Gender Differences in Recognition for Group Work

Heather Sarsons, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram*

May 7, 2020

Forthcoming in the *Journal of Political Economy*

Abstract

We study whether gender influences credit attribution for group work using observational data and two experiments. We use data from academic economists to test whether coauthorship matters differently for tenure for men and women. We find that conditional on quality and other observables, men are tenured similarly regardless of whether they coauthor or solo-author. Women, however, are less likely to receive tenure the more they coauthor. We then conduct two experiments that demonstrate that biases in credit attribution in settings without confounds exist. Taken together, our results are best explained by gender and stereotypes influencing credit attribution for group work.

* Sarsons, University of Chicago Booth (heather.sarsons@chicagobooth.edu); Gërxhani, European University Institute; Reuben, New York University Abu Dhabi and LISER; Schram, Amsterdam School of Economics and European University Institute. This paper subsumes a paper of the same title that Sarsons submitted to JPE in 2017. During the review process, Sarsons was asked by Editor Kamenica to add additional material to the paper in the form of an experiment. The experiment she designed and ran (Experiment 1 in this paper) had overlap with an experiment that Gërxhani and Schram had discussed with Sarsons in 2016 and with an experiment that Gërxhani, Reuben and Schram subsequently designed and ran (Experiment 2 in this paper). Gërxhani, Reuben and Schram wrote to Sarsons and the JPE after seeing Sarsons' paper that was, at the time, conditionally accepted at the JPE. At this point, Editor Kamenica turned the paper over to Editor List due to potential conflict issues (Sarsons was now his colleague). After some deliberations, the authors agreed to merge the projects. Editor List then had the new joint study peer reviewed. We thank both editors and four anonymous referees. Sarsons especially thanks Roland Fryer, Claudia Goldin, Larry Katz, David Laibson, and Amanda Pallais for their guidance and encouragement. We thank the editor, John List, and four anonymous referees. We also thank Mitra Akhtari, Amitabh Chandra, John Coglianese, Oren Danieli, Ellora Derenoncourt, Florian Ederer, Ben Enke, Raissa Fabregas, Nicole Fortin, Nickolas Gagnon, Peter Ganong, Edward Glaeser, Siri Isaksson, Emir Kamenica, Sara Lowes, Rob McMillan, Eduardo Montero, Gautam Rao, Alex Segura, Nihar Shah, Randolph Sloof, Peter Tu, Jeroen van de Ven, Justin Wolfers, and various conference and seminar participants for their helpful comments and suggestions.

1 Introduction

Do employers use gender when allocating credit for group work, particularly when individual contributions are unobserved? Organizations increasingly rely on group work for production (Lazear and Shaw, 2007), yet there is little empirical evidence documenting how credit for group work is allocated. Unless employers can perfectly observe each worker's contribution to the team's output, they must decide how to allocate credit without having full information as to what each worker did. This could leave room for demographic characteristics, such as gender, to influence the allocation of credit.

In this paper, we test whether uncertainty over an individual's contribution to a project leads to differential attribution of credit that contributes to the gender promotion gap. In many industries, women are not only hired at lower rates than men are, they are also promoted at lower rates.[1] This paper explores whether gender differences in credit for group work exist and whether they explain part of the promotion gap.

We primarily look at the tenure decisions of academic economists to test whether gender influences the allocation of credit for coauthored papers. Economics is a relevant setting as there is a large tenure gap between men and women, and because the amount of coauthoring has risen dramatically in recent years (Ginther and Kahn, 2004; Hammermesh, 2013). Using data from economists' CVs, we track individuals' career trajectories and compare whether the trajectory is different for individuals who coauthor versus solo-author, and whether there is a difference by gender.

Within economics, we find that men and women who solo-author most of their work have similar tenure rates conditional on a proxy for the quality of papers. However, an additional coauthored paper is correlated with an 8.2% increase in tenure probability for men but only a 5.6% increase for women. This gap is significantly less pronounced for women who coauthor with women, suggesting that the attribution of credit is related to the gender mix of coauthors. Furthermore, a man who coauthors is no less likely to receive tenure than a comparable man who solo-authors even though there is presumably more uncertainty as to how much work he did. A counterfactual exercise suggests that this difference in credit allocation explains 60% of the unconditional gender gap in tenure rates and 84% of the gap that remains after controlling for average paper quality, citations, tenure and PhD institution ranks, and field.

To ensure that we are not picking up on ability differences between men and women, we control for the quality of papers using both journal rankings and citations, allowing

[1] Blau and DeVaro (2007), for example, find that across jobs, women are less likely to be promoted than men conditional on worker's performance and ability ratings. In the UK, female managers are nearly 40% less likely to be promoted than male managers (Elmins et al., 2016).

for a comparison of men and women with similar research portfolios. The results are also robust to including other individual-level controls such as length of time to tenure and the seniority of one's coauthors, as well as tenure year, tenure institution, and primary field fixed effects.

We argue that these results are most consistent with a story of women receiving less credit for their joint work with men because of bias. To show this, we first use current CV and citation data to compare the productivity of men and women who did and did not receive tenure at the institution where they initially went up for tenure. While the estimates are imprecise, we find suggestive evidence that women who coauthor and are denied tenure produce more solo-authored papers that publish in high-ranking journals than men who are denied tenure. Data on citations show a similar result.

We then rule out several alternative explanations for the empirical patterns. For example, several papers have demonstrated that selection into coauthorship in economics is not random.[2] We test for selection into coauthorship and do not find any evidence that women coauthor with high ability or more senior men. We also look at the timing of coauthorship and find no evidence that women begin coauthoring if they have a slower start to their careers. The empirical patterns are also inconsistent with taste-basted discrimination.

Because the CV data do not allow us to rule out the possibility that women actually contribute less to papers that are coauthored with men, we conduct two experiments designed to test whether real or perceived differences in contributions drive credit allocation. In the first experiment, we first hire individuals to complete quizzes on topics that are either male or female-stereotyped. We then hire participants who act as "predictors" and are randomized into an individual treatment or a joint treatment. Predictors in the individual treatment are shown two individual's separate quiz scores while predictors in the joint treatment are shown the combined score of two individuals. They are then asked to predict the performance of each participant on future quizzes.

In the joint treatment, women are predicted to perform worse than their male counterparts for male-stereotyped quizzes, suggesting that predictors believe that women contributed less to the combined score (that is, they performed worse). However, if pairs performed a female-stereotyped quiz, women and men are given equal credit. To understand whether these results are driven by participants' beliefs about the ability distributions of men and women, we randomly provide some participants with the distribution of scores on the initial quiz by gender. Women appear to be given equal credit in the female-stereotyped quiz because participants view it as being gender-neutral. That is, they do not

[2] See, for example, Boschini and Sjögren (2007), Garcia and Sherman (2015), and Bikard et al (2015).

realize that women tend to outperform men. Showing participants the gender distribution of scores corrects this belief and women are then predicted to have a better performance in future female-stereotyped quizzes but it does not affect the predicted performance gap for women and men performing male-stereotyped tasks.

The second experiment is conducted in a more natural setting with human resources personnel. Following a similar design, we again test whether women are less likely than men to receive credit for good group performance. We additionally elicit the HR personnels' beliefs about male and female performance and find that differences in the allocation of credit are largely driven by differences in beliefs. We also find that male HR personnel are more likely to hire in favor of men, and women in favor of women.

This paper replicates and builds off of the results in Sarsons (2017), which shows the basic correlational patterns between paper composition and tenure. In this paper, we replicate the results using more data and then use the C.V. data and two experiments to establish a channel through which gender influences the allocation of credit. The paper also relates to a large literature seeking to understand difference in labor market outcomes between men and women. Factors such as productivity, personality and behavioural differences (such as competition aversion), and fertility preferences have been shown to explain some differences in career choice and progression.[3] In academia in particular, studies have pointed to both supply-side factors, including differences in subject matter interest (Dynan and Rouse, 1997) and the availability of role models (Hale and Regev, 2014; Carrell et al., 2010); demand-side factors, such as implicit bias (Milkman et al., 2015; Moss-Racusin et al., 2012); and institutional factors (Antecol et al., 2018). This paper directly tests whether the differential treatment of work output contributes to the gender gap, contributing to a literature documenting gender differences in men and women's behavior in teams (Coffman, 2014; Isaksson, 2019; Born et al., 2019).

The remainder of the paper is organized as follows. Section 2 describes the data and shows that a tenure gap exists between male and female economists. In Section 3, we show that the tenure gap closes as women produce more solo-authored papers but does not close as they produce more coauthored papers. Women have a consistently lower probability of tenure for each additional coauthored paper than men. We show that the results are robust to accounting for attrition, and to using different journal rankings and definitions of tenure. In Section 4, we argue that the results are in line with a story in which women receive less credit for joint work with men and discuss why women still coauthor with men despite the low returns to doing so. We also test alternative explanations of the

[3]There is a large literature documenting gender differences in productivity, attitudes toward different types of work, and family choices. See, for example, Niederle and Vesterlund (2007), Buser et al. (2014), Antecol et al. (2018), Ceci et al. (2014), Reuben et al. (2017), and Ginther and Kahn (2004).

relationship between coauthorship and tenure and argue that none can fully explain the observed empirical patterns. Section 5 discusses the design and results of the experiments. Section 6 discusses how we might expect coauthorships to evolve in the long-run, and section 7 concludes.

2 Data

To examine the relationship between paper composition and tenure, we construct a dataset using the CVs of economists who came up for tenure between 1985 and 2014 at one of the top 35 U.S. PhD-granting universities[4]. The academic progression documented in the CVs makes it possible to evaluate the relationship between an individual's research output and career progression. We can then compare the degree of collaborative work and reward for that work, and compare these results for men versus women.

2.1 Sample Selection and Data Overview

We include only PhD-granting institutions in the sample as tenure evaluation at these schools is primarily based on research output, of which we have a clear measure. Other institutions like liberal arts colleges place greater weight on teaching ability for tenure, something that we cannot measure. We exclude business and public policy schools for similar reasons.[5] It is reasonable to assume that the top 35 economics departments in the U.S. emphasize research which is measured by the number and quality of papers one produces.

One problem in collecting tenure information is that the CVs of individuals who went up for tenure, were denied it, and left to industry or government are difficult to find, leading to a sample selection problem. To deal with this issue, we collected historical faculty lists from 23 of the 35 schools and locate over 90% of faculty who had ever gone up for tenure at these 23 institutions. For the remaining 12 schools that did not have historical faculty lists available, we looked at the top 75 U.S. institutions, the top 5 Canadian institutions, and the top 5 European institutions to locate anyone who went up for tenure at a top 35 U.S. school and then moved to another institution. We also checked economists' CVs at the major Federal Reserve Boards and other large research institutes, such as Mathematica, in the U.S. While there might still be a sample selection problem, we show in Section

[4]The list of institutions are taken from the RePEc/IDEAS Economics Department rankings. The list of schools included can be found in Appendix C.

[5]Business and policy schools might also value teaching differently and put weight on different types of journals.

3.2.1 that the results are robust to using only the sample for which we have historical faculty lists.

From individuals' CVs, we code where and when they received their PhDs, their employment and publication history, and their primary and secondary fields. When looking at the relationship between publications and tenure in the main analysis, we only include papers that were published up to and including the year an individual goes up for tenure. Book chapters are not included in the paper count. In a robustness check, we include papers that were published one and two years after tenure.

To control for the quality of a person's publications, we primarily use the "AER equivalent" ranking measure developed by Kalaitzidakis et al. (2003). This measure converts journal publications into their equivalent number of American Economic Review papers.[6] Less than 10% of journal articles cannot be converted because the journal does not appear in the ranking. In these cases we give the publication a ranking of zero.[7]

Using the AER-equivalent measure instead of a list journal rank allows for different distances between journal ranks and for multiple journals to hold the same rank. For example, the top field journals can all hold the same rank. Other journal rankings force a ranking among these even though the journals might count the same amount toward tenure depending on one's field. For robustness, we replace this paper quality measure with the RePEc/IDEAS ranking of economics journals in Section 3.2.2.

Finally, we include citations, measured in 2015, of pre-tenure papers as a control variable. These citations were scraped from Google Scholar.

We supplement this dataset with results from a survey designed to measure individuals' beliefs about the returns to various types of papers. The survey also contains information on how frequently individuals present their papers. The exact questions and nature of the survey are discussed in greater detail in Section 4.

2.2 Construction of Tenure

To determine whether someone received tenure, we follow the guidelines on each school's website (as of 2015) as to when tenure decisions are made. The majority of schools require faculty to apply for tenure 7 years after their initial appointment. We therefore consider years 6-8 to be the "tenure window" in which someone applies for tenure to account for people who go up for tenure early or late (because of a leave of absence, for example).

[6]The American Economic Review is regarded as one of the top journals in economics. Most journal publications are therefore converted to be some fraction of an AER paper.

[7]If someone does not have any solo or coauthored papers, we set the relevant journal ranking to zero and include a dummy variable indicating that the individual has no solo (or coauthored) papers. This enables us to keep using the full sample.

We assume that an individual is denied tenure if s/he moves to a university ranked 5 positions below the initial institution during the tenure window. Similarly, we assume that an individual is denied tenure if s/he moves from academia to industry during the tenure window. Defining tenure in this way accounts for the fact that some people switch institutions 2-3 years after their initial appointment, not because they were denied tenure but for personal preferences, and that some people might choose to move to a comparable school around the time of tenure even though they were offered tenure at their original institution. For example, someone who moves from MIT to Harvard after 7 years was presumably offered tenure at MIT but chose to move to Harvard for other reasons.

As mentioned, a person who moves 5 or fewer years after his or her initial appointment is not assumed to have been denied tenure since s/he moved before the tenure window starts. If someone moves before the tenure window, we use the second institution they were at to determine tenure. For example, if a person's first job is at University A but s/he moves to University B after three years, we use University B as the tenure institution but do not start the tenure clock over. We do not restart the clock because the data shows that in over 80% of cases, the individual still appears to go up for tenure within 8 years of his or her appointment at the first institution. However, we do extend this tenure clock in a robustness check.

Individuals who move from an academic institution into industry before the tenure window are excluded from the sample.

2.3 Summary Statistics

Table 1 presents summary statistics of the data. Approximately 68% of the full sample received tenure, but this masks a stark difference between men and women. Only 52% of women received tenure while 73% of men did.

Total Papers, *Solo-authored*, and *Coauthored* are the number of papers in each group that an individual had published by the time of tenure. These publication counts do not include books or book chapters. Papers published in non-economics journals (such as a political science journal) are included but receive a ranking of 0 (the lowest ranking). The results are robust to excluding publications in non-economics journals.

There is no statistically significant difference in the number of papers that men and women produce. Panel B looks at differences in the quality of papers. Men are no more likely to publish their papers in "Top 5" journals (American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics, and The Review of Economic Studies) than women. The only statistically significant productivity difference is that men tend to publish their coauthored papers in slightly higher-ranking journals. Specifically, men's coauthored papers have an average ranking of 0.34 AER-equivalents while women's coauthored papers have an average ranking of 0.30 AER-equivalents. We therefore control for the quality of papers, measured using the AER-equivalent ranking as well as average citations, throughout the analysis.

Panel C displays differences in coauthoring patterns between men and women. *Number Unique CAs* is the number of unique coauthors an individual has had by tenure. Men and women have roughly the same number of coauthors but there are some differences in the types of people men and women coauthor with. For example, women are less likely to coauthor with senior faculty and more likely to coauthor with other assistant professors. This could in part be driven by the fact that they are also more likely to coauthor with other women, many of whom are also junior professors.

For illustrative purposes, we plot the number of women and men who have various combinations of solo and coauthored papers in Appendix Figure B1, as well as the average probability of receiving tenure for each paper combination in Appendix Figure B2. Most men and women have a similar combination of solo and coauthored papers. Appendix Figure B2 illustrates that individuals with a large number of either solo or coauthored papers are likely to receive tenure. However, Panel A suggests that women with a higher fraction of their papers that are solo-authored have a better chance of receiving tenure than women with a mix of solo and coauthored papers. We examine this claim formally in the next section.

3 Empirical Strategy and Results

3.1 Main Results

We show three main results. We first establish that a significant tenure gap exists between men and women. We then show that the gap becomes more pronounced the more women coauthor, and that women who solo-author all of their papers have comparable tenure rates to men. Finally, we show that the gender of a woman's coauthor matters. Women who coauthor with other women do not suffer a coauthor penalty.

3.1.1 The Tenure Gap

Figure 1 plots the coefficient $\hat{\beta}_1$ from estimating

$$T_{ifst} = \beta_1 TotPapers_i + \beta_2 TotPapers_i^2 + \gamma' Z_i + \theta_f + \theta_s + \theta_t + \epsilon_{ifst} \tag{1}$$

separately for men and women using OLS. The dependent variable, T_{ifst}, is an indicator that individual i in field f at school s receives tenure in year t. $TotPapers_i$ is the number of papers (both coauthored and solo-authored) individual i has at the time he or she went up for tenure. A quadratic in the number of papers is included to capture non-linearities in how publications matter for tenure. The vector of individual-level controls, Z_i, includes average journal rank (measured as average AER-equivalents), the log of total citations, the number of years it took i to go up for tenure, and the total number of coauthors on i's papers. Tenure institution (θ_s), tenure year (θ_t), and field fixed effects (θ_f) are also included as tenure standards likely vary over time and by field and department.

The figure shows that a significant tenure gap exists between men and women even after controlling for productivity, primary field, tenure institution, and tenure year. While an additional paper is correlated with a 13-16 percentage point increase in tenure probability for men and women, women are consistently 10-13 percentage points less likely to receive tenure than men conditional on having written the same number and quality of papers. The lower intercept for women could stem from tenure committees starting with a lower prior about women's ability. However, if all papers were clear signals of ability and tenure committees are Bayesian, we would expect the slope of the relationship between papers and tenure to be steeper for women. Put differently, if men and women received equal credit for papers, the coefficient on $TotPapers_i$ should be significantly larger for women than for men.

We provide a formal test for the difference in slopes for men and women in Column 1 of Table 2, where we present the estimates from

$$\begin{aligned} T_{ifst} = \beta_1 TotPapers_i + \beta_2 fem_i + \beta_3(TotPapers_i \times fem_i) + \beta_4 TotPapers_i^2 \\ + \gamma' Z_i + \theta_f + \theta_s + \theta_t + \epsilon_{ifst} \end{aligned} \tag{2}$$

This is similar to estimating equation 1 except that we interact total papers with a female dummy, fem_i rather than splitting the sample. There is no significant difference in the marginal benefit of an additional paper to men and women.

3.1.2 The Tenure Gap and Paper Composition

To test whether coauthored papers matter differently for men and women, we separate papers into those that are solo-authored and those that are coauthored and estimate

$$T_{ifst} = \beta_1 S_i + \beta_2(fem_i \times S_i) + \beta_3 CA_i + \beta_4(fem_i \times CA_i) + \delta_1 fem_i + \gamma' Z_i + \theta_f + \theta_s + \theta_t + \epsilon_{ifst} \tag{3}$$

using OLS. Here, S_i and CA_i are the number of solo-authored and coauthored papers an individual has at the time of tenure.

