Preferences for Job Tasks And Gender Gaps in the Labor Market

Madeleine Gelblum*

January 16, 2020

Please click here for the most recent version.

Abstract

Women and men work in markedly different jobs, leading to persistent occupational segregation by gender. This paper provides evidence that gender differences in how individuals value activities performed at work, termed job tasks, can help explain these sorting patterns. I conduct a hypothetical choice experiment to elicit workers' willingness to pay for a set of tasks that are more frequently performed by one gender than the other. The experimental scenarios ask participants to choose between two hypothetical jobs that differ in terms of pay and the amount of time spent on a gender-typical task, but are otherwise the same. I find significant gender differences in willingness to pay for three of the five tasks that I examine. Willingness to pay is significantly higher among participants who report spending more time on a task in their current job, suggesting that estimates are correlated with actual sorting behavior. I show that gender differences in preferences for the tasks that I investigate can account for a substantial portion of occupational segregation in the U.S. labor market.

*Gelblum: Harvard University. Email: mgelblum@g.harvard.edu. I am grateful to my advisors, David Deming, Claudia Goldin, Lawrence Katz and Amanda Pallais, for their guidance. I thank Lisa Abraham, David Autor, Christopher Avery, Oriana Bandiera, Alex Bell, Nicholas Bloom, Michela Carlana, Elijah de la Campa, Edward Glaeser, Joshua Goodman, Marie-Pascale Grimon, Helen Ho, Isabel Macdonald, Stephen Machin, Alan Manning, Guy Michaels, Jörn-Steffen Pischke, Maya Sen, Elie Tamer and many others for helpful feedback. This paper has also benefited from comments by seminar participants at Harvard University and the London School of Economics. This work has been supported by the Harvard University Multidisciplinary Program in Inequality and Social Policy, the Harvard Kennedy School Women and Public Policy Program, and the Lynde and Harry Bradley Foundation.

1 Introduction

Women and men work in markedly different jobs, leading to persistent occupational segregation by gender (Blau et al. 2013). Figure 1 shows that the median woman is employed in an occupation in which 70 percent of workers are female, while the median man works in an occupation that is 71 percent male.[1] Occupational segregation contributes to the gender wage gap (Blau and Kahn 2017) and may reflect an inefficient allocation of workers to jobs. Indeed, women remain under-represented in many high-paying professional occupations despite having higher levels of education than men (Goldin et al. 2006).[2]

Gender differences in how individuals value activities performed at work, termed *job tasks*, may help explain these sorting patterns. Research documents that women and men work in jobs that involve different activities (e.g. Lordan and Pischke 2018; Cortes et al. 2018). In addition, measures of tasks can account for a large fraction of the variation in the share of workers in an occupation who are female, termed the *female share*.[3] Workers may have preferences over tasks as in Rosen (1986), such that they are willing to accept lower wages in jobs that involve activities they enjoy, and must be compensated extra to perform activities they dislike. If women and men have different preferences over tasks, these valuations may contribute to occupational segregation and other gender gaps.

This paper examines whether preferences over job tasks differ by gender. I conduct a hypothetical choice experiment embedded in a survey to elicit workers' willingness to pay (WTP) for a set of *gender-typical* tasks that women perform more frequently than men, or vice versa. While it is likely that preferences for a task will affect sorting, women and men may differ in how frequently they perform a task for other reasons. In particular, women and men may have a comparative advantage in different activities (Baker and Cornelson 2018). Tasks may also be correlated with gender-based discrimination (Kuhn and Shen 2013), or with other amenities that women and men value differently, such as flexible work arrangements (Goldin 2014). Using observational data on wages and job choices to disentangle preferences for tasks from these other factors presents a challenge.[4]

[1] Figure 1 displays the distribution of the share of workers in an occupation who are female for currently employed individuals aged 18 and older, using data from the 2012-2016 American Community Survey (ACS).

[2] Goldin (2014) shows that gender differences in occupation can account for approximately a third of the gender wage gap even among college graduates employed full-time.

[3] Table 1 displays the R^2 and adjusted R^2 statistics from regressions of the occupational female share in the ACS on a set of occupation-level explanatory variables. Column 1 includes the mean log hourly wage, mean log hours of work per week, and share of workers with a college degree. Column 2 includes all *generalized work activity* and *work style* variables from the O*NET, a database of occupational characteristics. The adjusted R^2 of 0.77 based on the task measures in Column 2 is nearly twice as large as the adjusted R^2 of 0.40 based on the variables in Column 1.

[4] A large literature documents the difficulty of estimating compensating differentials due to unobserved human capital and amenities (e.g. Brown 1980; Bell 2019). Similarly, measuring labor market discrimination

By contrast, the hypothetical choice experiment allows me to identify task valuations by randomly assigning wage offers and specifying worker choice sets. Specifically, the experimental scenarios ask participants to choose between two hypothetical jobs that differ in terms of pay and the amount of time spent on a focal task, but are otherwise the same. Thus the difference in wage offers between the two jobs is known and is unrelated to worker skills or other attributes. In addition, participants are told that the two jobs are exactly the same in terms of schedule, co-workers, benefits and all other characteristics aside from the wage and time spent on the focal task. Therefore, participants are unlikely to view the two jobs as differing in terms of factors such as discrimination that are not directly connected to work activities.

The experiment elicits preferences for five conceptual categories of tasks that are performed in a broad range of jobs and are not tied to specific credentials. The task measures are based on variables from the O*NET, a database of occupational characteristics, that are correlated with the occupational female share in the American Community Survey (ACS). The variables that I incorporate into the experiment explain more than half of the variation in the female share, and capture nearly as much variation as the full set of task measures available in the O*NET. The selected measures include two female-typical tasks related to interpersonal activities - *helping and caring for others* and *working and communicating with others*. A third female-typical task, *documenting and recording information*, is important in many female-dominated jobs in health, education and social services. The male-typical task of *operating and repairing equipment* is consistent with the notion that men enjoy working with their hands or with machinery. A second male-typical task, *making decisions and solving problems*, is essential in many majority-male professional occupations.

I use participant choices from the experiment to estimate WTP for gender-typical tasks as a share of the wage, guided by a simple discrete choice model. I find that women are willing to pay significantly more than men for the female-typical tasks of *helping and caring for others* and *documenting and recording information.* In my preferred specification, women are willing to forgo 3.3 percentage points more than men as a share of their wage to work in a job in which they spend more time *helping and caring for others*, and are willing to give up 2.6 percentage points more than men to spend more time *documenting and recording information.* In addition, men have a significantly higher WTP than women - by 8 percentage points - for the male-typical task of *operating and repairing equipment.* I find no significant gender differences in WTP for the female-typical task of *working and communicating with others* or the male-typical task of *making decisions and solving problems.*

remains an empirical challenge (Blau and Kahn 2017; Altonji and Blank 1999). In both cases, a key issue is that a worker's outside options are not observed in conventional labor market data.

Furthermore, I find that for all tasks, WTP is significantly higher among those who report devoting a larger share of working hours to that activity in their current job, consistent with preferences for tasks affecting sorting decisions. In addition, WTP results are similar when I weight the experiment sample by gender, race, education and major occupation to match the nationally representative ACS. These findings suggest that the WTP estimates are predictive of real-world sorting decisions and may reflect preferences for a broader share of the labor market despite the fact that the experiment sample is not randomly selected.

In the final section of the paper, I examine the implications of the WTP results for gender gaps in the labor market. First, I document that observed gender differences in sorting on the job tasks that I investigate are substantial in both the experiment sample and the ACS. Sorting on these five tasks as measured by a segregation index can account for more than three quarters of occupational segregation by gender in the ACS.

Next, I calculate the gender differences in sorting that are predicted by my model given the preference estimates and wage differentials associated with the tasks. I find that these predicted differences in sorting can explain nearly 70 percent of the segregation attributable to the gender-typical tasks in the experiment sample, and more than 40 percent in the ACS. These results suggest that gender differences in preferences for the five tasks that I examine can account for approximately a third of occupational segregation in the U.S. labor market. This pattern of findings is robust to a variety of alternative specifications, including controlling for other job amenities and repeating the analysis in other data sources with better measures of human capital.

By contrast, the preference estimates appear to explain little of the gender wage gap, but results are not conclusive. This finding is largely driven by the fact that some of the tasks that I examine widen the wage gap while others narrow it, and these effects offset each other.[5] Therefore, estimates of the contribution of preferences for tasks to the gender wage gap are inherently sensitive to the choice of which activities to examine, as well as the magnitude of the wage differentials associated with tasks.

1.1 Related Literature

This paper contributes to a large literature on the determinants of occupational segregation and other gender gaps in the labor market. Several studies examine gender differences in preferences for job amenities and other workplace characteristics. Experimental evidence from laboratory and field settings suggests that women are less likely than men to choose

[5] In particular, the female-typical task of *helping and caring for others* is associated with a wage penalty that increases the magnitude of the gender wage gap, while the female-typical task of *documenting and recording information* offers a wage premium and thus decreases the magnitude of the gap.

competitive compensation schemes (Niederle and Vesterlund 2007; Flory et al. 2015), and more likely to select team-based pay (Kuhn and Villeval 2015).[6] Other research argues that women place greater value on flexible scheduling arrangements and working fewer hours due to household constraints (Goldin 2014; Wiswall and Zafar 2018; Mas and Pallais 2017; Goldin and Katz 2016; Wasserman 2019; Denning et al. 2019; Cortes and Pan 2019).[7]

Most relevant to this paper are a handful of studies in economics (Lordan and Pischke 2018; Cortes and Pan 2018; Fortin 2008; Grove et al. 2011) and a larger body of research in psychology (e.g. Su et al. 2009; Pinker 2008) suggesting that women have a greater preference than men for jobs that involve helping others or working with people rather than things. In particular, Lordan and Pischke (2018) summarize a large number of O*NET variables as three latent factors that they label *people*, *brains* and *brawn*, and show suggestive evidence that women have a relative preference for people compared with brawn jobs.[8]

The current project contributes to this prior literature by providing evidence that women have a higher WTP than men for helping tasks and a lower WTP for activities related to equipment and machinery, consistent with the notion of a gender difference in preferences for people versus things. In contrast to previous studies that have used observational methods or descriptive surveys, however, this paper offers the first set of experimental evidence that preferences for tasks as measured by WTP differ by gender.

This project also relates to research contending that women have a comparative advantage in performing interpersonal tasks relative to certain physical activities (Cortes et al. 2018; Baker and Cornelson 2018; Ngai and Petrongolo 2017; Borghans et al. 2014; Weinberg 2000; Beaudry and Lewis 2014; Bacolod and Blum 2010; Black and Spitz-Oener 2010; Welch 2000).[9] Much of this literature emphasizes the same stylized fact that motivates this project - that women and men work in jobs that involve different activities - but proposes the alternative interpretation that skills rather than preferences may account for these differences. Task-specific preferences and skills are likely to be correlated, and I cannot shed light on

[6]See Bertrand (2011) and Croson and Gneezy (2009) for a review of the literature on gender differences in preferences related to attitudes and personality traits.

[7]Studies also report evidence that workers prefer colleagues of the same gender (Pan 2015) and that women (men) are more likely to participate in a group activity that requires stereotypically female (male) topical knowledge (Coffman 2014), consistent with a role for norms and identity in explaining gender gaps (Akerlof and Kranton 2000).

[8]Lordan and Pischke (2018) find that women tend to have higher reported job satisfaction if they work in an occupation with higher people and lower brawn content; men exhibit a similar qualitative pattern of reported job satisfaction, but the magnitudes are smaller. The authors also ask a sample of secondary students to choose between pairs of occupations, and find that female students in particular are more likely to choose jobs with higher people content.

[9]In particular, Cortes et al. (2018) contend that a female advantage in social skills may explain women's increased representation in cognitive occupations in recent decades. Baker and Cornelson (2018) document evidence of gender differences in sorting on the spatial, motor and sensory skill requirements of jobs.

the process of preference and skill formation. However, this project assesses the extent to which gender differences in sorting on tasks can be explained by women and men responding differently to the same wage offer because of their task valuations, rather than women and men receiving different wage offers due to skill differences.

Finally, this project builds on recent studies that use hypothetical choice data to estimate preferences for job amenities (Mas and Pallais 2017; Wiswall and Zafar 2018; Maestas et al. 2018; Datta 2019). In particular, Mas and Pallais (2017) estimate WTP for flexible work arrangements and report similar results from field and hypothetical choice experiments, suggesting that a purely hypothetical approach can generate amenity valuations that are relevant for real-world decisions. Similarly, Wiswall and Zafar (2018) use a hypothetical choice experiment to assess WTP for several workplace attributes, and find that estimated preferences predict subsequent college major and job choices and can explain a meaningful share of the gender wage gap in their sample.

The remainder of the paper proceeds as follows. Section 2 discusses the gender-typical tasks that I examine. Section 3 describes the design of the hypothetical choice experiment. Section 4 lays out the model and econometric strategy, and Section 5 reports the results of the experiment. Section 6 discusses implications of the experimental results for gender gaps in the labor market. Finally, Section 7 concludes.

2 Gender-Typical Tasks

The goal of the hypothetical choice experiment is to elicit preferences for gender-typical tasks that women are more likely to perform than men, or vice versa. However, representative data on the frequency of task performance among U.S. workers are not available. Therefore, as a proxy for the concept of frequency I use information on the importance of tasks from the O*NET, a U.S. Department of Labor database of occupational characteristics.[10]

I focus on O*NET measures in the *generalized work activities* and *work styles* domains, which I interpret as providing information about conceptual categories of tasks that are performed in a broad range of jobs. Importantly, these task categories are not explicitly linked to formal educational credentials. To examine gender differences in these measures, which are reported at the occupation level, I link the O*NET variables to information on the share of workers in an occupation who are female. I use data on currently employed workers aged 18 and older from the 2012-2016 American Community Survey (ACS) to construct the

[10]The O*NET data are based primarily on surveys of workers in each occupation. The survey questions related to the task importance variables ask respondents, "How important is X to the performance of your current job?", where X is an occupational characteristic.

occupational female share.[11]

I consider O*NET work activities and work styles that are positively (negatively) correlated with the occupational female share to be measures of female-typical (male-typical) tasks. However, a large number of the work activities and work styles have statistically significant bivariate relationships with the female share, and many of the O*NET variables are correlated with each other. To select a set of these measures for inclusion in the experiment, I follow a hybrid quantitative and qualitative approach.

I estimate a series of ordinary least squares (OLS) regression models predicting the female share based on the O*NET variables. Specifically, within each O*NET domain I regress the female share on: 1) all tasks simultaneously, 2) a group of tasks with the most positive and most negative bivariate coefficients, and 3) a group of tasks that are rated as highly predictive using a random forest algorithm. I also repeat the regression analysis including controls for broad occupation cluster, mean log hourly wages and the share of workers with a college degree or more in each occupation.[12] I then search qualitatively for O*NET variables that are statistically significant and consistent in sign across specification, and combine some similar measures, yielding a final set of five measures. Appendix A1 provides further details on the process of task selection.

Table 2 displays the names of the three female-typical and two male-typical tasks that I include in the experiment, along with a list of the O*NET variable(s) on which each task is based. The designation of *working and communicating with others and displaying a cooperative attitude*[13] as a female-typical task is consistent with the notion that women may prefer jobs that involve working with people (Lordan and Pischke 2018) or have a comparative advantage in performing interpersonal activities (Cortes et al. 2018). The female-typical task of *helping and caring for others* also involves interpersonal interaction; the selection of this task supports evidence that care work is overwhelmingly performed by women (England 2005; Folbre, ed 2012). Similarly, labeling *operating, repairing and maintaining vehicles, devices or equipment*[14] as male-typical is consistent with the hypothesis that men prefer working with "things" (Su et al. 2009). The O*NET rates the female-typical task of *documenting and recording information* as highly important in many female-dominated occupations in healthcare, education and social services.[15] Finally, the male-typical task of *making decisions and solving problems* is ranked as important in a range of professional occupations that

[11] The ACS collects demographic and socioeconomic information from a random sample of the U.S. population and is administered annually by the U.S. Census Bureau.

[12] Tables A1, A2, A3 and A4 shows the results of this analysis.

[13] Hereafter, I refer to this task as *working and communicating with others.*

[14] Hereafter, I refer to this task as *operating and repairing equipment.*

[15] Figures A3 and A4 display the female share and gender-typical task levels by major occupation category.

remain majority male, such as physician, lawyer and many STEM jobs.[16]

In addition to the name of each task, I provide survey participants with a definition and examples, also displayed in Table 2. I use the O*NET documentation as a guide, but modify the examples to ensure that they do not reference only female-dominated or only male-dominated occupations. While the O*NET is designed to measure task differences across occupations, the hypothetical scenarios ask participants to consider changing the amount of time spent on a task while keeping other aspects of the job the same. The examples in Table 2 are therefore designed to enable participants to envision performing each task in a wide range of jobs.

Figure 2 displays mean levels of the five gender-typical tasks for women and men, using the O*NET measures on which the tasks are based.[17] The tasks are rescaled to reflect percentiles weighted by employment in the ACS, such that a value of 50 indicates the median task level. As expected, task levels differ dramatically by gender. Women work in jobs in which the mean percentiles of the female-typical tasks of *helping and caring for others*, *documenting and recording information* and *working and communicating with others* are 17, 11, and 22 percentage points higher, respectively, than the jobs in which men work. By contrast, women work in jobs in which the mean percentile of the male-typical task of *operating and repairing equipment* is 20 percentage points lower, compared with the jobs held by men.[18]

Table 1 displays the R^2 and adjusted R^2 statistics from OLS regressions of the female share on all 57 work activity and work style variables in Column 2, and the selected task measures in Column 3.[19] The adjusted R^2 based on all the O*NET variables in Column 2 is 0.77, indicating that these measures can explain a very large fraction of the variation in the female share. Furthermore, the adjusted R^2 of 0.67 in Column 3 suggests that the five selected tasks capture the majority of the female share variation, and more than 85 percent of the variation accounted for by the full set of work activities and work styles.

[16]Tables A5, A6, A7, A8 and A9 display the ten occupations with the highest and the ten occupations with the lowest levels of each task measure, along with the female share in each displayed occupation.

[17]For the selected tasks that combine multiple O*NET variables, I average the component variables to create a single measure.

[18]The mean percentile of *making decisions and solving problems* is 3.5 percentage points lower among women than men. However, Table A1 shows that the relationship between the female share and this task is larger in magnitude when controlling for other work activities.

[19]The O*NET variables are standardized to have a mean of zero and standard deviation of one.

3 Experiment Design

3.1 Survey Recruitment and Preliminary Questions

I recruit and compensate participants in the experimental survey using Amazon Mechanical Turk (MTurk). MTurk is a platform that enables researchers and others to pay individuals to perform online activities such as completing surveys.[20] Research suggests that samples recruited through MTurk have adequate psychometric properties such as internal consistency and test-retest reliability (Buhrmester et al. 2011), and that levels of measurement error are similar compared with representative survey samples (Snowberg and Yariv 2018).[21] I administered the survey to 1,931 participants over two rounds in June 2018,[22] restricting the sample to MTurk participants who are U.S. residents and who have an approval rating of at least 95 percent on the platform.[23]

The survey begins by asking about the participant's current employment status, hours of work per week, pay rate, industry and occupation.[24] If participants report that they are not currently employed, then all questions about the current job are modified to refer to the most recent job.[25] I also specify that these questions refer to work other than completing activities on the MTurk platform. Next, the survey asks participants to report the number of hours per week they spend in their current or most recent job performing each of the five gender-typical tasks.[26] After the hypothetical scenarios, the survey gathers information on gender, age, race, ethnicity and educational attainment.

[20] An MTurk requester posts a description of the activity, including the number of participants required, the pay, and typically the amount of time required. Participants complete the activity on a first come, first served basis. Requesters then review the work and approve or deny each submission.

[21] In addition, recent studies use MTurk to recruit samples for descriptive surveys of labor market activity (Abraham and Amaya 2018; Katz and Krueger 2019).

[22] Participants were paid $1.80 to take the survey, which required an average of 10 minutes to complete. Compensation was therefore approximately equal to the Massachusetts minimum wage in 2018. Participants completed the survey on the Qualtrics platform, and then submitted a unique completion code on MTurk to receive compensation.

[23] MTurk requesters can specify that participants meet certain requirements in order to be eligible to complete an activity. Participant location is self-reported on MTurk, but the Qualtrics platform collects data on the latitude and longitude of the respondent's device or IP address. I exclude from the analysis participants who were physically present outside of the U.S. while completing the survey.

[24] The response options for occupation and industry correspond to major occupation and industry categories using the ACS occupation and the North American Industry Classification System (NAICS) codes.

[25] In addition, I ask non-employed participants about the number of months since they last worked, and if they report that they have never worked I omit questions about the current or most recent job.

[26] The answer choices correspond to intervals that each represent 10 percent of the participant's total weekly hours of work. There is no restriction that the number of hours reported on the five tasks sum to the total number of hours worked per week, as participants likely spend a portion of their time on tasks that the survey does not ask about, and may also perform some of the gender-typical tasks concurrently.

3.2 Hypothetical Scenarios

The survey asks participants to consider a series of hypothetical scenarios, each associated with one of the five gender-typical tasks. In the scenarios, participants are asked to envision that they have been given a choice between two jobs that differ in terms of pay and the amount of time spent on the focal task. Participants are then asked to indicate which job they would prefer. I randomize the order in which the scenarios appear and the display order of the answer choices within each scenario.

In each hypothetical scenario, one job is randomly selected to offer the participant's wage in the current or most recent job, expressed using the pay period that the participant reports.[27] The other job offers a wage that is higher than the participant's current or most recent wage by a randomly selected percentage from the following set: 1) 0 percent, 2) 5 percent, 3) 10 percent, 4) 15 percent, or 5) 20 percent.[28]

I hypothesize that worker preferences over the amount of time spent on tasks are not finely tuned, such that it is easiest for participants to choose between two mutually exclusive bins, one of which includes their optimal allocation of hours to the focal task, and one of which does not. Therefore, participants are given a choice between spending less than a cutoff of C hours per week on the focal task in one job, and C or more hours per week on the focal task in the other job (termed the *high-task* job).

For a randomly selected half of the sample, C is equal to 10 percent of total hours worked per week in the participant's current or most recent job. This cutoff can be interpreted as defining the extensive margin of task performance, as it gives participants a choice between spending little or no time on an activity compared with at least some time. This design has the advantage of involving the same comparison for all tasks.

However, gender differences in preferences for tasks may be largest at cutoffs along the intensive margin of task performance. Therefore, in the other half of the sample I choose a cutoff percentage of time, P, separately for each task, to maximize the difference between the share of women and the share of men who report spending at least P percent of their on the focal task, in a pilot version of the survey.[29] The cutoff number of hours, C, is then

[27]Participants can report their pay on an hourly, weekly, bi-weekly, twice monthly, monthly or yearly basis. For participants who have never worked, I use a value of \$20 per hour for the wage rate in this job. Note that \$20/hour falls between the mean hourly wage of \$23.86 and median of \$17.81 for all U.S. workers, based on the 2016 Occupational Employment Statistics (OES) data.

[28]For *operating and repairing equipment*, the set of percentages is {0%, 5%, 10%, 15%, 20%, 25%}, based on pilot data suggesting that WTP for this task is larger in magnitude, compared with the other tasks.

[29]Specifically, the cutoff percentages are 20 percent of hours worked for *helping and caring for others* and *documenting and recording information*, 40 percent of hours worked for *working and communicating with others*, 0 percent of hours worked (which I operationalize as less than 1 hour versus at least one hour) for *operating and repairing equipment*, and 30 percent of hours worked for *making decisions and solving problems*.

equal to P percent of the hours worked per week in the participant's current or most recent job.[30]

Participants are told to assume that other than the wage and the amount of time spent on the focal task, the two jobs in each scenario are exactly the same in all other ways, including total hours worked per week, schedule, co-workers, benefits, and the set of activities they do when not performing the focal task. Emphasizing that other aspects of the job do not vary decreases the probability that participants will view time spent on the focal task as related to factors such as discrimination and the gender composition of co-workers that may be correlated with task performance in real-world settings. In order to clarify that the high-task job does not require working more hours in total, I specify that in both positions, participants would work the same number of hours per week as in their current or most recent job.[31]

The experiment aims to measure preferences over conceptual categories of tasks that may contribute to gender gaps in the labor market. Therefore, the hypothetical scenarios do not provide additional information about job context, such as occupation or industry, that would tie the WTP estimates to a specific set of detailed work activities that constitute the focal task in a given context, as well as to a specific set of counterfactual activities that cannot be performed at the same time as the focal task.

As a result, however, the WTP estimates may reflect the distribution of contexts and, in particular, counterfactual activities that participants envision. I assume that participants are most likely to imagine their own current or most recent job. Therefore, the counterfactuals imagined may differ by gender, given the well-documented gender differences in job choices. While I cannot rule out bias from this issue, I discuss evidence in Section 5 that suggests that it is unlikely to meaningfully affect results.

Figure 3 displays an example of the hypothetical scenarios shown to participants where the focal task is *working and communicating with others*, the cutoff number of hours is on the extensive margin, and the participant reports working 40 hours per week and pay of $20 per hour in the current job.

