Online Appendix for "A Cross-Cohort Analysis of Human Capital Specialization and the College Gender Wage Gap" Not for Publication

A1 Data Description

A1.1 Main ACS Samples

Our analysis is conducted using the 2014 to 2017 American Community Survey (ACS). The sample is restricted to include those who are not living in institutional group quarters, were born in one of the 50 U.S. states, have attained at least four years of college completion, and are age 23 to 67. We construct 5-year birth cohorts centered around the reported birth cohort. For example, the 1965 birth cohort includes those born between 1963 and 1967 (inclusive). When data for key demographic variables is missing, the ACS imputes values including age, sex, race, place of birth, educational attainment and undergraduate major. In the 2014 to 2017 ACS, 276,448 (2.2% of the 2014-2017 ACS) respondents have imputed educational attainment information and 196,379 respondents (1.6% of the 2014-2017 ACS) have imputed degree field information. We restrict our sample to include only those with non-imputed age, sex, race, origin, educational attainment and undergraduate major field information. We use inverse probability weighting to correct for non-response. In doing so, we preserve the age, sex race, and state of birth joint distribution. In total, our analysis sample of ACS respondents includes 1,718,330 individuals.

In our analysis, we proxy an hourly wage by dividing reported annual labor income by the reported usual hours the respondent worked in the previous year times the reported number of weeks the respondent worked during the previous year. As the weeks worked variable is an intervalled variable in the 2014-2017 ACS, we assign the midpoint of the category as the number of weeks worked. Nominal wages are converted to real 2018$. In all analyses including wages, we follow the conventional practices in the literature and restrict the sample to a set of people with well-measured wages: those who are employed civilians (excluding the self-employed) with non-missing annual labor income and strong attachment to the labor market defined as usually working at least 30 hours a week for a minimum of 27 weeks in the previous year. In calculating the potential wage indices by occupation and undergraduate major, we restrict the sample to white men in their peak wage years (ages 45 to 55) with well-measured wages. All analyses use log wages.

We use the variables degfield, degfield2, degfieldd, and degfield2d in IPUMS to identify both broad and detailed majors. IPUMS provides 176 detailed major codes for the 2014 to 2017 samples. In our analysis, we aggregate to 134 detailed major categories by subsuming very small major categories into larger categories. For example, we combined General Agriculture, Soil Science, and Miscellaneous Agriculture into one detailed agricultural major. Similarly, we combine Mathematics, Actuarial Science, and Mathematics and Computer Science into one detailed mathematics major. Our main analysis uses these detailed major categories. There are 29 broad major categories in our analysis. We use the broad major categories to describe trends in Figure 1 and describe major-to-occupation mappings in Figure 5.

Approximately 11 percent of the observations in our sample have dual majors. Our analysis requires a maximum of one major for each unit of observation. Thus, we assign a primary major to each person based on the maximum median potential wage in the two majors (based on white men aged 43-57 as described above). This assignment process relies on the assumption that agents will present their highest-wage major as their primary major in the labor market.

Figures A1 to A5 include a full listing of our detailed and broad major codes. Our data replication kit provides the code for our combination of majors.

We use a balanced panel of detailed occupation codes based on the 1990 Occupation codes and following the cross-walking strategy outlined by David Dorn which constructs a panel of 330 occupation codes.[28] In our analysis, we aggregate to 251 detailed occupation codes by subsuming very small occupation categories into larger categories. For people who are employed, the ACS reports occupation based on primary occupation. For people who are unemployed, the ACS reports occupation based on their most recent primary occupation in the last five years.

A2 Gender Differences in Wages and Employment Rates, Historical U.S. Censuses

As way of background, we measure time series trends in gender wage and employment gaps for young individuals using cross-sectional data from historical U.S. Censuses. We focus on cross-sections from the 1960, 1970, 1980, 1990, and 2000 U.S. Censuses as well as the 2010-2012 American Community Surveys (pooled). Within these data sets, we define the wages of those with at least a bachelor's degree similarly to our wage computation within the 2014-2017 ACS. Employment rates are measured as the individual currently working at least 30 hours per week (full-time).

The solid line in Figure A6 shows the difference in log wages between young working women and young working men with a bachelor's degree at decadal intervals using the historical Census and recent ACS data. We define "young" as individuals between the ages of 25 and 34. Given the decadal frequency and 10-year age range, each point on the lines in Figure A6 represents a different cohort of individuals. Young women with at least a bachelor's degree entering the labor market in 1960 (born between 1926 and 1935) earned average hourly wages that were 31 log points lower than their male counterparts. The gender wage gap for those with a bachelor's degree narrowed to 13 log points for the young cohort entering the labor market in 2010 (born between 1976 and 1985). The wage gap declined monotonically by a total of 18 log points for the cohorts entering the labor market between 1960 and 2010 with most of the decline occurring for cohorts entering the labor market between 1980 and 1990 (those born between 1956 and 1965). The results in Figure A6 show that substantial gender wage convergence occurred within the highly selected sample of those

[28] See https://www.ddorn.net/data.htm.

who attained a bachelor's degree.

The dashed line of Figure A6 shows gender convergence in full time employment rates among our sample of college graduates. We define full time employment as those individuals who report currently working at least 30 hours per week. Throughout all years, roughly 90 percent of college-educated young men reported working full time. For women, only about 35 percent of 25-34 year old female college graduates worked full time in 1960 – a gender employment rate gap of 35 percentage points. By 2010, roughly 80 percent of young educated women worked full time. For individuals with a bachelor's degree, there was a strong increase in women's wages and employment propensities relative to their male counterparts during the last half century. One goal of the paper is to assess how changing differences in undergraduate major choice is associated with changing gender wage and employment gaps across cohorts.