The results are presented in Table 2. An additional solo-authored paper is associated with a 9.7 percentage point increase in men's tenure rates and a 15.4 percentage point increase in women's tenure rates (who start from a lower base tenure rate). If the lower initial tenure rate for women is due to employers holding the belief that women are lower ability, it seems that the signals from solo papers begin to outweigh the employer's prior. This is consistent with a model in which employers start with a lower prior about women and update as they receive clear signals about a woman's ability, giving women full credit for this solo work. This is further discussed in the next section.

If coauthored papers are an unclear signal of ability, an employer must make a judgment call as to how much each coauthor contributed to the paper which could lead to differential attribution of credit. Indeed, we see that while an additional coauthored paper helps both men and women, men benefit much more than women. Men's tenure rates increase by 8.2 percentage points when they produce a coauthored paper whereas women's increase by 5.6 percentage points.

However, the fact that men benefit nearly as much from a coauthored paper as they do from a solo-authored paper is at odds with the story that employers are dividing credit for projects among authors. If employers do divide credit, not all men can get 100% of the credit, particularly for those papers coauthored with other men.[8] This result could point to an alternative mechanism. For example, if employers exhibit taste-based discrimination, they could use joint projects as an excuse to promote men over women. We discuss and test several such alternative stories in Section 4.

The relationship between paper composition and tenure is summarized in Figure 2. This figure plots the relationship between the fraction of an individual's papers that are solo-authored and tenure, controlling for the total number of papers, citations, journal quality, number of coauthors, and tenure institution, year, and field fixed effects. For men, it does not matter if one coauthors or solo-authors: tenure rates are comparable conditional on the quality of papers. Women who write all of their papers alone have similar tenure rates to men. However, women who coauthor all of their papers have an approximately 37% tenure rate, substantially lower than that of men who coauthor all of their papers (72%). The slope for women is $\hat{\beta} = 0.521$ and is statistically significant at the 1% level (s.e.=0.158).

[8]It could be the case that because tenure committees are evaluating one person, they always assume that the man they evaluate deserves full credit for the paper (and we do not see the amount of credit they would have given to the other man). It is impossible to evaluate such theories with these data.

3.1.3 Does Coauthor Gender Matter?

The probability of receiving tenure is not lower for all women who coauthor. In Table 3, we categorize coauthored papers into those written with only men, only women, or a mix of men and women:

$$\begin{aligned} T_{ifst} =& \beta_1 S_i + \beta_2 (fem_i \times S_i) + \beta_3 CAmale_i + \beta_4 (fem \times CAmale_i) + \beta_5 CAmix_i \\ &+ \beta_6 (fem \times CAmix_i) + \beta_7 CAfem_i + \beta_8 (fem_i \times CAfem_i) + \beta_9 fem_i \\ &+ \gamma' Z_i + \theta_f + \theta_s + \theta_t + \epsilon_{ifst} \end{aligned} \tag{4}$$

As before, S_i is the number of solo-authored papers individual i has at the time of tenure. $CAfem_i$ is the number of coauthored papers individual i has in which all of the coauthors are female. Similarly, $CAmale_i$ is the number of papers i has in which all of the coauthors are male and $CAmix_i$ is the number of papers i has in which the coauthors consist of men and women.

The estimated coefficients on the interaction terms show that the negative relationship between coauthoring and tenure for women is driven almost entirely by papers that are coauthored with men. While a coauthored paper with another man is associated with an 8.7 percentage point increase in tenure probability for a man, it is associated with a 3.1 percentage point increase in tenure probability for a woman.[9] An additional paper with a woman, however, is associated with an 11.6 percentage point increase in tenure probability for a woman. While this estimate is imprecise due to sample size, we can say that an additional coauthored paper with a woman has a more positive impact on tenure than an additional coauthored paper with a man. Any explanation as to why women have lower tenure rates than men when they coauthor must therefore be correlated with coauthor gender. The estimates are robust to including all of the control variables discussed earlier.

3.1.4 Counterfactual Analysis

We conduct a counterfactual analysis to estimate how much of the gender gap in tenure rates can be explained by the different treatment of coauthored papers. We first estimate

$$T_{ifst} = \beta_1 S_i + \beta_2 CA_i + \delta_1 fem_i + \gamma' Z_i + \theta_f + \theta_s + \theta_t + \epsilon_{ifst} \tag{5}$$

[9] These results again show the puzzling pattern that the amount of credit that is divided among male coauthors adds up to more than one.

and use the estimates to predict the probability of tenure, $\hat{T}_i$, for everyone in the sample. We then let the female dummy fem_i be 0 for everyone and predict tenure rates again (call this $\tilde{T}_i$). The difference $\hat{T}_i - \tilde{T}_i$ gives the gender gap in tenure rates conditional on all observable characteristics but not allowing for differences in the marginal impact of solo and coauthored papers for men and women.[10]

We then repeat this exercise using the estimates from equation 4, first letting the female dummy equal one and then predicting tenure rates again letting the female dummy (and therefore all of the interactions) equal zero. This second set of predicted tenure probabilities tells us what women's predicted tenure rate would be if their papers were treated in the same way that men's papers are treated.

The unconditional gender gap in tenure rates is 22 percentage points. The conditional gap in tenure rates from equation 5 is approximately 16 percentage points. Thus, observable characteristics such as differences in time to tenure and paper quality account for about 27% of the gap. The results from using equation 4 to predict tenure probabilities suggest that the gap would close by a further 13.5 percentage points if men and women's papers were treated similarly. The different assignment of credit thus accounts for approximately 60% of the unconditional tenure gap and 84% of the conditional gap.

3.2 Robustness Checks

One may be concerned that the results are a product of the types of productivity measures used or are affected by missing data. In this section, we show that the results are robust to using only the sample for which we have historical faculty lists, to using different journal rankings, to accounting for papers published shortly after tenure, and to using different measures of paper counts.[11]

3.2.1 Attrition

The results will be biased if the sample excludes individuals who are denied tenure and go into industry, government, or other institutions where we do not observe them. This would be particularly problematic if men who go to industry after being denied tenure disproportionately coauthored their papers. If this is true, we would be overestimating

[10] Interacting all variables except for the number of solo/coauthored papers with the female dummy does not substantially change the results.

[11] In Appendix Table A1, we also test whether the results vary by school rank and over time. The estimates suggest that the coauthoring penalty is driven largely by schools outside of the top 10, although the estimates are imprecise. The coauthorship penalty is also stronger in later years but again the estimates are imprecise.

the benefit of coauthoring for men. We would have a similar problem if women who go to industry after being denied tenure typically wrote solo-authored papers.

As discussed in Section 2.1, we attempted to find such individuals by searching institutions outside of the top 35 U.S. schools, federal reserves, and other research institutes. To further allay concerns about sample selection, we run the analysis on the sample for which we received historical faculty lists. These lists allow us to track who went up for tenure and find them even if they left academia. The results, presented in Column 1 of Table 4, do not change when run on the sample for which there should be very few missing observations. The coefficient on the $Female \times Coauthored$ interaction is significant only at the 10% level due to the smaller sample, but the direction and magnitude do not change.

3.2.2 Journal Rankings

In the main analysis, we use a flexible journal ranking that allows multiple journals to hold the same rank. However, while the economics profession largely agrees on what the "top" journals are, rankings of field journals or lower-tier journals have changed over time. In Columns 2-4 of Table 4, we show that the results are robust to using three alternative journal ranking metrics as controls.

In Column 2, we use the current RePEc-IDEAS journal ranking. This ranking forces a linear relationship between journals and tenure but also contains a larger number of journals. The main results do not change when using this ranking.

In Column 3, we allow journal rankings to change over time. We use historical rankings of economics journals (drawn from Laband and Piette, 1994, and combined with current rankings) and match each paper with its journal ranking at the time it was published. Using these rankings accounts for journals moving in rank over time as well as new journals being added. The coefficient on the $Female \times Coauthored$ interaction is slightly smaller but the same pattern persists. An additional coauthored paper is associated with an 8.1 percentage point increase in tenure probability for men and a 5.6 percentage point increase for women. In section 4, we also separate papers into "Top 5s" and "non-Top 5s".

Finally, in Column 4, we divide the AER-equivalent measure into deciles and control for the number of solo and coauthored papers an individual has in each decile. For example, if an individual publishes one solo-authored paper in the AER and another in the lowest-rank journal, she will have one paper in the tenth bin, one in the first bin, and zero in the others. Thus, instead of having a single coauthored or solo-authored paper rank control, we include ten variables controlling for the quality of an individual's solo-authored papers (the number of solo papers in each AER-equivalent bin) and ten variables controlling for the quality of an individual's coauthored paper (the number of coauthored papers in each AER-equivalent bin). Again, the results hold.

3.2.3 Tenure Definition

In the main analysis, we only consider papers that were published up to and including the year that an individual goes up for tenure. If an individual goes up for tenure in 1995, for example, papers published in 1996 are not included in the paper count even though they may have been "revise and resubmits" at the time of tenure. This could affect the results if men who coauthor have several promising unpublished papers at the time of tenure but women who coauthor do not, in which case we are not actually comparing people with similar publication records. In Columns 5 and 6 of Table 4, we include papers that are published one and two years after a person's tenure year in the paper count variables. The magnitude of the coefficients are smaller but the results do not change: women continue to benefit less from coauthored papers than men do.

3.2.4 Paper Count Variable

While we control for journal quality, the main independent variables (number of solo and coauthored papers) may not accurately reflect how tenure committees decide on tenure cases. For example, institutions might trade off the quantity and quality of papers in different ways. In Column 7 of Table 4, we use an alternative measure for the number of papers. Specifically, after converting each publication to its AER-equivalent, we add up the AER-equivalent measure to give the total number of "AERs" an individual has at the time of tenure. For example, if an individual published two solo-authored papers and one is worth 0.25 AERs and the other worth 0.8 AERs, the individual will have 1.05 solo-authored AERs at the time of tenure.

Again, the patterns are the same. An additional coauthored "AER" paper is correlated with an 8.9 percentage point increase in a man's tenure probability but a 5.3 percentage point increase in a woman's tenure probability.

3.3 Testing Against Other Disciplines and Coauthoring Conventions

Many disciplines use different coauthoring conventions, such as listing authors in order of contribution. However, these disciplines differ on several other dimensions, such as the fraction of women in the disciplines and what is most important for tenure (publications, grants, conference proceedings, etc.). In Appendix A, we conduct the same analysis for a sample of sociologists, a discipline that order authors by contribution. The sample and results are discussed in more detail in the Appendix, but we do not find evidence of women being penalized for coauthoring. What matters is being first author on a paper: being first author is correlated with a 5% increase in tenure probability for both men and women. Because sociology differs from economics in many ways, though, it is difficult to interpret whether these results suggest that ordering authors by contribution helps eliminate bias or whether the larger presence of women helps to eliminate it.

4 Channels and Explanations for Coauthorship

The previous section established three facts:

1. For very few papers, women have a lower tenure probability than men;
2. As women produce more solo-authored papers, their tenure probability converges to that of comparable men;
3. Women benefit less than men from work coauthored with men.

There are several explanations for these patterns. In this section, we argue that the results are most consistent with a story of women receiving less credit for their joint work with men rather than a story of women contributing less when they work with men. We assume that tenure committees begin with the prior that women are on average of lower ability than men, and that solo-authored papers provide a clear signal of one's ability whereas coauthored papers provide an unclear signal. Employers then misattribute credit for work produced by a man and a woman as the man is assumed to be of higher ability.

We test this argument by comparing the productivity of men and women who were denied tenure. We then explore and rule out several threats to this story. Specifically, we test for preference-based sorting, women receiving less exposure by presenting less, and taste-based discrimination. In the next section, we also present evidence from two experiments designed to completely shut down the possibility that women put in less effort when working with men, and find additional evidence that women receive less credit than men when they perform a stereotypically male task, or when they are evaluated by a man.

The claim that women receive less credit than men begs the question of why women would coauthor with men in the first place. We explore three potential explanations. First, we test whether women do not anticipate the penalty associated with coauthoring and therefore miscalculate the payoff to a coauthored paper. Second, we test whether low-ability women who may not be able to publish on their own select into coauthoring despite the costs. Finally, we examine whether women have slower starts to their careers and therefore coauthor with men as tenure approaches. Our evidence is most consistent with women being unaware of the cost of coauthoring. An additional explanation that we cannot test is that there are compensating differentials to coauthoring that are unrelated to ability.

4.1 Post-Tenure Decision Productivity Differences

If tenure committees hold the prior that women are lower ability than men and if solo-authored papers provide clear signals of ability, we will see differences in tenure rates for men and women with few publications. However, additional solo-authored publications of the same quality will have a larger marginal impact on a woman's tenure probability than a man's. As these clear signals begin to dominate the committee's prior, tenure rates between men and women will converge.

If committees are biased toward giving men more credit for work coauthored with women, we would expect to see the following. Assuming that there is some fixed amount of credit that can be given for a paper, a man will benefit more than a woman from joint work between them. In addition, both men and women will benefit more from their coauthored work with women than their coauthored work with men, as two men who coauthor will be assumed to have contributed similarly while a woman will be assumed to have contributed less.

These two claims largely play out in the data. Table 2 shows that the marginal solo-authored paper helps women more than it helps men as they start from a lower baseline tenure rate. Table 3 shows that men benefit the most from coauthoring with women (an increase in tenure probability of 9.7% when coauthoring with a woman vs. 8.7% when coauthoring with a man) although this difference is insignificant. Similarly, women benefit more from coauthoring with other women than with men. One result that is inconsistent with a story of credit allocation is the fact that the total amount of credit that can be allocated, at least when all coauthors are men, seems to add up to more than one. Men benefit as much from a coauthored paper as they do from a solo-authored paper, suggesting that tenure committees are either making a mistake when dividing credit (for example, each committee assumes that the male author under consideration for tenure at its school did most of the work), or that there is an alternative mechanism behind the results. In Section 4.2, we test several potential mechanisms.

We would see these same empirical patterns if women contribute less to projects that are joint with men. Comparing the productivity of men and women who were denied tenure helps to partially disentangle these two stories. If women who coauthor are given less credit, then women who coauthor and are denied tenure should on average be more productive than men who are denied tenure. If women who coauthor simply contribute less, we would not expect to see productivity differences between men and women who are denied tenure, or we should see women being less productive.[12]

We use two productivity measures to test whether women who coauthor and are denied tenure are more productive than men: the number of solo-authored AER-equivalents an individual publishes after the tenure decision and the log number of citations an individual has as of 2015.[13] Individuals who leave academia and do not publish after tenure are excluded from the AER-equivalent outcome sample, but including them and setting their number of post-tenure papers to zero does not change the results.

Table 5 shows the results from estimating

$$
\begin{aligned}
Y_{ifst} = \beta_1 fem_i + \beta_2 FracCA_{it} + \beta_3 T_i + \beta_4(fem_i \times FracCA_{it}) + \beta_5(fem_i \times T_i) \\
+ \beta_6(FracCA_{it} \times T_i) + \beta_7(FracCA_{it} \times T_i \times Fem_i) + \gamma' Z_i' + \theta_f + \theta_t + \theta_p + \epsilon_{ifst}
\end{aligned}
\tag{6}
$$

where the outcome variable Y_{ifst} is one of the two productivity measures described above and T_i is a tenure dummy. We include a post-tenure institution fixed effect, θ_p, to account for the fact that individuals will have access to different resources depending on where they go after the initial tenure decision.

Column 1 shows the results from estimating equation 6 with the number of solo-authored AER-equivalents as the outcome. Women who are denied tenure and coauthor have 0.4 more solo-authored AER-equivalents than men who are denied tenure and coauthor ($\hat{\beta}_2 + \hat{\beta}_4$). Column 2, which has log citations as the outcome variable, shows a similar pattern although the results are much noisier. Together, these results provide some suggestive evidence that these women receive less credit for joint projects.

4.2 Alternative Stories

There are other possible explanations for the above findings, not all of which can be tested with these particular data. Here we shed light on three standard and testable channels:

[12]This is the classic Becker outcomes test of discrimination. However, these statements involve several assumptions. First, they will not hold if there are differences in men's and women's reactions to being denied tenure that affect their productivity. Second, men and women are not differentially able to secure resources after being denied tenure. While we control for the rank of school that one ends up at, there could be differences in the quality of coauthors that one is able to get, the types of conferences one is accepted to, and so on. We control for an individual's post-tenure decision institution rank but other factors might differ between men and women. We fully shut down these channels in the experiments in Section 5.

[13]Citations were scraped from Google scholar in 2015. For the AER-equivalent outcome, we do not compare coauthored papers as these can reflect the ability of one's coauthors. Citation data includes both solo and coauthored papers as the data came in this structure.

preference-based sorting, women not claiming credit for their work, and taste-based discrimination.[14] The empirical patterns are inconsistent with all of the proposed explanations.

4.2.1 Preference-Based Sorting

If women prefer to coauthor with senior faculty, we could reasonably expect that women would have lower tenure rates. Assuming senior faculty are more likely to be credited for a paper, the fact that most senior faculty are men would drive the correlation between coauthoring with a man and tenure. That is, women receive less credit because they enjoy coauthoring with senior faculty and these senior faculty are predominantly male.

The basic summary statistics showed that women were not more likely to coauthor with senior faculty than men. However, we conduct an additional test as to whether coauthorship with senior faculty could be driving the results. We reestimate equation 3 but control for the fraction of a person's coauthors who are senior. The results are presented in Column 3 of Table 7. The seniority of women's coauthors does not explain the results. Controlling for seniority, an additional coauthored paper increases a man's probability of tenure by 8 percentage points but a woman's by 5 percentage points.

4.2.2 Women Not Claiming Credit for Papers

Women might be given less credit for their work if they are less likely to claim it as their own.[15] For example, if women present less frequently than men, people might associate a paper with the male coauthor who presents it more. The survey discussed in Section 4.3.1 also asked individuals how many times per year they present their work and whether they are more or less likely to present their coauthored papers than their coauthor. Panel B of Table 6 shows that men and women report the same likelihood of presenting their joint papers relative to their coauthors. Interestingly, though, women present their solo-authored papers fewer times per year than men do. It is possible that women do not "advertise" their work as much as men do and this leads to women receiving less recognition for their work in general. If this were true, though, women who solo author should also be less likely to receive tenure.

[14] We test for ability-based sorting below when we look at why women still choose to coauthor despite not receiving credit.

[15] Isaksson (2019) finds experimental evidence that women often claim less credit than men for their contributions to solving puzzles.