4 Model and Econometric Strategy

I use data from participant choices in the hypothetical scenarios to estimate WTP for gender-typical tasks, separately by gender, guided by a simple discrete choice model. As in Rosen

[30]For participants who report that they have never worked, the extensive margin cutoff C is equal to 4 hours, and the intensive margin cutoff is equal to P percent of 40 hours.

[31]For participants who have never worked, I specify that both jobs involve 40 hours of work per week.

(1986), I assume that workers derive utility from the wage and the amenities offered by a job.

In the hypothetical scenarios, participants are instructed to choose between two jobs that are the same except for pay and the amount of time spent on the focal gender-typical task. Thus participant i's utility from job $j \in \{1, 2\}$ related to task $k \in \{1, \ldots, 5\}$ can be expressed as

$$U_{ijk} = \alpha_i + \theta_k T_{jk} + \delta_k \ln w_{ijk} + \varepsilon_{ijk} \tag{1}$$

where α_i reflects participant-specific factors affecting utility from work, T_{jk} is an indicator for job j being the high-task job for gender-typical task k (described hereafter as the *high-task* k job or the job with a *high level of task* k), w_{ijk} is the wage offer for participant i in job j related to task k, θ_k and δ_k are preference parameters indexed by task, and ε_{ijk} is a worker-, job- and task-specific preference parameter. For convenience, let $j = 1$ index the high-task k job, such that $T_{1k} = 1$ and $T_{2k} = 0$, $\forall k$.

The parameter θ_k can be interpreted as reflecting preferences for the task k amenity, while δ_k reflects preferences over wages. The magnitude of θ_k relative to δ_k determines mean WTP for task k. The ε_{ijk} parameter shifts WTP for individual i. I assume that ε_{ijk} has a standard Extreme Value (EV) Type I distribution.[32]

An important benefit of the experimental setting is that w_{ijk} is observed for both jobs in the participant i's choice set; by contrast, standard survey and administrative data sources contain information only on realized wages. In addition, amenities or other characteristics of the hypothetical job that participant i envisions that are not directly affected by the amount of time spent on task k will not vary between jobs, and will be absorbed by the α_i term.

Participant i chooses the job with the high level of task k if

$$\begin{aligned} U_{i1k} &> U_{i2k} \\ \kappa_{ik} &< \theta_k + \delta_k \omega_{ik} \end{aligned} \tag{2}$$

where $\kappa_{ik} \equiv \varepsilon_{i2k} - \varepsilon_{i1k}$ and $\omega_{ik} \equiv \ln(w_{i1k}/w_{i2k})$ is the log difference in wage offers between the high-task and low-task jobs for individual i. The EV Type I assumption implies that κ_{ik} has a standard logistic distribution. Therefore, the parameters δ_k and θ_k can be estimated by logistic regression, where the outcome, y_{ik}, is an indicator for participant i choosing the high-task k job, and the predictors are an intercept term and ω_{ik}.[33]

[32] The EV Type I distribution describes the behavior of a random variable that is the maximum of some underlying sample. This distributional assumption can be motivated by the notion that preferences are shaped by repeated exposure to a task or bundle of tasks.

[33] This strategy follows the approach used in recent literature estimating preferences for job amenities

My primary hypothesis is that δ_k and θ_k differ by gender. I therefore estimate the model for the entire experiment sample and separately by gender, yielding coefficients $\hat{\delta}_{gk}$ and $\hat{\theta}_{gk}$, where $g \in \{a, f, m\}$ indexes all participants, women and men, respectively.

To derive an expression for WTP, note that the logistic distribution for κ_{ik} implies that for an individual of gender g, the probability of choosing the high-task k job is given by

$$\begin{aligned} \Pr(y_{ik} = 1) &= \mathrm{F}_\kappa(\theta_{gk} + \delta_{gk}\omega_{ik}) \\ &= \frac{1}{1 + \exp(-(\theta_{gk} + \delta_{gk}\omega_{ik}))} \end{aligned} \tag{3}$$

where $\mathrm{F}_\kappa(\cdot)$ is the CDF of κ_{ik}. Conditional on the parameters for gender g, this probability can be expressed as a function of ω_k:

$$\mathrm{H}_{gk}(\omega_k) \equiv \mathrm{F}_\kappa(\theta_{gk} + \delta_{gk}\omega_k). \tag{4}$$

For each value of ω_k, the share choosing the high-task job is the proportion of participants willing to pay $100 * (1 - \exp(\omega_k))$ percent of their wage to spend more time on task k.

At the mean and median of the κ_{ik} distribution for gender g, $\Pr(y_{ik} = 1) = 0.5$ and thus $\omega_k = -\theta_{gk}/\delta_{gk}$. Therefore, the mean WTP for more time spent on task k as a proportion of the wage among individuals of gender g is given by

$$\lambda_{gk} \equiv 1 - \exp\left(-\frac{\theta_{gk}}{\delta_{gk}} \right). \tag{5}$$

I estimate λ_{gk} for $g \in \{a, f, m\}$ and the gender difference in λ_{gk}:

$$\beta_k \equiv \lambda_{fk} - \lambda_{mk}, \tag{6}$$

using the Delta method to calculate robust standard errors.

I hypothesize that $\beta_k > 0$ if task k is female-typical and $\beta_k < 0$ if task k is male-typical. Furthermore, if $\delta_{fk} \approx \delta_{mk}$, then $\mathrm{H}_{fk}(\omega_k) > \mathrm{H}_{mk}(\omega_k)$ when $\beta_k > 0$ and $\mathrm{H}_{fk}(\omega_k) < \mathrm{H}_{mk}(\omega_k)$ when $\beta_k < 0$, for almost all values of ω_k. Therefore, the model predicts that if women (men) have a higher WTP for task k at the mean, then more women (men) will generally sort into high-task k jobs.

based on discrete choice experiments (Mas and Pallais 2017; Maestas et al. 2018).

5 WTP Results

5.1 Descriptive Statistics

This section presents descriptive statistics for the experiment sample recruited using the MTurk platform. Table 3 displays summary statistics for currently employed individuals in the experiment sample, overall and by gender, in Columns 1-3, and comparable statistics for employed individuals aged 18 and older in the 2012-2016 ACS in Columns 4-6.[34] I focus on the statistics for employed individuals because 90 percent of the experiment sample reports being currently employed, in contrast to 60 percent of individuals in the ACS.

The experiment sample is substantially younger (34 versus 42 years in Table 3), more likely to be White (72 versus 65 percent of the sample), and more educated, compared to the overall U.S. population as captured by the ACS. Specifically, 56 percent of the experiment sample reports having at least a bachelor's degree and 8 percent has a high school diploma or less, while 34 percent of the employed ACS sample falls in each of these categories. Experiment participants also report modestly lower hourly wages ($21/hour versus $25/hour in the ACS).[35]

The experiment sample is 53 percent female. Women in the experiment sample are substantially more likely than men to be White (77 percent versus 66 percent). In addition, women are less likely than men to be employed (89 versus 92 percent in Table A10), work fewer hours per week (36 versus 39 hours), and have lower hourly wages ($19/hour versus $24/hour).[36]

Figures 4 and 5 show the distribution of employment by major occupation and industry categories in the experiment sample compared with the ACS.[37] Computer and math, education, arts, sports and media, and business operations and finance occupations are substantially over-represented in the experiment sample, while managerial, production, maintenance and transportation, health practitioner and technician, and personal care and cleaning occupations are under-represented. Among industry categories, information, educational services,

[34] Table A10 displays comparable statistics for employed and non-employed individuals in both samples.

[35] In both samples, I exclude hourly wage observations that are less than $3 or greater than $200. Hourly wages in the ACS are calculated as annual earnings divided by annual hours of work, and are inflated to 2018 dollars.

[36] The comparable gender gaps in employment, hours of work and hourly wages have the same sign in the ACS. However, women are slightly less likely to have a college degree in the experiment sample (55 versus 58 percent in Table 3), but are more likely to have a college degree in the ACS (36 versus 32 percent). I estimate a version of the WTP analysis in which I weight the experiment sample to match the ACS by gender, race, educational attainment and major occupation, to ensure that factors such as the education distribution and gender differences in the racial and ethnic composition of the experiment sample are not driving results.

[37] The figures use data on the most recent job for participants in the experiment sample who are not currently employed, but restrict to employed workers in the ACS.

professional, scientific and technical services, finance and insurance, administrative services, and arts, entertainment and recreation are over-represented in the experiment sample, while jobs in health and social assistance, manufacturing, construction and extraction, public administration, and wholesale trade are under-represented.

Figure 6 displays the percentage of weekly hours worked that participants report spending on the five gender-typical tasks, in 10 percentage point intervals. The distributions reveal substantial heterogeneity across task; for example, only 1 percent of the sample spends no time *working and communicating with others*, while 54 percent spends no time *operating and repairing equipment*.

Figure 7 shows the distribution of time spent on tasks separately for women and men. It is clear that women spend substantially more time than men on the female-typical tasks of *helping and caring for others*, *working and communicating with others*, and *documenting and recording information*, and less time on the male-typical task of *operating and repairing equipment*. For example, women are 17 percentage points more likely than men to spend at least 50 percent of their time *helping and caring for others*, and 18 percentage points more likely than men to spend no time *operating and repairing equipment*.[38] These results suggest that the O*NET variables measuring task importance do capture information about the frequency of task performance, as hypothesized.

5.2 Baseline WTP

Table 4 reports mean WTP as a proportion of the wage for the jobs that involve spending more time on the gender-typical tasks (i.e. the high-task jobs), using data from the hypothetical choice experiment. This table is based on job choices from all participants, regardless of whether they face the extensive or intensive margin cutoffs in the hypothetical scenarios. Each cell in the first three rows of the table gives an estimate of λ_{gk} from a gender g- and task k-specific regression. The first, second and third rows display results for all participants, women and men, respectively, while the columns indicate task. The final row presents estimates of β_k, the female-to-male difference in WTP. Figures 8 and 9 plot the estimates of λ_{gk} and β_k, respectively.

In the sample as a whole, WTP is negative or close to zero for all gender-typical tasks. Specifically, the estimate in Column 1 of Table 4 indicates that participants must be compensated an additional 2.6 percent in order to be willing to work in a job with a high level of *helping and caring for others*. Similarly, participants must be paid approximately 2 percent

[38]The distribution of time spent on the male-typical task of *making decisions and solving problems* is fairly similar across genders, which is consistent with the finding in Figure 2 that the overall gender difference in the mean level of this task in the O*NET is relatively small.

more to spend more time *documenting and recording information* or *working and communicating with others*. The WTP estimate for *operating and repairing equipment* is largest in absolute value; participants require an additional 12 percent in pay to be willing to work in a job with a high level of this activity. Finally, WTP for *making decisions and solving problems* is very close to zero and insignificant.

It is striking that there is no task for which mean WTP is positive in the overall sample, despite the fact that for all tasks except *operating and repairing equipment*, over half the sample reports working in a high-task job.[39] Indeed, WTP among women in the second row is negative for all three female-typical tasks and statistically significant for two of the three (*documenting and recording information* and *working and communicating with others*), while WTP among men in the third row is negative and statistically significant for the male-typical task of *operating and repairing equipment*. It may be that many workers have substantial "white space" or downtime in their jobs during which they are essentially idle. Therefore, participants may interpret spending more time on any one activity as requiring greater effort because it reduces the downtime available to them.

The final row of Table 4 shows evidence of significant gender differences in WTP for three of the five tasks examined. Women are willing to pay 3.3 and 2.6 percentage points more than men for the female-typical tasks of *helping and caring for others* and *documenting and recording information*, respectively. Similarly, men's WTP for the male-typical task of *operating and repairing equipment* is 8.0 percentage points higher than the estimate for women. These significant gender differences are consistent with the hypothesis that $\beta_k > 0$ ($\beta_k < 0$) for female-typical (male-typical) tasks.

I find no significant gender differences in WTP for the female-typical task of *working and communicating with others* or the male-typical task of *making decisions and solving problems*. The finding for *making decisions and solving problems* suggests that the observed correlation between this task and the female share may be due to factors such as discrimination that constrain women from entering jobs that offer decision-making authority. The result for *working and communicating with others* is surprising given the large observed gender difference in this task in the O*NET in Figure 2.

Figures 10 and 11 plot the share of participants choosing the high-task job for each gender-typical task against the log difference in wage offers, for the entire sample and for women and men separately, along with the predicted probabilities from the logistic specification. It is clear that women are more likely to choose a job with a high level of *helping and*

[39] I define participants as working in a high-task k job if they report spending more time on task k in their current or most recent job than the cutoff number of hours in the scenario related to task k that they are shown.

caring for others, and men are more likely to choose a job with a high level of *operating and repairing equipment*, at nearly all wage differentials. By contrast, it is evident from the figures that there is little gender difference in job choices at any wage differential for *working and communicating with others* or *making decisions and solving problems*. Gender differences in choices related to *documenting and recording information* occur primarily with a wage differential of zero or when the high-task job offers a higher wage.

When the wage offer is the same in both jobs, women are 14 percentage points more likely than men to choose a job with a high level of *helping and caring for others*, 15 percentage points more likely than men to choose a job with a high level of *documenting and recording information*, and 24 percentage points less likely than men to choose a job with a high level of *operating and repairing equipment*. These differences are statistically significant.[40]

5.3 Heterogeneity and Robustness Checks

Table 5 reports WTP estimates separately for participants shown hypothetical scenarios with extensive margin cutoffs (Panel A) and intensive margin cutoffs (Panel B), as well as the differences between the estimates (Panel C).[41] Results for both sub-samples are qualitatively similar to those in Table 4, although WTP estimates are generally more negative on the intensive margin, consistent with participants having a distaste for spending a greater amount of time on the gender-typical tasks. The gender difference in WTP for *helping and caring for others* is also larger in magnitude on the intensive margin (5.9 percentage points versus an insignificant 1.4 percentage points on the extensive margin).

Table A11 displays WTP estimates separately for participants with at least a college degree and those without a college degree. Results are similar across groups and compared with baseline estimates in Table 4. The exception is that WTP for *making decisions and solving problems* is positive and significant among college workers and negative and significant among non-college workers, and the difference across groups is also significant (4.5 percentage points when pooling women and men). This finding is consistent with college-educated workers being more likely to hold jobs involving problem-solving and high-stakes decision-making.

Tables A12 and A13 show WTP estimates excluding participants who are inattentive and

[40]Women are also are 6 percentage points more likely to choose a job with a high level of *working and communicating with others* and 3 percentage points less likely to choose a job with a high level of *making decisions and solving problems*, but these differences are not significant.

[41]As described in Section 3.2, participants are asked to choose between a job in which they spend less than a cutoff number of hours, C, on the focal task, and a job in which they spend C or more hours on the focal task; the value of C depends on whether participant is assigned to the extensive or the intensive margin cutoffs.

who are not currently employed, respectively. I measure inattention by asking participants at the end of the survey to indicate the decisions they made in the hypothetical choice experiment for a randomly selected two of the five gender-typical tasks. I consider a participant to be inattentive if they answer either question incorrectly.[42] Results among attentive and employed participants are similar to baseline estimates in Table 4.

5.4 External Validity

If workers sort according to preferences for tasks, then individuals working in jobs that involve more time spent on a task are predicted to have a higher WTP for that activity. To test this hypothesis, Table 6 shows WTP for participants currently working in a high-task job (Panel A) versus those currently working in a low-task job (Panel B), with the difference between the estimates in Panel C. I designate participants as working in a high-task k job if the number of hours they report spending on task k in their current or most recent job is greater than the cutoff number of hours in the hypothetical scenario they face for that task.

Consistent with the prediction, WTP estimates for those currently in high-task jobs are higher than estimates for those in low-task jobs, for all tasks and both genders.[43] For example, workers who currently spend more *working and communicating with others* are willing to pay 7.6 percentage points more for this task than those who currently spend less time on this activity. In addition, overall WTP for *helping and caring for others* and *making decisions and solving problems* is positive (approximately 1.5 percent) and statistically significant among those currently spending more time on these tasks. These results indicate that WTP estimates are correlated with equilibrium labor market outcomes, suggesting that the hypothetical choice experiment measures task valuations with real-world relevance.

The qualitative pattern of gender differences in WTP in Table 6 matches the baseline results in Table 4. However, in some cases the differences are smaller in magnitude and lose significance.[44] This finding is not surprising, as one might expect gender differences to shrink or even disappear among workers who make the same sorting decisions and therefore are likely to have more homogeneous preferences, compared with the overall population.

The experiment sample is not statistically representative of the broader U.S. population, and Section 5.1 documents that the distributions of race, education, occupation and industry differ meaningfully between the experiment sample and the ACS. To assess how this selection

[42] I find that 18 percent of participants are inattentive using this definition.

[43] The differences in WTP between those in high-task and low-task jobs are statistically significant for all tasks when pooling women and men, and for nearly all tasks when considering women and men separately.

[44] Specifically, the gender difference in WTP for *helping and caring for others* is insignificant in both the high-task and low-task sub-samples, and the gender difference in *documenting and recording information* is insignificant in the low-task sub-sample.

may affect results, in Table 7 I repeat the WTP analysis, weighting the sample to match currently employed workers in the ACS by gender, race (White versus non-White), college degree receipt and major occupation. Results are similar to the unweighted estimates, although the gender difference in WTP for *helping and caring for others* is somewhat larger in magnitude (7.0 percentage points compared with 3.3 percentage points).

5.5 Interpretation

The goal of the experiment is to ensure that participants make choices only on the basis of preferences over pay and gender-typical tasks. Therefore, wages are randomly assigned and participants are instructed to assume that jobs are the same except for explicitly stated differences. However, it is possible that WTP estimates may still reflect task-specific skills, concerns about discrimination or preferences for job characteristics that participants view as correlated with tasks.

To investigate this possibility, at the end of the survey I ask participants about the motivations for their choices in a randomly selected two out of the five hypothetical scenarios. Table 8 displays responses by all participants (Column 1) and participants choosing the high- and low-task jobs (Columns 2 and 3, respectively), pooling across tasks.[45]

The two most common responses aside from the job offering better pay (cited by nearly 60 percent of participants) are that the chosen job sounds "more enjoyable/interesting," and that the chosen job would be a "better fit for my existing skills and abilities," each cited by 34 percent of participants. I interpret these responses to reflect choice motivations related to current preferences and skills, respectively. In addition, 18 percent of participants indicate that the chosen job would "allow me to strengthen or develop new skills," which suggests an investment motivation. The preference response has a correlation coefficient of 0.31 and 0.27 with the responses related to existing and new skills, respectively.

These results are consistent with the hypothesis that task-specific preferences and skills are correlated or even jointly determined. Workers may find it more interesting or enjoyable to perform tasks in which they have a productivity advantage, and may also invest in developing skills relevant to tasks they enjoy performing. I cannot provide insight into the process of task-specific preference and skill formation. To the extent that task-specific skills have a causal impact on preferences, I interpret this effect as a component of the preference parameter that I estimate.

Another possibility is that WTP estimates reflect career concerns related to task-specific skills. Specifically, participants may be willing to accept lower wages in a job in which they

[45] Figure A5 shows the wording of the response options available to participants. I randomize the order of responses and allow multiple entries.

have a comparative advantage because they believe they are more likely to be promoted or less likely to be terminated in that position. In addition, participants may be willing to pay to develop competencies that they believe will lead to higher wages in the future. While career concerns are likely to be small in a hypothetical setting, I cannot rule them out. However, Tables A14 and A15 show that WTP results are similar when restricting the sample to participants who do not cite a better fit for existing skills or developing new skills, respectively, as a motivation for their choice.

Table 8 also indicates that 23 percent of participants say that the chosen job would "require less effort." This answer is much more common among those choosing a low-task job (36 percent) than among those choosing a high-task job (8 percent),[46] suggesting that participants do indeed view jobs involving more time spent on a gender-typical task as requiring greater effort. This finding provides evidence that participants are focused on the difference in the amount of time spent on the focal task between the two scenarios, rather than the time spent on other activities. Therefore, it seems unlikely that gender differences in the counterfactual activities that participants envision are driving results.[47]

Finally, only between 5 and 10 percent of participants cite reasons for their choices related to identity (the chosen job would have "more people like me"), discrimination (participants would be "treated better" in the chosen job), and prestige. Therefore, it does not seem that beliefs about discrimination or gender identity are major factors affecting choices.

6 Implications for Gender Gaps

In this section, I examine the implications of the WTP estimates for gender gaps in the labor market. I focus on gender differences in job sorting on the five gender-typical tasks, occupational segregation, and the gender wage gap, using data from the experiment sample and the ACS.

[46]By contrast, all other response options are selected more often by participants choosing the high-task job.

[47]As an additional robustness check, in Table A16 I report estimates of the difference in differences in WTP across gender and tasks. If participants envision the same job involving the same set of activities for all scenarios, then the counterfactual tasks imagined will difference out in the comparison between tasks. The gender differences in WTP for all female-typical tasks compared with *operating and repairing equipment* are statistically significant, as is the difference between *helping and caring for others* and *making decisions and solving problems* and between *documenting and recording information* and *making decisions and solving problems.*

6.1 Observed Sorting and Segregation

I begin by documenting observed gender differences in sorting on the five gender-typical tasks in the experiment sample and the ACS. These observed differences provide a baseline against which I compare the gender differences in sorted that are predicted by the preference estimates from the experiment.

I define the gender difference in sorting on task k to be

$$Q_k \equiv p_{fk} - p_{mk} \tag{7}$$

where p_{gk} is the share of workers of gender $g \in \{f, m\}$ employed in a job that involves a high level of task k.

In the experiment sample, I consider workers to have a job with a high level of task k if they report that in their current or most recent job, they spend more time on task k than the cutoff number of hours in the hypothetical scenario for task k that they are shown. In the ACS, I use the occupation-level O*NET measures to classify jobs as high task or low task, as I have no data on the frequency of task performance. Specifically, I select a cutoff percentile for each task such that the share of workers in the ACS above the cutoff matches the share of workers in a high-task job in the experiment sample, and consider occupations to be high task if they fall above this percentile.[48]

The first column of Panel A in Table 10 reveals substantial gender differences in sorting (Q_k) in the experiment sample, as suggested by the gender-specific distributions of time spent on tasks in Figure 7. Women are 15.4, 10.5 and 8.0 percentage points more likely than men to work in jobs that involve a high level of the female-typical tasks of *helping and caring for others*, *documenting and recording information*, and *working and communicating with others*, respectively. By contrast women are 18 percentage points less likely than men to spend more time *operating and repairing equipment*. The gender gap in sorting on *making decisions and solving problems* is close to zero.[49]

The first column of Panel B in Table 10 shows a similar pattern of results in the ACS. Women are approximately 27.0, 14.1 and 25.9 percentage points more likely than men to work in jobs that involve high levels of *helping and caring for others*, *documenting and recording information*, and *working and communicating with others*, respectively, and 30.5

[48] The analyses in this section restrict the experiment sample to all participants with valid wages and information on race, ethnicity and education. The ACS sample used in this section includes all individuals in the 2012-2016 data who are aged 18 and older, currently employed, have valid non-zero earnings, and work in an occupation that can be matched to the O*NET. Appendix A1 provides further detail on the O*NET measures.

[49] In fact, women are approximately 1 percentage point more likely to be in a job with a high level of *making decisions and solving problems*, contrary to the hypothesis that this activity is male-typical.

percentage points less likely than men to hold a job with a high level of *operating and repairing equipment*.[50] These differences are somewhat larger in magnitude than those in the experiment sample, which is not surprising given that the O*NET measures used to construct the gender-typical tasks are selected on the basis of their robust correlation with the occupational female share.

I also calculate a task-based segregation index as in Duncan and Duncan (1955) that summarizes the gender differences in sorting on all tasks with a single statistic. Consider a set of job categories categories, indexed by $j \in \{1, \ldots, J\}$, where each category is defined by having a high or low level of each gender-typical task, such that $J = 2^5$. The index is then given by

$$I \equiv \frac{1}{2} \sum_{j=1}^{J} |p_{fj} - p_{mj}] \tag{8}$$

where p_{gj} is the proportion of workers of gender g sorting into job category j. The index is scaled such that a value of 0 indicates gender equality, while a value of 1 reflects total segregation.

The second column of Table 10 reports a value of 0.262 for the task-based segregation index in the experiment sample in Panel A, and a value of 0.389 in the ACS in Panel B. By comparison, Blau et al. (2013) calculate a gender segregation index in the 2009 ACS using the 2000 Census occupation codes, which include approximately 500 categories, and report a value of 0.510.[51] Thus more than three quarters (76 percent) of occupational segregation by gender can be accounted for by gender differences in sorting on 32 occupation categories defined by gender-typical tasks. This comparison provides a baseline estimate of how much preferences over the tasks I examine contribute to segregation under the assumption that observed gender differences in sorting on these tasks are entirely due to preferences.

6.2 Predicted Sorting and Segregation

Next, I calculate predicted gender differences in sorting on the gender-typical tasks, using the model and estimated preference parameters from the hypothetical choice experiment and wage differentials associated with the tasks in the experiment sample and ACS.

For the estimated preference parameters, I use coefficients from a pooled specification in which I stack the choice data from the experiment for all tasks and both genders and run a

[50] Again, the gender difference in sorting on *making decisions and solving problems* is close to zero; women are 1.1 percentage points less likely to work in jobs with a high level of this task.