A3 Construction of Potential Wage Indices

In Section 3, we define our potential wage index as:

$$I_c^{P,M} = \frac{\sum_{m=1}^{M} s_{female,c}^{m} \bar{Y}_{male}^{m}}{\sum_{m=1}^{M} s_{male,c}^{m} \bar{Y}_{male}^{m}} - 1 \tag{6}$$

In practice, we compute this index by running the following regression locally within 5-year birth cohort c:

$$\bar{Y}_{male_i}^{m} = \alpha + \beta Female_i + \Gamma X_i + \epsilon_i \tag{7}$$

where $Female_i$ is a dummy variable indicating whether the respondent i has self-reported as female and X_i is a vector of demographic characteristics: race, state of birth, masters attainment, doctorate attainment, and marital status. Our potential wage index, $I_c^{P,M}$, is defined as β from each local regression by 5-year birth cohort and therefore, measures the differential "potential" wage of women of cohort c given that the female distribution of major choice in a given cohort may differ from males in their cohort. The units of this index are differential potential log wage based on major choice. In computing the comparable index for occupation, we substitute detailed occupation code, o, for detailed major code, m.

A4 Robustness: Key Results

A4.1 Robustness Figure 2

In Figure 2 of the main text, we restrict the sample on which our gender similarity indices are built to all individuals with reported majors (for $I_c^{DD,M}$ and $I_c^{P,M}$) or to all individuals with reported occupations (for $I_c^{DD,O}$ and $I_c^{P,O}$). Some individuals with reported majors are not working during the 2014-2017 period. Likewise, some individuals with reported occupations are not currently working (given the ACS asks occupations for people who are currently not working but may have worked at some point in the prior 5 years). To see if including those who are currently not working bias our indices, we perform a robustness exercise by creating the respective indices restricted to a sample of individuals with strong attachment to the labor market as defined throughout our analysis (civilians who are not self-employed and report working for at least 30 hours a week for at least 27 weeks in the previous year). The results of this exercise are shown in Appendix Figure A7. The results in Appendix Figure A7 are nearly identical to the results in Figure 2 of the main text. This suggests that our results are insensitive to whether we include individuals with strong attachment to the labor market or all individuals when describing patterns of gender sorting in major (occupation) choice.

Another potential issue with the results in Figure 2 stems from the fact that the ACS only asks undergraduate major in recent years (from 2009 to 2017). When we compare patterns for different birth cohorts, we risk confounding cohort and age effects. It is unlikely that this is problematic for our results about the convergence of undergraduate majors given that major is likely fixed over an individual's life cycle. Occupations are not fixed over the lifecycle, so this presents a potential problem with respect to how we have described occupational segregation by gender in Figure 2. To address this and separate age and cohort effects on occupational segregation by gender, we use data from the 1980, 1990, and 2000 U.S. Censuses along with multiple waves of the American Community Survey to measure $I_c^{DD,O}$ for different birth cohorts at a constant age. The results are shown in Appendix Table A1. As with the results in the main text, birth cohort refers to 5-year birth cohorts centered around the birth year listed. Similarly, age refers to 5-year age ranges centered on the age listed. As seen in Appendix Table A1, age effects are not substantively biasing the main results shown in Figure 2 of the main text. Within each age range, we see large convergence in the occupation similarity between men and women across birth cohorts.

A4.2 Robustness Table 1, Panel A

Appendix Tables A7 and A8 show a series of robustness checks on the results shown in Panel A of Table 2 of the main text. Appendix Table A7 shows our key regression results without including our vector of demographic and time controls. Focusing on column 1 of the Table, the raw gender gap in wages among individuals without a bachelor's degree for our pooled sample was 26.8 log points. Including demographic controls as in column 1 of Table 2 of the main text, the gender gap only fell to 23.3 log points. The demographic controls only explain a small fraction of the gender wage gap among college graduates.

Appendix Table A8 shows the robustness results for Panel (a) of Figure 2 of the main text to the alternate classification of majors and occupations. In our base specification in the main text, we used detailed occupation and major codes when defining the potential wage variables Y_i^m and Y_i^o. In the top panel of Appendix Table A8, we use the broad occupation and major codes to define Y_i^m and Y_i^o. In the bottom panel, we omit Y_i^m and Y_i^o altogether from the regression and instead include a vector of dummy variables for each broad major (occupation). The results of these alternate specifications are nearly identical to the results shown in Table 2 of the main text. This suggests that most of the variation in explaining gender wage gaps arises from differences across (as opposed to within) the broad major and occupation controls.

A4.3 Table 1, Panel B

In Panel B of Table 2, we show the importance of occupational choice and undergraduate major in explaining gender wage gaps for the 1958-1967 and the 1978-1987 birth cohorts. Appendix Table A9 shows the same results for the 1948-1957 and the 1968-1977 birth cohorts.

A5 Hours Differences Across Occupations

As discussed in the main text, there is a large literature highlighting the fact that women are in occupations with lower annual hours worked relative to men. In this section, we explore these results in the context of our methodology. To guide our empirical work, we define the following two variables: $\bar{H}^m_{male}$ and $\bar{H}^o_{male}$. $\bar{H}^m_{male}$ is defined as the median log annual hours worked for native-born white men between the ages of 43 and 57 who graduated with major m (regardless of subsequent occupation in which they worked). This is the potential hours associated with a given major based on older male hours. $\bar{H}^o_{male}$ is defined as the median log annual hours worked for native-born, white men between the ages of 43 and 57 who currently work in occupation o (regardless of undergraduate major). This is the potential hours associated with a given occupation based on older male hours. We refer to these variables as our potential annual hours worked indices. Majors (occupations) where men work more on average will have higher levels of $\bar{H}^m_{male}$ ($\bar{H}^o_{male}$).

How similar are men and women with respect to their occupational choices based on potential annual hours worked? Appendix Figure A8 displays $I^{H,M}_c$ and $I^{H,O}_c$ for different cohorts. $I^{H,M}_c$ is our potential hours index based on male annual hours worked in different majors and is defined as $I^{H,M}_c = \frac{\sum_{m=1}^{M} s^m_{female,c}\bar{H}^m_{male}}{\sum_{m=1}^{M} s^m_{male,c}\bar{H}^m_{male}} - 1$. Like our potential wage indices in the main text, the only reason $I^{H,M}_c$ only differs from 0 if men and women inhabit different majors. Likewise, $I^{H,O}_c$ is our potential hours index based on male annual hours worked in different occupations and is defined as $I^{H,O}_c = \frac{\sum_{o=1}^{O} s^m_{female,c}\bar{H}^o_{male}}{\sum_{o=1}^{O} s^m_{male,c}\bar{H}^o_{male}} - 1$.