4.2.3 Taste-Based Discrimination

If some employers have a distaste for tenuring women, as in Becker (1971), we should see women who write solo-authored papers being denied tenure as well. If employers cannot plausibly deny a woman who solo-authored several well-published papers, however, they might be constrained to deny tenure only to those for whom they can make a reasonable case. If it can be argued that a woman who coauthors did little of the work, taste-based discrimination could help to explain the results as employers have an excuse for denying tenure to coauthoring women. However, as shown in Table 3, only women who coauthor with men have lower tenure rates. This would imply that employers have a particular distaste for tenuring women who coauthor with men, which seems unlikely.

4.3 Why Do Women Coauthor?

4.3.1 Ability-Based Sorting

Our results could be explained by ability-based sorting. For example, if coauthoring lowers the cost of producing a paper, but women know that they receive less credit for papers, high ability women might forego the cost savings and choose to work alone. They know they can produce high quality papers by themselves and send the employer a clearer signal of their ability. However, if low ability women can only produce high quality papers with the help of a high ability man, they might coauthor even if they receive less credit. High ability men will agree to coauthor with them if it reduces the cost of the paper without reducing the quality. Employers would then know that any woman coauthoring with a man is lower ability, leading them to rationally deny women who coauthor tenure.

In what follows, we test whether women anticipate receiving less credit, whether high ability women sort out of coauthoring with men, and whether men coauthor with women whose careers begin more slowly. To do so, we first present survey evidence suggesting that women do not know that the returns to coauthoring are lower than solo-authoring. We then show that women do receive some credit for papers that publish well, suggesting that employers might believe that there is some assortative matching. We also provide evidence that even when women tend to work with men who are slightly higher ability than themselves this unequal match does not explain the gender gap in tenure.

Survey Evidence on Knowledge of Returns to Coauthoring If women know that their returns to coauthoring with men are low, it is plausible that high ability women would choose to solo-author or only work with other women. Here we test whether women anticipate receiving less credit for collaborative work using a survey conducted with economists currently working at the top 35 U.S. economics departments. The survey was sent to all professors, regardless of rank, at these institutions and received a 32% response rate. The gender composition of the sample is representative of the profession today, with 89 respondents being female and 300 being male. In the survey, economists were asked the following question:

Suppose a solo-authored AER increases your chance of receiving tenure by 15%. For each of the following, please give an estimate of how much you think the described paper would increase your chance of receiving tenure.

Respondents then go through five types of papers (coauthored AER, coauthored AER with senior faculty, coauthored AER with junior faculty, solo-authored top field, and coauthored top field) and record their beliefs about the returns to these papers.[16]

In Table 6, we test the difference in the mean beliefs of men and women.[17] There is no statistically significant difference in the beliefs of men and women for any type of paper. Men believe that a coauthored AER will increase their chance of receiving tenure by 12.1%, and women by 12.2%. Women believe that there are slightly lower returns to AER papers coauthored with senior faculty (8.8% versus 9.1% for men), but the difference is not statistically significant. These results suggest that, in this context, women are unaware of the true returns to coauthoring.

Evidence on Sorting by Ability from CVs A second test of whether women know that they will receive less credit for papers and sort accordingly is to look at the correlation between propensity to coauthor and ability. We first test whether high ability women are less likely to coauthor than low ability women and then test for assortative matching among coauthors. We proxy for ability using the quality of journal that an individual's job market paper was published in. We assume that the job market paper is the first solo-authored paper an individual publishes after he or she graduates.[18]

If women anticipate discrimination, ability and the fraction of one's papers that are coauthored will be negatively correlated. High ability women should be less likely to

[16]We did not ask respondents about paper coauthored with men/women so that they would not be primed to think about gender.

[17]Because the survey was anonymous, the answers can not be linked to the CV data. We can therefore only test for differences in means without controls.

[18]Unfortunately, we have no data on how many job market papers are co-authored and how long it typically takes to publish a job market paper but this should not affect our results as long as there are not differences by gender.

coauthor. In Figure 3.A we plot the coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ from estimating

$$FracCA_{ifst} = \beta_1 a_i + \beta_2(fem_i \times a_i) + \beta_3 fem_i + \beta_4 TotPapers_i + \gamma' Z_i' + \theta_f + \theta_s + \theta_t + \epsilon_{ifst} \tag{7}$$

where $FracCA_{ifst}$ is the fraction of person i's papers that are coauthored and a_i is person i's ability (job market paper rank). If high ability women anticipate receiving less credit, we expect $\hat{\beta}_2 < 0$. In Figure 3.A, however, we see that ability is uncorrelated with the fraction of papers that are coauthored for both men and women: both estimates are precise zeros. There is no evidence that women along the ability distribution act strategically in their choice to coauthor versus solo author.

We also find no evidence that high ability women strategically coauthor with other women rather than men. Figure 3.B plots the results from equation 7 using the fraction of papers that are coauthored with women as the dependent variable. Women are more likely to coauthor with other women than men are but there is no sorting by ability.

While women do not seem to be sorting according to ability, it is possible that women tend to work with higher-ability or more prominent coauthors who then receive more credit for a paper. We test for this by correlating a person's ability with that of his or her coauthors. While we do not have the job market paper information for all coauthors in the dataset, we can see where the coauthors were working at the time the individual went up for tenure. As a measure of average coauthor ability, we take the average school rank of all of an individual's pre-tenure coauthors. For example, if i coauthors with j and k and j works at the 5th-ranked institution and k works at the 15th-ranked institution, the average ability of i's coauthors is 10.

We correlate i's ability with the average ability of her coauthors in Figure 4. The line of best fit is plotted controlling for number of coauthored and solo-authored publications, time until tenure, and field, institution, and tenure year fixed effects.

Men and women both sort positively on ability but women are more likely to collaborate with individuals at more highly-ranked institutions than men are. To see whether this explains the main results, we estimate

$$\begin{aligned} T_{ifst} = &\beta_1 S_i + \beta_2(fem_i \times S_i) + \beta_3 CA_i + \beta_4(fem_i \times CA_i) + \beta_5 rank_{iJ} \\ &+ \beta_6(CA_i \times rank_{iJ}) + \beta_7(fem_i \times CA_i \times rank_{iJ}) + \beta_8(fem_i \times rank_{iJ}) \\ &+ \beta_9 fem_i + \gamma' Z_i + \theta_f + \theta_s + \theta_t + \epsilon_{ifst} \end{aligned} \tag{8}$$

where $rank_{iJ}$ is the average institution rank of i's coauthors and all other variables are defined as before. The results are reported in Table 7. If men receive more credit because they are coauthoring with lower ability women, $\hat{\beta}_7$ should be negative. However, $\hat{\beta}_7$ is close to zero, indicating that the ability or prominence of one's coauthor is not driving the tenure gap for coauthoring women.

Returns to Top Papers For high ability women to receive no credit for their coauthored papers, employers would have to believe that there is no assortative matching by ability. Otherwise, employers would receive a signal that women who coauthor with high ability men are also high ability, and be more likely to promote them. Figure 4 shows that assortative matching does occur, but it is possible that employers do not recognize this. We test for this by looking at how credit for top 5 publications is allocated. If employers know that there is assortative matching, they should believe that women coauthoring with high-ability men are also likely to be high ability.

Table 8 shows the results from estimating

$$\begin{aligned} T_{ifst} =& \beta_1 TopS_i + \beta_2(fem_i \times TopS_i) + \beta_3 TopCA_i + \beta_4(fem_i \times TopCA_i) + \beta_5 NonTopS_i \\ &+ \beta_6 NonTopCA_i + \beta_7(fem_i \times NonTopS_i) + \beta_8(fem_i \times NonTopCA_i) + \beta_9 fem_i + \gamma' Z_i \\ &+ \theta_f + \theta_s + \theta_t + \epsilon_{ifst} \end{aligned} \tag{9}$$

where $TopS_i$ and $TopCA_i$ are the number of solo and coauthored papers that individual i has published in a top 5 journal. Similarly, $NonTopS_i$ and $NonTopCA_i$ are the number of solo and coauthored papers the individual has published in non-top 5 journals. In Table 8, the "nop-top 5" interaction terms are presented in the second column.

Power becomes an issue as (1) there are relatively few people publishing in the top 5 journals, and (2) cutting by gender means that there are even fewer women in each category.

Table 8 shows that coauthored papers published in a top 5 journal help women much more than those published in non-top 5 journals. Non-top 5 coauthored papers do not have any positive influence on women's tenure probability. It seems that employers receive some signal when a woman publishes her coauthored papers in top journals which is at odds with the hypothesis that only low ability women coauthor with men.

Overall, there is little evidence that ability-based sorting is driving the results.[19] If

[19]Garcia and Serman (2015) show that there could be selection into coauthorship driven by a desire to be first author on a paper (that is, depending on where you are in the alphabet relative to your coauthors). This would be an issue in this setting if, for example, men are more likely to be strategic than woman and are therefore more likely to be first author on a paper (which is correlated with having more citations). We test whether men are more likely to be first author on their papers than women and whether men have a "higher" author position overall. We find that men in our sample are first author 57% of the time while women are first author 55% of the time ($p = 0.907$).

anything, employers seem to recognize that high ability men and women might work together and are therefore more likely to grant these women tenure. However, their tenure rate is still lower than that of high ability men.

4.3.2 Timing of Coauthorship

It is possible that men offer to work with women who are struggling to publish. If this is the case, we should see women who have few publications in the early years of their appointment being more likely to coauthor with men. We test for this possibility by looking at differences in early publications and by testing whether women with a longer time lag between their initial appointment and first publication are more likely to coauthor with men.

Appendix Figure B3 descriptively shows the timing of publications for men and women, split by whether they received tenure at their initial tenure institution. More formally, we test whether women have fewer publications early in their careers by estimating

$$Y_{ifst} = \beta_1 Fem_i + \beta_2 T_{is} + \beta_3 (Fem_i \times T_{is}) + \beta_4 Papers_i + \beta_5 \bar{q}_i + \theta_f + \theta_s + \theta_t + \epsilon_{ifst} \quad (10)$$

where Y_{ifst} is the number of years between individual i's initial appointment and i's first post-appointment publication.[20] We test whether women who did not receive tenure had a longer publishing lag by interacting the female dummy term with an indicator for receiving tenure at school s, T_{is}. We control for the number of papers published pre-tenure ($Papers_i$) and the average quality of those papers ($\bar{q}_i$). All other variables are defined as before.

The results are presented in Table 9. Women who do not receive tenure do have a longer lag (approximately 0.5 years) between their first appointment and their first publication although the result is noisily estimated. We test whether women with a longer lag are more likely to coauthor with men by estimating

$$\begin{aligned} FracM_{ifst} = &\beta_1 Fem_i + \beta_2 T_{is} + \beta_3 (Fem_i \times T_{is}) + \beta_4 Y_i + \beta_5 (Fem_i \times Y_i) \\ &+ \beta_6 (Fem_i \times T_{is} \times Y_i) + \beta_4 Papers_i + \beta_5 \bar{q}_i + \theta_f + \theta_s + \theta_t + \epsilon_{ifst} \end{aligned} \quad (11)$$

where the outcome variable, Y_i in equation 10 is used as a regressor. If men bring women with a slow start to publishing onto their projects, we would expect to see $\hat{\beta}_5 > 0$.

The results, presented in Column 2 of Table 9, do not support the hypothesis that women who struggle to publish initially are more likely to begin publishing with men.

[20] We exclude papers that were published before the person's first appointment.

The coefficient on β_5 is negative, suggesting that women with a longer publishing lag are less likely to coauthor with men although this result is again insignificant.

Taken together, these results suggest that women coauthor with men at least in part because they do not anticipate the costs of doing so and not for reasons related to ability or preferences. In Section 6, we discuss what the long-run equilibrium behavior of men and women might be and the implications for efficiency of coauthorship.

5 Experimental Evidence

In the previous section, we provided suggestive evidence that factors like sorting and taste-based discrimination do not explain why women who coauthor with men are less likely to receive tenure. We instead argue that the results are most consistent with women receiving less credit for joint work with men. Specifically, because coauthored papers are an unclear signal of ability, women receive less credit for their joint work with men if they are believed to be of lower ability (Correll and Ridgeway, 2003). We cannot rule out, though, that real or perceived differences in effort explain the results. For example, tenure committees might hold the belief that women contribute less or provide lower effort when they work with men, regardless of their beliefs about a woman's ability. In addition, tenure committees might believe that low ability women choose to work with high ability men even if the empirical evidence suggests otherwise.

To shed light on whether different contributions to group work (or perceptions of differential contributions) and sorting are driving the results, we run two experiments designed to shut down these channels. The experiments also allow us to assess the role of beliefs about ability more directly. The first experiment is an artefactual experiment run on mTurk. The second is a framed field experiment for which we recruited individuals who work in human resources and whose job is to recruit personnel. Although these settings are different from academia, they provide additional evidence that gender plays a role in the allocation of credit due to differences in beliefs about the ability of men and women. The first does so in a relatively abstract setting with high control, while the second adds more context from the process of hiring candidates (see Harrison and List, 2004).

Both experiments consist of two incentivized parts. In the first step, workers are recruited to complete tasks individually. In the second step, designed to test whether people misallocate credit for joint work, another set of individuals are recruited to either predict how well the workers will do on a second set of related tasks (Experiment I) or to choose a worker to hire (Experiment II). In both experiments, we vary whether the predictors/hirers see workers' individual scores in the first task, or the sum of two individuals' scores.

5.1 Experiment I

The first experiment consists of two incentivized parts. In the first step, mTurk workers, henceforth referred to as "workers" are recruited to complete two related quizzes (Quiz 1 and Quiz 2).[21] We then recruit 506 mTurk participants, referred to as "predictors", to predict the Quiz 2 scores of a randomly-chosen man and a randomly-chosen woman on Quiz 2. Before making their predictions, the predictors are told that the workers completed the two quizzes on their own, and are shown information on Quiz 1. Specifically, they see the questions asked, the overall score distribution (not broken out by gender), and information about the Quiz 1 scores of the two workers they will be making predictions about. Predictors are then shown the Quiz 2 questions and are asked to estimate the score of both workers in Quiz 2. Predictors are paid a participation fee of $0.50 and receive $0.10 for each score they correctly predict. The instructions given to predictors are available in Appendix D. This experiment uses a 2x2x2 treatment design, described in detail below.

5.1.1 Treatments

Individual vs. Joint Scores Predictors are randomized into an Individual treatment or a Joint treatment. In the Individual treatment, predictors are shown the individual score of each of the two workers in Quiz 1. This treatment tests whether predictors correctly predict scores when they see a clear signal of each worker's ability. This parallels the solo-author paper analysis: if predictors correctly assign credit when they see a clear signal of ability, there should be no difference in how men and women are evaluated conditional on Quiz 1 scores.

In the Joint treatment, predictors are shown the *sum* of the scores of the two workers. For example, if worker A scored 3 out of 5 and worker B scored 4 out of 5, the predictor would see the score 7 out of 10 for that pair. Importantly, predictors are told that there was no interaction between workers: each worker completed the same quiz and was paid according to his or her individual score. Thus, predictors know that workers are randomly paired with a member of the opposite sex but worked independently and were individually incentivized. This treatment is designed to understand how predictors assign credit for performance when they cannot observe individual contributions, but in a setting where there is no selection into pairs (such as high-ability men working with

[21]Workers receive a participation fee of $0.30 plus $0.05 for each question they answer correctly. The quizzes contain five questions each and are available in Appendix D.

low-ability women) or free-riding. Therefore, the predictors' estimates should reflect only their beliefs about each worker's score and ability since they know that workers completed the quizzes individually and were individually incentivized. To draw a parallel between this treatment and the main analysis, the individual scores that make up the joint score can be thought of as each person's "contribution" to a group project that, in this case, is unaffected by selection or effort.

No-Information vs. Gender-Information To understand whether predictors' estimates are driven by (possibly incorrect) beliefs about ability or taste-based animus, we provided some predictors with information about the performance of men and women. In the No-Information treatment, the only aggregate information predictors receive is the overall score distribution. In the Gender-Information treatment, predictors are additionally shown the average score of male and female workers.[22]

If predictors exhibit taste-based animus, providing them with information about men and women's average performance will not change their predictions. In addition, comparing these treatments helps to understand whether differences in attribution are driven by incorrect beliefs about gender differences in performance. If participants hold mistaken beliefs about men and women's average performances, the gender information treatment should correct those beliefs, and the predictors should adjust their estimates accordingly.

Male-Stereotyped vs. Female-Stereotyped quizzes To evaluate whether differences in credit for joint work depends on the type of task that is being performed, workers in the Male-Stereotyped treatment completed math quizzes while workers in the Female-Stereotyped treatment completed grammar quizzes.

5.1.2 Results

The main experimental results are presented in Table 10, which shows how predictors' guesses vary based on the quiz-taker's gender and the treatment. Specifically, we estimate

$$Q2_{ij} = \beta_1 fem_i + \beta_2 D_j + \beta_3 (fem_i \times D_j) + \beta_4 Q1_i + \epsilon_{ij} \tag{12}$$

separately for the sample of individuals who took math quizzes (Columns 1 and 2) and grammar quizzes (Columns 3 and 4), and by joint/individual treatment. The outcome

[22]The overall score distribution was presented as a histogram. In the Gender-Information treatment, the histogram contained lines indicating the mean performance of men and women. See Figure D1 in Appendix D.

variable, $Q2_{ij}$, is predictor j's estimate of quiz-taker i's Quiz 2 score. An indicator for the quiz-taker being female, fem_i, is interacted with an indicator for the predictor being in the Gender-Information treatment, D_j. We also control for i's Quiz 1 score ($Q1_i$).

In the Individual treatment, there is no significant difference in men's and women's estimated performance in the math quiz. In this treatment, predictors base their estimations on the observed individual scores.[23] By contrast, in the Group treatment, where predictors see only the sum of a man's and a woman's Quiz 1 score, they predict that women scored less than men on the second math quiz. This mirrors the finding that women suffer a coauthor penalty when their contribution to a paper is unobserved but are not discriminated against when their contributions are observed, as in solo-authored papers. Predicting that the woman will do worse than the man in the Joint treatment suggests that predictors believe that the woman's first score was lower; that is, she is worse at the task and therefore contributed less to the joint score. Puzzlingly, showing predictors the mean scores of male and female workers does not change the predictions for the second math quiz. Together, these results suggest that predictors hold a prior that men are better than women at the math quiz, and the evidence that men are only slightly better does not affect this belief.

The results for the grammar quiz in Columns 3 and 4 suggest that the results are not driven by taste-based animus in which women are always penalized in collaborative situations. Here, women are not predicted to perform differently than men on the second quiz in both the Individual and Joint treatments. In addition, seeing that the mean grammar score of female workers is higher than that of male workers creates a gender difference in predicted scores in favor of women.

5.2 Experiment II

The second experiment was designed to study attribution of credit for joint work in a setting that more closely approximates a hiring scenario. In addition, this experiment allows us to test for gender differences based on the recruiter's gender, and to test more directly whether beliefs affect credit attribution.

