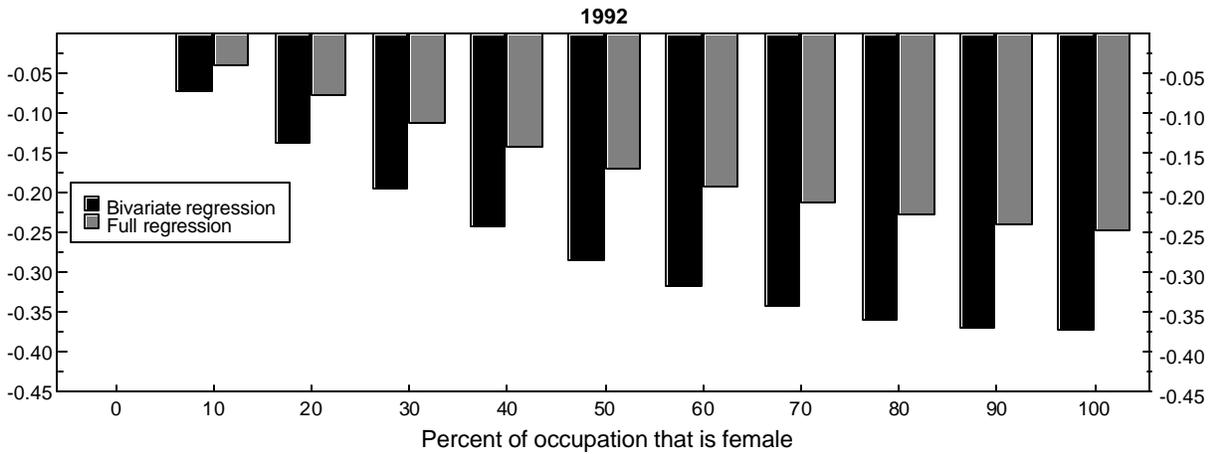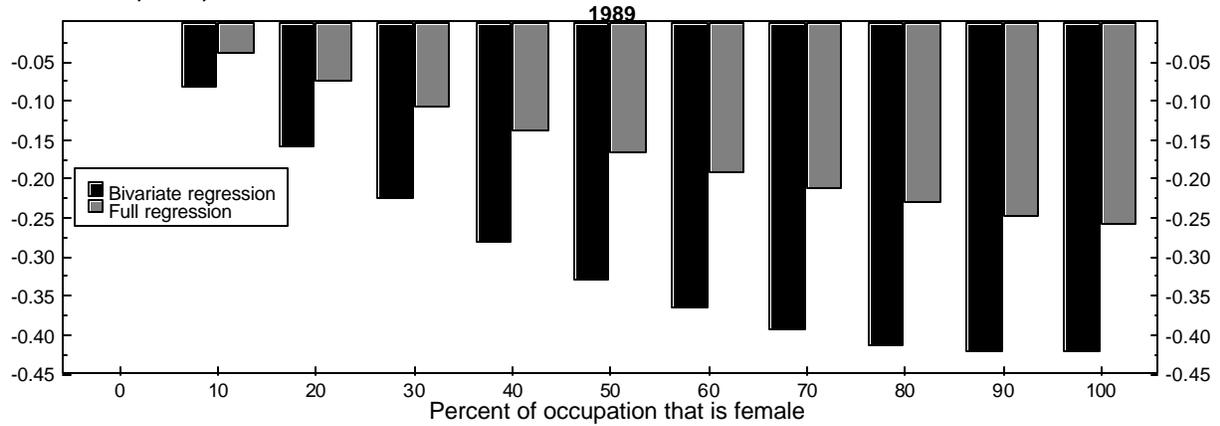
NOTE: The only control variable in the bivariate specification is the percent of an occupation that is female. The full specification contains all of the control variables (educational attainment, potential experience in 1989 and age in 1992 and 1999, dummy variables for region of residence, size of metropolitan area, black or other minority status, voluntary or involuntary part-time work status, marital status, membership in or coverage by a union contract, public sector employment, major industries (with private household service being the omitted category), and the percent of an individual's occupation that is made up of females.

SOURCE: Authors' tabulations from the Outgoing Rotation Group Files of the Current Population Survey.

Summary and conclusions

This article revisits the work of Johnson and Solon and others that estimate the portion of the gender gap that can be explained by the share of women in an occupation. Using microdata from the Outgoing Rotation Group files of the Current Population Survey, we first show that the share of women in an occupation is still one of the largest contributors to the gender wage gap. This occurs because women have a higher likelihood of working in female dominated jobs, which typically have below average wages. In fact, in 1999, our most recent data, the share of women is the largest contributor to the gender pay gap.

Without controlling for worker characteristics, we then estimate the size of the negative effect that employment in predominately female occupations has on the wages of men and women. We find that the relationship has weakened for both men and women since 1989. In 1999, a woman working in a predominately female occupation earned 25.9 percent less than a woman working in a predominately male occupation . The comparable figure for a man working in a predominately female occupation is 12.5 percent less than a man working in a predominately male occupation. One interpretation of these relationships is that they are the adverse consequences of occupational "crowding." Limited occupational choice creates crowding in particular occupations, putting downward pressure on the wages of men and women in these occupations. The limited occupational choices could be due to discrimination. However, others might argue that the crowding reflects the personal choices of some men and many women who want or need a job that fits their family obligations. The wage and percent female relationship reflects the presence of a compensating differential.