As seen from Appendix Figure A8, women choose majors and occupations associated with lower potential annual hours worked. The major and occupational choice of women have converged to that of men over time in a way that implies women and men are choosing majors and occupations with more similar hours requirements. For the most recent cohorts, women are choosing both majors and occupations where potential annual hours worked are roughly 2% lower than men. Consistent with the literature, we find that college-educated women are choosing occupations with lower annual hours worked. We contribute to the hours literature by introducing the fact that college-educated women are choosing undergraduate majors associated with lower annual hours worked. We also show that the gender similarity of occupations and majors based on potential hours has been converging over time.

Appendix Figure A9 shows gender differences in the mapping of majors to occupations where we measure occupations in units of potential annual hours worked $\bar{H}^o_{male}$. Appendix Figure A9 is otherwise analogous to Figure 5 of the main text. To measure gender differences in occupational choice in hours units conditional on undergraduate major, we define $I^{H,O}_c|m$ which just recalculates $I^{H,O}_c$ (as defined above) restricting the sample to those individuals that chose major m. Consider individuals who choose to major in Engineering (Panel A, solid line). Women from the 1950 birth cohort who majored in Engineering subsequently work in occupations that had potential hours worked that were 2% lower than otherwise similar males. That gap disappeared for women who majored in Engineering after the 1975 birth cohort. For all majors, the gender gap in potential annual hours worked of occupational choice conditional on major has fallen over time. Women are now choosing occupations that are more similar in hours worked to men, conditional on occupational choice.

Appendix Figure A10 summarizes the mapping of majors to occupations where we measure occupations in potential hours space. This figure is otherwise analogous to Figure 6 in the main text. Women from the 1975 birth cohort are in occupations – conditional on major choice – that have annual hours worked that are three percent lower than comparable men. As a reminder, occupational potential wage differences, conditional on major choice, were about 9 percent for this cohort. Some of the reason that women may be choosing occupations with lower wages is that those occupations also have lower annual hours worked.

A6 Wage Gap Decompositions

In this section, we discuss the findings of a decomposition exercise where we assess the contribution of the independent variables in our main estimation to the college gender wage gap. As with the estimations in Table 2 and Table A9, the sample is restricted to include those with strong attachment to the labor market. We begin by estimating locally within birth cohort log wage equations for men only where race, state of residence, and marital status are categorical variables. As with all other specifications, the independent variable for $Major$ is the potential log wage from major choice, $\bar{Y}_i^m$, and the independent variable for *Occupation* is the potential log wage from occupation choice, $\bar{Y}_i^o$. Entries in the "Log Points" column are the within-cohort $male - female$ differences in the mean of the corresponding variable multiplied by the within-cohort *male* log wage coefficients of the corresponding variable. Entries in the "% Explained" column are the "Log Points" entries divided by the within-cohort *Total Raw Gap*.

In our model specification, occupational specialization plays the largest role in explaining the college gender wage gap. This is true for all 10-year birth cohorts. In the oldest birth cohort (1948-1957), occupation explains 43.9% of the gender wage gap. For the youngest birth cohort (1978-1987), the importance of occupation declines by 7 percentage points explaining 36.9% of the gender wage gap.

The results in Table 2 and Table A9 show that major choice and occupation choice are *independently* related to the college gender wage gap. This finding is a contribution to the literature on the college gender wage gap and the role of pre-market specialization. In our decomposition exercise, we formally show that pre-labor market human capital specialization (major choice) has non-trivial importance in explaining the college gender wage gap. For the oldest birth cohort (1948-1957), major choice explains 17.6% of the college gender wage gap. For the youngest birth cohort (1978-1987), major choice explains 27.9% of the college gender wage gap. While much of the existing literature has focused on the role of human capital attainment with respect to the gender wage gap, our decomposition shows that human capital attainment *above and beyond a bachelors degree* (such as a graduate degree) explains considerably less of the college gender wage gap than both pre-market and market human capital specialization.

Finally, in thinking about the time series patterns, two findings are of particular interest. First, occupational specialization has become less important between the 1948-1957 and 1958-1967 birth cohorts and then mostly stabilized. For the 1948-1957 birth cohort, occupation explained 43.9% of the college gender wage gap. This fell to 38.0% for the 1958-1967 birth and was 37.5% and 36.9% for the 1968-1977 and 1978-1987 birth cohorts respectively. Second, college major has become increasingly important in explaining the gender wage gap for college graduates over time. It explained 10.3 percentage points *more* of the gap in the youngest (1978-1987) compared to the oldest (1948-1957) birth cohort.

Figure A1: List of Detailed and Broad Majors

Detailed Major	Broad Major
General Agriculture, Soil Science, Misc. Agriculture	Agriculture
Agriculture Production and Management	Agriculture
Animal Sciences	Agriculture
Food Science	Agriculture
Plant Science and Agronomy	Agriculture
Environmental Science	Environment and Natural Resources
Forestry	Environment and Natural Resources
Natural Resources Management	Environment and Natural Resources
Architecture	Architecture
Area, Ethnic, and Civilization Studies	Area, Ethnic, and Civilization Studies
Communications	Communications
Journalism	Communications
Mass Media	Communications
Advertising and Public Relations	Communications
Communication Technologies	Engineering
Computer and Information Systems	Computer and Information Systems
Computer Programming and Data Processing	Computer and Information Systems
Computer Science	Computer and Information Systems
Information Sciences	Computer and Information Systems
Computer Information Management and Security	Computer and Information Systems
Computer Networking and Telecommunications	Computer and Information Systems
Cosmetology Services and Culinary Arts	Cosmetology and Physical Fitness
General Education, School Counseling, Educational Administration and Supervision	Education Administration and Teaching
Elementary Education	Education Administration and Teaching
Mathematics Teacher Education	Education Administration and Teaching
Physical and Health Education Teaching	Education Administration and Teaching
Early Childhood Education	Education Administration and Teaching
Science and Computer Teacher Education	Education Administration and Teaching
Secondary Teacher Education	Education Administration and Teaching
Special Needs Education	Education Administration and Teaching
Social Science or History Teacher Education	Education Administration and Teaching
Teacher Education: Multiple Levels	Education Administration and Teaching
Language and Drama Education	Education Administration and Teaching
Art and Music Education	Education Administration and Teaching
Miscellaneous Education	Education Administration and Teaching