Before conducting the experiment, we collected individual characteristics from university students (henceforth referred to as job candidates), along with their performance in two incentivized real-effort tasks. The experiment itself was conducted with 479 actual human resource workers whose job is to recruit personnel. The HR workers, henceforth

[23]Women had a lower average score than men on the math quizzes (2.51/5 vs. 2.72/5) and a higher average score on the grammar quizzes (2.41 vs. 2.17). The distribution of scores on Quiz 1 are shown in Appendix Figure D1. This is the same figure that predictors are shown. If predictors are in the Gender-Information treatment, they also see the two lines indicating the mean male and female scores.

“recruiters”, were asked to choose job candidates for a task based on short resumes.

The “recruiters” complete an incentivized online survey. Each recruiter is sequentially shown three sets of four candidates’ resumes. Recruiters pick one candidate from each set and are paid according to the chosen candidate’s score in the real-effort task.[24]

After their choices, the recruiters’ belief about relative gender differences in ability is elicited. Specifically, they are asked to indicate the degree to which they think men or women are better at the real-effort task.[25] Therefore, Experiment II does not try to induce beliefs by providing information about scores, as in Experiment I. Instead, it elicits beliefs to observe whether beliefs are biased and to evaluate the extent to which individual beliefs affect the recruiters’ choices. The experiment uses a 2x2 design.

5.2.1 Treatments

Individual vs. Joint scores As in Experiment I, recruiters are randomized into an Individual or a Joint treatment. In the Individual treatment, recruiters see the individual scores of all four candidates in a set. In the Joint treatment, recruiters see two summed scores (the sum of candidate 1’s and 2’s score, and the sum of candidates 3’s and 4’s score). The sets are chosen such that one of the summed scores is obviously superior to the other to give recruiters a strong incentive to choose one of these two candidates. These superior candidate pairs always include a male and a female whose resumes are otherwise alike.[26] The pair of inferior candidates may vary on all characteristics but had much lower joint scores.

Search vs. Vocabulary tasks For the real-effort tasks, recruiters were randomized to pick candidates who performed a vocabulary task (finding words using a set of provided letters) or a numeric-search task (finding the highest numbers in each of two 10x10 matrices and adding them up, as in Weber and Schram, 2017). The tasks are described in more detail in Appendix E.[27] Compared to Experiment I, these tasks are arguably less stereotypical and have been shown to exhibit little to no gender difference in performance (Schram et

[24] Excluding the participation fee, recruiters earned an average of \$6 to complete the ten-minute experiment.

[25] Answers ranged from a difference in means of 4 or more points in favor of men to 4 or more points in favor of women. Choosing the correct answer is rewarded with \$1.50. The correct answer was calculated based on the actual scores of candidates in the tasks. For more details, see Appendix E.

[26] In addition to their scores, the resume of each candidate shows the candidate’s field of study, degree length (from three to five years), age, gender, and geographic region of origin. See Appendix E for more details and an example of a set of candidates.

[27] Recruiters received \$0.06 for each point in the vocabulary task by the chosen candidate or \$0.15 for each correct addition in the numerical search task.

al., 2019; Shurchkov, 2012). Because men and women perform similarly on these tasks, using them provides us with a stronger test of whether incorrect beliefs about performance drive credit allocation.

5.2.2 Results

When recruiters are informed about the individual scores, the candidate's gender does not affect the recruiters' choices, which are primarily determined by individual scores (for details, see Appendix E). This mirrors the result in the Individual treatment in Experiment I and the observation that women and men receive equal credit for solo-authored papers.

To investigate credit attribution in the Joint treatment, we use McFadden's random-utility model to explain the binary choice of whether or not to select a candidate under the restriction that only one out of four candidates can be chosen in a set (McFadden, 1974). More specifically, we assume that the utility of recruiter j from choosing candidate i in set k is given by

$$u_{jik} = \beta_1 fem_{ik} + \beta_2 (fem_{ik} \times Belief_j) + \beta_3 Score_{ik} + \gamma' Z_{ik} + \theta_{jk} + \epsilon_{jik}, \tag{13}$$

where fem_{ik} is an indicator that candidate i in set k is female, $Score_{ik}$ is candidate i's joint score in the task, and $Belief_j$ is recruiter j's belief about the difference in mean scores between men and women (constructed such that zero implies a belief of no gender differences in mean scores, and positive (negative) values imply a belief that men (women) are better). The vector of controls, Z_{ik}, include all the other elements of candidate i's resume, while θ_{jk} correspond to fixed effects for each recruiter-set combination. Recruiter i picks the candidate j that gives the highest utility in set k. The random variable ϵ_{jik} is assumed to have an extreme value distribution, which allows us to estimate the model using a conditional logistic regression. The estimation results are presented in Table 11 as odds ratios. Column 1 contains the results for the search task and Column 2 for the vocabulary task.

The results for all recruiters show that they are much more likely to pick candidates in pairs with a high joint score (i.e., from the superior pair in the set). On average, the gender of the candidate does not have a significant impact on the likelihood of being chosen in either task. However, this is no longer the case once recruiters are divided according to their gender. Columns 3 to 6 show the estimation results for male recruiters and Columns 7 to 10 for female recruiters. Columns 3 and 7 reveal that male recruiters are less likely to pick female candidates, though the odds ratio is not significantly different from one in the numeric-search task. This aligns with the results of the male-stereotyped quiz in Experiment I and the data on co-authorship in Economics, where men are given more credit for joint work. By contrast, Columns 7 and 9 show that female recruiters are significantly *more* likely to pick a female than a male candidate. We will return to this surprising result below. For both male and female recruiters, the task does not appear to matter much as the odds ratios for the female indicator are quite similar across the search and vocabulary tasks. Columns 4, 6, 8, and 10 show the estimation results once beliefs about gender differences in scores are introduced in the regressions. In all cases, recruiters who believe men are better than women at the task are significantly less likely to pick a female candidate (and vice-versa). Correcting for beliefs brings all the odds ratios of the female indicator closer to 1 (a significant effect for gender remains only for male recruiters in the vocabulary task), which suggests that the observed biases in credit attribution for joint work are largely mediated by the recruiters' beliefs about which gender is better.[28]

5.3 Discussion

While the experimental context is different from the academic context, the results provide evidence that, even after shutting down effort and selection channels, individuals make different inferences about men and women's contributions to a joint project that are rooted in beliefs.

Our two experiments differ along two dimensions: the stereotypical nature of the tasks used (higher in Experiment I) and the specificity concerning the hiring context (higher in Experiment II). In addition, we distinguish between the gender of the recruiter in Experiment II. Despite these differences, both experiments provide evidence of a bias against women when attributing credit to joint work. Experiment I shows that this can depend on the task under consideration, with women receiving less credit for stereotypically male tasks and men receiving less credit for stereotypically female tasks. Experiment II, on the other hand, shows that the credit-attribution bias can depend on the gender of the recruiter. Male recruiters exhibit this bias against women. Female recruiters, however, show the opposite bias by attributing more credit to women than to men for joint work. In both experiments, the differences can be explained by beliefs about which gender is better at a task.[29] The patterns observed in our two experiments can be used to re-evaluate the results on tenure decisions in economics. In fact, if we assume that such decisions are made

[28]Our data show (see Appendix E) that the beliefs of male recruiters are not significantly biased towards either gender in either task. By contrast, on average, female recruiters expect female candidates to do better than male candidates in both tasks.

[29]However, even after we account for beliefs, by providing information about Quiz 1 scores in Experiment I or controlling for measured beliefs in Experiment II, we observe too little credit is attributed to women by predictors in the math quiz of Experiment I and male recruiters in the vocabulary task of Experiment II. In other words, we find suggestive evidence that both beliefs about ability and taste-based animus can play a role.

primarily by men and that economics is seen as a stereotypically-male discipline, then our experimental results would predict the bias observed in attributing credit to co-authored papers.[30] Our experimental results further suggest that this bias is caused by (incorrect) beliefs about the male and female co-authors contributions to joint work.

6 How Will Coauthorships Evolve in the Long Run?

In section 4, we address various reasons why women coauthor with men even though they receive relatively little credit for the joint work. We concluded that the main explanation appears to be that many women are simply not aware of this biased credit attribution. Here, we consider whether the observed choices by coauthoring women and by tenure committees may constitute equilibrium behavior.

Though tenure committees repeatedly make decisions about female candidates, note that any particular decision involves a one-shot game of incomplete information with the candidate. A publication serves as a signal for the candidate's quality, and we assume that tenure committees prefer high quality candidates. For the committee to undervalue a female candidate's coauthored publication, they must believe that she contributed (too) little; that is, the publication does not signal high quality for her. This justifies giving the publication little weight in the tenure decision. If the candidate knows that this is how the tenure committee decides, she would either choose not to coauthor or contribute relatively little to coauthored papers.[31] As discussed in section 4, our data show no evidence of selection into coauthorship based on ability: high ability women are as likely to coauthor as low ability women. In subsection 4.1, we conduct "outcomes tests", comparing the productivity of men and women post-tenure. We find evidence that the women who do not receive tenure are more productive than men who do not receive tenure, suggesting that some bias is at play. Our experiments provide additional evidence of gender bias in credit attribution even when the contributions to a joint score are equal by design. We therefore conclude that the coauthorship choices documented in this paper are not a (long run) equilibrium.

As women become aware of the lack of credit they receive from coauthoring, particularly with men, their best response would be to opt out of coauthoring altogether or start coauthoring more with solely women (if tenure committees do not or are not perceived

[30]These results are also consistent with the lack of evidence of women being penalized for coauthoring in sociology (see Appendix A), a discipline with relatively more women and thus less likely stereotypically-male than economics.

[31]This is reminiscent of the Coate and Loury (1993) statistical discrimination model, where expectations of underinvestment in skills by a particular group are self-fulfilling in equilibrium.

to change their behavior). Assuming that co-authors are chosen at least partly to exploit synergies in expertise, a move towards best response behavior would then introduce inefficient choices of co-authors. Hence, the bias we observe in tenure committee decisions not only involves inequity, but may also lead to inefficiency if women start best responding to it.[32]

Alternatively, market forces might yield dynamics that change the way in which tenure committees value co-authored work. If high-quality female scholars are being undervalued, this creates an opportunity for departments that do not have a bias in the attribution to step in. Giving tenure to candidates who are subsequently successful might, in turn, force others to diminish their bias. This could occur, for example, because talented female scholars will have an incentive to seek jobs at unbiased departments.

At this stage we can only document that a credit-attribution bias in tenure decisions for economists exists. The future will show whether this leads to inefficient co-authorships or whether this bias will disappear over time.

7 Conclusion

Women receive tenure at significantly lower rates than men in many academic fields. As discussed in the introduction, this phenomenon is not exclusive to academia. Several explanations have been put forward for the gap, but it persists even after accounting for observable characteristics such as fertility preferences and productivity.

This paper proposes an alternative explanation. We argue that women receive less credit for group work when employers can not perfectly observe their contribution. When signals are noisy, employers have to infer each worker's ability or productivity. Coauthored papers provide employers with a noisy signal. The fact that women who work specifically with men receive tenure at lower rates than comparable women who work alone or with other women suggests that gender enters into the employer's inference process. However, when employers receive clear signals, men and women are treated similarly. For example, men and women receive the same amount of credit for solo-authored papers, which provide a clear signal of ability. Evidence from two experiments suggests that these results are not explained by sorting or differences in effort to group work. The experiments further suggest that this phenomenon is not specific to women as men also

[32] In Table 1, we see evidence that women do coauthor more with other women. In part, this might be due to some women strategically avoiding coauthoring with men. However, we cannot exclude other explanations such as compensating differentials for working with someone of one's own gender or gender-specific tastes for research topics.

suffer a penalty when working with women on a female-stereotyped task. Finally, the gender of the person assigning credit also influences credit attribution.

Being aware of this phenomenon is important in a world that is increasingly relying on group work for production. The tech industry, for example, prides itself on collaboration. In such male-dominated fields, however, group work could result in fewer women moving up the career ladder if credit is not properly attributed. The same could be true for men in female-dominated industries. The unequal attribution of credit would then contribute to and help maintain gender segregation in occupations.

References

[1] Antecol, Heather, Kelly Bedard, and Jenna Stearns. 2018. "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?" American Economic Review, 108 (9): 2420-41.

[2] Becker, Gary. 1971. *The Economics of Discrimination.* Chicago: Univ Chicago Press.

[3] Bikard, Michael, Fiona Murray, and Joshua Gans. 2015. "Exploring Trade-offs in the Organization of Scientific Work: Collaboration and Scientific Reward." Management Science, 61(7): 1473-1740.

[4] Blau, Francine D. and Jed DeVaro. 2007. "New Evidence on Gender Differences in Promotion Rates: An Empirical Analysis of a Sample of New Hires." Industrial Relations, 46(3): 511-550.

[5] Born, Andreas, Eva Ranehill, and Anna Sandberg. 2019. "A Man's World? The Impact of a Male Dominated Environment on Female Leadership." working paper.

[6] Boschini, Anne and Ana Sjögren. 2007. "Is Team Formation Gender Neutral? Evidence from Coauthorship Patterns." Journal of Labor Economics, 25(2): 325-365.

[7] Buser, Thomas, Muriel Niederle, and H Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." Quarterly Journal of Economics, 129 (3): 1409-47.

[8] Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." Quarterly Journal of Economics, 125(3): 1101-1144.

[9] Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams. 2014. "Women in Academic Science: A Changing Landscape." Psychological Science in the Public Interest, 15(3): 1-67.

[10] Coate, Stephen, and Glenn C. Loury. "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" The American Economic Review 83, no. 5 (1993): 1220-240.

[11] Coffman, Katherine Baldiga. 2014. "Evidence on Self-Stereotyping and the Contribution of Ideas." The Quarterly Journal of Economics, 129(4): 1625-1660.

[12] Correll, Shelley J. and Cecilia L. Ridgeway. 2003. "Expectation States Theory." Pages 29-51 in *The Handbook of Social Psychology*, ed. John Delamater. New York: Kluwer Academic Press.

[13] Dynan, Karen and Cecilia Rouse. 1997. "The Underrepresentation of Women in Economics: A Study of Undergraduate Economics Students." The Journal of Economic Education, 28(4): 350-368.

[14] Elmins, William, Robert Joyce, and Monica Costa Dias. 2016. "The Gender Wage Gap." Institute for Fiscal Studies Briefing Note 186.

[15] Garcia, Daniel and Joshua Sherman. 2015. "Norms and Team Formation: Evidence from Research Partnerships." Working paper.

[16] Ginther, Donna K. and Shulamit Kahn. 2004. "Women in Economics: Moving Up or Falling Off the Academic Career Ladder?" Journal of Economic Perspectives, 18(3): 193-214.

[17] Hale, Galina and Tali Regev. 2014. "Gender Ratios at Top PhD Programs in Economics." Economics of Education Review, 41: 55-70.

[18] Hammermesh, Daniel. 2013. "Six Decades of Top Economics Publishing: Who and How?" Journal of Economic Literature, 51(1): 162-172.

[19] Harrison, Glenn W. and John List. 2004. "Field Experiments" Journal of Economic Literature, 42(4): 1009-1055.

[20] Isaksson, Siri. 2019. "It Takes Two: Gender Differences in Group Work." Working paper.

[21] Kalaitzidakis, Pantelis, Theofanis Mamuneas, and Thanasis Stengos. 2003. "Rankings of Academic Journals and Institutions in Economics." Journal of the European Economic Assocation, 1(6): 1346-1366.

[22] Lazear, Edward and Kathryn Shaw. 2007. "Personnel Economics: The Economist's View of Human Resources." Journal of Economic Perspectives, 21(4): 91-114.

[23] McFadden, Daniel. 1973. "Conditional logit analysis of qualitative choice behavior." In P. Zarembka (ed.), *Frontiers in Econometrics*. New York: Academic Press.

[24] Milkman, K.L., M. Akinola, and D. Chugh. 2015. "What Happens Before? A Field Experiment Exploring How Pay and Representation Differentially Shape Bias on the Pathway into Organizations." Journal of Applied Psychology, 100(6): 1678-1712.

[25] Moss-Racusin, Corinne, John Dovidio, Victoria Brescoll, Mark Graham, and Jo Handelsman. 2012. "Science Faculty's Subtle Gender Biases Favor Male Students." PNAS, 109(41): 16474-16479.

[26] Niederle, Muriel and Lise Vesterlund. 2007. "Do Women Shy Away From Competition? Do Men Compete Too Much?" Quarterly Journal of Economics, 122(3): 1067-1101.

[27] Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. 2017. "Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender." The Economic Journal, 127 (604): 2153-86.

[28] Sarsons, Heather. 2017. "Recognition for Group Work: Gender Differences in Academia." AER Papers & Proceedings, 107(5): 141-145.

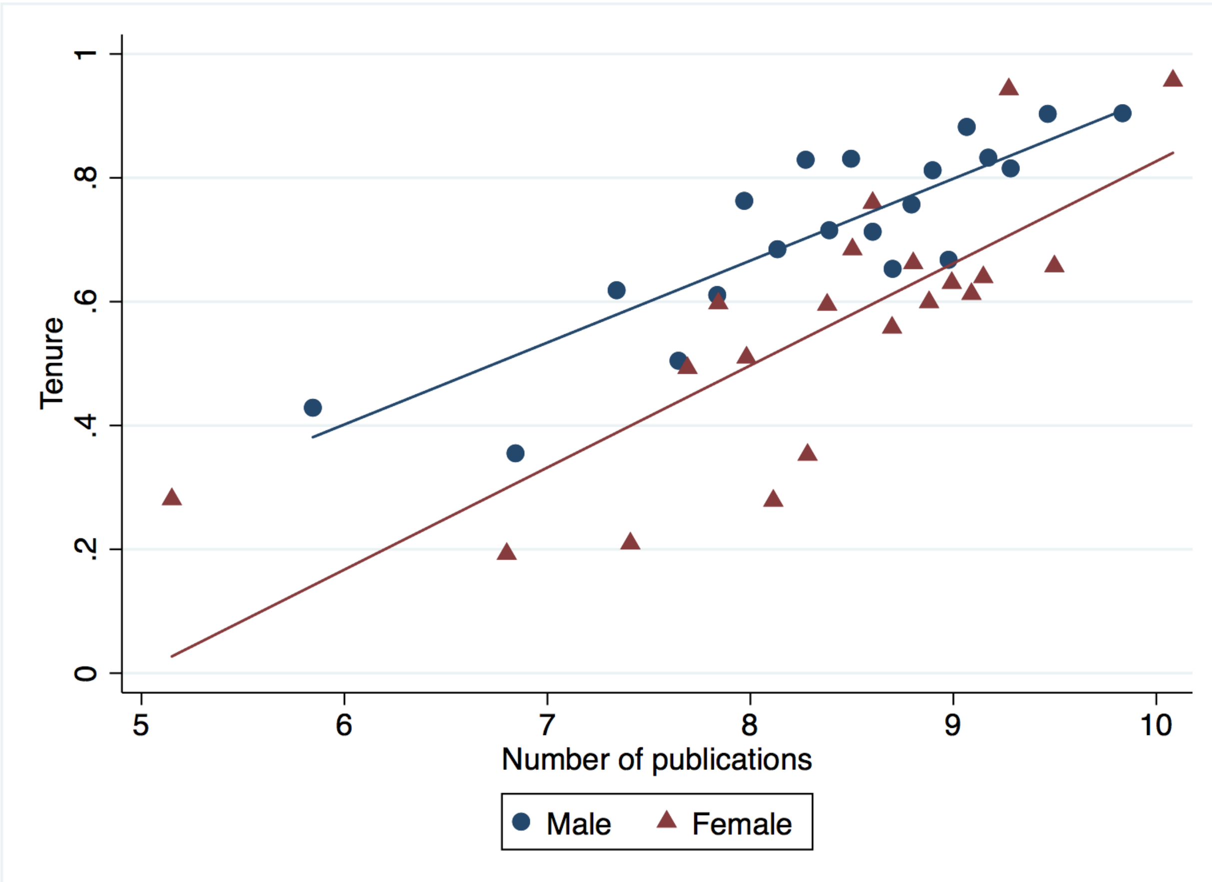
[29] Schram, Arthur, Jordi Brandts, and Klarita Gërxhani. 2019. "Social-status ranking: a hidden channel to gender inequality under competition." Experimental Economics, 22(2): 396-418

[30] Shurchkov, Olga. 2012. "Under Pressure: Gender Differences in Output Quality and Quantity Under Competition and Time Constraints." Journal of the European Economic Association, 10: 1189-1213.