[51] I calculate an occupational segregation index using currently employed individuals aged 18 and older in the 2012-2016 ACS data, and find a value of 0.497 using the full set of 478 occupation codes that appear in the data and a value of 0.498 using the 464 occupation codes that can be matched to the O*NET.

single logistic regression, clustering standard errors by participant. An observation in this regression is at the participant-by-scenario level, the outcome is an indicator for choosing the high-task job in that scenario, and the predictors are the log difference in wage offers between the high-task and low-task jobs in the scenario (ω_{ik}) and a set of indicators for the scenario relating to task k (T_{jk} for $k \in \{1, \ldots 5\}$). I allow the $\hat{\theta}_{gk}$ task coefficients to differ by gender, but estimate a single $\hat{\delta}$ coefficient for the entire sample.

Estimating a $\hat{\delta}$ coefficient that does not vary by task k allows me to calculate predicted probabilities of sorting into the job categories $j \in \{1, \ldots, 32\}$, which are necessary for calculating a predicted segregation index. Furthermore, I restrict $\hat{\delta}$ to be the same for women and men to ensure that predicted gender differences in sorting are driven only by estimated differences in preferences for tasks, rather than any gender difference in the coefficient on the wage differential.[52]

Table 9 shows WTP estimates based on the pooled specification. Results are very similar to the WTP estimates based on task-specific regressions and gender- and task-specific $\hat{\delta}_{gk}$ coefficients in Table 4.[53]

Using the $\hat{\delta}$ and $\hat{\theta}_{gk}$ coefficients from the pooled specification reported in Table 9 and the logistic parameterization, I calculate the share of workers of gender $g \in \{f, m\}$ who are predicted to sort into a high-task k job as a function of ω_k, the log difference in wage offers between high-task k and low-task k jobs, as in (4):

$$\hat{\mathrm{H}}_{gk}(\omega_k) \equiv \frac{1}{1 + \exp(-(\hat{\theta}_{gk} + \hat{\delta}\omega_k))}. \tag{10}$$

The predicted gender difference in sorting on task k is then

$$\hat{Q}_k \equiv \hat{\mathrm{H}}_{fk}(\hat{\omega}_k) - \hat{\mathrm{H}}_{mk}(\hat{\omega}_k) \tag{11}$$

[52] This specification can be motivated by a modification of (1) in which the utility of person i of gender $g(i) \in \{f, m\}$ from job $j \in \{1, \ldots, 32\}$ related to scenario c is given by

$$U_{ijc} = \alpha_i + \sum_{k=1}^{5} \theta_{g(i)k} T_{jk} + \delta \ln w_{ijc} + \varepsilon_{ijc} \tag{9}$$

where w_{ijc} is the wage offer for participant i in job j related to scenario c, $\theta_{g(i)k}$ is a task- and gender-specific preference parameter, δ is a preference parameter not indexed by task or gender, ε_{ijc} is a worker- , job- and scenario-specific preference parameter with a standard EV Type I distribution, and the other terms are as defined in (1). Note that in contrast to (1), in which $j \in \{1, 2\}$, here j indexes the job categories defined in Section 6.1.

[53] I also find similar results when I estimate WTP based on task-specific regressions in which the $\hat{\delta}_k$ coefficient differs by task but is restricted to be the same for women and men, and a specification in which all tasks are pooled but the $\hat{\delta}_g$ coefficient is allowed to differ by gender.

where $\hat{\omega}_k$ is an estimate of ω_k.

To generate $\hat{\omega}_k$, I estimate earnings equations of the form

$$\ln w_i = X_i'\beta + \xi_i \tag{12}$$

in the experiment sample and the ACS, where $\ln w_i$ is the log hourly wage of individual i, X_i is a vector of individual i's observable characteristics, and ξ_i is an individual-specific residual with expectation zero. The X_i vector includes a set of indicators for i's current job involving high levels of the gender-typical tasks, as well as controls for gender, race and ethnicity, education, potential experience and, in the ACS only, geography and year.[54] My estimate of ω_k is then

$$\hat{\omega}_k \equiv b_k \tag{13}$$

where b_k is the coefficient on the task k indicator from the OLS estimation of (12).

I interpret the b_k coefficients (termed *task wage differentials*) as equilibrium wage differentials associated with the gender-typical tasks. These differentials may reflect preferences for tasks as well as preferences for other job amenities correlated with tasks that affect labor supply decisions, producing compensating differences.[55] Search frictions and departures from perfect competition such as monopsony and bargaining may also contribute to the differentials.

The third column of Table 10 reports the task wage differentials (b_k) in the experiment sample (Panel A) and the ACS (Panel B). In both data sources, the female-typical task of *documenting and recording information* and the male-typical task of *making decisions and solving problems* are associated with substantial wage premiums (0.133 and 0.222, respectively, in the ACS), while the female-typical *helping and caring for others* offers a large wage penalty (−0.191 in the ACS). The wage differentials associated with *working and communicating with others* and *operating and repairing equipment*, by contrast, are quite small in

[54] I exclude hourly wage observations that are less than \$3 or greater than \$200 in 2018 dollars. In the ACS, I measure hourly wages as annual earnings divided by annual hours worked, inflate wages to 2018 dollars, and weight the regression by the ACS person weight. The race and ethnicity variables consist of mutually exclusive indicators for Black, Hispanic, and other race, with White non-Hispanic as the omitted category. The education variables consist of indicators for high school diploma or less in the experiment sample (less than a high school diploma and high school diploma in the ACS), some college, associate degree and graduate degree, with bachelor's degree as the omitted category. Potential experience is measured as age minus years of education minus six, restricted to be greater than or equal to zero, and the regression also includes the square of potential experience. In the ACS only, the geography variables consist of indicators for region and for metropolitan area status.

[55] In Rosen (1986), firms have a distribution of costs associated with offering an amenity, and the equilibrium wage differential equates the share of workers willing to pay that differential with the share of firms with a cost of offering the amenity that is equal to or less than the differential. By contrast, I view tasks as inherent to production in certain firms, but an amenity from the perspective of workers.

magnitude in both samples.

Column 4 reveals non-trivial predicted gender differences in sorting ($\hat{Q}_k$) based on the task wage differentials and preference estimates. In the experiment sample (Panel A), women are predicted to be 5.8, 6.0 and 2.7 percentage points more likely than men to work in jobs with high levels of the female-typical tasks of *helping and caring for others*, *documenting and recording information* and *working and communicating with others*, respectively. Column 5 shows that these predicted differences can account for 38, 57 and 33 percent, respectively, of observed sorting differences. In addition, women are predicted to be 16.7 percentage points less likely than men to work in jobs with a high level of the male-typical task of *operating and repairing equipment*, accounting for 94 percent of the observed sorting difference.[56]

In the ACS (Panel B), women are predicted to be 2.4, 5.4 and 2.7 percentage points more likely than men to work in jobs with high levels of *helping and caring for others*, *documenting and recording information*, and *working and communicating with others*, respectively, accounting for 9, 38 and 10 percent of observed differences in sorting. Women are also predicted to be 14.3 percentage points less likely than men to work in jobs with a high level of *operating and repairing equipment*, accounting for 47 percent of the observed difference.[57]

The predicted gender differences in sorting on *helping and caring for others* and *documenting and recording information* are somewhat smaller in magnitude in the ACS compared with the experiment sample because the task wage differentials in the ACS are farther from the overall mean WTP, where the gender difference in sorting is approximately maximized. Combined with the fact that observed sorting differences are larger in the ACS, these smaller predicted sorting differences result in preferences accounting for a lower share of actual sorting in the ACS compared with the experiment sample. It is not surprising that the WTP estimates are more predictive of sorting behavior in the sample used to elicit preferences than in the overall population.

I also calculate a value for the task-based segregation index, using the estimated preference parameters to predict gender differences in sorting into the task-based job categories. Specifically, I assume that worker i of gender g's choice of job category has a multinomial logistic distribution, and that the probability of sorting into category j can be expressed as

[56] Women are also predicted to be 1.1 percentage points less likely to work in jobs with a high level of *making decisions and solving problems*, although they are slightly more likely to report actually working in jobs involving high levels of this activity.

[57] Finally, women are predicted to be 0.7 percentage points less likely to work in jobs with a high level of *making decisions and solving problems*, explaining 60 percent of the small observed gender difference.

a function of ω_j, the log wage in job category j, as follows:

$$\hat{h}_{gj}(\omega_j) \equiv \frac{\exp(\hat{V}_{gj})}{\sum_{j'=1}^{J}\exp(\hat{V}_{gj'})} \tag{14}$$

where $\hat{V}_{gj} = \sum_{k=1}^{5}\hat{\theta}_{gk}T_{jk} + \hat{\delta}\omega_j$.[58] I estimate ω_j as the sum of the wage differentials for tasks that have a high level in job category j:

$$\hat{\omega}_j = \sum_{k=1}^{5} b_k T_{jk}. \tag{16}$$

Thus the value of the task-based segregation index based on predicted sorting is given by

$$\hat{I} = \frac{1}{2}\sum_{j=1}^{J}|\hat{h}_{fj}(\hat{\omega}_j) - \hat{h}_{mj}(\hat{\omega}_j)|. \tag{17}$$

The predicted segregation index has a value of 0.179 using wage differentials from the experiment sample, which is 68 percent of the magnitude of the observed segregation index calculated in that sample. In the ACS, the predicted segregation index has a value of 0.158, which is approximately 41 percent of the index value of 0.389 based on observed sorting and about 31 percent of the index value reported by Blau et al. (2013). This result suggests that preferences for the gender-typical tasks I examine can account for a degree of segregation that is nearly a third of overall occupational segregation by gender in the U.S. labor market.

6.3 Robustness Checks for Sorting and Segregation

The results in Table 10 suggest that gender differences in sorting on the gender-typical tasks are substantial, both in the experiment sample and the ACS, and that preferences for these tasks can account for a meaningful share of the observed sorting differences and overall occupational segregation by gender.

[58] The multinomial logistic parameterization can be motivated by a model in which the utility of person i of gender $g(i) \in \{f, m\}$ from job category $j \in \{1, \ldots, 32\}$ is given by

$$U_{ij} = \alpha_i + \sum_{k=1}^{5}\theta_{g(i)k}T_{jk} + \delta\omega_j + \varepsilon_{ij} \tag{15}$$

where ε_{ij} is a worker- and job-specific preference parameter with a standard EV Type I distribution. See Cameron and Trivedi (2005) for a derivation. This model differs from (9) in that ε_{ij} is indexed only by i and job category j, and not by scenario c. However, note that the predicted probability of sorting into a high-task k job ($\hat{Q}_k$) given by (10) and (11) is equal to the probability calculated by summing $\hat{h}_{gj}(\hat{\omega}_j)$ over all job categories j that have a high level of task k.

This qualitative pattern of results is robust to a variety of alternative specifications, including restricting the sample by age range, restricting the sample to full-time workers, repeating the analysis separately for individuals with and without a college degree, and adding controls for major industry category into the wage equation in (12) for the ACS sample.[59] Findings are also comparable when I define the indicators for participants being in a high-task job using the extensive or intensive margin cutoffs, rather than the cutoff that each participant faces.[60]

A potential limitation of the ACS is that hourly wages must be constructed by dividing annual earnings by annual hours worked, increasing the likelihood of measurement error.[61] Therefore, I repeat the analysis using the Current Population Survey Merged Outgoing Rotation Group (CPS MORG) data, which includes information on hourly wages for workers paid by the hour, and weekly earnings for others. Panel A of Table A17 documents patterns of gender differences in observed and predicted sorting in the CPS MORG data that are similar to results in the ACS.[62]

One concern is that the task wage differentials may reflect preferences for other job amenities that are systematically bundled with the gender-typical tasks.[63] Panel B of Table A17 repeats the analysis in the ACS including a set of measures for amenities that the literature suggests may be differentially valued by women compared with men.[64] The b_k

[59]Specifically, results in the ACS are very similar when restricting to workers aged 18 to 64 and excluding those who are self-employed, in institutional group quarters, or in the military, and when restricting to workers aged 25-64 who are employed full-time, defined as working at least 35 hours per week and 27 weeks per year. Results are qualitatively similar for college and non-college workers, but the magnitude of the observed and predicted gender differences in sorting are much larger for the less educated group, especially in the ACS, consistent with the fact that occupational segregation is greater among non-college workers (Blau et al. 2013).

[60]In these versions of the analysis, I also use preference parameters estimated based on the extensive or intensive margin sub-samples, respectively.

[61]Furthermore, annual weeks worked is constructed by multiplying usual hours of work per week by weeks worked in the previous year. Beginning in 2008, ACS respondents report weeks worked within intervals; therefore, the precise number of weeks worked is imputed using the 2005-2007 ACS, introducing additional measurement error and potential for bias.

[62]See Appendix A1 for further details on sample and variable construction in the CPS MORG data.

[63]The fact that the tasks may be correlated with other job amenities does not necessarily imply that the task wage differentials are biased, but it does suggest that they cannot be interpreted as compensating differentials for the tasks alone, and may also explain the disparity between predicted and observed sorting levels for the overall sample. For example, 52 percent of workers in the experiment sample report working in jobs with a high level of *helping and caring for others*, although this task is associated with a wage penalty of -0.089 and the WTP estimates imply that workers must be compensated an additional 2.6 percent, on average, to be willing to spend more time on this activity. One potential explanation for this finding is that jobs with a high level of *helping and caring for others* have other amenities that workers value.

[64]Specifically, the specification includes mean log hours worked per week by occupation, which Denning et al. (2019) find can account for a large portion of the gender wage gap, a set of five O*NET measures that Goldin (2014) uses as a proxy for time flexibility (the work activity *establishing and maintaining interpersonal relationships* and the work context variables *time pressure*, *contact with others*, *structured versus unstructured*

coefficients do indeed decrease in magnitude, but results on predicted sorting and segregation are similar to the baseline ACS specification in Table 10.[65]

Another concern is that the wage differentials may be correlated with unobserved human capital. A large literature on compensating differentials has noted that estimates tend to be "wrong-signed," i.e. amenities that are expected to be considered desirable by most or all workers are associated with wage premiums, suggesting that individuals in jobs with these attributes have unobservably higher skills (e.g. Brown 1980). Tasks are not a clear amenity or disamenity, so ability bias may be less of a concern in this context. Nevertheless, to investigate this possibility I repeat the analysis using data from the Panel Study of Income Dynamics (PSID), which includes information on actual work experience, and the National Longitudinal Surveys of Youth 1979 and 1997 (NLSY79 and NLSY97), which can be used to construct measures of cognitive, non-cognitive and social skills.[66] Panels C, D and E of Table A17 show that patterns of observed and predicted gender differences in sorting and segregation in the PSID, NLSY79 and NLSY97, respectively, are qualitatively similar to those in the ACS and experiment sample.

Thus incorporating additional measures of general human capital into the analysis does not appear to meaningfully change the pattern of findings. I cannot rule out bias from task-specific human capital.[67] However, it is important to note that the model predicts some gender difference in sorting regardless of the exact value of the estimate for ω_k. Figure A6 plots predicted gender differences in sorting by task as a function of ω_k, using the $\hat{\delta}$ and $\hat{\theta}_{gk}$ coefficients.[68] It is clear that the estimated preference parameters imply meaningful sorting differences for the three tasks with significant gender differences in WTP for a wide range of values of ω_k.

---

work, and *freedom to make decisions*), and the O*NET work context variable *level of competition*. The O*NET measures are standardized to have a mean of zero and standard deviation of one.

[65] Indeed, the predicted gender differences in sorting on *helping and caring for others* and *documenting and recording information* are larger compared with the baseline specification.

[66] I follow Blau and Kahn (2017) in constructing a measure of actual experience in the PSID. In the NLSY79 and NLSY97, I use scores on the Armed Forces Qualifying Test (AFQT) to capture cognitive skill, and follow the methodology of Deming (2017) to create measures for non-cognitive and social skill. See Appendix A1 for further details on sample and variable construction in the PSID, NLSY79 and NLSY97.

[67] Workers can generally be expected to sort into jobs based on comparative advantage in performing the activities required, as in Roy (1951). As discussed above, task-specific preferences and skills are likely to be correlated, and workers sorting into a high-task k job may therefore have both a high WTP and a comparative aptitude for task k. Without further assumptions, however, it is not clear how task-based sorting on comparative advantage is likely to affect the task wage differentials.

[68] The vertical lines show the estimated task wage differentials ($\hat{\omega}_k = b_k$) in the ACS from Table 11.

6.4 Gender Wage Gap

Lastly, I assess the contribution of observed and predicted gender differences in sorting to the gender wage gap. The female-to-male log hourly wage gap explained by observed gender differences in sorting on task k is given by

$$\mathbb{WG}_{expl,k} \equiv Q_k b_k. \tag{18}$$

The total wage gap explained by the five gender-typical tasks is the sum of the task-specific gaps:

$$\mathbb{WG}_{expl,tot} \equiv \sum_{k=1}^{5} Q_k b_k. \tag{19}$$

The task-specific and total gender wage gaps that can be explained by predicted gender differences in sorting simply replace Q_k in (18) and (19) with $\hat{Q}_k$, as follows:

$$\hat{\mathbb{WG}}_{expl,k} \equiv \hat{Q}_k b_k \tag{20}$$

$$\hat{\mathbb{WG}}_{expl,tot} \equiv \sum_{k=1}^{5} \hat{Q}_k b_k. \tag{21}$$

For female-typical tasks, I expect that $Q_k > 0$, so that a wage premium associated with task k ($b_k > 0$) explains a positive differential that serves to narrow the female-to-male wage gap, while a wage penalty ($b_k < 0$) explains a negative differential that serves to widen the wage gap. Similarly, for male-typical tasks, I expect that $Q_k < 0$, so that $b_k < 0$ ($b_k > 0$) explains a positive (negative) wage gap. Note that if the sign of $Q_k b_k$ differs across tasks, these gaps may partially or fully offset each other.

Table 11 displays results on the gender wage gap in the experiment sample (Panel A) and the ACS (Panel B). The first three columns repeat the estimates of observed gender differences in sorting (Q_k), the task wage differentials (b_k) and predicted gender differences in sorting ($\hat{Q}_k$), for reference. Column 4 shows that some of the task-specific wage gaps based on observed sorting ($Q_k b_k$) are non-trivial in magnitude. In particular, the substantial wage penalty associated with *helping and caring for others* implies that this task can explain a negative wage gap of -0.014 in the experiment sample and -0.052 in the ACS. By contrast, the wage premium associated with *documenting and recording information* implies that this task explains a positive, and thus "wrong-signed," wage gap of 0.012 in the experiment sample and 0.019 in the ACS.[69] However, this positive and negative gap largely cancel out, leading

[69] In both samples, the fact that the task wage differentials for *working and communicating with others* and *operating and repairing equipment* are quite small in magnitude means that these tasks contribute relatively little to the wage gap despite substantial gender differences in sorting. Conversely, *making decisions and*

to a total explained wage gap that is close to zero (-0.003) in the experiment sample and negative but modest in size (-0.017) in the ACS. By contrast, I document sizable raw wage gaps of -0.199 and -0.183 in the experiment sample and the ACS, respectively. Column 5 reveals a similar pattern based on predicted gender differences in sorting. The estimated preference differences over the tasks I examine can explain gender wage gaps of close to zero (-0.002 in the experiment sample and 0.005 in the ACS).

These findings suggest that preferences for job tasks do not contribute meaningfully to the gender wage gap. However, I only examine five task categories, and given that certain tasks tend to widen the gap while others tend to narrow it, the sum of the task-specific gaps is inherently sensitive to the choices of which activities to include in the analysis. Results are also likely to depend on the precise values of the estimated task wage differentials and how tasks are measured. In particular, I elicit preferences over the binary options of spending more versus less time on a task, and it may be that valuations of time spent on tasks measured more finely may contribute more to the gender wage gap.

To illustrate this possibility, I estimate a series of decompositions of the gender wage gap, as follows:

$$\bar{w}_f - \bar{w}_m = \underbrace{(\bar{X}_f - \bar{X}_m)'b}_{\textit{Explained}} + \underbrace{b_{fem}}_{\textit{Residual}} \tag{22}$$

where $\bar{w}_g$ is the mean log hourly wage for workers of gender $g \in \{f, m\}$, $\bar{X}_g$ is a vector of the mean levels for gender g of the observable characteristics in X_i from (12), with the exception of the control for gender, b is a vector of coefficients on X_i from the OLS estimation of (12), and b_{fem} is the coefficient on the female indicator from this regression. The first component of the gap is the portion *explained* by gender differences in mean observable characteristics, sometimes called the endowments effect, while the second component is a *residual* portion.[70]

Table 12 displays the results of this decomposition in the ACS. Column 1 reports the total, residual and explained wage gap based on a vector of observable characteristics that includes race and ethnicity, education, potential experience, geography, and year. These controls explain a positive female-to-male wage gap of 0.033, despite the total raw wage gap of -0.183, implying a residual gap of -0.216. In other words, women are predicted to earn about 3 percent more than men on the basis of their level of education and other basic observable characteristics. The specification in Column 2 adds the full set of occupation

solving problems does not contribute meaningfully to the gender wage gap because sorting differences are small, although this task provides a large wage premium.

[70] Note that (22) is a version of a Oaxaca-Blinder decomposition (Blinder 1973; Oaxaca 1973) where the coefficients used to weight the difference in observable characteristics are from a pooled regression, rather than a specification including only women or only men, as is commonly the case.

codes available in the ACS into the X_i vector.[71] The occupation indicators explain a gender wage gap of -0.056, about 31 percent of the total gap, and reduce the residual gap by 34 percent, to -0.142, compared with Column 1. It is clear that choice of occupation plays a meaningful role in the gender wage gap.

In Column 3, I replace the occupation indicators with the five gender-typical tasks, measured as continuous variables standardized to have a mean of zero and standard deviation of one. These continuous task variables can explain a gender wage gap of -0.034, about 19 percent of the raw gap, and reduce the residual gap by about 21 percent compared with Column 1. Finally, Column 4 includes the binary task measures that I use in the sorting and wage gap analyses in Tables 10 and 11. As already noted, these indicators can explain a gender wage gap of -0.017, approximately 9 percent of the total gap, and they reduce the residual portion of the gap by approximately 12 percent, from -0.216 to -0.190.

Thus while the continuous task measures capture more than 60 percent of the explanatory power of the full set of occupation controls, the binary variables capture only 30 percent. These decomposition results suggest that the contribution of preferences for job tasks to the gender wage gap remains an open question.

7 Conclusion

This paper examines gender differences in preferences for job tasks and the implications for gender gaps in the labor market. I estimate WTP for a set of gender-typical tasks performed more frequently by one gender than the other, using a hypothetical choice experiment embedded in a survey. I find that women have a significantly higher WTP than men for the female-typical tasks of *helping and caring for others* and *documenting and recording information*, and a significantly lower WTP for the male-typical task of *operating and repairing equipment*. WTP is higher for participants who currently spend more time on the focal tasks, suggesting that the estimates are correlated with labor market outcomes.

I find that the estimated preference parameters can explain a meaningful portion of observed gender differences in sorting on the gender-typical tasks, in both the experiment sample and the ACS. Indeed, gender differences in WTP for tasks can account for approximately a third of overall occupational segregation by gender in the ACS. By contrast, preferences for the gender-typical tasks I examine appear to explain little of the gender wage gap, but this finding may be sensitive to which tasks are selected and how they are measured.

These results suggest that gender differences in valuations of work activities have a role

[71] There are 464 codes in the 2012-2016 ACS that match to the O*NET and appear in the analysis sample for the decomposition.

to play in explaining sorting patterns in the labor market, as hypothesized. Task-specific skills and discrimination may also be important, however, and I am not able to examine how preference formation may be related to these other factors. Further research is needed to understand the extent to which current task valuations are caused by inherent gender differences in preferences rather than past resource constraints and gender norms.

References

Abraham, Katharine and Ashley Amaya, "Probing for informal work activity," Working Paper 24880, National Bureau of Economic Research, Cambridge, MA 2018.

Acemoglu, Daron and David Autor, "Skills, tasks and technologies: Implications for employment and earnings," in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4B, Elsevier, 2011, pp. 1043–1171.

Akerlof, George and Rachel Kranton, "Economics and identity," *Quarterly Journal of Economics*, 2000, *115* (3), 715–753.

Altonji, Joseph G. and Rebecca M. Blank, "Race and gender in the labor market," in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3, Elsevier, 1999, pp. 3143–3259.

Altonji, Joseph, Prashant Bharadwaj, and Fabian Lange, "Changes in the characteristics of American youth: Implications for adult outcomes," *Journal of Labor Economics*, 2012, *30* (4), 783–828.

Autor, David H and David Dorn, "The growth of low-skill service jobs and the polarization of the US labor market," *American Economic Review*, 2013, *103* (5), 1553–1597.

Bacolod, Marigee P. and Bernardo S. Blum, "Two sides of the same coin: U.S. "residual" inequality and the gender gap," *Journal of Human Resources*, 2010, *45* (1), 197–242.

Baker, Michael and Kirsten Cornelson, "Gender-based occupational segregation and sex differences in sensory, motor, and spatial aptitudes," *Demography*, 2018, *55* (5), 1749–1775.