Figure A2: List of Detailed and Broad Majors (continued)

Detailed Major	**Broad Major**
Engineering: General, Military Technologies, Metallurgical, Biomedical, Geological and Geophysical, Mining and Mineral, Naval Architecture and Marine, Nuclear, Petroleum	Engineering
Aerospace Engineering	Engineering
Biological Engineering	Engineering
Chemical Engineering	Engineering
Civil and Architectural Engineering	Engineering
Computer Engineering	Engineering
Electrical Engineering, Electrical Engineering Technology, Electrical and Mechanic Repairs and Technologies	Engineering
Engineering Mechanics, Physics, and Science	Engineering
Environmental Engineering	Engineering
Industrial and Manufacturing Engineering, Precision Production and Industrial Arts	Engineering
Mechanical Engineering	Engineering
Miscellaneous Engineering	Engineering
Engineering Technologies	Engineering
Engineering and Industrial Management	Engineering
Industrial Production Technologies	Engineering
Mechanical Engineering Related Technologies	Engineering
Miscellaneous Engineering Technologies	Engineering
Linguistics, and Comparative Language and Literature	Linguistics and Foreign Languages
French, German, Latin and Other Common Foreign Languages	Linguistics and Foreign Languages
Other Foreign Languages	Linguistics and Foreign Languages
Family and Consumer Sciences	Family and Consumer Sciences
Pre-Law and Legal Studies, Court Reporting	Law
English Language and Literature	English and Literature
Composition and Speech	English and Literature
Liberal Arts	Liberal Arts and Humanities
Humanities	Liberal Arts and Humanities
Library Science	Education

Figure A3: List of Detailed and Broad Majors (continued)

Detailed Major	**Broad Major**
Biology, Misc. Biology, Pharmacology, Botany, Neuroscience, Genetics	Biology and Life Sciences
Biochemical Sciences	Biology and Life Sciences
Molecular Biology	Biology and Life Sciences
Ecology	Biology and Life Sciences
Microbiology	Biology and Life Sciences
Physiology	Biology and Life Sciences
Zoology	Biology and Life Sciences
Mathematics, Actuarial Science, Mathematics and Computer Science	Math and Statistics
Applied Mathematics	Math and Statistics
Statistics and Decision Science	Math and Statistics
Interdisciplinary,Multi-Disciplinary, Intercultural and International Studies	Multi-Disciplinary Studies (General)
Nutrition Sciences	Multi-Disciplinary Studies (General)
Physical Fitness, Parks, Recreation, and Leisure	Cosmetology and Physical Fitness
Philosophy and Religious Studies	Philosophy and Theology
Theology and Religious Vocations	Philosophy and Theology
Physical Sciences, Astronomy and Astrophysics, Geosciences, Nuclear, Industrial Radiology, and Biological Technologies	Physical Sciences
Atmospheric Sciences and Meteorology	Physical Sciences
Chemistry	Physical Sciences
Geology and Earth Science	Physical Sciences
Oceanography	Physical Sciences
Physics	Physical Sciences
Materials Science and Materials Engineering	Engineering
Multi-disciplinary or General Science	Physical Sciences
Psychology, Cognitive Science and Biopsychology, Social Psychology	Psychology
Educational Psychology	Psychology
Clinical Psychology	Psychology
Counseling Psychology	Psychology
Industrial and Organizational Psychology	Psychology
Miscellaneous Psychology	Psychology

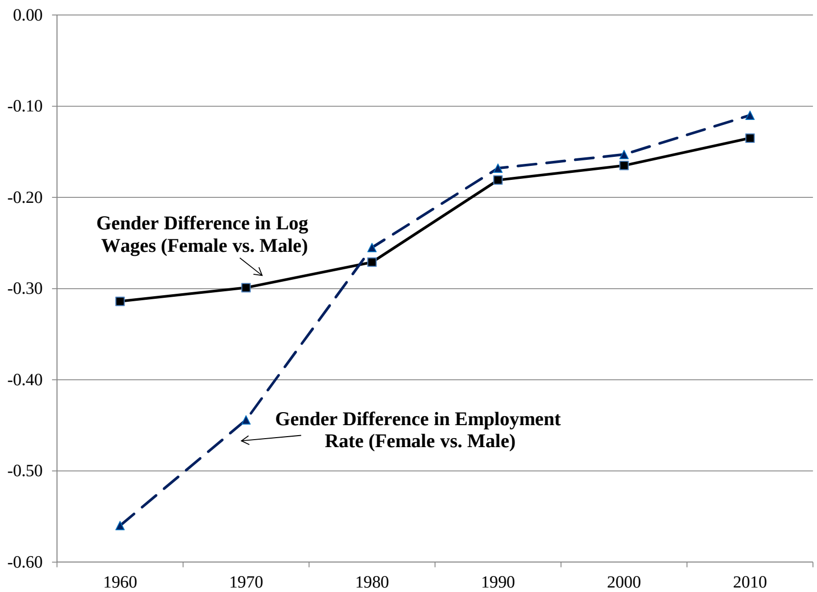
Figure A4: List of Detailed and Broad Majors (continued)

Detailed Major	**Broad Major**
Criminal Justice and Fire Protection	Criminal Justice and Fire Protection
Public Administration	Public Affairs, Policy, and Social Work
Public Policy	Public Affairs, Policy, and Social Work
Human Services and Community Organization	Public Affairs, Policy, and Social Work
Social Work	Public Affairs, Policy, and Social Work
General Social Sciences	Social Sciences
Economics, Agricultural Economics, Business Economics	Social Sciences
Anthropology and Archeology	Social Sciences
Criminology	Social Sciences
Geography	Social Sciences
International Relations	Social Sciences
Political Science and Government	Social Sciences
Sociology	Social Sciences
Miscellaneous Social Sciences	Social Sciences
Construction Services	Construction Services
Transportation Sciences and Technologies	Construction Services
Fine Arts, Commercial Art and Graphic Design, Film, Video and Photographic Arts, Studio Arts, Miscellaneous Fine Arts	Fine Arts
Drama and Theater Arts, Music, Visual and Performing Arts	Fine Arts
Art History and Criticism	Fine Arts
General Medical and Health Services	Nursing, Medical and Health Sciences
Communication Disorders Sciences and Services	Nursing, Medical and Health Sciences
Health and Medical Administrative Services	Nursing, Medical and Health Sciences
Medical Assisting Services	Nursing, Medical and Health Sciences
Medical Technologies Technicians	Nursing, Medical and Health Sciences
Health and Medical Preparatory Programs	Nursing, Medical and Health Sciences
Nursing	Nursing, Medical and Health Sciences
Pharmacy, Pharmaceutical Sciences, and Administration	Nursing, Medical and Health Sciences
Treatment Therapy Professions	Nursing, Medical and Health Sciences
Community and Public Health	Nursing, Medical and Health Sciences
Miscellaneous Health Medical Professions	Nursing, Medical and Health Sciences