[31] Weber, Matthias and Arthur Schram. 2017. "The Non-Equivalence of Labour Market Taxes: A Real-Effort Experiment." The Economic Journal, 127(604): 2187-2215.

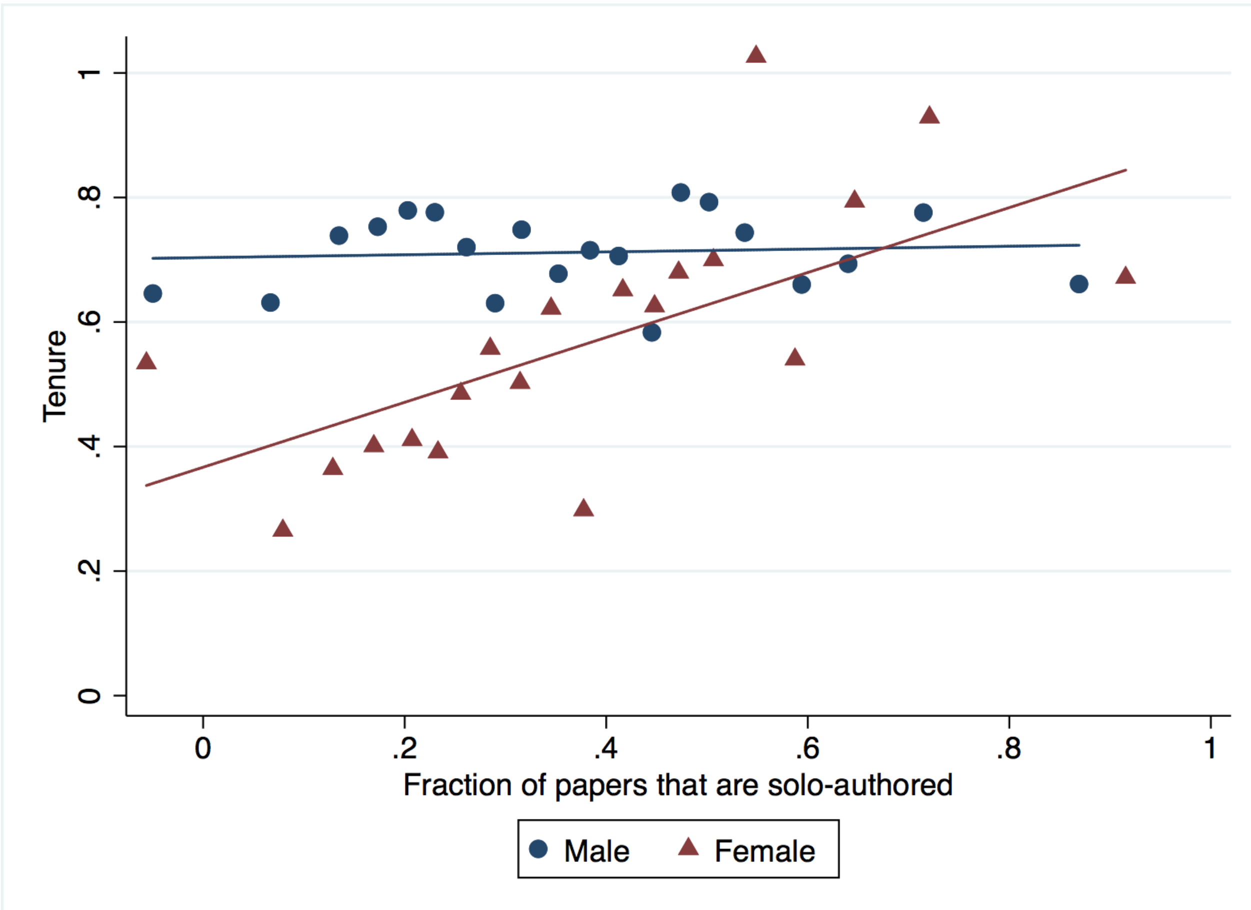
Figures

FIGURE 1: TOTAL PAPERS AND TENURE



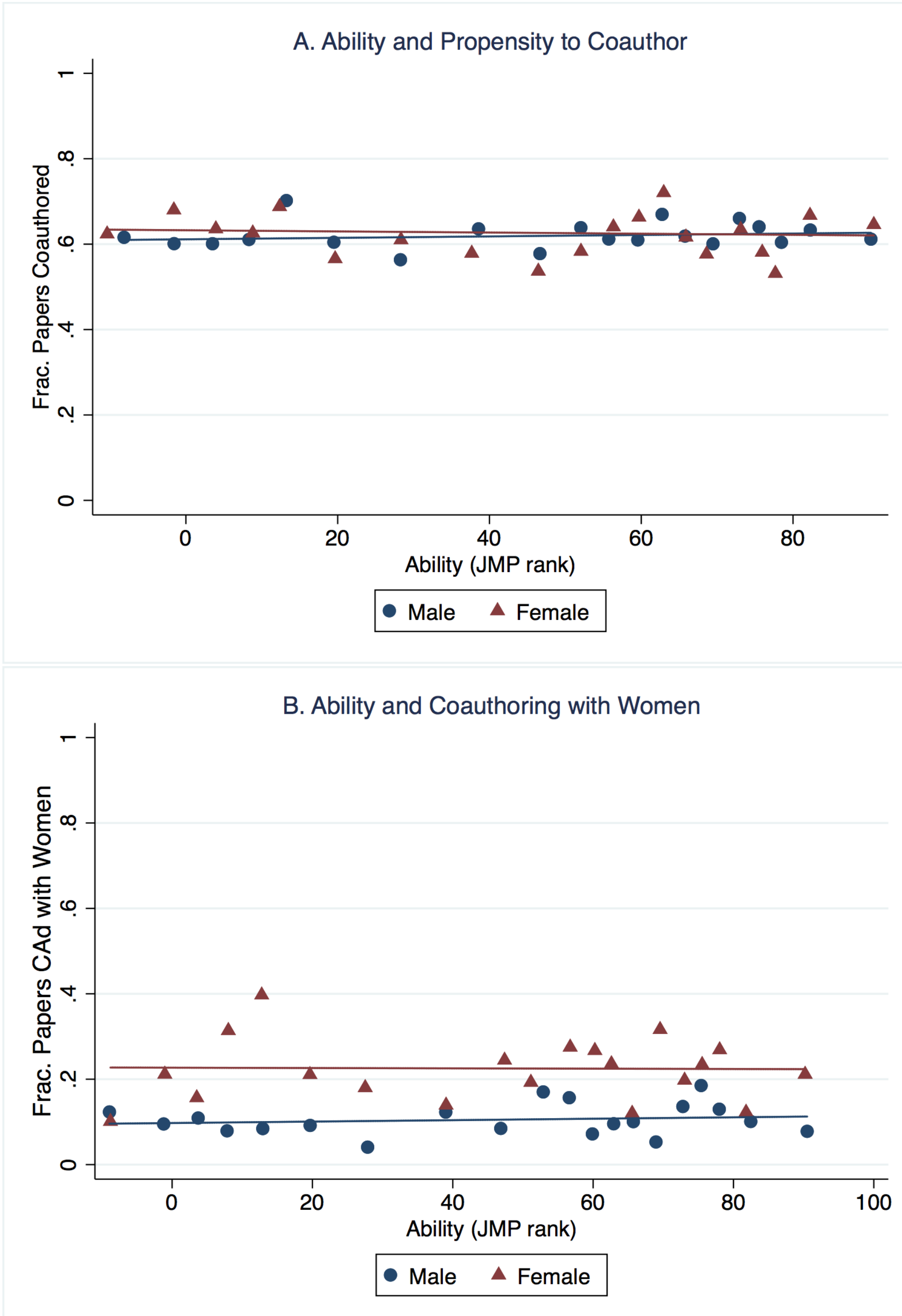
Notes: This binned scatterplot shows the correlation between the total number of publications an individual has at the time they go up for tenure and the probability of receiving tenure. The y-variable, tenure, is a binary variable that equals one if an individual received tenure at their initial institution of employment. For more details on how the tenure variable is constructed, see Section 2. To construct the plot, tenure is first residualized with respect to the following controls: number of years it took to go up for tenure, average journal rank of pre-tenure publications, log citations, total coauthors, and tenure school, tenure year, and field fixed effects. The x-variable, number of publications, is then divided into twenty equal-sized groups. Within each of these groups, we plot the mean of the y-variable (tenure) residuals against the mean of the x-variable (also within each bin). We then add back the unconditional mean of Tenure to help with the interpretation of the line of best fit. The lines of best fit are estimated using the full sample (N=621) and have slopes of $\beta = 0.132$ (s.e. = 0.016) for men and $\beta = 0.165$ (s.e. = 0.043) for women.

FIGURE 2: RELATIONSHIP BETWEEN PAPER COMPOSITION AND TENURE



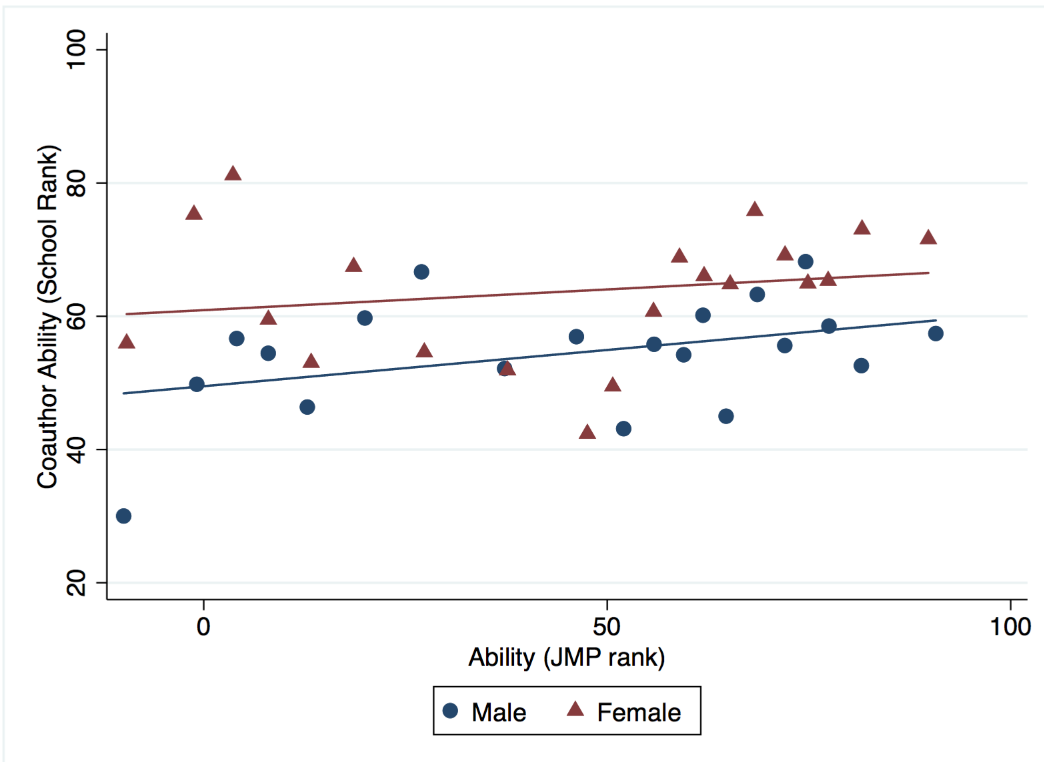
Notes: This figure is a binned scatterplot of the correlation between tenure and the fraction of an individual's papers that are solo-authored, split by gender. The y-variable is a binary variable indicating whether an individual received tenure. To construct the plot, tenure is first residualized with respect to the following controls: total number of papers an individual published by the time of tenure, number of years it took to go up for tenure, average journal rank of pre-tenure publications, log citations, total coauthors, and tenure school, tenure year, and field fixed effects. The x-variable, fraction of papers that are solo-authored, is then divided into twenty equal-sized groups. Within each of these groups, we plot the mean of the y-variable (tenure) residuals against the mean of the x-variable (also within each bin). We then add back the unconditional mean of Tenure to help with the interpretation of the line of best fit. The line of best fit using OLS is shown separately for men and women. The lines of best fit are estimated using the full sample (N=621) and have slopes of $\beta = 0.521$ (s.e. = 0.158) for women and $\beta = 0.023$ (s.e. = 0.748) for men.

FIGURE 3: ABILITY AND SORTING



Notes: This binned scatterplot shows the correlation between an individual's ability and the propensity to coauthor (Panel A) or the propensity to coauthor with women (Panel B). The outcome variable in Panel A is the fraction of an individual's papers that were published by tenure that are coauthored. The outcome variable in Panel B is the fraction of an individual's pre-tenure papers that are coauthored with only women. We proxy for an individual's ability with the rank of the journal in which the individual's job market paper was published. The plot is constructed as described in Figure 1 with the y-variable residualized on the following controls before plotting: total solo and coauthored papers, the number of years it took to go up for tenure, log citations, and tenure school, tenure year, and field fixed effects. The lines of best fit using OLS are shown separately for men and women. The estimates for Fig. 3A are $\beta = -0.0001$ (s.e. = 0.0003) for women and $\beta = 0.0002$ (s.e. = 0.0002) for men. The estimates for Fig. 3B are $\beta = -0.00004$ (s.e. = 0.0008) for women and $\beta = 0.0002$ (s.e. = 0.0003) for men.

FIGURE 4: ASSORTATIVE MATCHING



Notes: This binned scatterplot shows the correlation between an individual's ability, proxied by the journal in which their job market paper is published, and their coauthor's ability, proxied by the average school rank of their coauthors. The school rank of coauthors are measured at the time that individual i went up for tenure. School rankings are taken from IDEAS/RePEc. The plot is constructed as described in Figure 1 with the y-variable residualized on the following controls before plotting: total solo and coauthored papers, the number of years it took to go up for tenure, log citations, and tenure school, tenure year, and field fixed effects. The line of best fit using OLS is shown separately for men and women. The lines of best fit are estimated on the full sample and have slopes of $\beta = 0.062$ (s.e. = 0.091) for women and $\beta = 0.109$ (s.e. = 0.056) for men.

Tables

TABLE 1: SUMMARY STATISTICS

	Full	Male	Female	p-value
Panel A:				
Tenure	0.68	0.73	0.52	0.001
	(0.47)	(0.44)	(0.50)	
Years to tenure	6.8	6.6	7.3	0.001
	(1.6)	(1.6)	(1.8)	
Total papers	8.3	8.4	8.0	0.262
	(3.9)	(4.1)	(3.3)	
Solo-authored	3.0	3.0	3.0	0.879
	(2.4)	(2.4)	(2.3)	
Coauthored	5.3	5.4	5.0	0.189
	(3.6)	(3.7)	(3.1)	
Panel B:				
Top 5 Solo	0.67	0.66	0.68	0.900
	(0.98)	(0.99)	(0.92)	
Top 5 Coauthored	1.3	1.3	1.2	0.570
	(1.4)	(1.4)	(1.4)	
AER Equivalent:				
Solo Pubs.	0.34	0.34	0.33	0.500
	(0.24)	(0.23)	(0.25)	
Coauthored Pubs.	0.33	0.34	0.30	0.039
	(0.20)	(0.21)	(0.18)	
Panel C				
Number Unique CAs	4.52	4.55	4.47	0.767
	(2.79)	(2.78)	(2.83)	
Frac. coauthors who are:				
Full Professor	0.46	0.47	0.41	0.052
	(0.35)	(0.33)	(0.38)	
Associate Professor	0.16	0.15	0.16	0.810
	(0.24)	(0.23)	(0.28)	
Assistant Professor	0.25	0.23	0.28	0.060
	(0.24)	(0.22)	(0.30)	
Graduate Student	0.017	0.015	0.021	0.239
	(0.067)	(0.056)	(0.095)	
Female	0.13	0.094	0.270	0.001
	(0.23)	(0.179)	(0.309)	
Observations	644	501	143	

This table displays the average tenure rate, pre-tenure productivity, and pre-tenure authorship patterns of men and women who went up for tenure at one of top 35 U.S. economics departments between 1985 and 2014. The top 35 institutions are taken determined according to the RePEc/IDEAS economics department rankings. In Panel A, Tenure is an indicator that equals one if an individual was promoted to associate or full professor 6-8 years after his or her initial appointment. Years to tenure is the number of years between an individual's PhD graduation year and the year s/he went up for tenure. All paper counts are measured as the number of papers an individual had published at the time of tenure. Top 5 Solo/Coauthored are the number of publications an individual had published in one of the top 5 economics journals: AER, QJE, Econometrica, JPE, and ReStud. *AER Equivalent* is a measure that converts an individual's publications into the number of AER-equivalent publications they correspond to. For more details on this variable, see Section 2. Number Unique CAs is the number of different coauthors an individual had published with by the time s/he went up for tenure. Coauthor positions (full, associate, assistant, and graduate student) are the positions an individual's coauthors had at the time that individual went up for tenure.