Beaudry, Paul and Ethan Lewis, "Do male-female wage differentials reflect differences in the return to skill? Cross-city evidence from 1980-2000," *American Economic Journal: Applied Economics*, 2014, *6* (2), 178–194.

Bell, Alex, "Job amenities & earnings inequality," Job Market Paper, Harvard University 2019.

Bertrand, Marianne, "New perspectives on gender," in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4B, Elsevier, 2011, pp. 1543–1590.

Black, Sandra and Alexandra Spitz-Oener, "Explaining women's success: Technological change and the skill content of women's work," *Review of Economics and Statistics*, 2010, *92* (1), 187–194.

Blau, Francine D. and Lawrence M. Kahn, "The gender wage gap: Extent, trends, and explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

Blau, Francine, Peter Brummund, and Albert Liu, "Trends in occupation segregation by gender 1970-2009: Adjusting for the impact of changes in the occupational coding system," *Demography*, 2013, *50* (471-492).

Blinder, A. S., "Wage discrimination: Reduced form and structural estimates," *Journal of Human Resources*, 1973, *4*, 436–455.

Borghans, Lex, Bas ter Weel, and Bruce A Weinberg, "People skills and the labor-market outcomes of underrepresented groups," *Industrial & Labor Relations Review*, 2014, *67* (2), 287–334.

Breiman, Leo, "Random forests," *Machine Learning*, 2001, *45* (5-32).

Brown, Charles, "Equalizing differences in the labor market," *Quarterly Journal of Economics*, 1980, *94* (1), 113–134.

Buhrmester, Michael, Tracy Kwang, and Samuel Gosling, "Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?," *Perspectives on Psychological Science*, 2011, *6* (1), 3–5.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics*, Cambridge, UK: Cambridge University Press, 2005.

Coffman, Katherine Baldiga, "Evidence on self-stereotyping and the contribution of ideas," *Quarterly Journal of Economics*, 2014, *129* (4), 1625–1660.

Cortes, Guido Matias, Nir Jaimovich, and Henry Siu, "The end of men and the rise of women in the high-skilled labor market," Working Paper 24274, National Bureau of Economic Research, Cambridge, MA 2018.

Cortes, Patricia and Jessica Pan, "Occupation and gender," in Susan L. Averett, Laura M. Argys, and Saul D. Hoffman, eds., *Oxford Handbook of Women and the Economy*, Oxford University Press, 2018.

_ **and** _ , "When time binds: Substitutes for household production, returns to working long hours, and the skilled gender wage gap," *Journal of Labor Economics*, 2019, *37* (2), 351–398.

Croson, Rachel and Uri Gneezy, "Gender differences in preferences," *Journal of Economic Literature*, 2009, *47* (2), 448–474.

Datta, Nikhil, "Willing to pay for security: A discrete choice experiment to analyse labour supply preferences," Working Paper 1632, Centre for Economic Performance, London School of Economics and Political Science, London 2019.

Deming, David J., "The growing importance of social skills in the labor market," *Quarterly Journal of Economics*, 2017, *132* (4), 1593–1640.

Denning, Jeffrey, Brian Jacob, Lars Lefgren, and Christian vom Lehn, "The return to hours worked within and across occupations: Implications for the gender wage gap," Working Paper 25739, National Bureau of Economic Research, Cambridge, MA 2019.

England, Paula, "Emerging theories of care work," *Annual Review of Sociology*, 2005, *31*, 381–399.

Flory, Jeffrey, Andreas Leibbrandt, and John List, "Do competitive workplaces deter female workers? A large-scale natural field experiment on job-entry decisions," *Review of Economic Studies*, 2015, *82* (1), 122–155.

Folbre, Nancy, ed., *For Love and Money: Care Provision in the United States*, New York: Russell Sage Foundation, 2012.

Fortin, Nicole, "The gender wage gap among young adults in the United States: The importance of money versus people," *Journal of Human Resources*, 2008, *43* (4), 884–918.

Goldin, Claudia, "A grand gender convergence: Its last chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

_ **and Lawrence Katz**, "The most egalitarian of all professions: Pharmacy and the evolution of a family-friendly occupation," *Journal of Labor Economics*, 2016, *34* (3), 705–746.

_ **,** _ **, and Ilyana Kuziemko**, "The homecoming of American college women: The reversal of the college gender gap," *Journal of Economic Perspectives*, 2006, *20* (4), 133–156.

Grove, Wayne, Andrew Hussey, and Michael Jetter, "The gender pay gap beyond human capital: Heterogeneity in noncognitive skills and in labor market tastes," *Journal of Human Resources*, 2011, *46* (4), 827–874.

Katz, Lawrence and Alan Krueger, "Understanding trends in alternative work arrangements in the United States," Working Paper 25425, National Bureau of Economic Research, Cambridge, MA 2019.

Kuhn, Peter and Kailing Shen, "Gender discrimination in jobs ads: Evidence from China," *Quarterly Journal of Economics*, 2013, *128* (1), 287–336.

_ **and Marie Claire Villeval**, "Are women more attracted to co-operation than men?," *Economic Journal*, 2015, *125* (582), 115–140.

Lordan, Grace and Jorn-Steffen Pischke, "Does Rosie like riveting? Male and female occupational choices," Working Paper, London School of Economics and Political Science 2018.

Maestas, Nicole, Kathleen Mullen, David Powell, Till von Wachter, and Jeffrey Wenger, "The value of working conditions in the United States and implications for the structure of wages," Working Paper 25204, National Bureau of Economic Research, Cambridge, MA 2018.

Mas, Alexandre and Amanda Pallais, "Valuing alternative work arrangements," *American Economic Review*, 2017, *107* (12), 3722–3759.

Mullainathan, Sendhil and Jann Spiess, "Machine learning: An applied econometric approach," *Journal of Economic Perspectives*, 2017, *31* (2), 87–106.

Ngai, Rachel and Barbara Petrongolo, "Gender gaps and the rise of the service economy," *American Economic Journal: Macroeconomics*, 2017, *9* (4), 1–44.

Niederle, M and L Vesterlund, "Do women shy away from competition? Do men compete too much?," *Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

Oaxaca, R., "Male-female wage differentials in urban labor markets," *International Economic Review*, 1973, *14*, 693–709.

Pan, Jessica, "Gender segregation in occupations: The role of tipping and social interactions," *Journal of Labor Economics*, 2015, *33* (2), 365–408.

Pinker, Susan, *The sexual paradox: Troubled boys, gifted girls and the real difference between the sexes*, New York: Macmillan, 2008.

Rosen, Sherwin, "The theory of equalizing differences," in Orley C. Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics*, Vol. 1, Elsevier, 1986, pp. 641–692.

Roy, A.D., "Some thoughts on the distribution of earnings," *Oxford Economic Paper*, 1951, *3* (2), 135–146.

Snowberg, Erik and Leeat Yariv, "Testing the waters: Behavior across participant pools," Working Paper 24781, National Bureau of Economic Research, Cambridge, MA 2018.

Su, Rong, James Rounds, and Patrick Armstrong, "Men and things, women and people: A meta-analysis of sex differences in interests," *Psychological Bulletin*, 2009, *135* (6), 859–884.

Wasserman, Melanie, "Hours constraints, occupational choice and fertility: Evidence from medical residens," Working Paper, UCLA Anderson School of Management 2019.

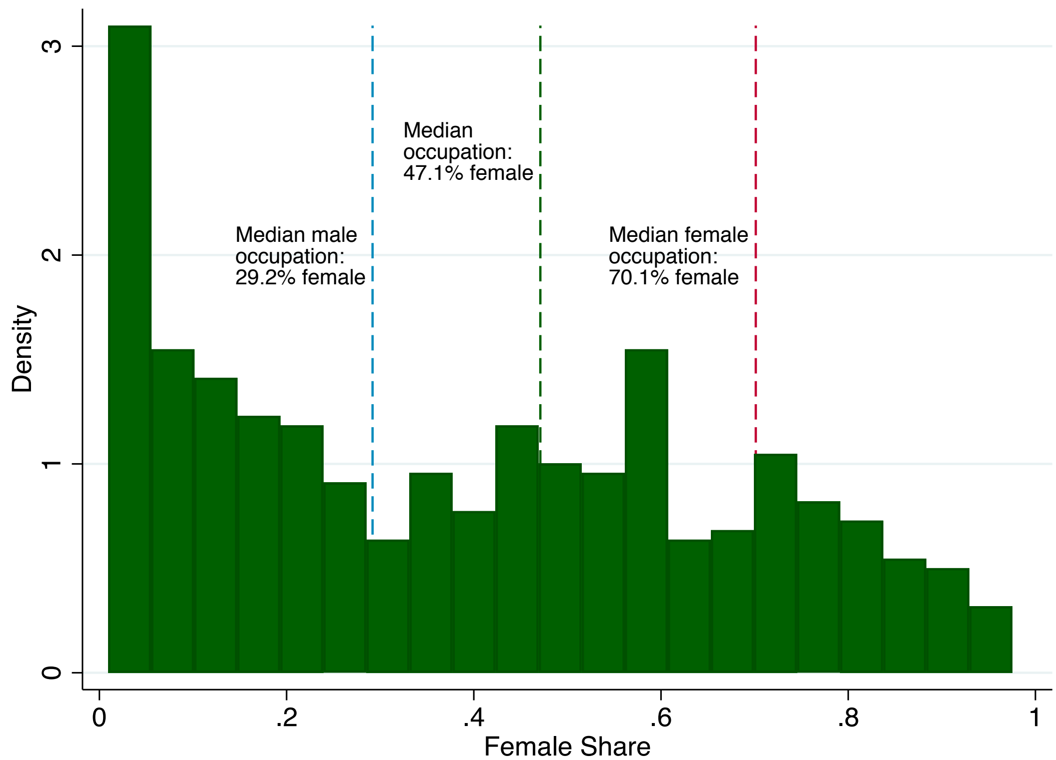
Weinberg, Bruce A, "Computer use and the demand for female workers," *Industrial and Labor Relations Review*, 2000, *53* (2), 290–308.

Welch, Finis, "Growth in women's relative wages and inequality among men: One phenomenon or two?," *American Economic Review*, 2000, *90* (2), 444–449.

Wiswall, Matthew and Basit Zafar, "Preference for the workplace, investment in human capital, and gender," *Quarterly Journal of Economics*, 2018, *133* (1), 457–507.

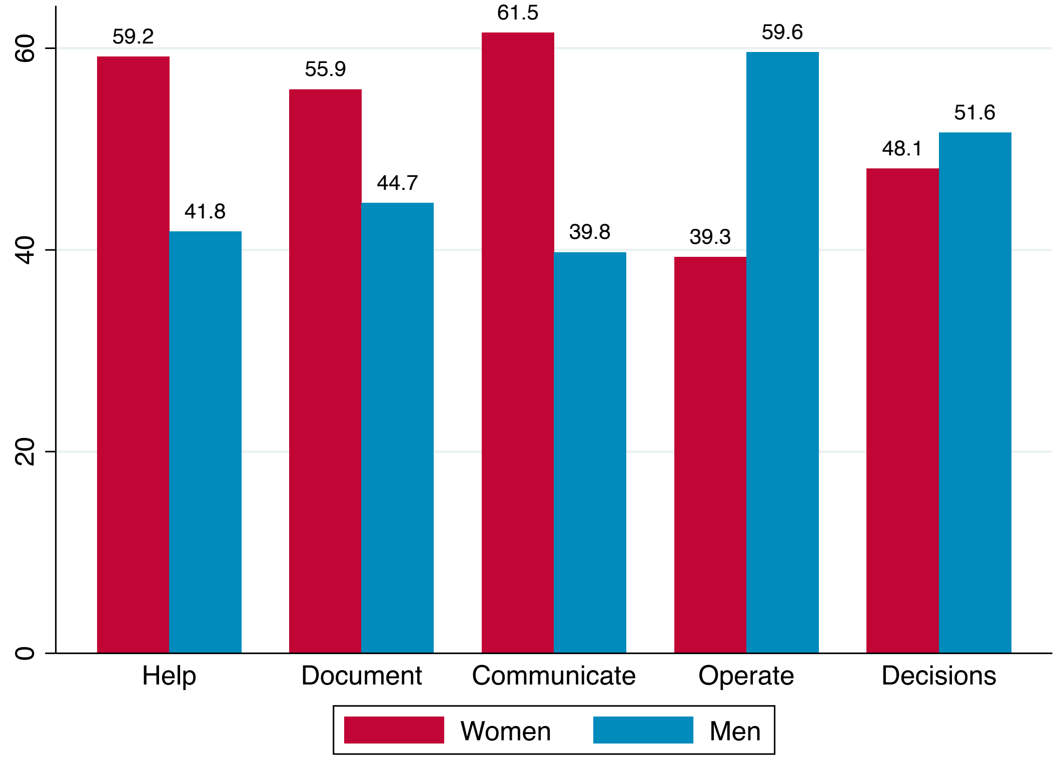
Figures

Figure 1: Distribution of Occupational Female Share



Notes. This figure shows the distribution of the share of workers in an occupation who are female using data from currently employed individuals aged 18 and older in the 478 occupations available in the 2012-2016 ACS. The dashed lines indicate the median occupational female share for (from left to right) male workers, all workers and female workers.

Figure 2: Task Percentiles by Gender



Notes. This figure shows mean levels of the five gender-typical tasks included in the experiment, separately by gender, using the O*NET work activities and work styles listed in Table 2. Tasks based on multiple O*NET variables are constructed by averaging the component variables. Each task is measured in percentiles weighted by currently employed individuals aged 18 and older in the 2012-2016 ACS.

Figure 3: Example of Hypothetical Scenario

Imagine that you have been given a choice between the following two jobs that differ in terms of pay and the amount of time you spend **working and communicating with others and displaying a cooperative attitude**.

The jobs are exactly the same in all other ways, including schedule, co-workers, benefits, and the set of activities you do when you are not working and communicating with others and displaying a cooperative attitude.

Which job would you prefer?

- A job in which you spend **less than 4 hours** per week **working and communicating with others and displaying a cooperative attitude**. You work 40 hours per week in total.

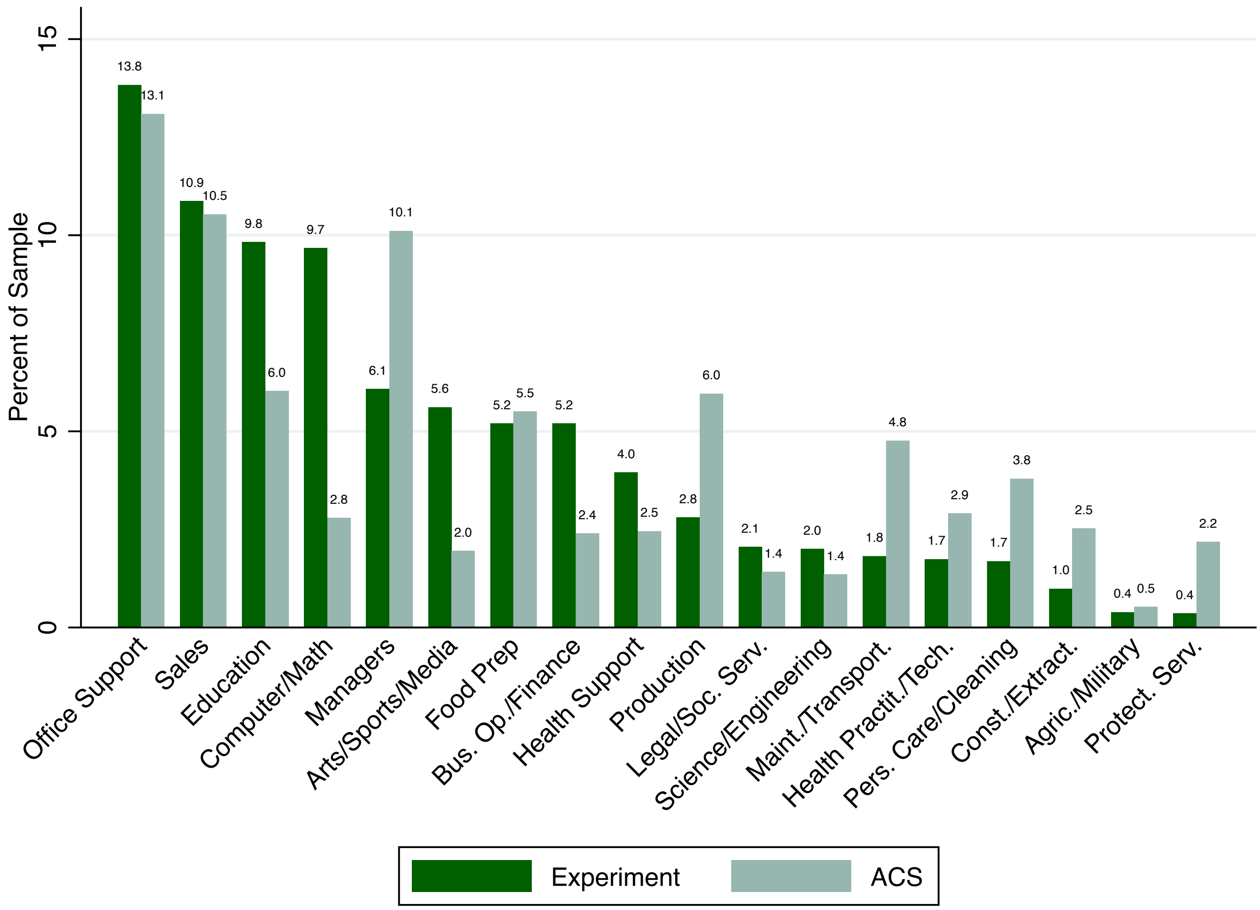
  You are paid **$24 per hour**.

- A job in which you spend **4 hours or more** per week **working and communicating with others and displaying a cooperative attitude**. You work 40 hours per week in total.

  You are paid **$20 per hour**.

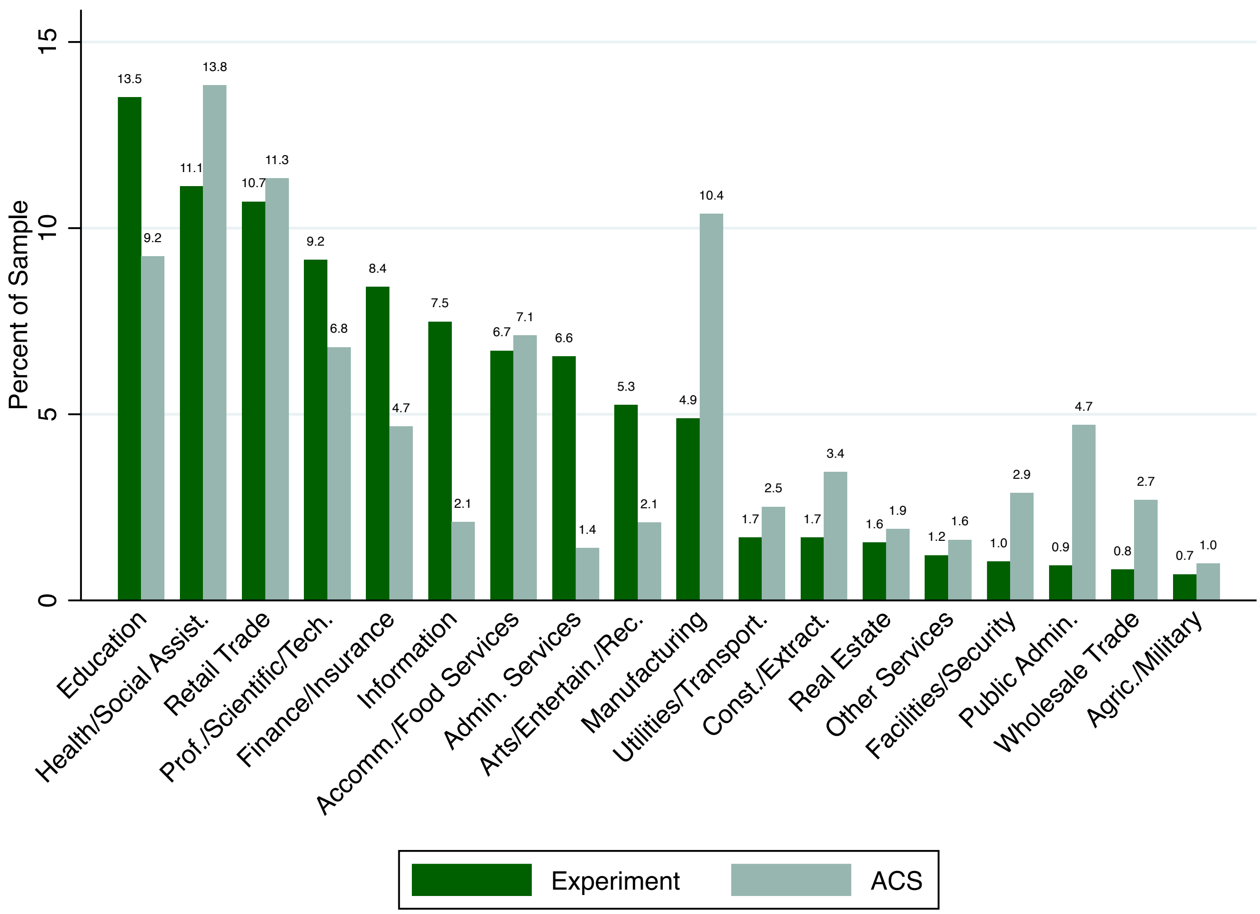
Notes. This figure shows an example of a scenario from the hypothetical choice experiment related to the task *working and communicating with others* for a participant who reports working 40 hours per week and being paid $20/hour in the current or most recent job.

Figure 4: Occupation Categories in Experiment Sample vs. ACS



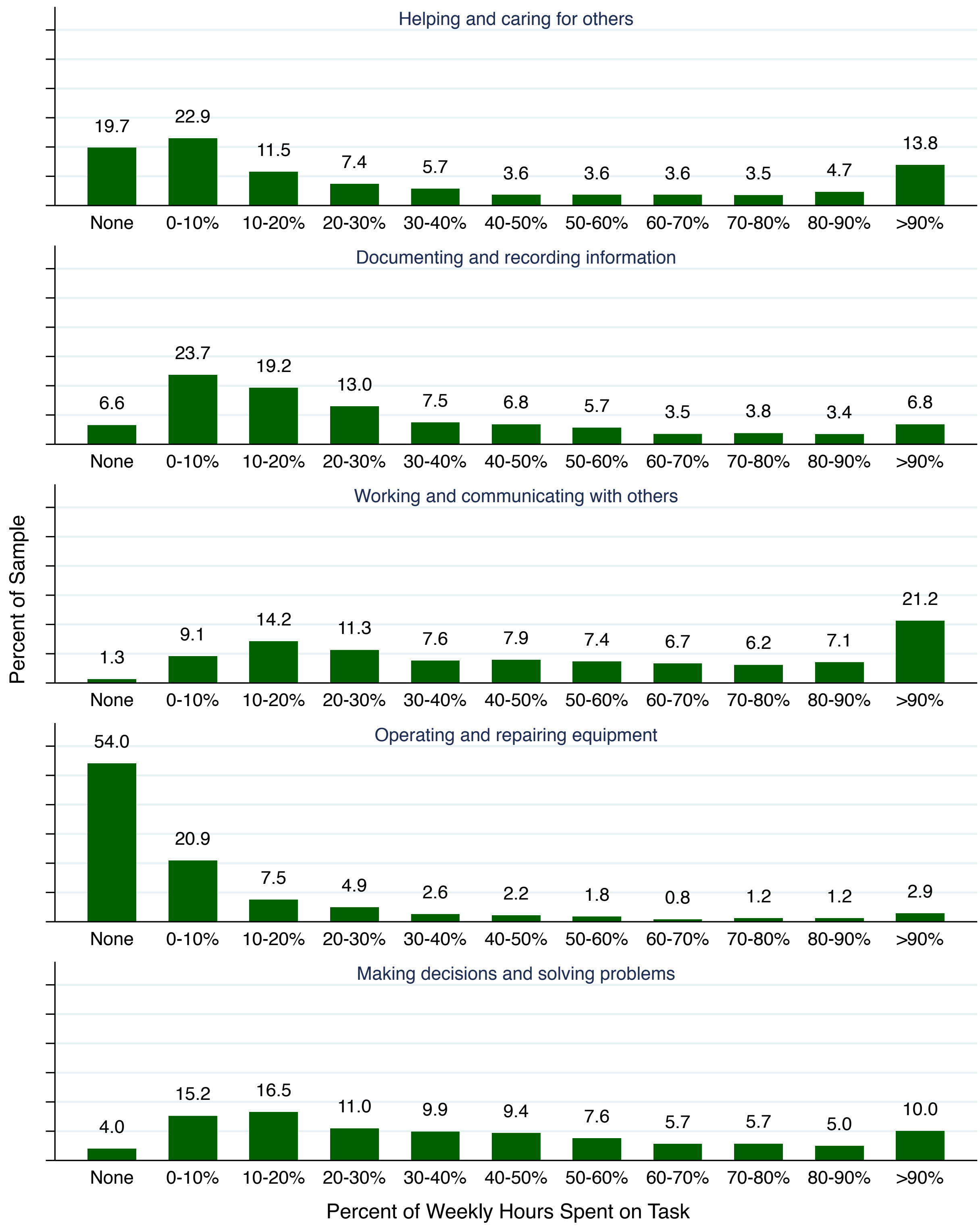
Notes. This figure shows the share of the experiment sample compared with the 2012-2016 ACS in each occupation category. The data from the experiment sample include occupation in the most recent job for participants who are not currently employed. The data from the ACS include only currently employed individuals aged 18 and older. Occupation categories are based on the two-digit occupation codes in the ACS, with some additional aggregation.