Figure A5: List of Detailed and Broad Majors (continued)

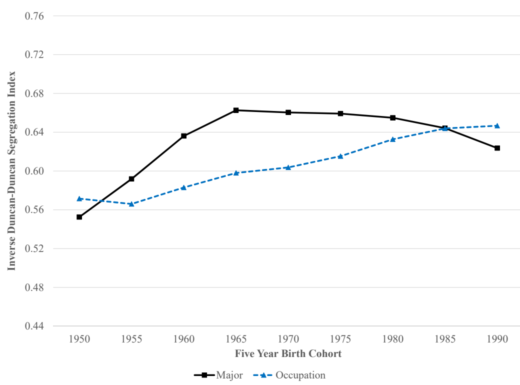
Detailed Major	Broad Major
General Business	Business
Accounting	Business
Business Management and Administration	Business
Operations, Logistics and E-Commerce	Business
Marketing and Marketing Research	Business
Finance	Business
Human Resources and Personnel Management	Business
International Business	Business
Hospitality Management	Business
Management Information Systems and Statistics	Business
Miscellaneous Business and Medical Administration	Business
History	History
United States History	History

Figure A6: Gender Differences in Log Wages and Employment Rates, Individuals Aged 25-34 with Bachelor's Degree

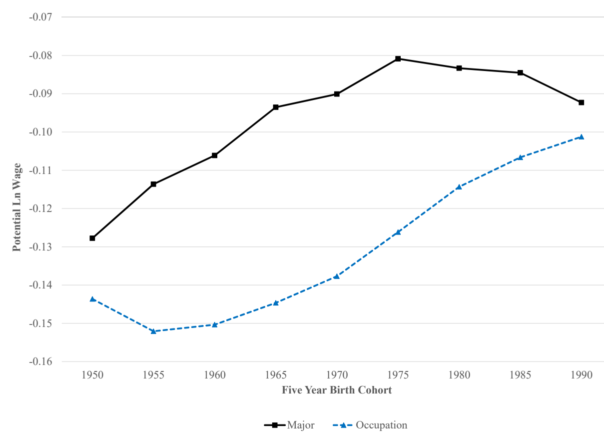


Notes: Figure shows the gender gap in log wages (solid line) and the gender gap in employment rates (dashed line) for individuals aged 25-34 with a bachelor's degree in different Census years. The 2010 data refers to a pooled sample of ACS respondents from sample years 2010, 2011, and 2012. Individual wages are self-reported annual labor income divided by self-reported annual hours worked. The employment rate is the fraction of respondents who report working full-time. Differences are measured as female data minus male data. See text for additional details.

Figure A7: Gender Similarity in Major Choice and Occupation by Cohort, Strongly Attached Sample



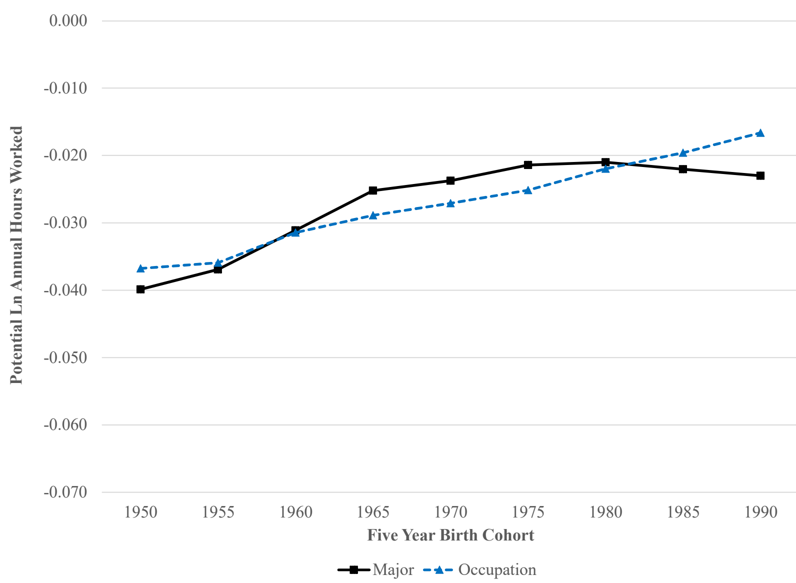
Panel A: Segregation Index Index



Panel B: Potential Wage Index

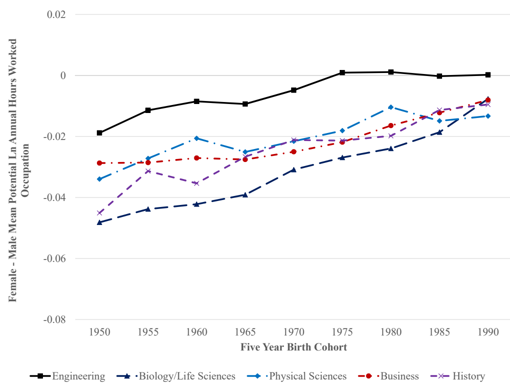
Notes: Figure plots the inverse segregation index (left panel) and potential wage index (right panel) for different cohorts conditioning on strong attachment to the labor market. The solid line in each panel show the indices for major choice. The dashed line in each panel show the indices for occupation. Data from the 2014-2017 ACS and is restricted to those with at least a bachelor's degree. See text for additional details.