TABLE 2: RELATIONSHIP BETWEEN PAPERS & TENURE

	Outcome Variable: Tenure				
	(1)	(2)	(3)	(4)	(5)
Sample:	Full	Full	Full	Female	Male
Total papers	0.142***				
	(0.016)				
Fem x Papers	-0.005				
	(0.012)				
Solo-authored		0.094***	0.097***	0.196***	0.095***
		(0.013)	(0.019)	(0.055)	(0.024)
Fem x Solo		0.048***	0.057***		
		(0.018)	(0.015)		
Coauthored		0.085***	0.082***	-0.031	0.090***
		(0.016)	(0.014)	(0.054)	(0.016)
Fem x Coauthored		-0.030*	-0.026*		
		(0.016)	(0.015)		
Total coauthors	-0.005	0.001	0.003	0.025	-0.001
	(0.004)	(0.005)	(0.005)	(0.016)	(0.006)
Total Papers Sq	-0.004***				
	(0.001)				
Solo Papers Sq		-0.005***	-0.005***	-0.007	-0.005**
		(0.001)	(0.002)	(0.005)	(0.002)
Coauthored Sq		-0.003***	-0.003***	0.000	-0.003***
		(0.001)	(0.001)	(0.003)	(0.001)
Log Citations	0.059***	0.031**	0.065***	0.098*	0.058***
	(0.012)	(0.012)	(0.013)	(0.054)	(0.016)
AER Equiv. Ranking	0.533***				
	(0.116)				
AER Equiv. Solo		0.139*	0.331***	0.321	0.416***
		(0.071)	(0.069)	(0.206)	(0.091)
AER Equiv. CA		0.201**	0.325***	0.486*	0.280***
		(0.099)	(0.073)	(0.248)	(0.089)
Female	-0.135	-0.166	-0.205*		
	(0.105)	(0.121)	(0.109)		
Tenure Inst. FE	Y	N	Y	Y	Y
Tenure Year FE	Y	N	Y	Y	Y
Field FE	Y	N	Y	Y	Y
Observations	625	629	621	139	482
R-squared	0.417	0.287	0.425	0.521	0.421

This table shows the relationship between publications and tenure. The dependent variable, Tenure, is binary and indicates whether an individual received tenure 6-8 years after being hired at the initial tenure institution. Total papers is the number of papers an individual published by the time s/he went up for tenure. Solo-authored and Coauthored are the number of solo or coauthored papers s/he had published at the time of tenure. AER Equiv. Ranking, AER Equiv. Solo, and AER Equiv. CA are journal quality measures described in Section 2. Total coauthors is the number of coauthors an individual had on the papers s/he had published by the time of tenure. Tenure length is the number of years it took the individual to go up for tenure. Citations are from Google Scholar and measured in 2015. The equations are estimated using a linear probability model. Bootstrapped standard errors are clustered by tenure institution and reported in parentheses. (*=p<0.10, **=p<0.05 ,***=p<0.01)

TABLE 3: COAUTHOR GENDER

	(1)	× Female
Solo-authored	0.093***	0.049**
	(0.019)	(0.015)
CA with only fem CAs	0.097***	0.019
	(0.024)	(0.020)
CA with only male CAs	0.087***	-0.056***
	(0.015)	(0.015)
Pubs. with M and F CAs	0.087**	0.033
	(0.026)	(0.042)
Female	-0.156	
	(0.101)	
Total coauthors	-0.001	
	(0.004)	
Log Citations	0.064***	
	(0.014)	
AER Equiv. CA	0.332***	
	(0.073)	
AER Equiv. Solo	0.328***	
	(0.065)	
Tenure Inst. FE	Yes	
Tenure Year FE	Yes	
Field FE	Yes	
Observations	621	

This table presents the results of one regression where the variables that are interacted with Female (a dummy indicating that the researcher is a woman) are displayed in the right-hand column. *Papers with only fem CAs* is the number of publications an individual has in which all coauthors are female. Similarly, *Papers with only male CAs* and *Papers with male and fem CAs* are the number of publications with only male coauthors and with a mix of male and female coauthors respectively. Controls for tenure length; quadratics in the number of papers; and tenure institution, year, and field fixed effects are also included. The equations is estimated using a linear probability model. Bootstrapped standard errors are reported in parentheses and are clustered by tenure institution. (*=p<0.10, **=p<0.05 ,***=p<0.01)

TABLE 4: ROBUSTNESS CHECKS

	Faculty List Sample	Journal Rankings			Publication Count		Total AERs
		RePEc	Over Time	AER Bins	Tenure +1	Tenure +2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Solo-authored	0.115***	0.083***	0.078***	0.078***	0.058***	0.038***	0.060***
	(0.024)	(0.018)	(0.017)	(0.019)	(0.017)	(0.013)	(0.019)
Fem x Solo	0.050**	0.055***	0.052***	0.045***	0.044***	0.055***	0.091*
	(0.019)	(0.015)	(0.015)	(0.015)	(0.014)	(0.013)	(0.024)
Coauthored	0.092***	0.079***	0.081***	0.069***	0.038***	0.031***	0.089***
	(0.023)	(0.015)	(0.016)	(0.020)	(0.011)	(0.008)	(0.011)
Fem x Coauthored	-0.032*	-0.026*	-0.025*	-0.029*	-0.022*	-0.013	-0.036*
	(0.019)	(0.014)	(0.014)	(0.015)	(0.013)	(0.013)	(0.019)
Years to Tenure	-0.046***	-0.054***	-0.055***	-0.051***	-0.046***	-0.045***	-0.050***
	(0.009)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.011)
Total Coauthors	-0.001	0.004	0.004	-0.004	0.008*	0.010**	0.003
	(0.007)	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.005)
Log Citations	0.056***	0.070***	0.074***	0.079***	0.072***	0.074***	0.094***
	(0.017)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.016)
CA Paper Rank	0.310***	0.003**	0.003***		0.345***	0.345***	
	(0.084)	(0.001)	(0.001)		(0.081)	(0.081)	
Solo Paper Rank	0.457***	0.002*	0.004***		0.291***	0.299***	
	(0.077)	(0.001)	(0.001)		(0.069)	(0.065)	
Female	-0.158	-0.193*	-0.197**	-0.175*	-0.197*	-0.280***	-0.199**
	(0.128)	(0.102)	(0.099)	(0.100)	(0.105)	(0.110)	(0.048)
Observations	369	621	621	621	621	621	621

The dependent variable in all columns is an indicator for receiving tenure. Column 1 restricts the sample to those schools we received a historical faculty list from. Column 2 uses RePEc journal rankings as the paper quality measure. The ranking used can be found at https://ideas.repec.org/top/top.journals.all.html. Column 3 uses historical journal rankings from RePEc to allow for rankings to change over time and to account for new journals entering. Column 4 controls for the number of papers within each of 10 AER "bins". For this analysis, the AER-equivalent measure is divided into deciles. For each individual, we then add up the number of solo and coauthored papers within each decile and include the number of papers in each bin as controls. In Columns 5 and 6, we include papers that were published one and two years after an individual went up for tenure in the paper counts. In Column 7, we use the AER Equivalent measure of journal ranking to calculate the total number of AER equivalents (solo and coauthored) an individual had at the time of tenure. We use this measure in place of the solo and coauthored paper counts (the main independent variables). All regressions control for a quadratic in the number of papers and tenure length, as well as tenure institution, tenure year, and field fixed effects. Bootstrapped standard errors are reported in parentheses and are clustered by tenure institution. (*=p<0.10, **=p<0.05 ,***=p<0.01)

TABLE 5: FUTURE PRODUCTIVITY

Outcome Var:	Post Tenure Solo AER Equivalents	Log Citations
	(1)	(2)
	Poisson	OLS
Fraction Coauthored	-1.45***	0.533
	(0.500)	(0.390)
Female	-0.232	-0.151
	(0.380)	(0.414)
Female × Frac. Coauthored	1.057*	0.742
	(0.576)	(0.660)
Tenured	0.194	0.496*
	(0.352)	(0.289)
Tenured× Frac. Coauthored	0.002	0.408
	(0.006)	(0.486)
Female × Tenured	0.185	0.210
	(0.528)	(0.509)
Fem × Tenured × Frac. Coauthored	-0.991	-0.740
	(1.131)	(0.769)
Top 5 Coauthored	0.013**	
	(0.007)	
Total papers		0.071***
		(0.015)
Tenure Inst. FE	N	Y
Post-Tenure Inst. FE	Y	N
Tenure Year FE	Y	Y
Field FE	Y	Y
Observations	621	621

Column 1 shows the results from estimating equation 6 using a zero-inflated Poisson model, where the outcome variable is the number of solo-authored AER equivalents an individual published after the tenure decision (measured as of 2017). "Top 5 Coauthored" is the number of coauthored AER equivalents the individual published after tenure. Post-tenure institution is the institution the individual went to following the tenure decision. For people who received tenure, this the same as the tenure institution. Column 2 shows the results from estimating the same equation using OLS where log citations is the outcome variable. Citations are measured in 2015. Robust standard errors are reported in parentheses and are clustered at the tenure institution or post-tenure institution level. (*=p<0.10, **=p<0.05 ,***=p<0.01)

TABLE 6: SURVEY RESULTS

	(1) Men	(2) Women	(3) p-value
Panel A: Beliefs about Returns to Papers			
Coauthored AER	12.1	12.2	0.939
Coauthored AER, Sr. Faculty	9.1	8.8	0.528
Coauthored AER, Jr. Faculty	13.3	13.4	0.796
Solo Top Field	8.0	8.2	0.669
Coauthored Top Field	6.3	6.8	0.223
Panel B: Frequency of Presenting Papers			
Times Presented	3.1	2.2	0.071
Present More Freq. than CA	0.37	0.44	0.203
Observations	300	89	

This table presents the mean responses for men and women to the following survey questions: Panel A: "Suppose a solo authored AER increases your chance of receiving tenure by 15 percent. By how much do you think each of the following increases your change of receiving tenure?" Panel B: "How many times per year do you typically present your solo-authored papers? Are you more or less likely than your coauthors to present a joint paper?" *Present More Freq. than CA* is the fraction of respondents who reported that they are more likely than their coauthors to present a joint paper. The survey was conducted with a sample of academic economists currently working at a top 35 U.S. economics department. Respondents were anonymous.

TABLE 7: ACCOUNTING FOR SORTING

	Dep. Variable: Tenure		
	(1)	(2)	(3)
Solo-authored	0.086***	0.084***	0.087***
	(0.017)	(0.017)	(0.018)
Fem x Solo	0.064***	0.067***	0.060***
	(0.017)	(0.018)	(0.018)
Coauthored	0.087***	0.089***	0.087***
	(0.014)	(0.014)	(0.016)
Fem x Coauthored	-0.032*	-0.032*	-0.032**
	(0.016)	(0.016)	(0.016)
Female	-0.220	-0.348*	-0.187
	(0.121)	(0.133)	(0.127)
Rank Difference	0.001		
	(0.002)		
Fem × Rank Difference	-0.001		
	(0.002)		
Avg. Coauthor Rank		-0.002	
		(0.001)	
Fem × Avg. Coauthor Rank		0.003	
		(0.002)	
Frac. Full Prof.			-0.035
			(0.076)
Fem × Frac. Full Prof.			0.026
			(0.068)
Observations	595	595	595

The dependent variable in all columns is an indicator for receiving tenure. Column (1) shows the relationship between solo and coauthored papers and tenure when controlling for the difference between individual i's institution rank and the average institution rank of his or her coauthors. Column (2) controls for the average institution rank of an individual's coauthors, and column (3) controls for the fraction of an individual's coauthors who are full professors. Only coauthors that an individual coauthored with up until tenure are included. All regressions control for tenure length, journal rank (AER equivalent measure), and log citations. They also include tenure institution, tenure year, and field fixed effects. The sample size is smaller in this analysis because individuals with no coauthors are excluded. (*=p<0.10, **=p<0.05 ,***=p<0.01)

TABLE 8: PAPER SPLIT BY TOP 5

Dep Var: Tenure		
	(1)	
	Top 5	Non-Top 5
Solo	0.067***	0.033***
	(0.019)	(0.007)
Coauthored	0.086**	0.031***
	(0.016)	(0.007)
Fem x Solo	0.020	0.055**
	(0.037)	(0.019)
Fem x Coauthored	-0.007	-0.035**
	(0.031)	(0.017)
Female	-0.171	
	(0.108)	
Total coauthors	-0.002	
	(0.005)	
Years to tenure	-0.049***	
	(0.008)	
Log Citations	0.079***	
	(0.012)	
Tenure Inst. FE	Y	
Tenure Year FE	Y	
Field FE	Y	
Observations	621	
R-squared	0.415	

This table presents the results from estimating equation 9. The results in the able are from this single regression, but solo and coauthored papers are split into those published in the top 5 journals (Column 1) and journals below the top 5 (Column 2). Top 5 papers are those published in the American Economic Review, Econometrica, the Journal of Political Economy, Quarterly Journal of Economics, or the Review of Economic Studies. The dependent variable is an indicator for receiving tenure. The regression includes tenure institution, tenure year, and field fixed effects. Robust standard errors are clustered by tenure institution and reported in parentheses. (*=p<0.10, **=p<0.05 ,***=p<0.01)

TABLE 9: TIMING OF COAUTHORSHIP WITH MEN

	Years to First Publication (1)	Fraction of Papers with Men (2)
Female	0.054	0.062
	(0.186)	(0.080)
Tenure	-0.002	0.060
	(0.130)	(0.047)
Female × Tenure	-0.151	-0.258**
	(0.237)	(0.089)
Years to 1st Pub.		0.016
		(0.015)
Fem × Years to 1st Pub.		-0.013
		(0.032)
Tenure × Years to 1st Pub.		-0.030
		(0.018)
Fem × Tenure × Years to 1st Pub.		0.017
		(0.041)
Total papers	-0.119***	0.009*
	(0.015)	(0.003)
AER Equiv.	-0.380	0.251**
	(0.344)	(0.081)
School FE	Y	Y
Tenure Year FE	Y	Y
Primary Field FE	Y	Y
Observations	603	594

This table tests whether there are gender differences in the timing of an individual's first publication (Column 1) and whether women who take a longer time to publish their first paper are more likely to coauthor with men (Column 2). The outcome variable in Column 1 is the number of years it takes an individual to publish his or her first paper after graduating, and is measured as the year of the individual's first publication minus the year of the individual's initial faculty appointment. Articles published before the first appointment (i.e. during graduate school) are not counted. The outcome variable in Column 2 is the fraction of an individual's papers published by tenure that are coauthored with men. The independent variable, $Yearsto1stPub$ is the outcome variable in Column 1. Both regressions include tenure institution, tenure year, and field fixed effects. Robust standard errors are reported in parentheses. (*=p<0.10, **=p<0.05 ,***=p<0.01)

TABLE 10: EXPERIMENT I PREDICTED SCORE BY QUIZ TYPE

Dep. Var.: Predicted Quiz 2 Score	Math		Grammar	
	Ind.	Joint	Ind.	Joint
	(1)	(2)	(3)	(4)
Female	0.111	-0.243**	-0.021	0.103
	(0.081)	(0.118)	(0.071)	(0.114)
Gender-Information	0.145	0.076	-0.241**	-0.456***
	(0.097)	(0.134)	(0.112)	(0.130)
Female × Gender-Information	-0.108	-0.106	0.194	0.738***
	(0.115)	(0.154)	(0.131)	(0.153)
Quiz 1 Score	0.735***	0.020	0.725***	0.016
	(0.057)	(0.051)	(0.066)	(0.053)
Constant	0.137	3.432***	0.289	3.246***
	(0.212)	(0.361)	(0.248)	(0.386)
Observations	250	266	231	262
Predictors	125	133	116	131
R-squared	0.298	0.041	0.239	0.139

This table presents the results from Experiment I in which participants predict how well an individual did on a math or grammar quiz based on that individual's performance on an earlier quiz. Columns 1 and 2 show the results for the math quiz and Columns 3 and 4 show the results from the grammar quiz. In the experiment, participants were randomized into the Individual treatment, where participants saw each individual's score on a previous quiz (Columns 1 and 3), or the Joint treatment, where participants saw the sum of two individuals' scores (Columns 2 and 4). Gender-Information is a dummy indicating that participants were told the average quiz scores of all men and women. (*=p<0.10, **=p<0.05 ,***=p<0.01)

Table 11: Experiment II Odds Ratios of Being Picked by Task and Recruiter Gender

Dep. Var.:	All recruiters		Male recruiters				Female recruiters			
Picked by recruiter	Search	Vocab.	Search		Vocabulary		Search		Vocabulary	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female	1.120	1.041	0.890	0.922	0.750**	0.723**	1.341**	1.157	1.390**	1.237
	(0.108)	(0.099)	(0.136)	(0.142)	(0.104)	(0.098)	(0.167)	(0.152)	(0.181)	(0.178)
Female $\times$ Belief				0.694***		0.724***		0.760***		0.844*
				(0.072)		(0.072)		(0.081)		(0.086)
Joint Score	1.173***	1.032***	1.151***	1.149***	1.036***	1.036***	1.190***	1.189***	1.032***	1.032***
	(0.026)	(0.003)	(0.036)	(0.037)	(0.005)	(0.005)	(0.036)	(0.037)	(0.005)	(0.005)
Candidate resumé controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Set $\times$ recruiter fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3144	2172	1368	1368	996	996	1776	1776	1176	1176
Recruiters	262	181	114	114	83	83	148	148	98	98

This table presents the results from Experiment II in which human resource recruiters pick one candidate out of four for a search or vocabulary task based on short resumes. Columns 1, 3-4, and 7-8 show the results for the search task and Columns 2, 5-6, and 9-10 show the results for the vocabulary task. Results are shown separately depending on the recruiter's gender: all recruiters in Columns 1-2, male recruiters in Columns 3-6, and female recruiters in Columns 7-10. All regressions include fixed effects for each set-recruiter combination and controls for other variables in the candidates' resumes. Results are presented as odds ratios. Standard errors are clustered at the recruiter level. (*=p<0.10, **=p<0.05 ,***=p<0.01)

Appendix A Additional Tables

TABLE A1: RESULTS BY INSTITUTION AND YEAR

	Panel A: Tenure Institution		
Institution Rank:	Top 10	Top 20	Top 35
	(1)	(2)	(3)
Solo-authored	0.031***	0.053***	0.039**
	(0.006)	(0.018)	(0.018)
Coauthored	0.035**	0.052***	0.023***
	(0.013)	(0.016)	(0.007)
Fem x Coauthored	0.002	-0.048**	-0.048*
	(0.027)	(0.020)	(0.026)
Fem x Solo	0.074*	0.071*	0.104***
	(0.035)	(0.037)	(0.035)
Female	-0.471*	-0.048	-0.245
	(0.247)	(0.173)	(0.243)
Observations	211	157	155
	Panel B: Tenure Year		
Tenure Year:	1985-1995	1996-2005	2006-2014
	(1)	(2)	(3)
Solo-authored	0.034***	0.043**	0.033*
	(0.010)	(0.021)	(0.019)
Coauthored	0.018*	0.049***	0.047***
	(0.010)	(0.011)	(0.015)
Fem x Coauthored	0.011	-0.047**	-0.053*
	(0.041)	(0.022)	(0.027)
Fem x Solo	0.145***	0.079***	0.054
	(0.037)	(0.029)	(0.042)
Female	-0.787***	-0.219	-0.003
	(0.275)	(0.160)	(0.202)
Observations	141	157	215

Panel A shows the relationship between coauthoring and tenure by tenure institution rank. Schools are divided into the top 10, top 20, and top 35 departments, according to the RePEc rankings. All regressions include the following controls: time until tenure, number of coauthors, log citations, solo and coauthored journal rankings, and tenure year and field fixed effects. Panel B shows the relationship splitting the sample by time period. The year groups are the years that an individual went up for tenure. All regressions include the following controls: time until tenure, number of coauthors, log citations, solo and coauthored journal rankings, and tenure rank and field fixed effects. (*=p<0.10, **=p<0.05 ,***=p<0.01)

Sociology Results

The sociology sample consists of randomly sampled faculty at the top 20 sociology PhD-granting departments in the U.S.[33] There are 250 sociologists in the sample, 40% of whom are female. Summary statistics are presented in Table A2. There is no statistically significant difference between men and women's tenure rates (with the mean tenure rate being 76%) although men seem to publish more solo-authored articles than women.