Figure 5: Industry Categories in Experiment Sample vs. ACS



Notes. This figure shows the share of the experiment sample compared with the 2012-2016 ACS in each industry category. The data from the experiment sample include industry in the most recent job for participants who are not currently employed. The data from the ACS include only currently employed individuals aged 18 and older. Industry categories are based on the two-digit North American Industry Classification System (NAICS) codes, with some modifications.

Figure 6: Task Distributions - Entire Sample

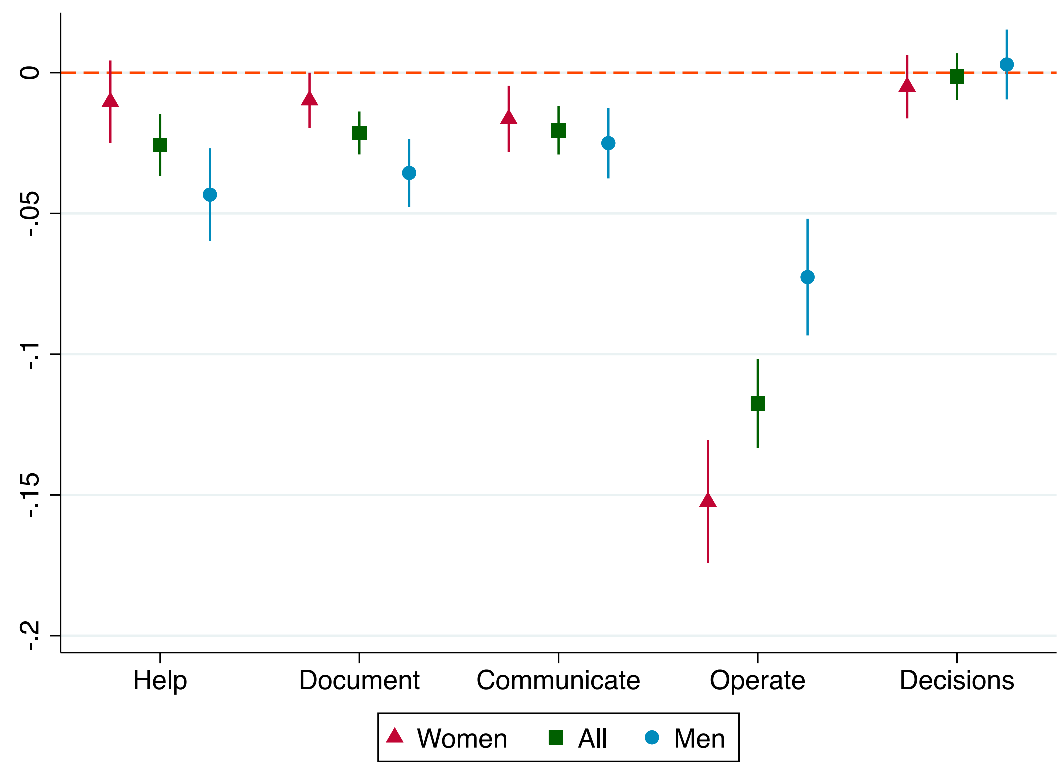


Notes. This figure shows the distribution of time spent on each gender-typical task in the current or most recent job among experiment participants.

Figure 7: Task Distributions - By Gender

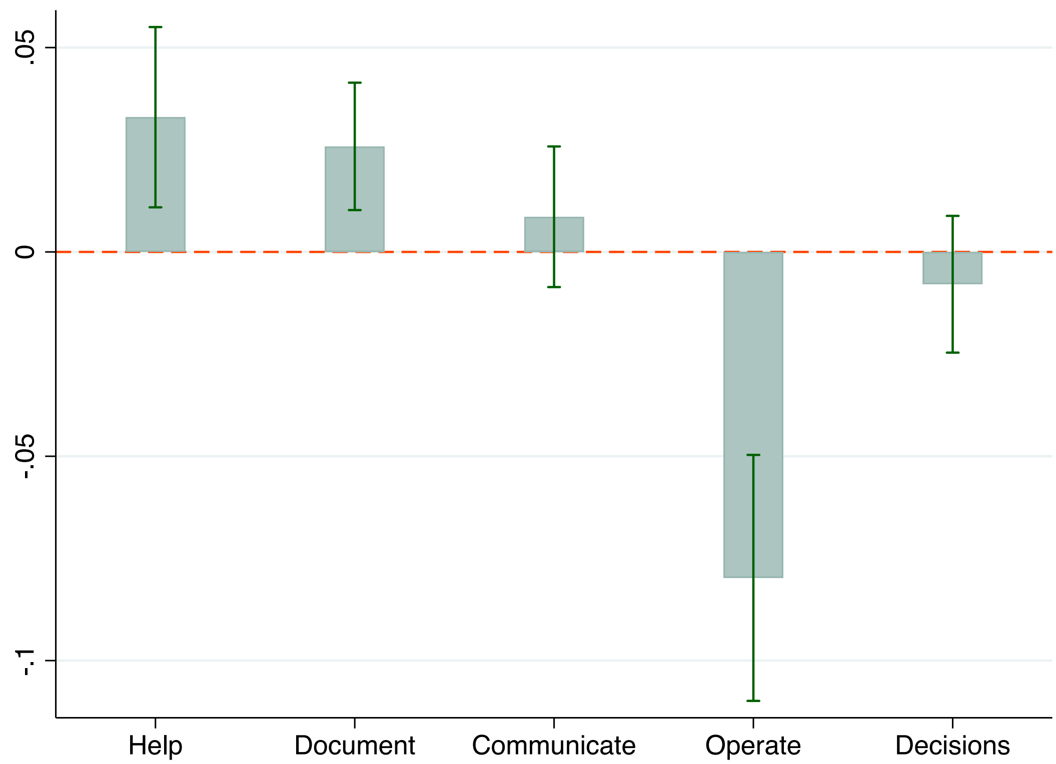
Notes. This figure shows the distribution by gender of time spent on each gender-typical task in the current or most recent job among experiment participants.

Figure 8: WTP for Tasks as Share of Wage



Notes. This figure plots the estimates from Table 4 of WTP for spending more time on the gender-typical tasks as a proportion of the wage for the entire sample (λ_{ak}), women (λ_{fk}) and men (λ_{mk}), using choice data from the experiment.

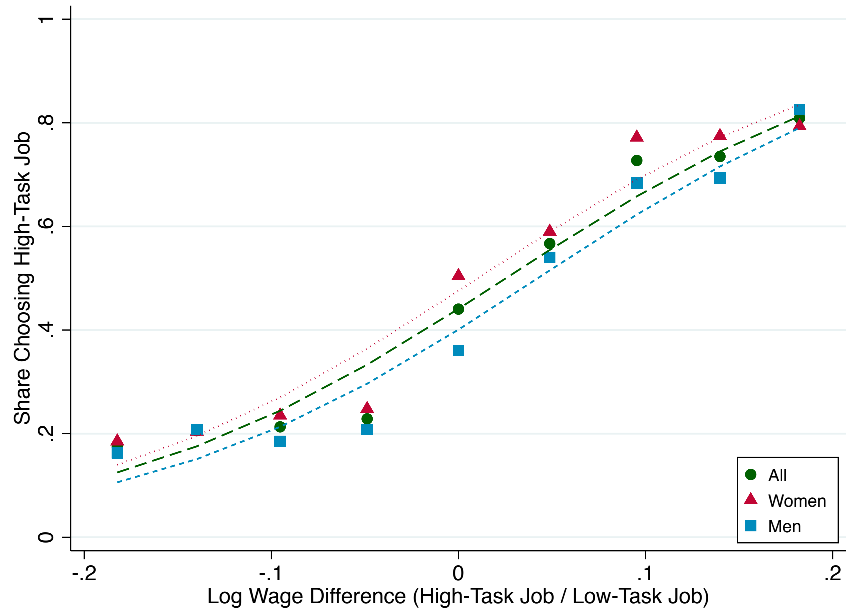
Figure 9: WTP for Tasks: Gender Difference (W−M)



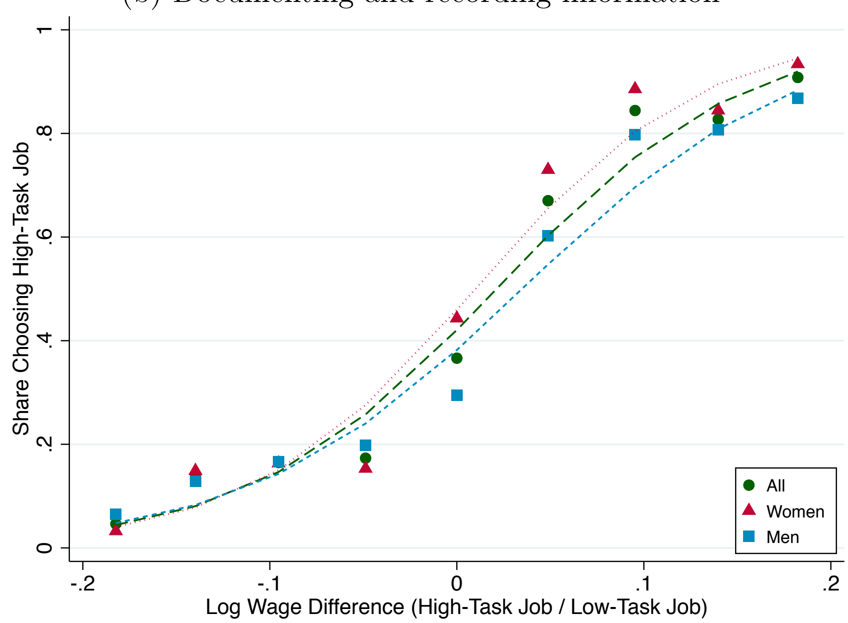
Notes. This figure plots the estimates from Table 4 of the female-to-male difference in WTP (β_k), using choice data from the experiment.

Figure 10: Job Choices - Female-Typical Tasks

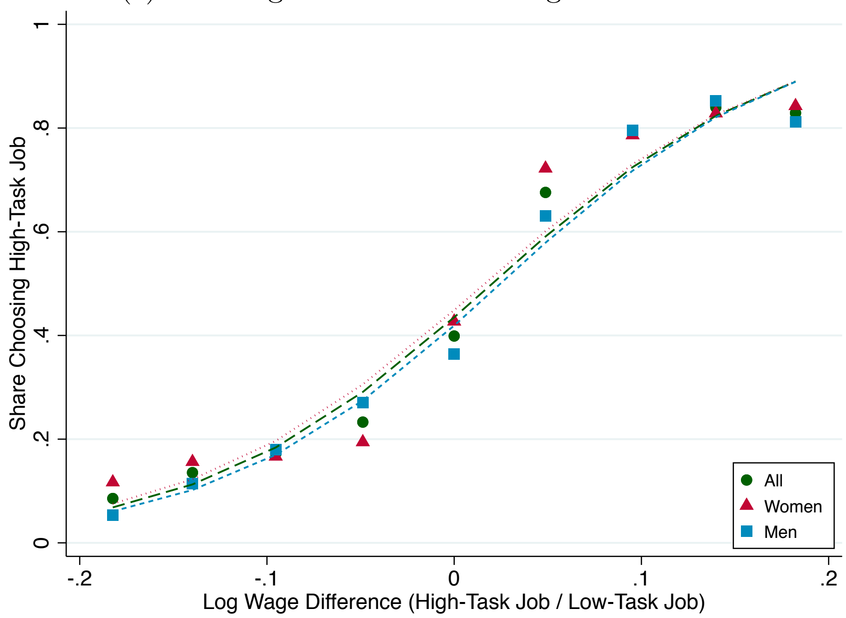
(a) Helping and caring for others



(b) Documenting and recording information



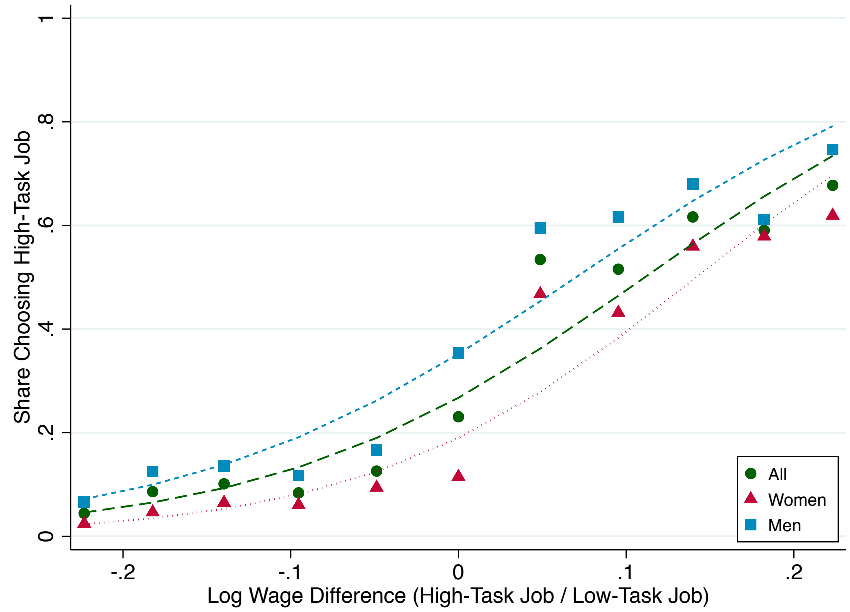
(c) Working and communicating with others



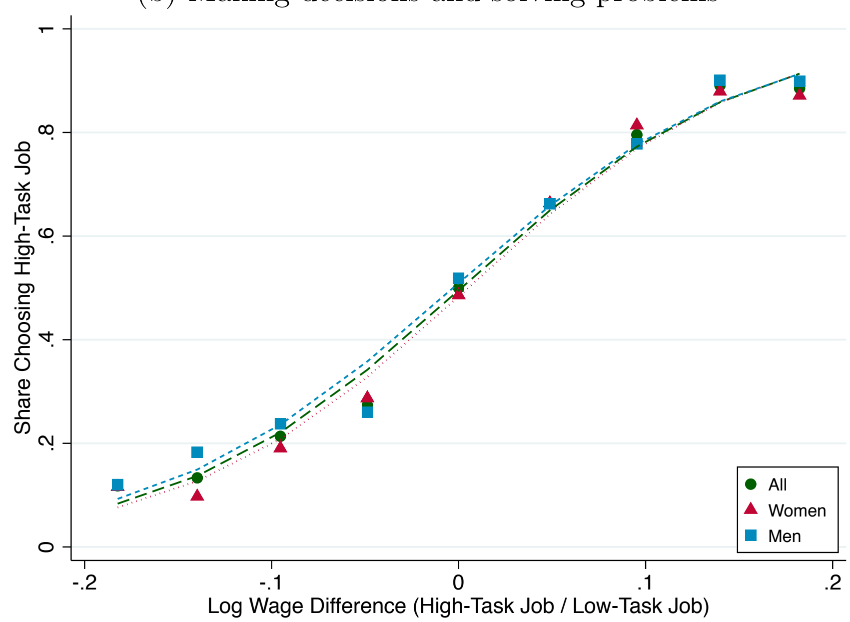
Notes. This figure shows the share of all experiment participants, women and men choosing the high-task job in the hypothetical scenarios as a function of the log difference in wage offers, for female-typical tasks. The high-task job is the job that involves spending more than the cutoff amount of time on the focal task.

Figure 11: Choice of High-Task Job - Male-Typical Tasks

(a) Operating and repairing equipment



(b) Making decisions and solving problems



Notes. This figure shows the share of all experiment participants, women and men choosing the high-task job in the hypothetical scenarios as a function of the log difference in wage offers, for male-typical tasks. The high-task job is the job that involves spending more than the cutoff amount of time on the focal task.

Tables

Table 1: Predicting the Occupational Female Share

	Controls	All Tasks	Selected Tasks
R^2	0.405	0.802	0.672
Adjusted R^2	0.401	0.774	0.668
N	464	464	464

Notes. This table shows the R^2 and adjusted R^2 statistics and the number of observations from a series of OLS regressions in which the outcome is the share of currently employed individuals aged 18 and older who are female in each occupation in the 2012-2016 ACS. In Column 1, the predictors are the mean log hourly wage, the mean log usual hours of work per week, and the share of workers with at least a college degree in each occupation, also from the ACS. In Column 2, the predictors are all 57 variables in the work activities or work styles domains of the O*NET. In Column 3, the predictors are the O*NET work activities and work styles listed in Table 2 that are selected for inclusion in the experiment. The O*NET variables are standardized to have a mean of zero and standard deviation of one. For the selected tasks that combine multiple O*NET variables, I average the component variables to create a single measure and re-standardize. The regressions include the 464 ACS occupations that can be matched to the O*NET.

Table 2: Gender-Typical Tasks

Female-Typical Tasks

Helping and caring for others

Definition: Providing personal assistance, medical attention, emotional support, or other personal care to people such as co-workers, customers, or patients.

Examples: Helping a co-worker complete an assignment, assisting a customer in finding a product, or caring for injured people in a hospital.

*O*NET measure(s)*:

Assisting and Caring for Others (work activity)

Documenting and recording information

Definition: Entering, transcribing, recording, storing, or maintaining information in written or electronic form.

Examples: Recording the weights of trucks that use highways, documenting proceedings in a court room, or maintaining information about a patient's health.

*O*NET measure(s)*:

Documenting/Recording Information (work activity)

Working and communicating with others and displaying a cooperative attitude

Definition: Generally working with others rather than alone and being pleasant and good-natured with others on the job.

Examples: Meeting with co-workers to discuss a project, answering a client's questions over the phone, or facilitating a workshop.

*O*NET measure(s)*:

Social Orientation (work style)
Cooperation (work style)

Male-Typical Tasks

Operating, repairing and maintaining vehicles, devices or equipment

Definition: Running, navigating, servicing, repairing, adjusting, or testing vehicles, machines, devices, moving parts, or equipment.

Examples: Driving a car or truck, adjusting the settings on a medical device, or repairing a circuit board.

*O*NET measure(s)*:

Operating Vehicles, Mechanized Devices, or Equipment (work activity)
Repairing and Maintaining Mechanical Equipment (work activity)
Repairing and Maintaining Electronic Equipment (work activity)

Making decisions and solving problems

Definition: Analyzing information and evaluating results to choose the best solution and solve problems.

Examples: Selecting the menu options for a cafeteria, choosing a location for a retail store, or finalizing the budget for a school.

*O*NET measure(s)*:

Making Decisions and Solving Problems (work activity)

Table 3: Summary Statistics - Employed Only

	Experiment			ACS		
	All	Women	Men	All	Women	Men
Female	0.526	1.000	0.000	0.472	1.000	0.000
Age	34.4	35.1	33.5	42.0	42.0	42.1
White	0.719	0.769	0.663	0.648	0.645	0.651
Black	0.080	0.073	0.087	0.110	0.126	0.096
Hispanic	0.111	0.079	0.147	0.161	0.147	0.174
Other Race	0.090	0.079	0.103	0.080	0.082	0.079
HS or less	0.084	0.086	0.082	0.337	0.293	0.376
Some college	0.229	0.220	0.239	0.238	0.247	0.230
Associate's degree	0.124	0.144	0.101	0.090	0.104	0.077
Bachelor's degree	0.412	0.385	0.443	0.212	0.224	0.201
Graduate degree	0.151	0.164	0.135	0.123	0.131	0.116
Hours per week	37.6	36.4	39.0	39.3	36.6	41.7
Wage (hourly)	21.01	18.68	23.67	25.05	22.17	27.67
N	1,742	917	825	7,031,598	3,367,987	3,663,611

Notes. This table shows summary statistics in the experiment sample compared with the 2012-2016 ACS, restricting to currently employed participants in the experiment sample and currently employed individuals aged 18 and older in the ACS. Statistics in the ACS are weighted by the ACS person weight.

Table 4: WTP for Tasks as Share of Wage

	Help	Document	Communic.	Operate	Decisions
All	-0.026**	-0.021**	-0.020**	-0.118**	-0.001
	(0.006)	(0.004)	(0.004)	(0.008)	(0.004)
Women	-0.010	-0.010*	-0.016**	-0.152**	-0.005
	(0.007)	(0.005)	(0.006)	(0.011)	(0.006)
Men	-0.043**	-0.036**	-0.025**	-0.073**	0.003
	(0.008)	(0.006)	(0.006)	(0.011)	(0.006)
Diff (W-M)	0.033**	0.026**	0.009	-0.080**	-0.008
	(0.011)	(0.008)	(0.009)	(0.015)	(0.009)
N	1,931	1,931	1,931	1,931	1,931

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table shows estimates of WTP for spending more time on the gender-typical tasks as a proportion of the wage for the entire sample (λ_{ak}), women (λ_{fk}) and men (λ_{mk}), and the female-to-male difference in WTP (β_k), using choice data from the experiment.

Table 5: WTP for Tasks - Extensive vs. Intensive Margin

(a) Extensive Margin

	Help	Document	Communic.	Operate	Decisions
All	-0.015^{+}	-0.016**	0.003	-0.127**	0.006
	(0.007)	(0.005)	(0.006)	(0.012)	(0.006)
Women	-0.008	-0.004	0.006	-0.155**	0.001
	(0.011)	(0.006)	(0.009)	(0.016)	(0.009)
Men	-0.022*	-0.030**	-0.000	-0.089**	0.011
	(0.010)	(0.008)	(0.009)	(0.017)	(0.009)
Diff (W-M)	0.014	0.026*	0.006	-0.065**	-0.009
	(0.015)	(0.010)	(0.012)	(0.024)	(0.012)
N	953	953	953	953	953

(b) Intensive Margin

	Help	Document	Communic.	Operate	Decisions
All	-0.038**	-0.028**	-0.042**	-0.109**	-0.008
	(0.009)	(0.006)	(0.006)	(0.011)	(0.006)
Women	-0.013	-0.016*	-0.037**	-0.150**	-0.010
	(0.011)	(0.008)	(0.009)	(0.015)	(0.008)
Men	-0.071**	-0.043**	-0.048**	-0.058**	-0.005
	(0.015)	(0.010)	(0.009)	(0.014)	(0.009)
Diff (W-M)	0.059**	0.027*	0.011	-0.092**	-0.005
	(0.018)	(0.013)	(0.012)	(0.020)	(0.012)
N	978	978	978	978	978

(c) Difference (Extensive − Intensive)

	Help	Document	Communic.	Operate	Decisions
All	0.023*	0.012	0.045**	-0.018	0.014+
	(0.011)	(0.008)	(0.009)	(0.016)	(0.009)
Women	0.004	0.013	0.042**	-0.004	0.011
	(0.015)	(0.010)	(0.012)	(0.022)	(0.011)
Men	0.049**	0.013	0.047**	-0.031	0.016
	(0.018)	(0.013)	(0.013)	(0.022)	(0.013)
Diff (W-M)	-0.045+	-0.000	-0.005	0.027	-0.004
	(0.024)	(0.016)	(0.018)	(0.031)	(0.017)
N	1,931	1,931	1,931	1,931	1,931

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table shows WTP estimates for experiment participants who face extensive margin cutoffs in the hypothetical scenarios in Panel A, WTP estimates for participants who face intensive margin cutoffs in Panel B, and the difference between the estimates in Panel C.

Table 6: WTP for Tasks - Currently in High-Task vs. Low-Task Job

(a) Currently in High-Task Job

	Help	Document	Communic.	Operate	Decisions
All	0.015*	-0.010*	-0.001	-0.047**	0.014**
	(0.007)	(0.005)	(0.005)	(0.011)	(0.005)
Women	0.020*	0.001	0.003	-0.085**	0.013+
	(0.008)	(0.006)	(0.006)	(0.016)	(0.007)
Men	0.007	-0.026**	-0.005	-0.019	0.014+
	(0.011)	(0.008)	(0.007)	(0.014)	(0.007)
Diff (W-M)	0.013	0.027**	0.008	-0.067**	-0.001
	(0.014)	(0.010)	(0.010)	(0.021)	(0.010)
N	988	1,156	1,408	694	1,272

(b) Currently in Low-Task Job

	Help	Document	Communic.	Operate	Decisions
All	-0.075**	-0.042**	-0.077**	-0.157**	-0.030**
	(0.010)	(0.007)	(0.010)	(0.011)	(0.007)
Women	-0.063**	-0.034**	-0.086**	-0.181**	-0.039**
	(0.015)	(0.011)	(0.016)	(0.015)	(0.010)
Men	-0.084**	-0.049**	-0.068**	-0.116**	-0.019
	(0.012)	(0.010)	(0.013)	(0.016)	(0.012)
Diff (W-M)	0.021	0.014	-0.018	-0.065**	-0.020
	(0.019)	(0.015)	(0.020)	(0.022)	(0.015)
N	935	767	515	1,229	651

(c) Difference (High-Task − Low-Task)

	Help	Document	Communic.	Operate	Decisions
All	0.090**	0.033**	0.076**	0.109**	0.044**
	(0.012)	(0.009)	(0.011)	(0.015)	(0.009)
Women	0.083**	0.036**	0.089**	0.095**	0.052**
	(0.017)	(0.012)	(0.017)	(0.022)	(0.012)
Men	0.091**	0.023+	0.063**	0.097**	0.033*
	(0.017)	(0.013)	(0.015)	(0.021)	(0.014)
Diff (W-M)	-0.008	0.013	0.026	-0.002	0.019
	(0.024)	(0.018)	(0.023)	(0.031)	(0.018)
N	1,923	1,923	1,923	1,923	1,923

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table shows WTP estimates for experiment participants who report currently working in a high-task job in Panel A, WTP estimates for participants who report currently working in a low-task job in Panel B, and the difference between the estimates in Panel C. I designate participants as working in a high-task k job if the number of hours they report spending on task k in their current or most recent job is greater than the cutoff number of hours in the hypothetical scenario they face for that task.