Figure A8: Gender Similarity in Major and Occupation by Cohort, Measuring Occupations by Potential Hours Worked

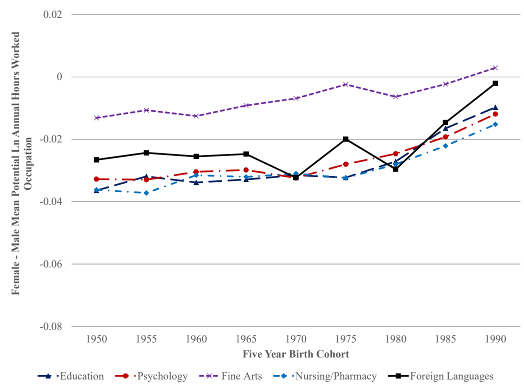


Notes: Figure plots the gender similarity index for different cohorts based on potential hours worked. The solid line in each panel shows the indices based on major choice. The dashed line shows the indices based on subsequent occupation. Data from the 2014-2017 ACS and is restricted to those with at least a bachelor's degree. See text for additional details.

Figure A9: Within-Major Gender Differences in Potential Hours by Occupation, by Gender and Cohort



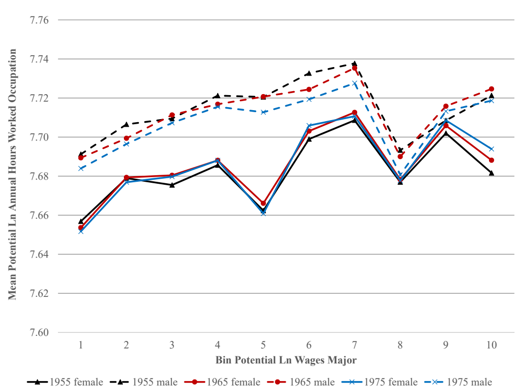
Panel A: Male Dominated Majors



Panel B: Female Dominated Majors

Notes: These figures show the trends in $I_c^{H,O|m}$ conditional on having graduated with major m. Panel A are male-dominated majors. Panel B are female-dominated majors. As with the left panel of Figure 2, potential wage in an occupation, $\bar{H}^o_{male}$, is computed using only annual hours worked of native white males 43-57 who are working full time in the 2014-2017 ACS.

Figure A10: Mapping of Potential Wages by Major to Potential Hours by Occupation, by Gender and Cohort



Panel A: Levels

Panel B: Differences

Notes: These figures show the mapping between major and occupation. On the x-axes, we have binned majors based on $\bar{Y}^m_{male}$, the log wage deciles of native, white men age 43 to 57. On the y-axis in Panel A, we report $I_c^{H,O|d,m}$, the mean log potential occupational hours worked within these deciles described separately by gender and cohort. In Panel B, the y-axis reports female - male differences in $I_c^{H,O|d,m}$ for two of the cohorts.

Table A1: Robustness of Trends in Inverse Duncan-Duncan Index for Gender Occupational Similarity, Controlling for Age Effects

Cohort	30	35	40	45	50	55
1925						0.358
1930					0.368	
1935				0.380		0.439
1940			0.405		0.475	
1945		0.435		0.499		0.517
1950	0.496		0.503		0.517	
1955		0.522		0.520		0.551
1960	0.609		0.555		0.559	0.576
1965		0.592		0.565	0.583	
1970	0.621		0.581	0.582		
1975		0.605	0.591			
1980	0.625	0.607				
1985	0.632					

Note: This table computes the inverse Duncan-Duncan index for gender similarity in occupational sorting ($I_c^{DD,O}$) for different birth cohorts and age ranges. See main text for construction of the index. Cohorts are five year birth cohorts centered around the birth cohort listed. Age are five year age ranges centered around the age listed. Data come from the 1980, 1990, and 2000 U.S. Censuses as well as various years of the American Community Survey.

Table A2: Dispersion in Occupational Choice Conditional on Major, 1968-77 Birth Cohort

	Herfindahl-Herschman Index $HHI^{m}_{g,c}$	
Broad Major	Men	Women
Agriculture	0.10	0.08
Environment and Natural Resources	0.10	0.09
Architecture	0.26	0.21
Area, Ethnic, and Civilization Studies	0.09	0.10
Communications	0.12	0.12
Computer and Information Sciences	0.18	0.13
Cosmetology Services and Physical Fitness	0.10	0.10
Education Administration and Teaching	0.29	0.48
Engineering	0.16	0.13
Linguistics and Foreign Languages	0.09	0.11
Family and Consumer Sciences	0.11	0.15
Law	0.13	0.14
English and Literature	0.09	0.11
Liberal Arts and Humanities	0.09	0.14
Biology and Life Sciences	0.12	0.10
Math and Statistics	0.11	0.13
Multi-Disciplinary Studies (General)	0.09	0.11
Philosophy and Theology	0.10	0.09
Physical Sciences	0.09	0.08
Psychology	0.09	0.11
Criminal Justice and Fire Protection	0.19	0.10
Public Affairs, Policy, and Social Work	0.13	0.18
Social Sciences	0.11	0.10
Construction Services	0.29	0.25
Fine Arts	0.08	0.09
Nursing, Medical and Health Sciences	0.25	0.42
Business	0.16	0.14
History	0.10	0.11

Note: Table shows occupational dispersion within major category for men and women born between 1968 and 1977 for different majors. Specifically, this table reports $HHI^{m}_{g,c}$ from the 2014-2017 ACS. We use broad major and broad occupation categories. Values closer to 0 reflect more dispersion.