TABLE A2: SOCIOLOGY SUMMARY STATISTICS

	Men	Women	p-value
Tenure	0.752	0.776	0.547
	(0.433)	(0.419)	
Total papers	12.15	10.18	0.033
	(7.808)	(5.726)	
Total coauthored	6.409	5.959	0.567
	(6.641)	(4.999)	
Solo papers	5.745	4.224	0.003
	(4.451)	(2.892)	
Time to tenure	7.584	7.520	0.686
	(1.607)	(1.724)	
Books	0.779	0.571	0.139
	(1.185)	(0.799)	
Observations	150	100	

This table presents summary statistics for the full sample of sociologists and separately for men and women. All paper and book count variables (*Total Papers*, *Solo-authored*, *Coauthored*, and *Top 5*s) are the number of papers or books an individual had published at the time of tenure.

To test whether men and women are treated differently, we reestimate equation 3 using a probit model but include measures of the number of papers that researcher i is first author on. The results are presented in Table A3. We include the number and fraction of papers a researcher is first author on in Columns 1 and 2 respectively, along with female dummy interaction terms.

[33] Ranking from U.S. News Education

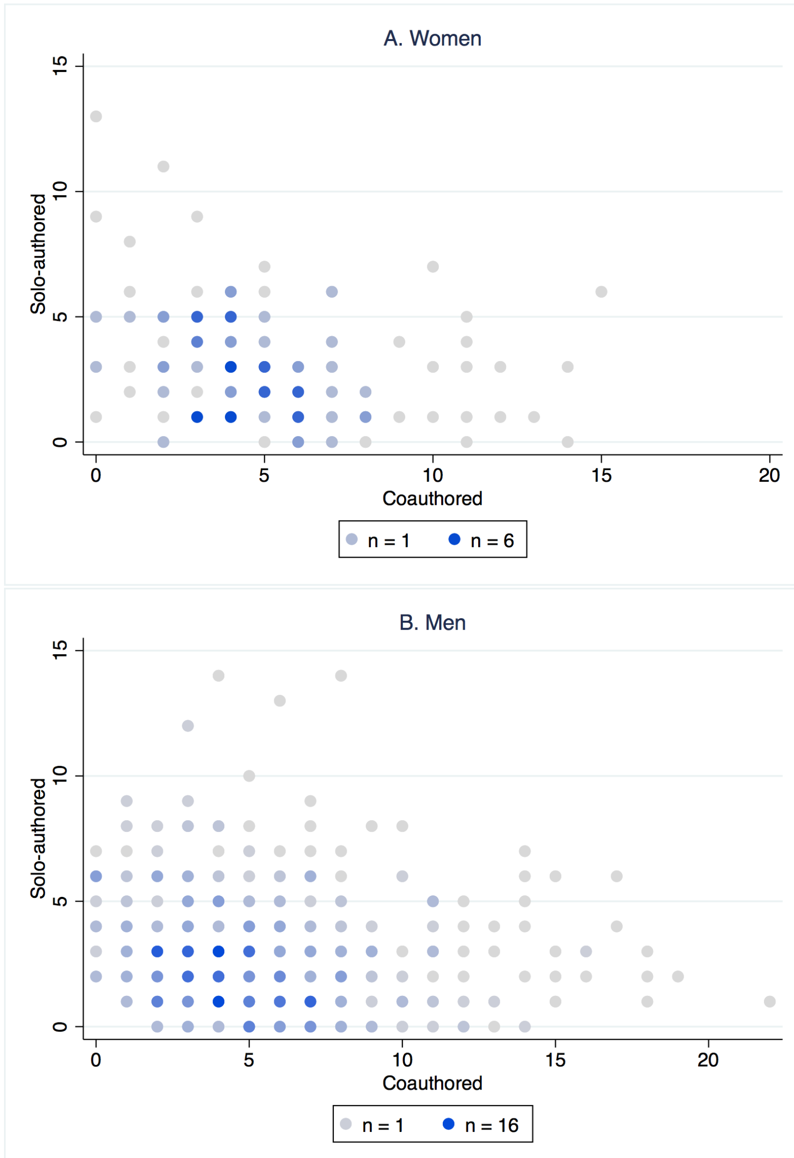
TABLE A3: SOCIOLOGY: PAPERS AND TENURE

Dep Var: Tenure	Probit (1)	Probit (2)
Total first author	0.050**	
	(0.017)	
Fem x First Author	0.026	
	(0.040)	
Fraction first author		0.403***
		(0.043)
Fem x Frac. First Author		-0.042
		(0.172)
Solo papers	0.008	0.000
	(0.006)	(0.006)
Fem x Total Solo	0.002	0.007
	(0.011)	(0.011)
Total Coauthored	-0.010*	0.009
	(0.004)	(0.007)
Fem x Total CA	-0.020	0.001
	(0.017)	(0.015)
Books	0.063*	0.058
	(0.032)	(0.035)
Book chapters	0.007	0.005
	(0.013)	(0.012)
Female	0.026	0.010
	(0.114)	(0.163)
School FE	Yes	Yes
Tenure Year FE	Yes	Yes
Observations	237	209

This table shows the relationship between the number and types of papers an individual publishes and tenure for a sample of sociologists. The dependent variable is a binary variable indicating whether the individual received tenure 6-7 years after being hired at the initial tenure institution. *Total first author* is the number of papers an individual is first author on while *Fraction first author* is the fraction of an individual's papers that s/he was first author on. The equations are estimated using a probit model and the marginal probabilities calculated at the mean are displayed. Standard errors, reported in parentheses, are clustered by tenure institution. (*=p<0.10, **=p<0.05 ,***=p<0.01)

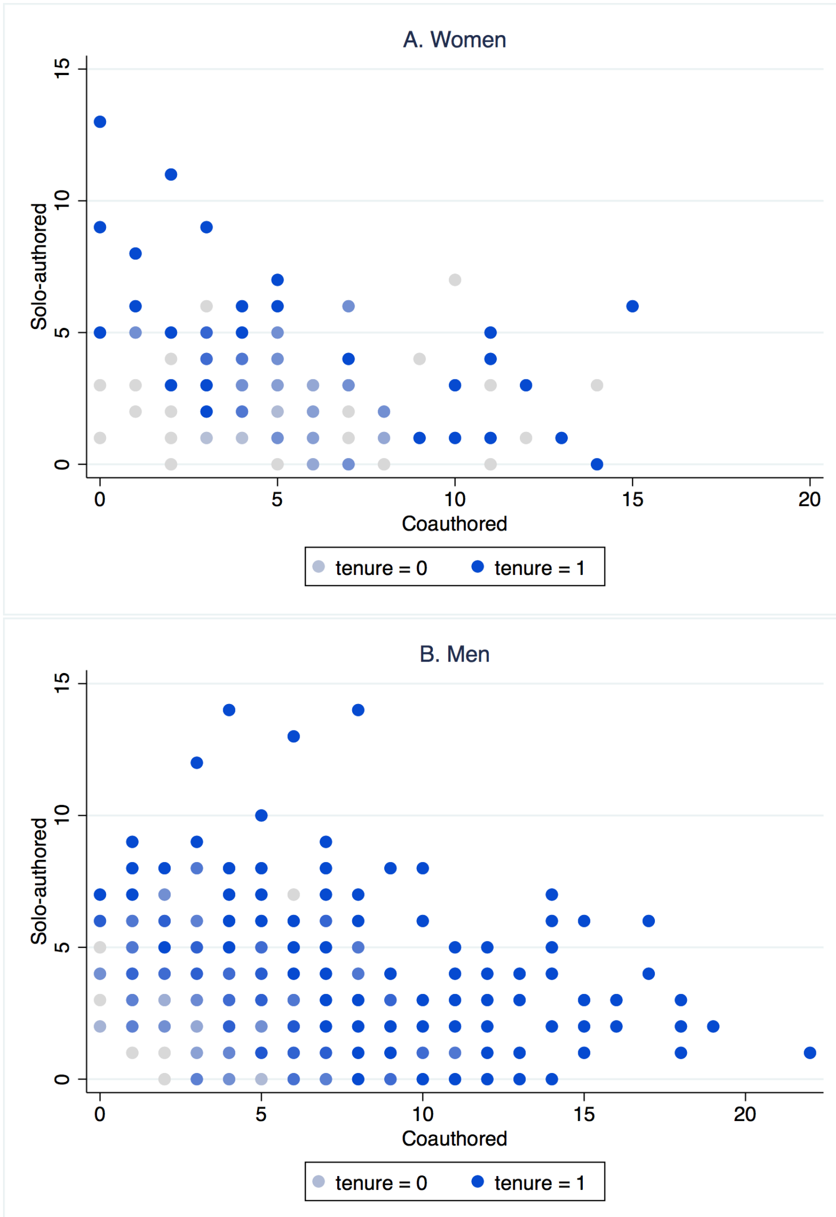
Appendix B Additional Figures

FIGURE B1: DISTRIBUTION OF PAPER COMBINATIONS



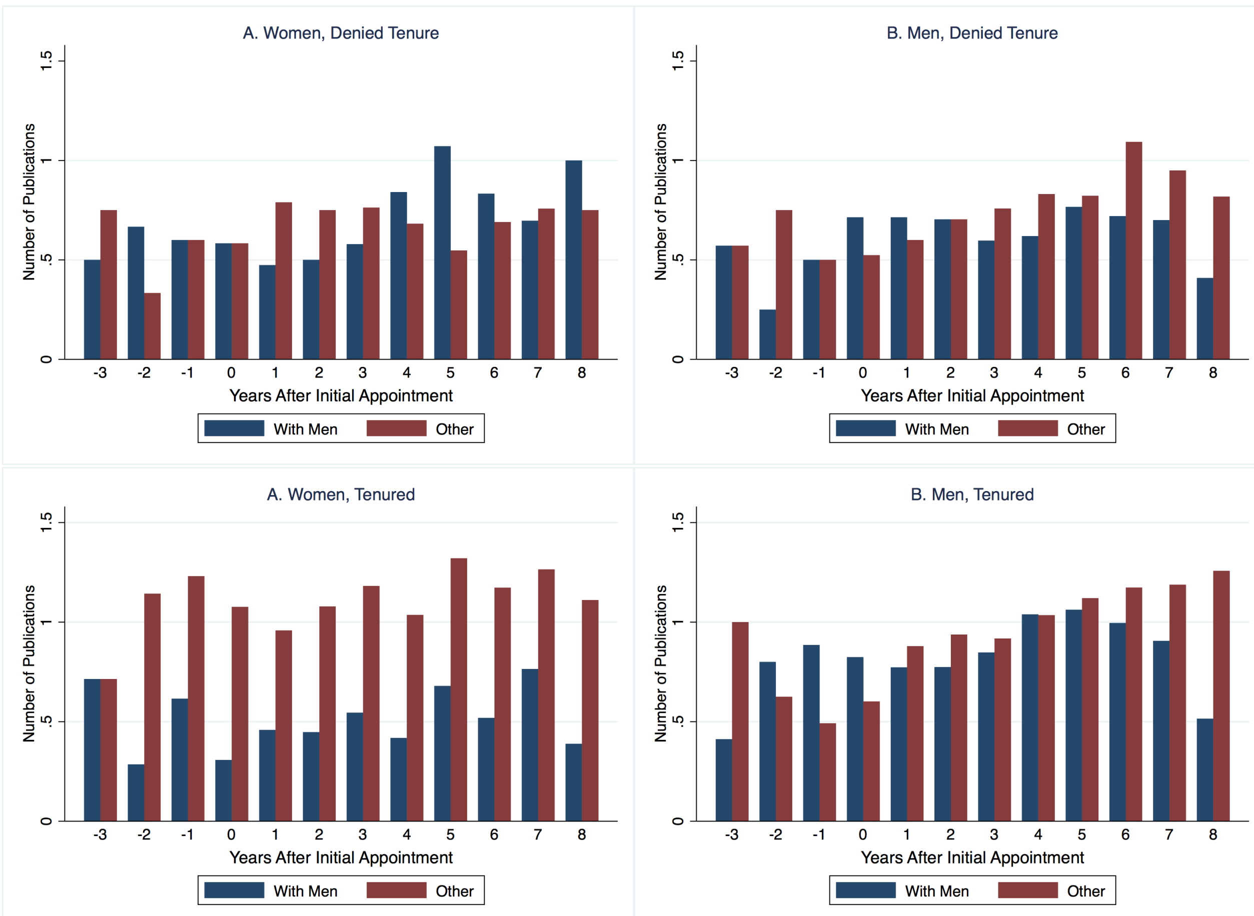
Notes: This figure shows the number of women (Panel A) and men (Panel B) who had various combinations of solo and coauthored papers at the time of tenure. Each dot represents a specific combination of papers with the number of coauthored papers measured on the x-axis and the number of solo-authored papers measured on the y-axis. The shading of the dots represents how many individuals had that combination of papers at the time they went up for tenure, with darker shades indicating a larger number of individuals with that combination. In the legend, "n" is the minimum and maximum number of individuals who have a specific paper combination. Panel A is constructed using the full sample of women (N=143) and Panel B is constructed using the full sample of men (N=501).

Figure B2: Tenure Probabilities by Paper Combinations



Notes: This figure plots the unconditional tenure probability for women (Panel A) and men (Panel B) who have various combinations of papers at the time they go up for tenure. Coauthored papers are counted along the x-axis and solo-authored papers are counted along the y-axis. A darker shade indicates a higher probability of receiving tenure. For example, if a dot is the darkest shade, it indicates that individuals with that combination of solo and coauthored papers receives tenure with probability one. Panel A is constructed using the full sample of women (N=143) and Panel B is constructed using the full sample of men (N=501).

FIGURE B3: TIMING OF PUBLICATIONS



Notes: This figure shows the average number of publications an individual has in the years surrounding his or her initial appointment as an assistant professor. Year 0 is the year that the individual begins working at his/her tenure institution (tenure institutions are defined in Section 2). The blue bars represent publications that are coauthored with men. The red bars represent all other publications (either solo-authored or coauthored with women). Panels A and B show the timing of publications for women and men who were denied tenure. Panels C and D show the timing of publications for women and men who received tenure.

Appendix C Institutions List

Received faculty list: Brown, Columbia, Cornell, Duke, Harvard, Michigan State University, New York University, Northwestern, Ohio State University, Penn State, Rutgers, Stanford, UC Berkeley, UC Davis, UC San Diego, UCLA, University of Virginia, University of Maryland, University of Michigan, University of Minnesota, University of Pennsylvania, University of Wisconsin-Madison

No faculty list: Boston College, Boston University, California Institute of Technology, Georgetown, MIT, Princeton,University of Southern California, University of Chicago, University of Texas - Austin, University of Rochester, Vanderbilt, Yale

Appendix D Experiment I

This section provides additional information for Experiment I. As mentioned in the main body of the paper, the first experiment was conducted with participants from the mTurk online platform. First, 80 participants were recruited to play the role of "workers" and perform two five-question quizzes (21 men and 19 women completed the math quizzes while 23 men and 17 women completed the grammar quizzes). Workers received a participation fee of $0.30 plus $0.05 for each question they answer correctly. On average, workers earned $0.55. The quizzes used are provided below.

For the main part of the experiment, 505 participants were recruited to predict the scores of one randomly-chosen male worker and a randomly-chosen female worker in a task. Predictors were paid a participation fee of $0.50 and received $0.10 for each score they correctly predicted. The number of predictors in each treatment was as follows: 242 recruiters were assigned to the Individual treatment, of which 120 were assigned to the No-Information treatment (62 for math quizzes and 58 for grammar quizzes) and 122 to the Gender-Information treatment (63 for math quizzes and 59 for grammar quizzes), and 264 recruiters were assigned to the Joint treatment, of which 138 were assigned to the No-Information treatment (70 for math quizzes and 68 for grammar quizzes) and 126 to the Gender-Information treatment (63 for math quizzes and 63 for grammar quizzes).

D.1 Quizzes used

Grammar Quiz 1

1. The storm prevented .... on a picnic.

   (a) us to going (b) us going (c) us to go (d) us from going

2. A man's concept of liberty is different from ....

   (a) a woman's (b) womens (c) a woman (d) woman's

3. .... hour went by before we received .... invitation

   (a) an/an (b) a/a (c) an/a (d) a/an

4. When a subordinate clause is followed by the main clause, what is required?

   (a) a dash (b) a semi-colon (c) a period (d) a comma

5. .... are used around a relative clause that defines the noun it follows.

   (a) Only commas (b) No commas (c) Semi-colons (d) Quotation marks

Grammar Quiz 2

1. I am dizzy and need to .... down

   (a) lie (b) lay (c) lye (d) go lay

2. Which of these is not an article?

   (a) The (b) A (c) It (d) An

3. His idea is .... mine

   (a) different to (b) different from (c) different than (d) different then

4. Adverbs can modify which of the following?

   (a) nouns (b) adjectives (c) pronouns (d) none of the above

5. .... did you bump into?

   (a) Who (b) Whose (c) Who's (d) Whom

Math Quiz 1

1. Which of the following is a subset of {b,c,d}?

   (a) { } (b) {a} (c) {1,2,3} (d) {a,b,c}

2. A man's regular pay is $3 per hour up to 40 hours. Overtime is twice the payment for regular time. If we was paid $168, how many hours overtime did he work?

   (a) 8 (b) 16 (c) 28 (d) 48

3. 3 4/5 expressed as a decimal is

   (a) 3.40 (b) 3.45 (c) 3.50 (d) 3.80

4. Which of the following is the highest common factor of 18, 24, and 36?

   (a) 6 (b) 18 (c) 36 (d) 72

5. Given that a and b are integers, which of the following is not necessarily an integer?

   (a) $2a - 5b$ (b) a^7 (c) b^a (d) ab

Math Quiz 2

1. Items bought by a trader for \$80 are sold for \$100. The project expressed as a percentage of cost price is

   (a) 2.5% (b) 20% (c) 25% (d) 50%

2. A man bought a shirt at a sale. He saves \$30 on the normal price when he paid \$120 for the shirt. What was the percentage discount on the shirt?