Table 7: WTP for Tasks - Weighted to Match ACS

	Help	Document	Communic.	Operate	Decisions
All	-0.039**	-0.029**	-0.028**	-0.099**	-0.008
	(0.010)	(0.007)	(0.008)	(0.013)	(0.006)
Women	-0.004	-0.015+	-0.025**	-0.138**	-0.013+
	(0.011)	(0.008)	(0.009)	(0.017)	(0.007)
Men	-0.074**	-0.043**	-0.030*	-0.055**	-0.003
	(0.018)	(0.011)	(0.012)	(0.018)	(0.010)
Diff (W-M)	0.070**	0.027*	0.005	-0.083**	-0.010
	(0.021)	(0.014)	(0.015)	(0.024)	(0.012)
N	1,728	1,728	1,728	1,728	1,728

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table shows WTP estimates for currently employed individuals in the experiment sample that have been weighted to match currently employed workers aged 18 and older in the 2012-2016 ACS by gender, race (White versus non-White), college degree receipt, and major occupation category.

Table 8: Reasons for Choices

	All	High-Task	Low-Task
Offers better pay	0.594	0.632	0.564
More enjoyable/interesting	0.338	0.433	0.265
Better fit for skills	0.338	0.381	0.305
Develop new skills	0.177	0.281	0.097
Require less effort	0.234	0.075	0.356
More people like me	0.076	0.107	0.052
Would be treated better	0.066	0.071	0.061
More prestigious	0.054	0.074	0.039
N	3,862	1,683	2,179

Notes. This table shows the share of the experiment sample citing each of the reasons listed as a motivation for choices made in the hypothetical scenarios. Each participant was asked to indicate reasons for the choices made in scenarios relating to a randomly selected two out of the five gender-typical tasks. Column 1 shows reasons cited by all participants, Column 2 shows reasons cited bt participants who chose the high-task job in that scenario, and Column 3 shows reasons cited by participats who chose the low-task job.

Table 9: WTP for Tasks - All Tasks Pooled, Single δ Parameter

	Help	Document	Communic.	Operate	Decisions
Women	-0.009	-0.012*	-0.017**	-0.143**	-0.006
	(0.007)	(0.006)	(0.006)	(0.009)	(0.006)
Men	-0.039**	-0.039**	-0.027**	-0.061**	0.003
	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)
Diff (W-M)	0.030**	0.027**	0.010	-0.082**	-0.009
	(0.010)	(0.009)	(0.009)	(0.012)	(0.009)
N	9,655	9,655	9,655	9,655	9,655

Notes. Standard errors clustered by participant. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table shows WTP estimates from a pooled specification in which choices are stacked and all tasks are included in a single regression. The $\hat{\delta}$ coefficient on the log difference in wage offers is also restricted to be the same for women and men in this specification.

Table 10: Sorting and Segregation Explained by Tasks

(a) Experiment

	Q	I	b	$\hat{Q}$	$\hat{Q}/Q$	$\hat{I}$	$\hat{I}/I$
Help	0.154		-0.089	0.058	0.375		
Document	0.105		0.115	0.060	0.566		
Communic.	0.080		-0.014	0.027	0.332		
Operate	-0.179		0.008	-0.167	0.938		
Decisions	0.007		0.168	-0.011	-1.608		
Index		0.262				0.179	0.684
N			1,785				

(b) ACS

	Q	I	b	$\hat{Q}$	$\hat{Q}/Q$	$\hat{I}$	$\hat{I}/I$
Help	0.270		-0.191	0.024	0.089		
Document	0.141		0.133	0.054	0.380		
Communic.	0.259		0.050	0.027	0.104		
Operate	-0.305		-0.017	-0.143	0.470		
Decisions	-0.011		0.222	-0.007	0.595		
Index		0.389				0.158	0.405
N			6,419,869				

Notes. This table shows observed and predicted gender differences in sorting on the gender-typical tasks and task-based segregation in the experiment sample (Panel A) and the 2012-2016 ACS (Panel B). Q ($\hat{Q}$) is the observed (predicted) gender difference in the share of workers sorting into the high-task job. b is the coefficient on the task from the wage regression. I ($\hat{I}$) is the value of the segregation index based on observed (predicted) sorting.

Table 11: Wage Gap Explained by Tasks

(a) Experiment

	Q	b	$\hat{Q}$	Qb	$\hat{Q}b$
Help	0.154	-0.089	0.058	-0.014	-0.005
Document	0.105	0.115	0.060	0.012	0.007
Communic.	0.080	-0.014	0.027	-0.001	-0.000
Operate	-0.179	0.008	-0.167	-0.001	-0.001
Decisions	0.007	0.168	-0.011	0.001	-0.002
Total Explained by Tasks				-0.003	-0.002
Total Wage Gap		-0.199			
N		1,785			

(b) ACS

	Q	b	$\hat{Q}$	Qb	$\hat{Q}b$
Help	0.270	-0.191	0.024	-0.052	-0.005
Document	0.141	0.133	0.054	0.019	0.007
Communic.	0.259	0.050	0.027	0.013	0.001
Operate	-0.305	-0.017	-0.143	0.005	0.002
Decisions	-0.011	0.222	-0.007	-0.003	-0.001
Total Explained by Tasks				-0.017	0.005
Total Wage Gap		-0.183			
N		6,419,869			

Notes. This table shows the gender wage gaps explained by observed and predicted gender differences in sorting on the gender-typical tasks in the experiment sample (Panel A) and the 2012-2016 ACS (Panel B). Qb ($\hat{Q}b$) is the female-to-male log hourly wage gap explained by the observed (predicted) gender difference in sorting on the task.

Table 12: Wage Gap Decomposition in ACS

	(1)	(2)	(3)	(4)
Summary:				
Total Wage Gap	-0.183	-0.183	-0.183	-0.183
Residual Gap	-0.216	-0.142	-0.170	-0.190
Explained Gap	0.033	-0.041	-0.014	0.007
Explained By:				
Occupation		-0.056		
Continuous Tasks			-0.034	
Binary Tasks				-0.017
N	6,419,869	6,419,869	6,419,869	6,419,869

Notes. This table shows the results of a decomposition of the female-to-male log hourly wage gap in the 2012-2016 ACS. All columns include controls for race and ethnicity, education, potential experience, geography, and year. Column 2 also includes occupation indicators. Column 3 includes the gender-typical tasks measured as continuous variables standardized to have a mean of zero and standard deviation of one. Column 4 includes the binary measures of the gender-typical tasks that I use in Tables 10 and 11.

A1 Data Appendix

A1.1 Selecting Gender-Typical Tasks

This section describes the process of selecting the gender-typical job tasks that are examined in the hypothetical choice experiment.

ACS Data

I use data from the Integrated Public Use Microdata five-year ACS sample for 2012-2016 to measure the occupational female share. I restrict the pooled sample to workers aged 18 and older who are currently employed and are not classified as unpaid family workers. I calculate the proportion of workers who are female in each of the ACS occupation codes, weighting by the ACS person weight.

I also use the ACS data to measure the mean log hourly wage in each occupation, the mean usual hours of work per week in each occupation, and the share of workers in each occupation with at least a four-year college degree. To calculate hourly wages, I divide wage and salary income from the last 12 months by annual hours of work. I measure annual hours of work as usual hours of work per week multiplied by weeks worked in the last 12 months. Beginning in 2008, ACS respondents report weeks worked in intervals (1-13 weeks, 14-26 weeks, 27-39 weeks, 40-47 weeks, 48-49 weeks and 50-52 weeks). Therefore, I use data from the 2005-2007 ACS to impute the actual number of weeks worked in the 2012-2016 sample as mean weeks worked among respondents within each interval in the earlier period. I inflate wages to 2018 dollars using the Consumer Price Index from the Bureau of Labor Statistics (BLS), and exclude hourly wages observations that are less than \$3 and greater than \$200.

O*NET Data

I use data from the O*NET 21.1, released in November 2016, to measure the frequency of task performance by occupation. I focus on variables from the *generalized work activities* and *work styles* domains, which include 41 and 16 elements, respectively, that I interpret as providing information about categories of job tasks. The surveys on which the O*NET data are based ask workers to indicate the importance of each work activity and work style element to the performance of their current job, on a scale of 1 to 5. The published O*NET data report the mean importance value for each element in each occupation.

The O*NET surveys also asks respondents to report the level of each work activity required to perform their current job, on a scale of 0 to 7. I use measures of the importance of work activities rather than the level for consistency across domains, and because I interpret importance to map more closely onto the notion of the frequency with which an activity is performed. By contrast, the examples provided to guide workers in answering the questions about the level of an activity seem explicitly intended to capture information about worker skills and education requirements. For instance, the example for level two of *making decisions and solving problems* is “determine the meal selection for a cafeteria,” while the level five example is “make the final decision about a company’s 5-year plan.” I conduct robustness tests in which I use the level value and the level value multiplied by the importance value (with the variables first rescaled to fall between 0 and 10) for the work activity measures, and find similar results.

Occupational Crosswalk

The O*NET 21.1 uses occupation codes based on the BLS 2010 Standard Occupation Classification (SOC), but with some additional detailed codes. To link the O*NET variables to the female share data from the ACS, I first collapse the data to the mean value within each SOC code. I then use a crosswalk available from the Census Bureau to map each SOC code to an ACS occupation code and collapse the data to the mean value within each ACS code, weighting by the number of workers employed in each SOC occupation in 2012-2016, based on the BLS Occupational Employment Statistics (OES) data. The O*NET data can be matched to 464 ACS occupation codes, out of a total of 478 codes in the ACS data. After applying the crosswalk, I standardize the O*NET measures to have a mean of zero and a standard deviation of one.

Random Forest Analysis

As part of the process of selecting tasks, I use a random forest algorithm (Breiman 2001) to identify O*NET work activities and work styles that are highly predictive of the occupational female share. The outcome in this analysis is the female share, and the predictors are the O*NET variables. I conduct the analysis separately for the two O*NET domains. I choose the tuning parameters via cross-validation, following Mullainathan and Spiess (2017). Specifically, I use eight-fold cross-validation to select the number of trees (400, 500, 600 or 700), the minimum node size to which each tree is grown (3, 5, 7, 10 or 15), and the proportion of predictors available to be chosen at each internal node in each tree (0.2, 0.3 or 0.4).

While the primary purpose of the random forest technique is predictive, the algorithm also calculates two types of variable importance scores, which rank each predictor according to its contribution to prediction accuracy. The first type of variable importance score is calculated using the *permutation method*, which estimates the improvement in prediction accuracy out-of-sample produced by including a variable in the algorithm. Specifically, the permutation method importance score for predictor X_j is calculated as follows: For each tree t, randomly permute the values of X_j in the out-of-bag (OOB) data not used to construct the tree, such that variable X_j becomes pure noise. Calculate the mean squared prediction error (MSPE) generated by 1) running the true OOB data down tree t, and 2) running the permuted OOB data down tree t. Subtract the first MSPE from the second. Average these differences over all trees t in the forest, then divide by the standard deviation of the difference.

The second type of variable importance score is calculated using the *node purity method*, which yields the improvement in prediction accuracy in-sample produced by including a particular variable in the algorithm. Specifically, the node purity importance score for predictor X_j is calculated as follows: For each tree t and each internal node d at which X_j is chosen as the splitting variable, compute the MSPE for cases that reach node d generated by 1) splitting on X_j, with no further branches in the tree, and 2) a constant prediction at node d. Subtract the first MSPE from the second. Sum these differences over all nodes d in tree t, then average over all trees t in the forest.

I designate a variable as highly predictive if it is among the top-scoring variables using either importance score method. Specifically, for the analysis using the work activities domain, I classify a variable as highly predictive if its importance score places it among the top ten variables using either method. For the analysis using the work styles, I classify a variable as highly predictive if it is ranked among the top four variables using either importance score. This methodology results in 11 highly predictive work activities and 4 highly predictive work styles.

Figures A1 and A2 display the two types of variable importance scores generated by the random forest analysis for the O*NET work activities and work styles. The importance scores are standardized to have a mean of zero and a standard deviation of one within each domain, and are sorted according to the permutation scores. The measures designated as highly predictive are shown in dark green.

Selecting Tasks

As described in the main text, I use a hybrid quantitative and qualitative approach to select tasks to include in the experiment. Tables A1 and A2 show the results of the quantitative analysis for work activities and work styles, respectively. First, I regress the female share on each O*NET measure, standardized to have a mean of zero and standard deviation of one, and rank the coefficients from most positive to most negative (Column 1). Next, I estimate multivariate regressions in which I regress the female share on all O*NET variables in the work activities or work styles domain (Column 2). In Column 3, I regress the female share on the six work activities (two work styles) with the most positive bivariate coefficients and the six work activities (two work styles) with the most negative bivariate coefficients. Finally, in Column 4 I regress the female share on the variables designated as highly predictive by the random forest algorithm.

The results demonstrate that a large proportion of the O*NET measures (29 of 41 work activities and 12 of 16 work styles) have a significant bivariate relationship with the female share. It is also clear that many of the O*NET variables are correlated with each other, as the coefficients change substantially between the bivariate regressions in Column 1 and the multivariate regression in Column 2.

Tables A3 and A4 show that results are similar when I repeat this set of regression analyses including controls for occupation cluster, mean log hourly wages in each occupation and the share of workers with a college degree or more in each occupation in the ACS. The occupation clusters, which are constructed to resemble the broad categories in Acemoglu and Autor (2011), are: 1) managerial, professional and technical occupations ("professional"), 2) sales and office and administrative support occupations ("sales and clerical"), 3) production, construction, extraction, transportation, and installation, maintenance and repair occupations ("blue collar"), 4) healthcare support, protective service, food preparation, building and grounds cleaning and maintenance, and personal care occupations ("service"), and 5) agriculture and military occupations. I also find similar results when I repeat the entire analysis, including the random forest algorithm, using the female share based only on workers with and only on workers without a college degree, and using work activity variables that indicate the level rather than importance value and the level multiplied by the importance value, as described above (results not shown).

I follow a qualitative approach to select a final set of gender-typical tasks. Specifically, I look for work activities and work styles that are statistically significant and consistent in sign across multiple specifications in Tables A1, A2, A3 and A4. I focus on measures that are highly ranked based on the bivariate OLS coefficients, highly predictive in the random forest algorithm, or ideally both. I eliminate some measures that might be difficult for participants to understand, specifically *drafting, layinging out, and specifying equipment* and *estimating quantifiable characteristics*.

The names of the O*NET measures that I choose are displayed in bold in Tables A1, A2, A3 and A4. In some cases, I combine measures of related activities into a single gender-typical task that I include in the experiment, as documented in Table 2. In particular, I combine the work activities *operating vehicles, mechanized devices, or equipment*; *repairing and maintaining mechanical equipment*; and *repairing and maintaining electronic equipment*.

I also combine the work styles *social orientation* and *cooperation*.

Descriptive Results for Selected Tasks

Figures A3 and A4 display the female share (scaled from 0 to 100) and the mean levels of each of the selected gender-typical tasks by major occupation group. For the tasks that combine multiple O*NET variables, I average the variables to create a single measure.

The tasks are rescaled to represent percentiles weighted by employment in each occupation in the 2012-2016 ACS, including all currently employed individuals aged 18 and older. Thus the weighted mean of each task across all occupations is 50. It is clear that the female-typical (male-typical) tasks are rated as more important in majority-female (majority-male) occupation categories, as expected. For example, *helping and caring for others* is rated as most important in female-dominated health, personal care, social services and education occupations, as well as male-dominated protective service occupations. By contrast, *operating and repairing equipment* is rated as most important in male-dominated occupations, including construction, maintenance, transportation and production.

Tables A5, A6, A7, A8 and A9 display the ten occupations with the highest and the ten occupations with the lowest levels of each selected task measure, standardized to have a mean of zero and standard deviation of one, along with the female share in each displayed occupation.

A1.2 Implication for Gender Gaps

This section describes the data used in Table A17, which shows robustness checks for the results on the implications of the WTP estimates for gender differences in sorting and segregation.

In all data sources described below, I inflate wages to 2018 dollars using the Consumer Price Index from the Bureau of Labor Statistics (BLS), and exclude hourly wages observations that are less than \$3 and greater than \$200. In addition, I use the O*NET measures described above to classify occupations as involving a high or low level of the gender-typical tasks. In each dataset, I rescale the O*NET measures to reflect percentiles weighted by employment, and choose a cutoff percentile such that the share of workers in a high-task k job in that dataset matches the share in a high-task job in the experiment sample.

CPS MORG Data

The CPS is a monthly household survey conducted by the U.S. Census Bureau that is designed to represent the civilian, non-institutional U.S. population. Surveyed households are interviewed for four consecutive months, then excluded from the sample for eight months, then interviewed for an additional four consecutive months. Respondents are only asked to report weekly and hourly earnings in their fourth and eighth months in the survey; the files on these "outgoing" households comprise the MORG data.

The analysis in Panel A of Table A17 uses data from the 2012-2018 CPS MORG. I restrict the sample to currently employed individuals aged 18-64 who are not self-employed, have valid wages, and work in an occupation that can be matched to the O*NET. To construct the wage measure, I use data on hourly wages for workers who report being paid hourly, and weekly earnings divided by usual hours of work per week for others. I multiply weekly earnings values by 1.5 for individuals who report a top-coded weekly earnings value of $2,884.61 and exclude allocated wage observations, following Acemoglu and Autor (2011). The CPS occupation codes are very similar to the ACS codes, with some additional detailed occupations. Therefore, I use the crosswalk between the SOC codes and the ACS occupation codes to link the O*NET measures to the CPS MORG data, after adjusting the CPS codes to match the ACS codes.

The regression specification that I use to estimate the task wage differentials in the CPS MORG includes controls for race and ethnicity, educational attainment, potential experience and its square, year, region, metropolitan area status and union coverage. The race and ethnicity variables consist of mutually exclusive indicators for Black, Hispanic, and other race, with White non-Hispanic as the omitted category. The education variables consist of indicators for less than a high school diploma, high school diploma, some college, associate degree and graduate degree, with bachelor's degree as the omitted category. Potential experience is measured as age minus years of education minus six. Union coverage is measured as an indicator for being a member of a union or covered by a union or employee association contract. I weight the regression by the CPS earnings weight.

PSID Data

The PSID is a longitudinal survey that follows a representative sample of U.S. households first surveyed in 1968 and their descendants. The analysis in Panel C of Table A17 uses data from the 2007, 2009 and 2011 PSID waves that are available in the replication package for Blau and Kahn (2017). The sample in the replication package is restricted to observations on individuals aged 25-64 who are not self-employed and have valid data on hourly wages. I also exclude individuals who are missing data on region of residence. The PSID data use occupation codes from the 2000 Census, which I link to the ACS codes and thus to the O*NET using a set of aggregated Census/ACS occupation codes that are consistent over time. These codes were developed by Autor and Dorn (2013) and updated by Deming (2017) and for this project.

The regression specification that I use to estimate the task wage differentials in the PSID includes measures constructed by Blau and Kahn of years of full-time work experience, years of part-time work experience, and the squares of full-time and part-time experience. The authors define full-time work as at least 1,500 hours per year, and part-time work as less than 1,500 hours but greater than zero hours. To construct the work experience measures, the authors use the longitudinal structure of the data, and impute missing values as necessary.

The wage regression also includes controls for race and ethnicity, educational attainment, year, region and metropolitan area status. The race and ethnicity variables consist of mutually exclusive indicators for Black, Hispanic, and other race, with White non-Hispanic as the omitted category. The educational attainment variables consist of years of education and indicators for having exactly a bachelor's degree and exactly a graduate degree. I weight the regression by the PSID family weight.

NLSY79

The NLSY79 is a longitudinal study of a nationally representative sample of individuals who were aged 14 to 22 in 1979, the first year of the survey. The analysis in Panel D of Table A17 uses data from the 1979 to 2016 waves of the NLSY79. I restrict the sample to observations on individuals aged 18-64 with valid data on hourly wages, region of residence, educational attainment, and cognitive, non-cognitive and social skills. The NLSY79 uses occupation codes from the 1970 Census for the waves conducted in 1979 to 1981, occupation codes from the 1980 Census for the 1982 to 2000 waves, and occupation codes from the 2000 Census for the 2002 to 2016 waves. I link the Census occupation codes to the ACS codes and the O*NET using the set of aggregated Census/ACS occupation codes developed by Autor and Dorn (2013).

The regression specification that I use to estimate the task wage differentials in the NLSY79 includes measures of cognitive, non-cognitive and social skills that I construct following Deming (2017). Specifically, I measure cognitive skill using a standardized version of scores on the Armed Forces Qualifying Test (AFQT) that have been adjusted by Altonji et al. (2012) to be comparable across the NLSY79 and NLSY97 (see further information on the NLSY97 below). The measure of non-cognitive skill consists of the standardized mean of scores on the Rotter Locus of Self-Control and Rosenberg Self-Esteem Scale. Finally, the measure of social skill comprises the standardized mean of sociability in adulthood (self-reported in 1985), sociability at age 6 (self-reported retrospectively in 1985), the number of clubs the respondent participated in during high school, and an indicator for whether the respondent participated in high school sports.

The wage regression also controls for race and ethnicity, educational attainment, potential experience and its square, year, region, metropolitan area status, and residence in an urban versus rural area. The race and ethnicity variables consist of mutually exclusive indicators for Black and Hispanic, with White non-Hispanic as the omitted category. The educational attainment variables consist of years of education and indicators for having exactly a bachelor's degree and exactly a graduate degree. I weight the regression using a weight variable constructed by Altonji et al. (2012).

NLSY97

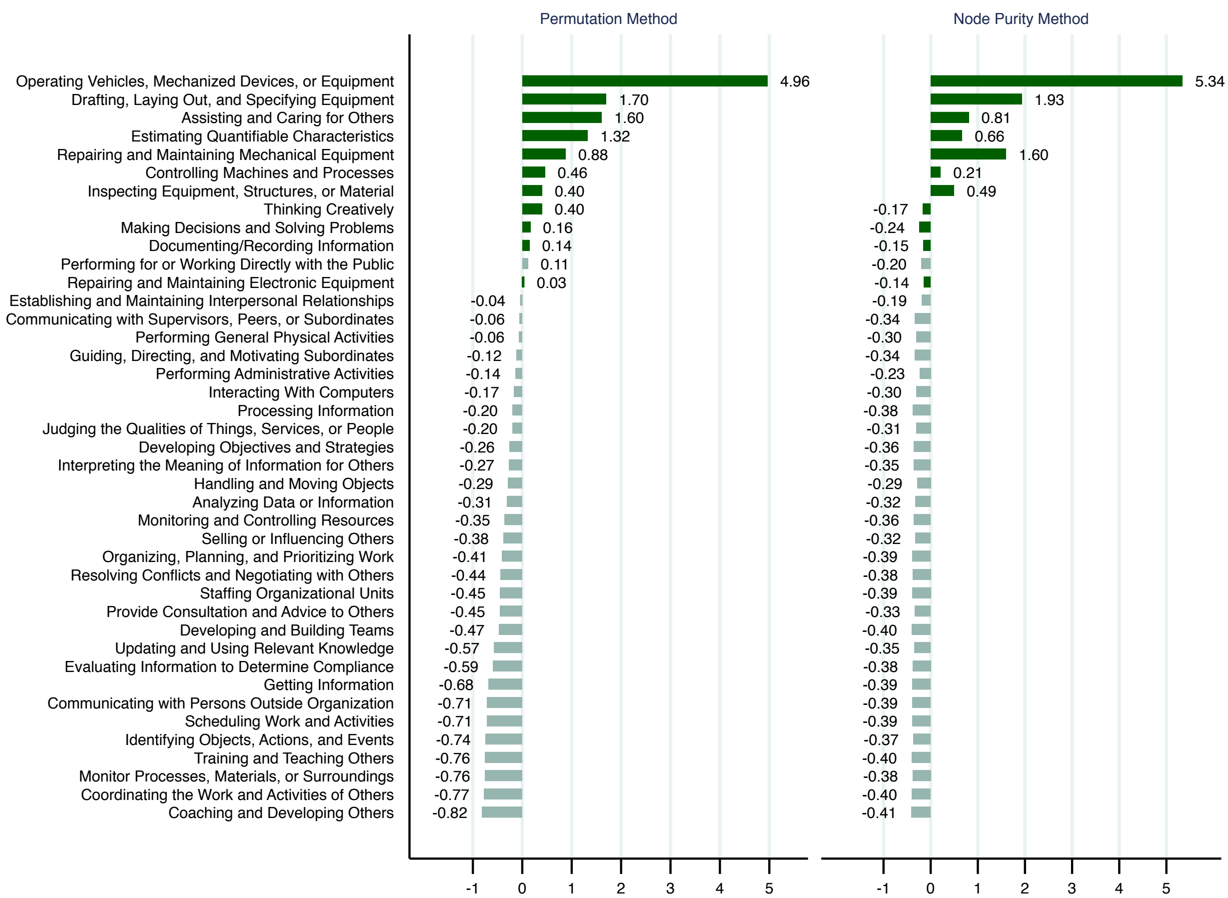
The NLSY97 is a longitudinal survey of a nationally representative sample of individuals aged 12 to 16 at the end of 1996 who were first interviewed in 1997. The analysis in Panel E of Table A17 uses data from the 1997 to 2015 waves of the NLSY97. I restrict the sample to observations on individuals aged 18-64 with valid data on hourly wages, region of residence, educational attainment, and cognitive, non-cognitive and social skills. The NLSY97 uses occupation codes from the 2000 Census. I link the Census occupation codes to the ACS codes and the O*NET using the set of aggregated Census/ACS occupation codes developed by Autor and Dorn (2013).