Table A3: Major to Occupation Mapping Measure, 1968-1977 Birth Cohort

Detailed Major	**Female-Male Mean Potential Ln Wage Occupation**
Zoology	-0.233
Early Childhood Education	-0.200
Microbiology	-0.193
Miscellaneous Psychology	-0.186
Linguistics, and Comparative Language and Literature	-0.170
International Business	-0.169
Nutrition Sciences	-0.164
Interdisciplinary,Multi-Disciplinary Studies	-0.163
Clinical Psychology	-0.154
Health and Medical Administrative Services	-0.154
Pre-Law and Legal Studies, Court Reporting	-0.152
Miscellaneous Social Sciences	-0.150
Cosmetology Services and Culinary Arts	-0.149
Biochemical Sciences	-0.146
Educational Psychology	-0.145
Family and Consumer Sciences	-0.142
Biology, Misc. Biology, Pharmacology, Botany, Neuroscience, Genetics	-0.139
Public Policy	-0.134
Psychology, Cognitive Science and Biopsychology, Social Psychology	-0.131
Molecular Biology	-0.128
Physiology	-0.126
Physical Sciences, Astronomy and Astrophysics, Geosciences, Nuclear, Industrial Radiology, and Biological Technologies	-0.125
Criminology	-0.122
Miscellaneous Business and Medical Administration	-0.118
Mathematics, Actuarial Science, Mathematics and Computer Science	-0.117
Computer Programming and Data Processing	-0.117
Liberal Arts	-0.117
Environmental Engineering	-0.116
Marketing and Marketing Research	-0.115
French, German, Latin and Other Common Foreign Languages	-0.115
Health and Medical Preparatory Programs	-0.114
Communication Disorders Sciences and Services	-0.113
Community and Public Health	-0.113
Humanities	-0.112
Criminal Justice and Fire Protection	-0.110
General Medical and Health Services	-0.110
General Business	-0.109

Table A4: Major to Occupation Mapping Measure, 1968-1977 Birth Cohort (continued)

Detailed Major	**Female-Male Mean Potential Ln Wage Occupation**
Food Science	-0.107
Sociology	-0.105
Engineering Mechanics, Physics, and Science	-0.105
Physics	-0.102
Transportation Sciences and Technologies	-0.101
Chemistry	-0.100
Public Administration	-0.099
Area, Ethnic, and Civilization Studies	-0.099
Computer Information Management and Security	-0.097
Science and Computer Teacher Education	-0.096
Finance	-0.094
Industrial Production Technologies	-0.093
Political Science and Government	-0.092
Information Sciences	-0.092
General Social Sciences	-0.089
Geology and Earth Science	-0.088
Materials Science and Materials Engineering	-0.088
Business Management and Administration	-0.088
Economics, Agricultural Economics, Business Economics	-0.088
Geography	-0.085
Miscellaneous Education	-0.084
Computer Science	-0.084
Teacher Education: Multiple Levels	-0.083
Oceanography	-0.083
Treatment Therapy Professions	-0.082
Computer and Information Systems	-0.082
English Language and Literature	-0.080
Operations, Logistics and E-Commerce	-0.080
Biological Engineering	-0.080
General Education, School Counseling, Educational Administration and Supervision	-0.076
Ecology	-0.076
Miscellaneous Health Medical Professions	-0.075
Communications	-0.074
Elementary Education	-0.072
History	-0.072
Plant Science and Agronomy	-0.072
Management Information Systems and Statistics	-0.071

Table A5: Major to Occupation Mapping Measure, 1968-1977 Birth Cohort (continued)

Detailed Major	**Female-Male Mean Potential Ln Wage Occupation**
Secondary Teacher Education	-0.071
Accounting	-0.067
Anthropology and Archeology	-0.065
Art History and Criticism	-0.064
Industrial and Organizational Psychology	-0.064
Drama and Theater Arts, Music, Visual and Performing Arts	-0.063
Physical Fitness, Parks, Recreation, and Leisure	-0.062
Chemical Engineering	-0.061
Environmental Science	-0.061
Social Work	-0.058
Art and Music Education	-0.058
Civil and Architectural Engineering	-0.055
International Relations	-0.055
Hospitality Management	-0.054
Human Services and Community Organization	-0.053
Computer Networking and Telecommunications	-0.052
Language and Drama Education	-0.052
Journalism	-0.051
Communication Technologies	-0.051
Human Resources and Personnel Management	-0.050
Statistics and Decision Science	-0.048
Mechanical Engineering	-0.047
Advertising and Public Relations	-0.044
Fine Arts, Commercial Art and Graphic Design, Film, Video and Photographic Arts, Studio Arts, Miscellaneous Fine Arts	-0.043
Industrial and Manufacturing Engineering, Precision Production and Industrial Arts	-0.042
Engineering Technologies	-0.042
Architecture	-0.042
Other Foreign Languages	-0.039
Philosophy and Religious Studies	-0.038
Medical Technologies Technicians	-0.038
Electrical Engineering, Electrical Engineering Technology, Electrical and Mechanic Repairs and Technologies	-0.034
Applied Mathematics	-0.034
Mass Media	-0.031
Intercultural and International Studies	-0.030
Mechanical Engineering Related Technologies	-0.029
Aerospace Engineering	-0.028
Physical and Health Education Teaching	-0.027

Table A6: Major to Occupation Mapping Measure, 1968-1977 Birth Cohort (continued)

Detailed Major	**Female-Male Mean Potential Ln Wage Occupation**
Atmospheric Sciences and Meteorology	-0.027
Mathematics Teacher Education	-0.026
Special Needs Education	-0.026
Miscellaneous Engineering Technologies	-0.025
Theology and Religious Vocations	-0.023
Social Science or History Teacher Education	-0.022
General Agriculture, Soil Science, Misc. Agriculture	-0.020
Computer Engineering	-0.017
Nursing	-0.015
Agriculture Production and Management	-0.012
Animal Sciences	-0.004
Engineering: General, Military Technologies, Metallurgical, Biomedical, Geological and Geophysical, Mining and Mineral, Naval Architecture and Marine, Nuclear, Petroleum	-0.003
Pharmacy, Pharmaceutical Sciences, and Administration	-0.002
Natural Resources Management	-0.001
United States History	0.002
Construction Services	0.002
Composition and Speech	0.005
Forestry	0.024
Engineering and Industrial Management	0.026
Miscellaneous Engineering	0.027
Counseling Psychology	0.036
Misc. Biology	0.043

Table A7: Major Choice, Occupation Choice and Gender Gaps in Wages and Employment

(a) Log Wage Regressions, No Controls

	Log Wages				Employment Rate	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.268	-0.189	-0.160	-0.135	-0.075	-0.066
	(0.007)	(0.006)	(0.005)	(0.004)	(0.003)	(0.003)
$\bar{Y}_i^m$		0.827		0.355		0.091
		(0.017)		(0.014)		(0.005)
$\bar{Y}_i^o$			0.862	0.793		
			(0.011)	(0.009)		
Controls	No	No	No	No	No	No
R^2	0.04	0.10	0.23	0.23	0.01	0.01

Note: This table is a robustness check on panel (a) of Table 1 with no demographic or time controls. Sample size for columns 1-4 is 2,270,392. Sample size for columns 5-6 is 3,428,990.