If true, then the estimates need to be adjusted for not only labor supply factors that proxy for personal preferences, but also for labor demand and institutional factors (for example, government employment and unions) that are correlated with both wages and the share of women in an occupation. To remove the potential biases and identify their roles in explaining the wage-percent female relationship, we utilize the decomposition in Johnson and Solon that identifies the observable sources of the relationship between wages and the share of women in an occupation.[15] In practice, this amounts to decomposing the difference between the relationship between wages and percentage of women that excludes worker characteristics and the relationship between wages and percentage of women that includes worker characteristics.

We find that, for men, the most important measurable factor in each year is the industry of employment as opposed to personal characteristics such as education, age, and region of residence. It is well known that particular industries pay more than others, and our results show that these industries, on average, have higher concentrations of men. The opposite occurs for women. Education and age are the most important factors for explaining the wage- and percent-female relationship. If education and age capture personal preferences that influence occupational choice, then solely focusing on expanding occupational choice will result in a small narrowing of the gender wage gap.[16] This is because, as shown in our Oaxaca-Blinder decompositions, even after we add the concentration of women in an occupation to the model, the overall gender wage gap is still largely unexplained.

Notes

[1] *Highlights of Women's Earnings,* Report no. 943 (Bureau of Labor Statistics, May 2000).

[2] See for example, Elaine Sorensen, "Measuring the Pay Disparity Between Typically Female Occupations and Other Jobs: A Bivariate Selectivity Approach," *Industrial and Labor Relations Review*, July 1989, 624–39; George Borjas, *Labor Economics* (New York, McGraw-Hill, 1996); and Barbara Bergmann, "The Effect on White Incomes of Discrimination in Employment," *Journal of Political Economy*, 79, March-April 1971, pp. 294–313.

[3] George Johnson and Gary Solon, "Estimates of the Direct Effects of Comparable Worth Policy," *The American Economic Review*, Dec. 1986, pp. 1117–25.

[4] Multicollinearity was not a problem because the concentration of women within an occupation was documented at the three-digit level, providing more than 490 occupation categories.

[5] The Outgoing Rotation Group Files represent one-fourth of the sample.

[6] Further, the CPS does not include more detailed measures of skill plus a wide range of unmeasurable human capital characteristics. However, numerous studies have controlled for gender differences in school choice, type of degree, choice to enter higher paying occupations, childcare, work history, school performance, and job-setting measures. Collectively, they find that a large unexplained gap remains. See, for example, Mary E. Corcoran, Paul N. Courant, and Robert G. Wood, "Pay Differences among the Highly Paid: The Male-Female Earnings Gap in Lawyers' Salaries, "*Journal of Labor Economics*, July 1993, pp. 417–41 and Catherine J. Weinberger, "Mathematical College Majors and the Gender Gap in Wages, "*Industrial Relations*, July 1999, pp. 407–13.

[8] Others have used non-CPS data sets to control for gender differences in school choice, type of degree, and choice to enter higher paying occupations. For example, Corcoran, Courant, and Wood, "Pay Differences," 1993, examined pay differentials between male and female lawyers, and found that even after controlling for childcare, work history, school performance, and job-setting measures,

one-third to one-fourth of the gender gap was left unexplained. Weinberger, "Mathematical College Majors," 1999, estimates gender gaps between men and women who acquired technical majors in college. While she found that the lower mathematical content of the college major of women explains a good deal of the gender gap, she found that almost 9 percent of the gap was left unexplained.

[9] Further, instead of asking survey respondents their years of schooling, the survey bracketed years of schooling below 12th grade and asked respondents to report their highest degree of completion.

[10] See, for example, Paula England, *Comparable Worth: Theories and Evidence* (New York, Aldine De Gruyter, 1992); Paula England, "The Failure of Human Capitol Theory to Explain Occupational Sex Segregation, "*Journal of Human Resources*, vol. 17, 1982, pp. 358–70; David Macpherson, and Barry Hirsch, "Wages and Gender Composition: Why do Women's Jobs Pay Less?" *Journal of Labor Economics*, vol. 13, no. 3, 1995, pp. 426–71; Borjas, *Labor Economics*, 1996; and Barbara Bergmann, *In Defense of Affirmative Action* (New York, Basic Books, 1996).

[11] A benefit of using the CPS is that it contains data on a large nationwide sample and includes detailed occupational data in addition to the more commonly found earnings, employment, human capital, and demographic information. One of the primary drawbacks of the CPS is that it does not include a measure of actual experience. Potential experience must be generated from age minus years of schooling minus six. This may be done until 1990. After that time, changes in the education question in the CPS make it impossible to accurately measure the number of years an individual spent in school. Estimating potential experience is more likely to overestimate the experience of women, as they are more likely to have taken time out of the labor force than are men.

[12] These estimates of the gender gap are created from the difference of the logarithm of the mean earnings of men and women. Estimates by the Bureau of Labor Statistics are created from a special method of finding the median earnings of men and women using $50 increments centered around units of $50. The results of these two methods agree in 1989 and 1992, but differ by 5 percentage points in 1999. When the gender gap in 1999 is estimated using these data by taking the difference of the median earnings of men and women, the result is similar to that obtained by BLS.

[13] Johnson and Solon, "Estimates of the Direct Effects of Comparable Worth Policy," 1986.

[14] The full results of these regressions are available upon request from the authors.

[15] Johnson and Solon, "Estimates of the Direct Effects of Comparable Worth Policy," 1986.

[16] To illustrate this point with published BLS data, we computed the wage gap between male and female median weekly earnings in the three-digit occupations that pay more than the U.S. median weekly earnings of $576 in 2000. The average wage gap for this sample is 33 percent, with only one occupational gap below 10 percent. The significance of these gaps is that they are for women who chose to go into the highest paying U.S. occupations, yet they experience pay gaps that are just as large and larger than the overall U.S. gender pay gap.