The regression specification that I use to estimate the task wage differentials in the NLSY97 includes measures of cognitive, non-cognitive and social skills that I construct following Deming (2017). Specifically, I measure cognitive skill using a standardized version of scores on the Armed Forces Qualifying Test (AFQT) that have been adjusted by Altonji et al. (2012), as described above. The measure of non-cognitive skill is the standardized mean of the self-reported personality traits organization, conscientiousness, dependability, thoroughness, trust, and discipline. Lastly, the measure of social skill comprises the standardized mean of the self-reported personality traits extraversion and animation (i.e. a negative score on a measure of being reserved or quiet).

The wage regression also controls for race and ethnicity, educational attainment, potential experience and its square, year, region, metropolitan area status, and residence in an urban versus rural area. The race and ethnicity variables consist of mutually exclusive indicators for Black and Hispanic, with White non-Hispanic as the omitted category. The education variables consist of indicators for less than a high school diploma, high school diploma, some college, associate degree and graduate degree, with bachelor's degree as the omitted category. I weight the regression using a weight variable constructed by Altonji et al. (2012).

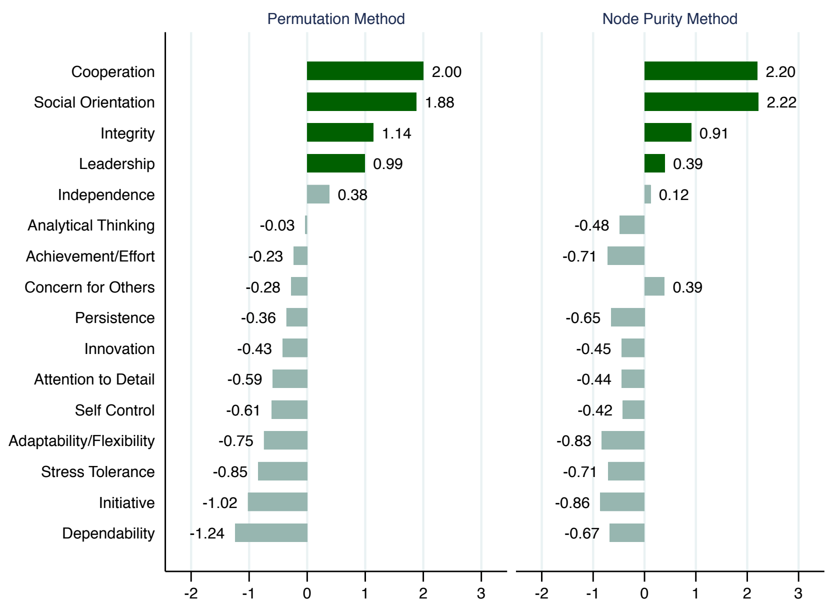
Appendix Figures

Figure A1: Variable Importance Scores - Work Activities



Notes. This figure displays the variable importance scores for the O*NET work activities generated by the random forest analysis. The importance scores are standardized to have a mean of zero and a standard deviation of one and are sorted according to the permutation scores. The measures designated as highly predictive are shown in dark green.

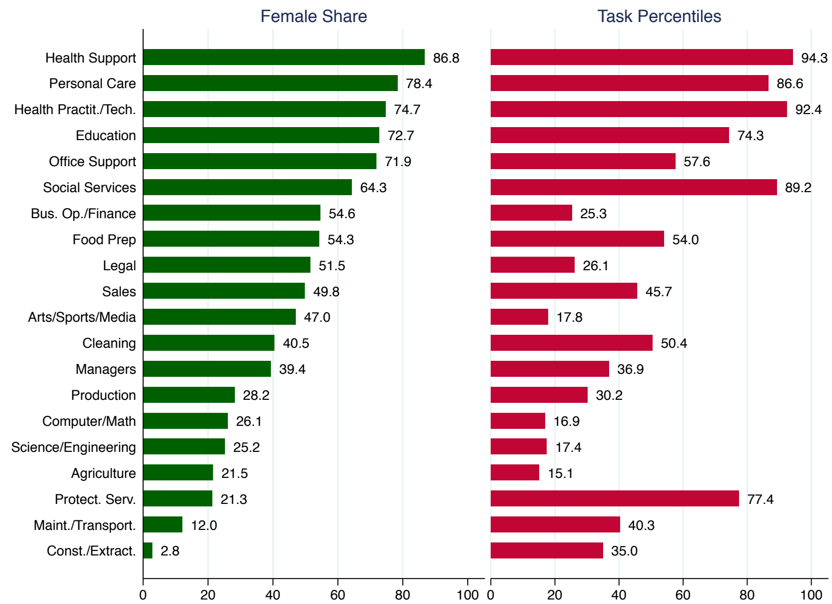
Figure A2: Variable Importance Scores - Work Styles



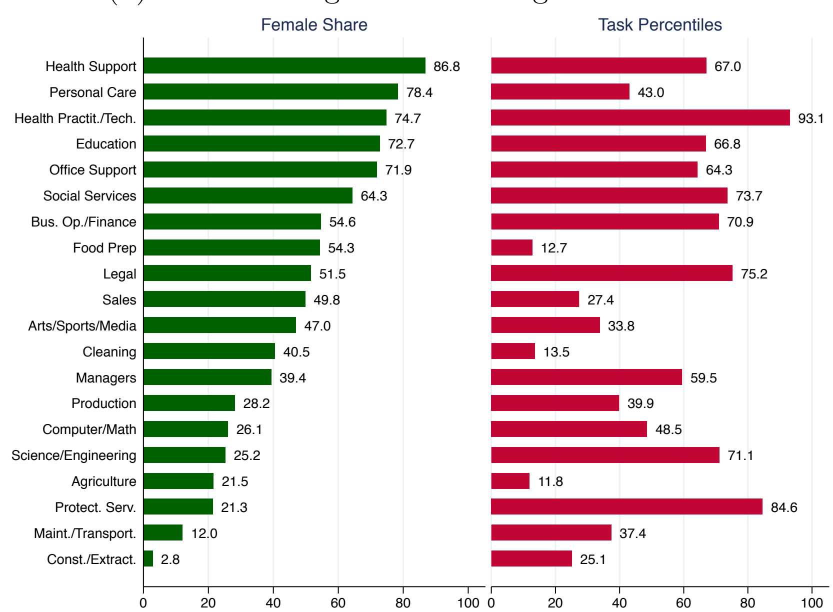
Notes. This figure displays the variable importance scores for the O*NET work styles generated by the random forest analysis. The importance scores are standardized to have a mean of zero and a standard deviation of one and are sorted according to the permutation scores. The measures designated as highly predictive are shown in dark green.

Figure A3: Female-Typical Tasks and Female Share by Occupation

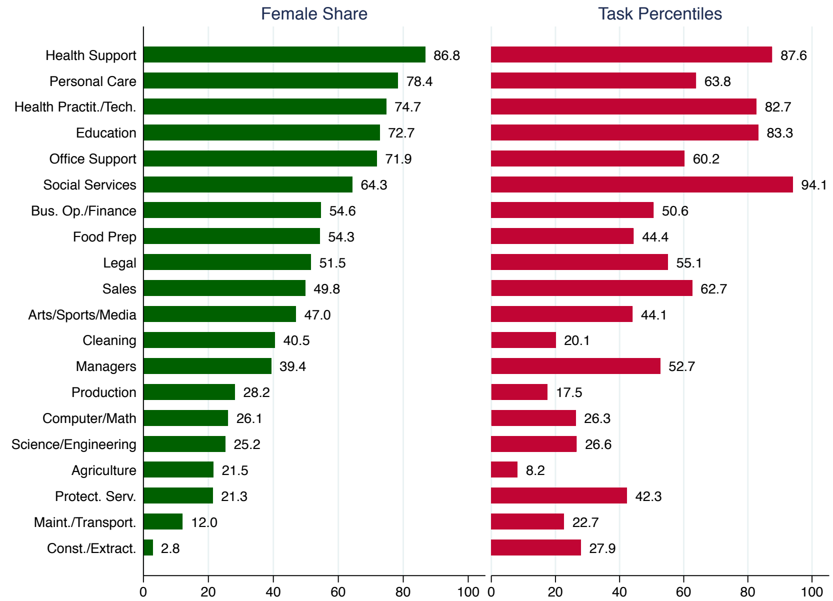
(a) Helping and Caring for Others



(b) Documenting and Recording Information



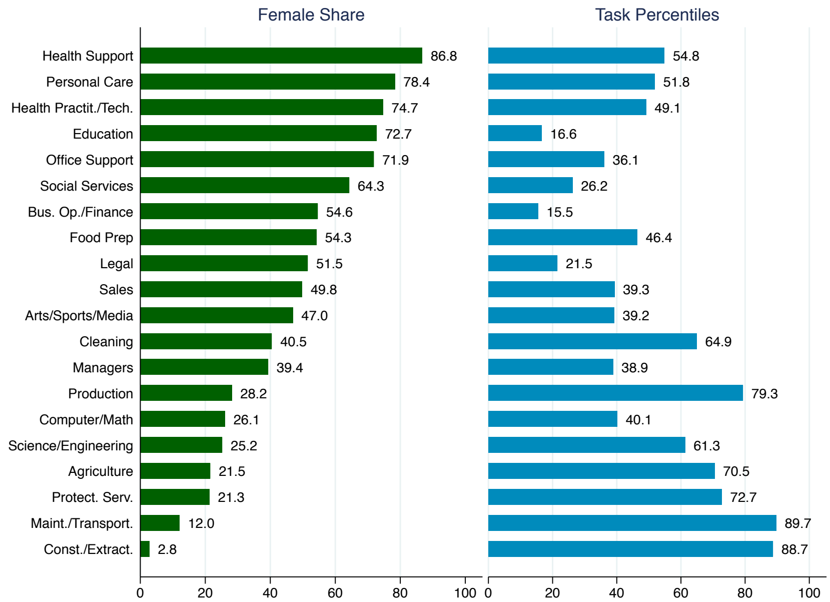
(c) Working and Communicating with Others



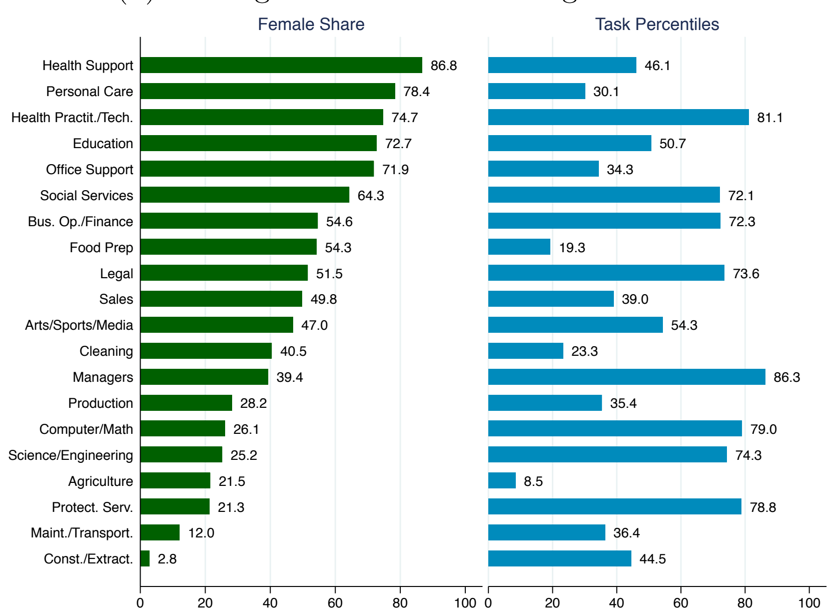
Notes. This figure displays the mean occupational female share in the 2012-2016 ACS (scaled from 0 to 100) and the mean levels of the selected female-typical tasks measured using the O*NET variables listed in Table 2, by major occupation group. Tasks based on multiple O*NET variables are constructed by averaging the component variables. Each task is rescaled to reflect percentiles weighted by employment in the 2012-2016 ACS.

Figure A4: Male-Typical Tasks and Female Share by Occupation

(a) Operating and Repairing Equipment



(b) Making Decisions and Solving Problems



Notes. This figure displays the mean occupational female share in the 2012-2016 ACS (scaled from 0 to 100) and the mean levels of the selected male-typical tasks measured using the O*NET variables listed in Table 2, by major occupation group. Tasks based on multiple O*NET variables are constructed by averaging the component variables. Each task is rescaled to reflect percentiles weighted by employment in the 2012-2016 ACS.

Figure A5: Reasons for Choices - Survey Text

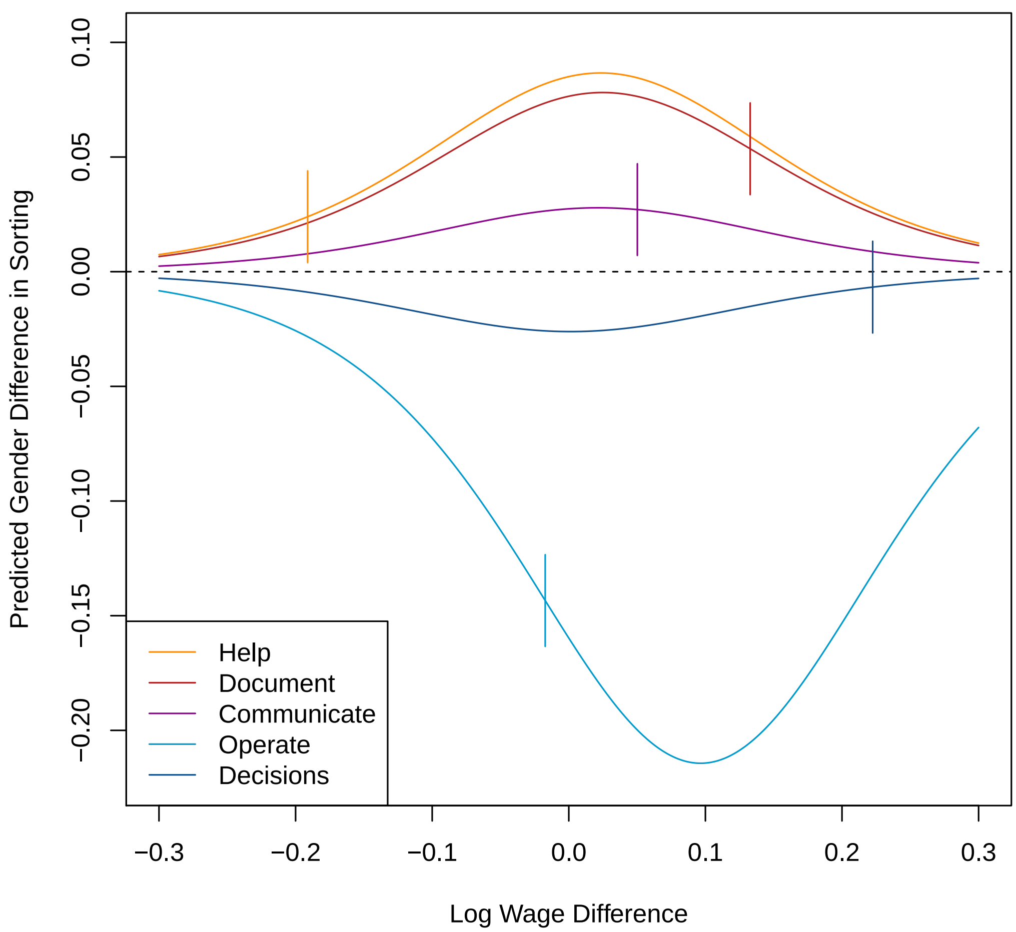
Why did you choose this job?

Select all that apply.

- This job is a better fit for my existing skills and abilities
- This job is more prestigious
- This job offers better pay
- There would be more people like me in this job
- This job sounds more enjoyable and/or interesting
- This job would allow me to strengthen or develop new skills
- This job would require less effort
- I would be treated better in this job
- Other reason (please specify):

Notes. This figure displays the wording of the response options available to experiment participants in the questions about the reasons for the choices made in the hypothetical scenarios. The order of responses is randomized, and participants may make multiple entries.

Figure A6: Predicted Sorting Differences by ω_k



Notes. This figure plots the predicted gender differences in sorting on the gender-tyipcal tasks ($\hat{Q}_k$) as a function of ω_k, the log difference in wage offers between the high-task and low-task jobs, using the coefficients from the specification in Table 9. The vertical lines show the estimated task wage differentials ($\hat{\omega}_k = b_k$) in the ACS from Table 11.

Appendix Tables

Table A1: Female Share Regressed on Work Activities

	(1)	(2)	(3)	(4)
Establishing and Maintaining Interpersonal Relationships	0.125**	0.002	-0.030*	
	(0.011)	(0.017)	(0.013)	
Performing Administrative Activities	0.104**	0.027*	0.028*	
	(0.012)	(0.013)	(0.012)	
Assisting and Caring for Others	0.104**	0.070**	0.083**	0.088**
	(0.011)	(0.015)	(0.013)	(0.010)
Performing for or Working Directly with the Public	0.094**	0.023	0.012	
	(0.012)	(0.015)	(0.010)	
Interacting With Computers	0.086**	0.010	-0.025	
	(0.013)	(0.018)	(0.017)	
Documenting/Recording Information	0.077**	0.016	0.031*	0.053**
	(0.013)	(0.014)	(0.013)	(0.009)
Resolving Conflicts and Negotiating with Others	0.067**	0.001		
	(0.013)	(0.016)		
Communicating with Persons Outside Organization	0.065**	-0.013		
	(0.012)	(0.016)		
Interpreting the Meaning of Information for Others	0.061**	0.027		
	(0.012)	(0.017)		
Organizing, Planning, and Prioritizing Work	0.055**	0.036*		
	(0.013)	(0.016)		
Getting Information	0.045**	-0.001		
	(0.013)	(0.016)		
Communicating with Supervisors, Peers, or Subordinates	0.044**	0.013		
	(0.013)	(0.015)		
Processing Information	0.040**	0.000		
	(0.013)	(0.020)		
Updating and Using Relevant Knowledge	0.033**	-0.012		
	(0.012)	(0.018)		
Selling or Influencing Others	0.026*	0.005		
	(0.012)	(0.012)		
Developing and Building Teams	0.020	0.020		
	(0.013)	(0.019)		
Coaching and Developing Others	0.014	-0.007		
	(0.013)	(0.023)		
Staffing Organizational Units	0.013	-0.010		
	(0.012)	(0.015)		
Analyzing Data or Information	0.012	0.008		
	(0.011)	(0.021)		
Provide Consultation and Advice to Others	0.006	-0.029*		
	(0.013)	(0.014)		
Training and Teaching Others	-0.003	0.029^{+}		
	(0.014)	(0.015)		

Continued on next page

Table A1 – *Continued from previous page*

	(1)	(2)	(3)	(4)
Scheduling Work and Activities	-0.004	0.005		
	(0.012)	(0.015)		
Developing Objectives and Strategies	-0.006	-0.024		
	(0.013)	(0.019)		
Thinking Creatively	-0.011	-0.004		0.001
	(0.012)	(0.016)		(0.012)
Evaluating Information to Determine Compliance	-0.012	-0.002		
	(0.013)	(0.013)		
Coordinating the Work and Activities of Others	-0.018	-0.020		
	(0.012)	(0.016)		
Judging the Qualities of Things, Services, or People	-0.024^{+}	0.014		
	(0.013)	(0.012)		
Monitoring and Controlling Resources	-0.027*	-0.005		
	(0.012)	(0.015)		
Guiding, Directing, and Motivating Subordinates	-0.028*	-0.033^{+}		
	(0.012)	(0.019)		
Making Decisions and Solving Problems	-0.031*	-0.041*		-0.059**
	(0.012)	(0.017)		(0.012)
Identifying Objects, Actions, and Events	-0.050**	-0.042**		
	(0.012)	(0.014)		
Monitor Processes, Materials, or Surroundings	-0.068**	0.040**		
	(0.012)	(0.014)		
Performing General Physical Activities	-0.104**	-0.007		
	(0.012)	(0.025)		
Handling and Moving Objects	-0.110**	0.037		
	(0.011)	(0.026)		
Estimating Quantifiable Characteristics	-0.121**	-0.010		-0.017
	(0.011)	(0.015)		(0.011)
Repairing and Maintaining Electronic Equipment	-0.137**	-0.005	0.004	-0.004
	(0.009)	(0.015)	(0.014)	(0.014)
Controlling Machines and Processes	-0.149**	-0.008	0.039*	0.024
	(0.010)	(0.023)	(0.019)	(0.018)
Inspecting Equipment, Structures, or Material	-0.160**	-0.027	-0.041*	-0.016
	(0.010)	(0.020)	(0.016)	(0.017)
Drafting, Laying Out, and Specifying Equipment	-0.167**	-0.035**	-0.060**	-0.044**
	(0.010)	(0.011)	(0.009)	(0.010)
Repairing and Maintaining Mechanical Equipment	-0.179**	-0.031	-0.052*	-0.045*
	(0.008)	(0.021)	(0.021)	(0.022)
Operating Vehicles, Mechanized Devices, or Equipment	-0.192**	-0.128**	-0.137**	-0.129**
	(0.008)	(0.016)	(0.014)	(0.013)
N		464	464	464
R^2		0.768	0.699	0.726
Adjusted R^2		0.745	0.691	0.719

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table displays coefficients from regressions of the occupational female share in the 2012-2016 ACS on O*NET variables from the work activities domain, standardized to have a mean of zero and standard deviation of one. Column 1 shows coefficients from bivariate regressions of the female share on each variable separately; the variables in this table are sorted by these coefficients, in descending order. Column 2 shows coefficients from the multivariate regression of the female share on all work activities. Column 3 shows coefficients from the female share regressed on the six work activities with the most positive bivariate coefficients and the six work activities with the most negative bivariate coefficients. Column 4 shows coefficients from the female share regressed on the variables designated as highly predictive by a random forest algorithm, as described in Appendix A1. Selected tasks are displayed in bold.

Table A2: Female Share Regressed on Work Styles

	(1)	(2)	(3)	(4)
Cooperation	0.139**	0.086**	0.125**	0.095**
	(0.012)	(0.022)	(0.020)	(0.020)
Social Orientation	0.131**	0.108**	0.106**	0.093**
	(0.011)	(0.022)	(0.020)	(0.017)
Concern for Others	0.124**	0.000		
	(0.011)	(0.019)		
Integrity	0.114**	0.078**		0.082**
	(0.012)	(0.017)		(0.013)
Self Control	0.103**	-0.056**		
	(0.013)	(0.021)		
Independence	0.092**	0.060**		
	(0.013)	(0.013)		
Adaptability/Flexibility	0.091**	0.026		
	(0.012)	(0.020)		
Dependability	0.090**	-0.021		
	(0.014)	(0.019)		
Stress Tolerance	0.086**	0.011		
	(0.013)	(0.020)		
Attention to Detail	0.052**	0.024		
	(0.013)	(0.016)		
Achievement/Effort	0.043**	0.042+		
	(0.012)	(0.025)		
Initiative	0.041**	0.040		
	(0.012)	(0.026)		
Persistence	0.018	-0.061*		
	(0.012)	(0.024)		
Innovation	0.004	-0.040*		
	(0.012)	(0.017)		
Leadership	0.003	-0.121**	-0.137**	-0.136**
	(0.012)	(0.016)	(0.015)	(0.013)
Analytical Thinking	-0.001	-0.038*	0.033*	
	(0.011)	(0.019)	(0.014)	
N		464	464	464
R^2		0.505	0.396	0.441
Adjusted R^2		0.487	0.391	0.436

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table displays coefficients from regressions of the occupational female share in the 2012-2016 ACS on O*NET variables from the work styles domain, standardized to have a mean of zero and standard deviation of one. Column 1 shows coefficients from bivariate regressions of the female share on each variable separately; the variables in this table are sorted by these coefficients, in descending order. Column 2 shows coefficients from the multivariate regression of the female share on all work styles. Column 3 shows coefficients from the female share regressed on the two work styles with the most positive bivariate coefficients and the two work styles with the most negative bivariate coefficients. Column 4 shows coefficients from the female share regressed on the variables designated as highly predictive by a random forest algorithm, as described in Appendix A1. Selected tasks are displayed in bold.