Table A8: Major Choice, Occupation Choice and Gender Gaps in Wages and Employment, Alternative Specifications Using Broad Major and Occupations

(a) Log Wage and Employment Rate Regressions with Broad Major and Occupation Potential Wage Indices, Pooled Cohorts

	Log Wages				Employment Rate	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.232	-0.172	-0.167	-0.141	-0.088	-0.083
	(0.006)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
$\bar{Y}_i^m$		0.832		0.461		0.061
		(0.022)		(0.015)		(0.004)
$\bar{Y}_i^o$			0.750	0.668		
			(0.013)	(0.010)		
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.22	0.26	0.33	0.34	0.13	0.13

(b) Log Wage and Employment Rate Regressions with Flexible, Broad Major and Occupation Dummies, Pooled Cohorts

	Log Wages				Employment Rate	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.232	-0.169	-0.168	-0.143	-0.088	-0.083
	(0.006)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)
Major dummies	No	Yes	No	Yes	No	Yes
Occupation dummies	No	No	Yes	Yes	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.22	0.27	0.34	0.35	0.13	0.13

Note: This table is a robustness check on the main results in Panel (a) of Table 1 using two alternate ways to control for occupation and major choice. In Panel (a) of this table, we include as independent variables measures of potential wages determined by the broad major and occupation choice instead of detailed major and occupation choice. In panel (b), we include as independent variables vectors of broad major dummies and occupation dummies instead of our potential wage controls. Sample size for panel A columns 1-4 is 2,256,630. Sample size for panel A columns 5-6 is 3,428,990. Sample size for panel B columns 1-4 is 2,256,630. Sample size for panel B columns 5-6 is 3,428,990.

Table A9: Major Choice, Occupation Choice and Gender Gaps in Wages and Employment

(a) Log Wage Regressions, Older Cohorts

	1948-1957 Birth Cohorts			1958-1967 Birth Cohorts		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.291	-0.163	-0.130	-0.322	-0.198	-0.168
	(0.009)	(0.005)	(0.005)	(0.008)	(0.005)	(0.004)
$\bar{Y}_i^m$			0.366			0.411
			(0.018)			(0.016)
$\bar{Y}_i^o$		0.886	0.819		0.909	0.823
		(0.016)	(0.013)		(0.015)	(0.012)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.12	0.30	0.31	0.13	0.32	0.33

(b) Log Wage Regressions, Younger Cohorts

	1968-1977 Birth Cohorts			1978-1987 Birth Cohorts		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
$Female_i$	-0.271	-0.169	-0.144	-0.155	-0.093	-0.065
	(0.008)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
$\bar{Y}_i^m$			0.410			0.443
			(0.015)			(0.010)
$\bar{Y}_i^o$		0.850	0.766		0.599	0.513
		(0.014)	(0.011)		(0.008)	(0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.14	0.31	0.33	0.13	0.25	0.27

Note: The specifications in this table are the same as the specifications shown in Panel (b) of Table 2. Columns 4-6 from this table are exactly the same as the results in Panel (b) of Table 2. The new results in this table are in columns 1-3 of both panels which show the results for alternate birth cohorts. Sample size for panel (a) columns 1-3 is 331,678. Sample size for panel (a) columns 4-6 is 533,348. Sample size for panel (b) columns 1-3 is 543,452. Sample size for panel (b) columns 4-6 is 614,106.

Table A10: Wage Decompositions: Explanatory Variables

(a) Older Cohorts

Variable	1948-1957 Birth Cohort (Log Points)	(% Explained)	1958-1967 Birth Cohort (Log Points)	(% Explained)
Race	0.0008	0.25%	0.0011	0.32%
State	0.0004	0.14%	0.0002	0.06%
Marital Status	0.0287	9.19%	0.0164	4.74%
Masters	-0.0048	-1.55%	-0.0034	-0.99%
Doctorate	0.0056	1.80%	0.0022	0.63%
Major	0.0549	17.57%	0.0547	15.81%
Occupation	0.1371	43.86%	0.1314	37.97%
Year	-0.0006	-0.19%	-0.0001	-0.04%
Explained	0.2221	71.07%	0.2025	58.51%
Unexplained	0.0904	28.93%	0.1436	41.49%
Total Raw Gap	0.31		0.35	

(b) Younger Cohorts

Variable	1968-1977 Birth Cohort (Log Points)	(% Explained)	1978-1987 Birth Cohort (Log Points)	(% Explained)
Race	0.0004	0.13%	0.0001	0.04%
State	0.0001	0.03%	0.0001	0.05%
Marital Status	0.0146	5.03%	0.0002	0.11%
Masters	-0.0080	-2.76%	-0.0064	-4.09%
Doctorate	0.0007	0.25%	0.0000	0.03%
Major	0.0444	15.27%	0.0438	27.85%
Occupation	0.1090	37.53%	0.0579	36.86%
Year	-0.0003	-0.09%	0.0001	0.03%
Explained	0.1609	55.39%	0.0956	60.87%
Unexplained	0.1296	44.61%	-0.0615	39.13%
Total Gap	0.29		0.16	

Note: Sample restrictions and cohorts consistent with Table A9. In these estimations, race, state of residence, and marital status are categorical variables instead of flexible dummies. This does not affect our main results and is only for ease in decomposition and display. As with all other specifications, the independent variable for *Major* is $\bar{Y}_i^m$, and the independent variable for *Occupation* is $\bar{Y}_i^o$. Entries in the "Log Points" column are the within-cohort $male - female$ differences in the mean of the corresponding variable multiplied by the within-cohort *male* log wage coefficients of the corresponding variable. Entries in the "% Explained" column are the "Log Points" entries divided by the within-cohort Total Raw Gap. The Total Raw Gap differs from the $Female_i$ in Column (1) of Table A9 in that it is the raw gender wage gap with no controls and the gender wage gap displayed as the coefficient for $Female_i$ in Column (1) of Table A9 includes demographic controls.