Table A3: Work Activities with Controls

	(1)	(2)	(3)	(4)
Establishing and Maintaining Interpersonal Relationships	0.062**	0.003	0.001	
	(0.012)	(0.016)	(0.013)	
Performing Administrative Activities	0.048**	0.020	0.010	
	(0.010)	(0.012)	(0.010)	
Assisting and Caring for Others	0.079**	0.068**	0.075**	0.075**
	(0.011)	(0.014)	(0.012)	(0.010)
Performing for or Working Directly with the Public	0.024*	0.010	-0.003	
	(0.011)	(0.014)	(0.009)	
Interacting With Computers	0.032*	-0.000	-0.002	
	(0.015)	(0.019)	(0.017)	
Documenting/Recording Information	0.063**	0.020	0.055**	0.063**
	(0.012)	(0.014)	(0.012)	(0.010)
Resolving Conflicts and Negotiating with Others	0.017	-0.005		
	(0.011)	(0.016)		
Communicating with Persons Outside Organization	-0.022*	-0.003		
	(0.011)	(0.015)		
Interpreting the Meaning of Information for Others	0.047**	0.024		
	(0.014)	(0.016)		
Organizing, Planning, and Prioritizing Work	0.022+	0.030+		
	(0.012)	(0.016)		
Getting Information	0.045**	0.007		
	(0.012)	(0.016)		
Communicating with Supervisors, Peers, or Subordinates	0.027*	0.009		
	(0.011)	(0.015)		
Processing Information	0.027*	-0.014		
	(0.014)	(0.020)		
Updating and Using Relevant Knowledge	0.027+	0.010		
	(0.014)	(0.017)		
Selling or Influencing Others	-0.032**	0.002		
	(0.010)	(0.012)		
Developing and Building Teams	0.007	0.022		
	(0.010)	(0.018)		
Coaching and Developing Others	0.006	-0.011		
	(0.010)	(0.021)		
Staffing Organizational Units	-0.015	-0.009		
	(0.010)	(0.014)		
Analyzing Data or Information	0.001	0.021		
	(0.017)	(0.021)		
Provide Consultation and Advice to Others	-0.015	-0.011		
	(0.013)	(0.014)		
Training and Teaching Others	0.007	0.011		
	(0.010)	(0.014)		

Continued on next page

Table A3 – *Continued from previous page*

	(1)	(2)	(3)	(4)
Scheduling Work and Activities	-0.027**	0.000		
	(0.010)	(0.014)		
Developing Objectives and Strategies	-0.040**	-0.021		
	(0.012)	(0.018)		
Thinking Creatively	-0.053**	-0.011		-0.003
	(0.011)	(0.015)		(0.012)
Evaluating Information to Determine Compliance	0.037**	0.008		
	(0.010)	(0.013)		
Coordinating the Work and Activities of Others	-0.010	-0.010		
	(0.010)	(0.016)		
Judging the Qualities of Things, Services, or People	-0.010	0.009		
	(0.010)	(0.011)		
Monitoring and Controlling Resources	-0.038**	-0.012		
	(0.010)	(0.014)		
Guiding, Directing, and Motivating Subordinates	-0.020*	-0.019		
	(0.010)	(0.018)		
Making Decisions and Solving Problems	-0.023+	-0.025		-0.014
	(0.012)	(0.016)		(0.012)
Identifying Objects, Actions, and Events	-0.008	-0.034*		
	(0.010)	(0.014)		
Monitor Processes, Materials, or Surroundings	0.017+	0.037**		
	(0.010)	(0.014)		
Performing General Physical Activities	-0.030+	-0.009		
	(0.016)	(0.025)		
Handling and Moving Objects	-0.026	0.022		
	(0.017)	(0.026)		
Estimating Quantifiable Characteristics	-0.067**	-0.013		-0.012
	(0.011)	(0.014)		(0.010)
Repairing and Maintaining Electronic Equipment	-0.046**	-0.012	-0.007	-0.010
	(0.011)	(0.014)	(0.014)	(0.013)
Controlling Machines and Processes	-0.046**	0.004	0.023	0.020
	(0.017)	(0.023)	(0.018)	(0.017)
Inspecting Equipment, Structures, or Material	-0.055**	-0.016	-0.014	-0.007
	(0.015)	(0.019)	(0.015)	(0.016)
Drafting, Laying Out, and Specifying Equipment	-0.084**	-0.029**	-0.037**	-0.031**
	(0.009)	(0.011)	(0.008)	(0.009)
Repairing and Maintaining Mechanical Equipment	-0.099**	-0.022	-0.044*	-0.040*
	(0.015)	(0.020)	(0.019)	(0.019)
Operating Vehicles, Mechanized Devices, or Equipment	-0.132**	-0.120**	-0.129**	-0.127**
	(0.012)	(0.017)	(0.014)	(0.013)
N		464	464	464
R^2		0.795	0.769	0.772
Adjusted R^2		0.772	0.759	0.763

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table displays coefficients from regressions of the occupational female share in the 2012-2016 ACS on O*NET variables from the work activities domain, standardized to have a mean of zero and standard deviation of one. Column 1 shows coefficients from regressions of the female share on each variable separately. Column 2 shows coefficients from the regression of the female share on all work activities. Column 3 shows coefficients from the female share regressed on the six work activities with the most positive bivariate coefficients and the six work activities with the most negative bivariate coefficients in Table A1. Column 4 shows coefficients from the female share regressed on the variables designated as highly predictive by a random forest algorithm, as described in Appendix A1. Selected tasks are displayed in bold. All regressions control for broad occupation cluster, mean log hourly wages and the share of workers with a college degree or more in each occupation in the ACS.

Table A4: Work Styles with Controls

	(1)	(2)	(3)	(4)
Cooperation	0.079**	0.032+	0.066**	0.059**
	(0.011)	(0.017)	(0.015)	(0.015)
Social Orientation	0.074**	0.069**	0.076**	0.069**
	(0.011)	(0.018)	(0.016)	(0.015)
Concern for Others	0.078**	0.026		
	(0.010)	(0.018)		
Integrity	0.068**	0.005		0.037**
	(0.012)	(0.015)		(0.012)
Self Control	0.061**	-0.010		
	(0.010)	(0.018)		
Independence	0.063**	0.054**		
	(0.010)	(0.011)		
Adaptability/Flexibility	0.054**	0.011		
	(0.012)	(0.016)		
Dependability	0.054**	-0.016		
	(0.011)	(0.014)		
Stress Tolerance	0.048**	0.003		
	(0.011)	(0.017)		
Attention to Detail	0.056**	0.046**		
	(0.010)	(0.011)		
Achievement/Effort	0.029*	0.031		
	(0.013)	(0.020)		
Initiative	0.021+	0.019		
	(0.013)	(0.019)		
Persistence	0.002	-0.048*		
	(0.013)	(0.019)		
Innovation	-0.004	-0.041**		
	(0.011)	(0.013)		
Leadership	-0.001	-0.071**	-0.088**	-0.086**
	(0.011)	(0.014)	(0.013)	(0.013)
Analytical Thinking	0.004	0.002	0.021	
	(0.014)	(0.016)	(0.014)	
N		464	464	464
R^2		0.689	0.632	0.638
Adjusted R^2		0.674	0.624	0.630

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table displays coefficients from regressions of the occupational female share in the 2012-2016 ACS on O*NET variables from the work styles domain, standardized to have a mean of zero and standard deviation of one. Column 1 shows coefficients from regressions of the female share on each variable separately. Column 2 shows coefficients from the regression of the female share on all work styles. Column 3 shows coefficients from the female share regressed on the two work styles with the most positive bivariate coefficients and the two work styles with the most negative bivariate coefficients in Table A2. Column 4 shows coefficients from the female share regressed on the variables designated as highly predictive by a random forest algorithm, as described in Appendix A1. Selected tasks are displayed in bold. All regressions control for broad occupation cluster, mean log hourly wages and the share of workers with a college degree or more in each occupation in the ACS.

Table A5: Occupations with Highest and Lowest Levels -

Helping and Caring for Others

Rank	Occupation	Female Share
1	Nurse anesthetists	0.592
2	Nurse practitioners and nurse midwives	0.911
3	Podiatrists	0.248
4	Licensed practical and licensed vocational nurses	0.887
5	Registered nurses	0.898
6	Dental assistants	0.942
7	Nursing, psychiatric, and home health hides	0.869
8	Recreational therapists	0.765
9	Medical assistants	0.914
10	Personal care aides	0.836
455	Fence erectors	0.017
456	Surveying and mapping technicians	0.093
457	Record keeping weighers, measurers, checkers, and samplers	0.462
458	Miscellaneous legal support workers	0.730
459	Economists	0.334
460	Computer programmers	0.218
461	Proofreaders and copy markers	0.712
462	Automotive glass installers and repairers	0.022
463	Actuaries	0.348
464	Financial analysts	0.388

Notes. This table displays the ten occupations ranked the highest and the ten occupations ranked the lowest in terms of the level of the task *helping and caring for others*, along with the share of workers who are female in each occupation based on the 2012-2016 ACS. Tasks are measured using the O*NET variables listed in Table 2 and are standardized to have a mean of zero and standard deviation of one.

Table A6: Occupations with Highest and Lowest Levels -

Documenting and Recording Information

Rank	Occupation	Female Share
1	Licensed practical and licensed vocational nurses	0.887
2	Nurse anesthetists	0.592
3	Nurse practitioners and nurse midwives	0.911
4	Registered nurses	0.898
5	Medical assistants	0.914
6	Physical therapists	0.696
7	Personal care aides	0.836
8	Optometrists	0.420
9	Medical transcriptionists	0.940
10	Respiratory therapists	0.645
455	Structural iron and steel workers	0.027
456	Shoe and leather workers	0.313
457	Sewing machine operators	0.743
458	Drywall installers, ceiling tile installers, and tapers	0.025
459	Food preparation workers	0.586
460	Transportation attendants, except flight attendants	0.586
461	Pressers, textile, garment, and related materials	0.667
462	Graders and sorters of agricultural products	0.690
463	Laundry and dry-cleaning workers	0.595
464	Tire builders	0.089

Notes. This table displays the ten occupations ranked the highest and the ten occupations ranked the lowest in terms of the level of the task *documenting and recording information*, along with the share of workers who are female in each occupation based on the 2012-2016 ACS. Tasks are measured using the O*NET variables listed in Table 2 and are standardized to have a mean of zero and standard deviation of one.

Table A7: Occupations with Highest and Lowest Levels -

Working and Communicating with Others

Rank	Occupation	Female Share
1	Flight attendants	0.776
2	Dental assistants	0.942
3	Elementary and middle school teachers	0.790
4	Actors	0.432
5	Reservation and transportation ticket agents and travel clerks	0.595
6	Meeting, convention, and event planners	0.767
7	Respiratory therapists	0.645
8	Social workers	0.810
9	Special education teachers	0.854
10	Counselors	0.715
455	Couriers and messengers	0.164
456	Printing press operators	0.204
457	Metal furnace operators, tenders, pourers, and casters	0.076
458	Library technicians	0.755
459	Computer hardware engineers	0.160
460	Machinists	0.044
461	Motion picture projectionists	0.186
462	Graders and sorters of agricultural products	0.690
463	Economists	0.334
464	Computer, automated teller, and office machine repairers	0.111

Notes. This table displays the ten occupations ranked the highest and the ten occupations ranked the lowest in terms of the level of the task *working and communicating with others*, along with the share of workers who are female in each occupation based on the 2012-2016 ACS. Tasks are measured using the O*NET variables listed in Table 2 and are standardized to have a mean of zero and standard deviation of one.

Table A8: Occupations with Highest and Lowest Levels -

Operating and Repairing Equipment

Rank	Occupation	Female Share
1	Elevator installers and repairers	0.011
2	Heating, air conditioning, and refrigeration mechanics and installers	0.012
3	Aircraft mechanics and service technicians	0.053
4	Computer, automated teller, and office machine repairers	0.111
5	Coin, vending, and amusement machine servicers and repairers	0.172
6	Electrical power-line installers and repairers	0.011
7	Automotive service technicians and mechanics	0.015
8	Heavy vehicle / mobile equipment service technicians and mechanics	0.012
9	Electrical and electronics repairers	0.070
10	Machinery maintenance workers	0.037
455	Insurance underwriters	0.654
456	Financial analysts	0.388
457	Judicial law clerks	0.543
458	Operations research analysts	0.485
459	Management analysts	0.420
460	Actuaries	0.348
461	Economists	0.334
462	Compensation and benefits managers	0.770
463	Brokerage clerks	0.703
464	Compensation, benefits, and job analysis specialists	0.767

Notes. This table displays the ten occupations ranked the highest and the ten occupations ranked the lowest in terms of the level of the task *operating and repairing equipment*, along with the share of workers who are female in each occupation based on the 2012-2016 ACS. Tasks are measured using the O*NET variables listed in Table 2 and are standardized to have a mean of zero and standard deviation of one.

Table A9: Occupations with Highest and Lowest Levels -

Making Decisions and Solving Problems

Rank	Occupation	Female Share
1	Nurse anesthetists	0.592
2	Social and community service managers	0.681
3	Actuaries	0.348
4	Lawyers, and judges, magistrates, and other judicial workers	0.363
5	Air traffic controllers and airfield operations specialists	0.216
6	Management analysts	0.420
7	Physicians and surgeons	0.353
8	Operations research analysts	0.485
9	Biomedical and agricultural engineers	0.144
10	Chief executives and legislators	0.245
455	Pressers, textile, garment, and related materials	0.667
456	Roasting, baking, and drying machine operators and tenders	0.315
457	Textile knitting and weaving machine setters, operators, and tenders	0.598
458	Postal service mail carriers	0.392
459	Miscellaneous personal appearance workers	0.854
460	Shoe and leather workers	0.313
461	Miscellaneous agricultural workers including animal breeders	0.208
462	Cleaners of vehicles and equipment	0.151
463	Interviewers, except eligibility and loan	0.796
464	Graders and sorters of agricultural products	0.690

Notes. This table displays the ten occupations ranked the highest and the ten occupations ranked the lowest in terms of the level of the task *making decisions and solving problems*, along with the share of workers who are female in each occupation based on the 2012-2016 ACS. Tasks are measured using the O*NET variables listed in Table 2 and are standardized to have a mean of zero and standard deviation of one.

Table A10: Summary Statistics - Employed and Non-Employed

	Experiment			ACS		
	All	Women	Men	All	Women	Men
Female	0.534	1.000	0.000	0.514	1.000	0.000
Age	34.4	35.0	33.7	47.0	47.9	46.0
White	0.720	0.770	0.663	0.650	0.648	0.652
Black	0.079	0.070	0.090	0.119	0.124	0.114
Hispanic	0.109	0.077	0.146	0.152	0.147	0.157
Other Race	0.091	0.083	0.101	0.079	0.081	0.076
HS or less	0.091	0.092	0.089	0.411	0.393	0.430
Some college	0.238	0.232	0.245	0.235	0.239	0.230
Associate's degree	0.124	0.143	0.103	0.078	0.086	0.069
Bachelor's degree	0.400	0.372	0.433	0.176	0.180	0.171
Graduate degree	0.146	0.161	0.130	0.101	0.101	0.100
Employed	0.902	0.889	0.917	0.602	0.553	0.654
N	1,931	1,031	900	12,330,760	6,393,549	5,937,211

Notes. This table shows summary statistics in the experiment sample compared with the 2012-2016 ACS, including all experiment participants and all individuals in the ACS aged 18 and older.

Table A11: WTP for Tasks - College vs. Non-College Workers

(a) College degree or more

	Help	Document	Communic.	Operate	Decisions
All	-0.023**	-0.021**	-0.014*	-0.126**	0.019**
	(0.008)	(0.005)	(0.006)	(0.011)	(0.006)
Women	-0.011	-0.009	-0.014+	-0.160**	0.020**
	(0.011)	(0.006)	(0.008)	(0.016)	(0.008)
Men	-0.037**	-0.036**	-0.016+	-0.088**	0.017+
	(0.010)	(0.009)	(0.008)	(0.014)	(0.009)
Diff (W-M)	0.026+	0.028**	0.002	-0.072**	0.004
	(0.015)	(0.011)	(0.011)	(0.022)	(0.012)
N	1,050	1,050	1,050	1,050	1,050

(b) Less than college degree

	Help	Document	Communic.	Operate	Decisions
All	-0.029**	-0.021**	-0.027**	-0.107**	-0.026**
	(0.009)	(0.006)	(0.007)	(0.012)	(0.006)
Women	-0.009	-0.010	-0.019*	-0.144**	-0.035**
	(0.011)	(0.008)	(0.009)	(0.015)	(0.009)
Men	-0.055**	-0.033**	-0.038**	-0.051**	-0.015+
	(0.014)	(0.009)	(0.011)	(0.015)	(0.009)
Diff (W-M)	0.045**	0.022+	0.019	-0.093**	-0.020
	(0.017)	(0.012)	(0.014)	(0.022)	(0.013)
N	872	872	872	872	872

(c) Difference (College − Non-College)

	Help	Document	Communic.	Operate	Decisions
All	0.006	-0.001	0.013	-0.019	0.045^{**}
	(0.011)	(0.008)	(0.009)	(0.016)	(0.009)
Women	-0.002	0.002	0.006	-0.016	0.056^{**}
	(0.015)	(0.010)	(0.012)	(0.022)	(0.012)
Men	0.018	-0.004	0.022^{+}	-0.037^{+}	0.032^{*}
	(0.017)	(0.012)	(0.013)	(0.021)	(0.013)
Diff (W-M)	-0.020	0.006	-0.017	0.020	0.024
	(0.023)	(0.016)	(0.018)	(0.031)	(0.017)
N	1,922	1,922	1,922	1,922	1,922

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table shows WTP estimates for experiment participants with a four-year college degree or more in Panel A, WTP estimates for participants with less than a college degree in Panel B, and the difference between the estimates in Panel C.

Table A12: WTP for Tasks - No Inattention

	Help	Document	Communic.	Operate	Decisions
All	-0.026**	-0.017**	-0.025**	-0.119**	0.000
	(0.006)	(0.004)	(0.004)	(0.008)	(0.004)
Women	-0.010	-0.005	-0.021**	-0.151**	-0.003
	(0.008)	(0.005)	(0.006)	(0.011)	(0.006)
Men	-0.043**	-0.033**	-0.030**	-0.079**	0.005
	(0.008)	(0.006)	(0.006)	(0.010)	(0.007)
Diff (W-M)	0.033**	0.028**	0.009	-0.072**	-0.008
	(0.011)	(0.008)	(0.009)	(0.015)	(0.009)
N	1,588	1,588	1,588	1,588	1,588

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table shows WTP estimates excluding participants who are inattentive. Inattention is measured by asking participants at the end of the survey to indicate the decisions they made in the hypothetical choice experiment for a randomly selected two of the five gender-typical tasks. A participant is considered inattentive if they answer either question incorrectly.

Table A13: WTP for Tasks - Employed

	Help	Document	Communic.	Operate	Decisions
All	-0.025**	-0.021**	-0.021**	-0.119**	-0.000
	(0.006)	(0.004)	(0.005)	(0.009)	(0.004)
Women	-0.008	-0.007	-0.018**	-0.156**	-0.006
	(0.008)	(0.005)	(0.006)	(0.012)	(0.006)
Men	-0.043**	-0.036**	-0.023**	-0.074**	0.006
	(0.008)	(0.006)	(0.007)	(0.011)	(0.007)
Diff (W-M)	0.035**	0.029**	0.005	-0.082**	-0.012
	(0.012)	(0.008)	(0.009)	(0.016)	(0.009)
N	1,742	1,742	1,742	1,742	1,742

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$
This table shows WTP estimates excluding participants who are not currently employed.

Table A14: WTP for Tasks - No Existing Skills Motivation

	Help	Document	Communic.	Operate	Decisions
All	-0.024*	-0.047**	-0.033**	-0.060**	-0.017*
	(0.010)	(0.007)	(0.008)	(0.010)	(0.007)
Women	-0.016	-0.032**	-0.028*	-0.080**	-0.017^{+}
	(0.015)	(0.009)	(0.011)	(0.013)	(0.010)
Men	-0.032*	-0.060**	-0.039**	-0.036*	-0.017
	(0.015)	(0.011)	(0.012)	(0.015)	(0.011)
Diff (W-M)	0.015	0.028*	0.010	-0.043*	-0.000
	(0.021)	(0.014)	(0.016)	(0.020)	(0.015)
N	461	544	523	496	532

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table shows WTP estimates excluding data from hypothetical scenarios for which participants cite a better fit for existing skills as a motivation for their choice. This table also restricts the data to the randomly selected two tasks for which each participant is asked to give a reason for their choice.

Table A15: WTP for Tasks - No Develop New Skills Motivation

	Help	Document	Communic.	Operate	Decisions
All	-0.045**	-0.034**	-0.049**	-0.130**	-0.024**
	(0.011)	(0.006)	(0.008)	(0.015)	(0.007)
Women	-0.028*	-0.018*	-0.039**	-0.167**	-0.024*
	(0.014)	(0.008)	(0.010)	(0.021)	(0.010)
Men	-0.070**	-0.049**	-0.062**	-0.078**	-0.025*
	(0.020)	(0.009)	(0.013)	(0.019)	(0.011)
Diff (W-M)	0.042+	0.031*	0.022	-0.089**	0.001
	(0.024)	(0.012)	(0.016)	(0.028)	(0.015)
N	617	672	632	659	598

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table shows WTP estimates excluding data from hypothetical scenarios for which participants cite developing new skills as a motivation for their choice. This table also restricts the data to the randomly selected two tasks for which each participant is asked to give a reason for their choice.

Table A16: WTP for Tasks - Difference in Differences

Help − Operate	0.113^{**}
	(0.019)
Document − Operate	0.106^{**}
	(0.017)
Communicate − Operate	0.088^{**}
	(0.018)
Help − Decisions	0.041^{**}
	(0.014)
Document − Decisions	0.034^{**}
	(0.012)
Communicate − Decisions	0.017
	(0.012)
N	1,931

Notes. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

This table shows the difference in differences in the WTP estimates across gender and tasks. The point estimates correspond to the difference across tasks in the estimates of β_k reported in the last row of Table 4.

Table A17: Sorting and Segregation - Robustness Checks

(a) CPS MORG

	Q	I	b	$\hat{Q}$	$\hat{Q}/Q$	$\hat{I}$	$\hat{I}/I$
Help	0.276		-0.222	0.017	0.063		
Document	0.147		0.131	0.054	0.369		
Communic.	0.264		0.045	0.027	0.104		
Operate	-0.312		-0.011	-0.149	0.478		
Decisions	-0.006		0.238	-0.006	0.956		
Index		0.389				0.161	0.413
N			648,149				

(b) Additional Amenities in ACS

	Q	I	b	$\hat{Q}$	$\hat{Q}/Q$	$\hat{I}$	$\hat{I}/I$
Help	0.270		-0.087	0.058	0.216		
Document	0.141		0.065	0.074	0.524		
Communic.	0.259		0.016	0.028	0.107		
Operate	-0.305		-0.039	-0.122	0.402		
Decisions	-0.011		0.079	-0.021	1.887		
Index		0.389				0.154	0.395
N			6,419,869				

(c) PSID

	Q	I	b	$\hat{Q}$	$\hat{Q}/Q$	$\hat{I}$	$\hat{I}/I$
Help	0.268		-0.153	0.034	0.128		
Document	0.197		0.127	0.056	0.282		
Communic.	0.286		0.014	0.028	0.097		
Operate	-0.311		-0.068	-0.097	0.312		
Decisions	-0.072		0.244	-0.005	0.075		
Index		0.451				0.127	0.282
N			17,576				

(d) NLSY79

	Q	I	b	$\hat{Q}$	$\hat{Q}/Q$	$\hat{I}$	$\hat{I}/I$
Help	0.279		-0.127	0.043	0.156		
Document	0.122		0.161	0.044	0.358		
Communic.	0.257		0.022	0.028	0.108		
Operate	-0.353		0.035	-0.190	0.538		
Decisions	-0.095		0.106	-0.018	0.191		
Index		0.400				0.192	0.479
N			170,211				

(e) NLSY97

	Q	I	b	$\hat{Q}$	$\hat{Q}/Q$	$\hat{I}$	$\hat{I}/I$
Help	0.282		-0.157	0.033	0.118		
Document	0.100		0.163	0.043	0.432		
Communic.	0.304		-0.036	0.025	0.082		
Operate	-0.312		0.025	-0.182	0.582		
Decisions	-0.045		0.177	-0.010	0.232		
Index		0.380				0.185	0.488
N			63,230				

Notes. This table shows observed and predicted gender differences in sorting on the gender-typical tasks and task-based segregation in the CPS MORG (Panel A), the ACS with additional job amenities included in the wage regression (Panel B), the PSID (Panel C), the NLSY79 (Panel D), and the NLSY97 (Panel E). See Appendix A1 for details of sample and variable construction in the additional datasets.

Q ($\hat{Q}$) is the observed (predicted) gender difference in the share of workers sorting into the high-task job. b is the coefficient on the task from the wage regression. I ($\hat{I}$) is the value of the segregation index based on observed (predicted) sorting.