   (a) 20 (b) 25 (c) 33.33 (d) 80

3. How many subsets does {a,b,c,d,e} have?

   (a) 2 (b) 4 (c) 10 (d) 32

4. What is the median of the given data: 13, 16, 12, 14, 19, 14, 13, 14

   (a) 14 (b) 19 (c) 12 (d) 14.5

5. In coordinate geometry, what is the equation of the x-axis?

   (a) $y = 0$ (b) $x = y$ (c) $x = 0$ (d) $y = 1$

D.2 Instructions

Below are the instructions for the Joint and Gender-Information treatments. Instructions for the Individual and No-Information treatments are almost identical and are available upon request.

Instructions screen 1

INSTRUCTIONS: Please read all the way through.

This project seeks to understand how well individuals can predict a person's future performance on a task based on his/her past performance.

We recruited a group of people to complete two math [grammar] quizzes. Each quiz had five questions. Participants had one minute to complete each quiz. In what follows, we will show two participants' scores from the first quiz. We then ask you to predict each participant's score on the second quiz. We will provide you with some basic information on each individual.

You will be paid \$0.50 for your participation but will also be paid a bonus of \$0.10 if you correctly guess a participant's score on the second quiz.

Instructions screen 2

We will first show you the distribution of scores on the first quiz. Each bar represents the fraction of people who obtained that score. For example, 30% of people scored 4/5 on the first quiz. The average score of female participants (2.5/5) is shown by the solid line. The average score of male participants (2.8/5) is shown by the dashed line.

Instructions screen 3

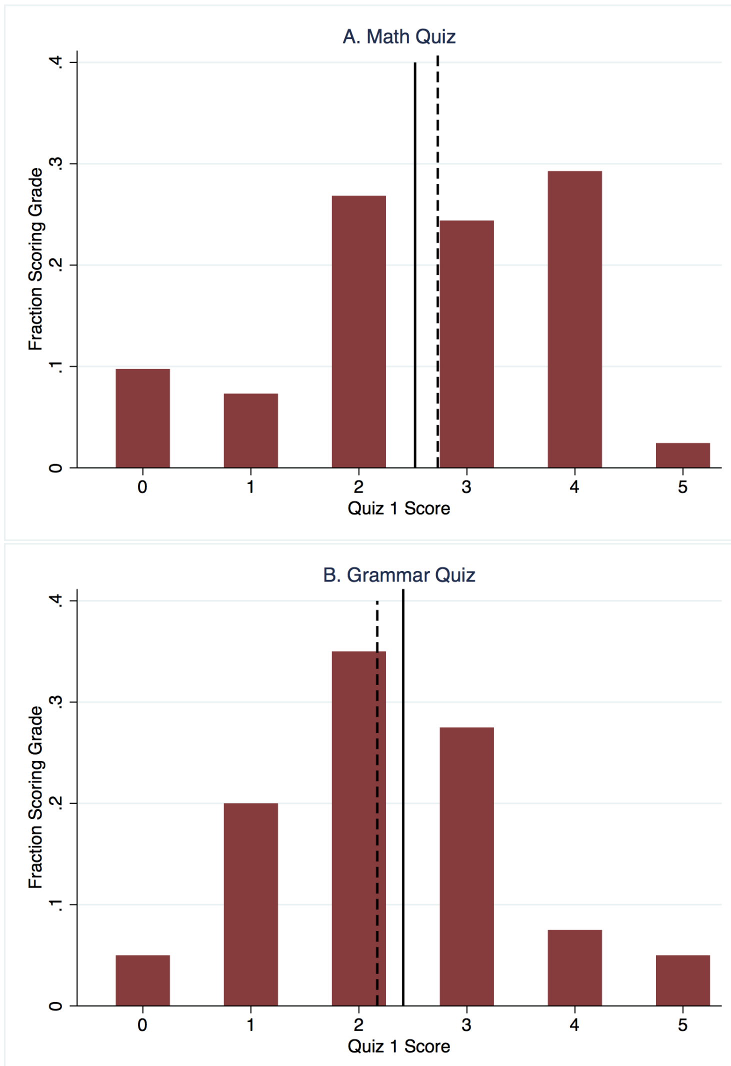
Below we are showing you a team's score on Quiz 1. Recall that each team member worked on the questions independently. We then take the sum of the two scores and assign it to the team. For example, if Person A scored 3/5 and Person B scored 4/5, their team score would be 7/10. We provide you with some basic demographic information about each team member.

Based on the team's performance, please predict each individual's score on Quiz 2. You can view each quiz by clicking on the link below.

Histograms

The histograms seen by recruiters containing the distribution of scores are seen below in Figure D1.

FIGURE D1: DISTRIBUTION OF SCORES IN QUIZ 1



Notes: These bar graphs show the distribution of scores on first math and grammar quizzes. The lines mark the means score of men (dashed line) and women (solid lines). The experiment participants who predicted scores saw these distributions with or without the lines, depending on whether they were in the Gender-Information treatment.

Appendix E Experiment II

This section provides additional information and analysis for Experiment II.

E.1 Candidates

Before running Experiment II, the sets of candidates are constructing using data from students who took part in laboratory experiments run in Bologna and Abu Dhabi.[34] In Bologna, 68 students completed one of the two tasks (16 men and 20 women completed the search task while 12 men and 20 women completed the vocabulary task), while in Abu Dhabi, 90 students completed both tasks (42 men and 48 women). Students were paid according to their performance in the tasks.

Vocabulary Task

Students are asked to solve Word-in-a-Word puzzles. They are given 'large' words, one at a time. The task is to find smaller Italian (Bologna) or English (Abu Dhabi) words that can be formed out of the letters of the large word. The task lasts 15 minutes. There is a maximum of 24 large words and participants can freely move to the next word at any time, but cannot return to previous words. The following rules apply: (i) words must consist of four letters or more, (ii) each letter of the large word can only be used once, (iii) proper nouns (names, etc.) are not allowed, and (iv) plurals and verb conjugations are allowed. Points are awarded to submitted words of n letters according to the following rules: (i) each word found in a dictionary adds $(n-3)$ points to the score, (ii) words not found in a dictionary subtract $(n-3)$ points from the score, (iii) words that are too short subtract 1 point from the score; and (iv) words submitted more than once have no impact on the score. Points were converted to cash at an exchange rate of 0.10 euros per point in Bologna (around \$0.11 per point) and 1 Emirati dirham per point in Abu Dhabi (around \$0.27 per point).

Search Task

Students are shown two 10x10 matrices. Each cell is filled with a two-digit number. The task is to find the highest number in each matrix, add these up, and enter the sum. Each correct answer increases the score by one point. After entering a number, a new pair of

[34]We thank BLESS for allowing us to use their facilities in Bologna. The experimental software used in Bologna was developed in PHP-MySQL with the help of Ailko van Veen and Joep Sonnemans, and was later adapted for the use in Qualtrics by Manu Muñoz. In Abu Dhabi, the experiment was run using zTree.

matrices appear, irrespective of whether the sum is correct. The task lasts 15 minutes. Points were converted to cash at an exchange rate of 0.50 euros for every point in Bologna (around \$0.55 per point) and 4 Emirati dirhams per point in Abu Dhabi (around \$1.09 per point).

Resumes

In addition to performing the tasks, students in Bologna and Abu Dhabi answered a few questions about their demographics and studies. This information and their scores are used to construct eight sets of "candidates" for each task. Each set consists of the resumes of four candidates. The resume of each candidate includes information about their score in the real effort task as well as their field of study, degree length (from three to five years), age, gender, and geographic region of origin. The score is shown for each candidate in the Individual treatment or as sums of two pairs of candidates in the Joint treatment. An example from the Vocabulary task treatment is provided in Figure E1. The other sets of this treatment and those of the Search task treatment are available upon request.

FIGURE E1: EXAMPLE OF ONE SET OF CANDIDATE RESUMES

Individual treatment

	Student 1	Student 2	Student 3	Student 4
Type of degree	4-year degree	4-year degree	4-year degree	5-year degree
Field of study	Social sciences	Social sciences	Engineering	Law studies
Age	19	19	19	24
Gender	Male	Female	Male	Female
Region of nationality	North America	North America	South Asia	European Union
Score in word task	27 points	13 points	14 points	0 points

Joint treatment

	Student 1	Student 2	Student 3	Student 4
Type of degree	4-year degree	4-year degree	4-year degree	5-year degree
Field of study	Social sciences	Social sciences	Engineering	Law studies
Age	19	19	19	24
Gender	Male	Female	Male	Female
Region of nationality	North America	North America	South Asia	European Union
Score in word task	Student 1 + Student 2 = 40 points		Student 3 + Student 4 = 14 points	

Note that the sets are constructed such that the summed score of one pair of candidates in the Joint Treatment is obviously better to that of the other pair (e.g., candidates 1 and 2 in Figure E1). The candidate pairs with the high score always include a male and a female whose resumes are otherwise alike. Specifically, the field of study, degree length, and geographic region of origin is always identical while age is allowed to vary but within a narrow range. The characteristics of the pair of candidates with lower joint scores are permitted to vary. This design is used to mask the purpose of the study to recruiters by giving them multiple characteristics to base their decision on, while at the same time keep these characteristics constant within the relevant pair of candidates.

E.2 Procedures

For Experiment II, human resource workers from the United States and India were recruited from Qualtrics' panel of participants to complete an incentivized online experiment.[35] Only respondents who are involved in their firm's hiring decisions and those that passed a set of attention checks are considered. Respondents who complete the experiment receive a participation fee set by Qualtrics plus additional incentives based on their choices. In total, 479 human resource workers (212 in the U.S. and 267 in India) took part in the experiment as "predictors".

Predictors are randomly assigned to the Vocabulary task treatment or the Search task treatment, and subsequently, to the Individual treatment (top example in Figure E1) or the Joint treatment (bottom example in Figure E1). The number of predictors in each treatment was as follows: 281 predictors were assigned to the Search task treatment, of which 19 were assigned to the Individual treatment (10 men and 9 women) and 262 to the Joint treatment (114 men and 148 women), and 198 predictors were assigned to the Vocabulary task treatment, of which 17 were assigned to the Individual treatment (6 men and 11 women) and 181 to the Joint treatment (83 men and 98 women). More predictors were assigned to the Joint treatment because that is the treatment of interest.

Predictors first complete a simplified version of the task they are assigned to and earn \$0.06 per point in the vocabulary task or \$0.15 per point in the Search task. Thereafter, in the main part of the experiment, each predictor sees three sets of four candidates and is required to pick one student from each set. The sets are shown sequentially and are picked at random from the eight constructed sets. The picked students' scores are paid

[35]Throughout the paper, we pool the data from the U.S. and India. However, our results are unaffected by further controlling for the recruiter's country. Running regressions like the ones in Table 11 including an interaction between the gender dummy fem_{ik} and a country indicator results in insignificant coefficients for the interaction term.

out to the predictor at a rate of $0.06 per point in the Vocabulary task or $0.15 per point in the Search task. Finally, predictors are asked whether they think that men or women are better at the task they have participated in. Responses are in five categories and choosing the correct answer (based on the students' actual scores in the task) is rewarded with $1.50. Instructions for the experiment are provided below.

E.3 Instructions

Below are the instructions for the Joint treatment with the Search task. Instructions for the Individual treatment and the Vocabulary task are very similar and are available upon request.

Instructions welcome screen

Thank you for taking part in this survey! The survey will take around 20 minutes to complete. We would like to see how people make choices when they have to select someone based on task performance. We will explain this in much more detail later.

You will be compensated for participating in this survey in the usual way. In addition, you may make *extra earnings*, depending on the answers you give and choices you make. How you can make extra earnings will be made clear in subsequent instructions. All extra earnings you make will be calculated in US dollars. Your total earnings in dollars will be paid to you as panel points in the usual manner. Once again, these extra earnings come on top of your compensation for participating.

Your decisions in the study are private and anonymous. They will not be linked to your name in any way. We are interested in your own decisions. We kindly request that you do not communicate with other people while taking part in the study.

The study consists of three parts. Part 2 will be explained after you have finished Part 1 and Part 3 will be explained after you have finished Part 2. Next, we will explain Part 1.

Instructions part 1 screen

In this first part, we ask you to do a simple addition task with which you can earn money.

When you start, you will see two matrices on the screen. Each matrix has 6 rows and 6 columns and is filled with randomly generated numbers. Your task is to find the largest number in each of the two matrices and then to add them up. We will give you an example below.

For each correct addition, you will receive $0.15. You will have five minutes to do this task. Irrespective of whether your answer is correct or incorrect, a new pair of matrices will appear after you enter your answer. This means that, for each pair, you have only one attempt to provide the correct answer. At the top of the screen you can see how many correct answers you have so far.

As mentioned, you will have five minutes in total. You will see the time that remains in the upper right corner of the screen. You will be allowed at most 40 addition attempts. This is much more than anyone can actually add up.

After you have finished reading these instructions, you will see a link. Click on this link to complete the addition task. Note that the addition task will open a new window in your browser. Once you have completed the task, you will be given a code. You will need this code to complete the study and receive your payment. Please write it down. If you accidentally close the window, you can click on the link again and it will show you the code.

Perform the addition task: Below is a *10-digit number*. Please write it down and then click on the link to perform the addition task. When you click on the link, a *new window* will appear where you will have to enter your 10-digit code. Note that if you enter the wrong code, we won't be able to pay you for your performance in Part 1.

Once you are done with the task, you will receive a password. You will have to come back to this page and enter the password below. This will confirm that you have completed the addition task.

Instructions part 2 screen

Before we instruct you about Part 2, we would like to inform you of the following:

Between 2016 and 2017, a large number of university students from all over the world performed an addition task like the one you have just performed.

There are two differences between your addition task and the one performed by the university students: students faced larger matrices (10x10 instead of 6x6) and were given more time to perform the task (15 min instead of 5 min). These changes were made for you to be able to experience the same task without taking too much of your time. However, despite these changes, the nature of the task remains the same. This means that your experience with the task should give you a sense of what is needed to do well.

Your choices in Part 2: We will present to you three different sets of profiles describing some of the characteristics of students who did this previous task. Each set contains profiles of four different students. For each set, we would like you to *choose one student*.

Your choice gives you money. More precisely, you will receive *$0.15 for each correct addition performed by the student you choose* when he or she did the task.

Because we will give you three sets of profiles to choose a student from, you need to make a choice three times. This means you will earn money three times. Note that once you have made a choice you won't be able to go back and change it.

The profiles we give you will contain background information about the students. Specifically, their age, field of study, gender, type of university degree they purse, and the region of the world they come from. We will also give you an indication of the score obtained by the students when they did the task. However, you will not be told each student's own score. Instead, we have grouped the students in pairs. Below is an example of how a set of four students will be presented. [Here the instructions included an example similar to the ones in Figure E1]

Please continue to make your three choices.

Instructions decision screen

Examine the profiles closely and choose one student. Remember, you will receive $0.15 for each correct addition performed by the student you choose when he or she did the task. The profiles of four different students are below.

Instructions part 3 screen

In Part 3, we ask you to estimate whether female students or male students were better in the previously-described task. More precisely, we calculated the average score of all female students and the average score of all male students who participated in the task across all the regions of the world. We ask you to estimate whether females or males scored better on average by answering the question below. If you estimate correctly, we you will earn an additional $1.50. I estimate that:

- Female students are much better (the average score of female students is 4 more than that of male students)
- Female students are slightly better (the average score of female students is between 1 and 3.99 more than that of male students)
- Male and female students are about the same (the average score of male and female students differs by less than 1)
- Male students are slightly better (the average score of male students is between 1 and 3.99 more than that of female students)

- Male students are much better (the average score of male students is 4 more than that of female students)

E.4 Additional analysis

This subsection contains the additional analysis of Experiment II that could not be included in the main body of the paper due to space constraints.

Individual Treatment

Table E1 shows results from analyzing recruiters' choices in the Individual treatment. Like in the main body of the paper, we use McFadden's random-utility model to explain the choice of whether or not to select one candidate out of four in each set. Columns 1 to 4 contain the results for the Individual treatment (Columns 1 and 2 for male recruiters and Columns 3 and 4 for female recruiters) and Columns 5 to 8 for the Joint treatment for comparison (Columns 5 and 6 for male recruiters and Columns 7 and 8 for female recruiters). The regressions include data form the search and vocabulary tasks to have enough independent observations. The only difference between this specification and that in the paper is that instead of the candidates' joint score, we use an indicator for having a dominated score in a set (i.e., not being the candidate with the highest score in the Individual treatment or not being one of the two candidates in the pair with the highest joint score in the Joint treatment). The estimation results are presented as odds ratios.

In contrast to the Joint treatment, the results show that in the Individual treatment, the gender of the candidate does not have a significant impact on the likelihood of being chosen irrespective of the gender of the recruiter. Moreover, the lower odds ratio for the indicator of the dominated score shows that, compared to the Joint treatment, recruiters focus relatively more on scores when making a decision in the Individual treatment. Finally, one can also see that including the recruiters' beliefs concerning the mean scores of men and women has a smaller effect in the Individual treatment vis-à-vis the Joint treatment, implying that Joint evaluation makes these beliefs a more important part of the decision.

Beliefs

The recruiters were asked to report their belief about the difference in mean scores of men and women in either the search or the vocabulary task. Answers were given in a five categories, which we code as: (-2) women much better (women's mean score is more than

TABLE E1: EXPERIMENT II ODDS RATIOS OF BEING PICKED

Dep. Var.:	Individual treatment				Joint treatment			
Picked by	Male recruiters		Female recruiters		Male recruiters		Female recruiters	
recruiter	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	1.445	1.597	1.045	1.036	0.856	0.857	1.276***	1.105
	(0.465)	(0.537)	(0.474)	(0.457)	(0.082)	(0.081)	(0.109)	(0.104)
Female × Belief		0.504*		0.965		0.717***		0.793***
		(0.190)		(0.238)		(0.053)		(0.060)
Highest score	0.016***	0.014***	0.080***	0.080***	0.256***	0.254***	0.139***	0.138***
	(0.014)	(0.014)	(0.030)	(0.030)	(0.034)	(0.034)	(0.020)	(0.020)
Observations	192	192	240	240	2364	2364	2952	2952
Recruiters	16	16	20	20	197	197	246	246

This table presents results from Experiment II. Columns 1-4 show the results for the Individual treatment and Columns 5-8 for the Joint treatment. Results are shown separately depending on the recruiter's gender: male recruiters in Columns 1-2 and 5-6, and female recruiters in Columns 3-4 and 7-8. All regressions include fixed effects for each set-recruiter combination and controls for other variables in the candidates' resumes. Results are presented as odds ratios. Standard errors clustered on recruiters. (*=p<0.10, **=p<0.05 ,***=p<0.01)

4 points larger), (-1) women slightly better (women's mean score is between 1 and 3.99 points larger), (0) about the same (mean scores differ by less than 1 point), (1) men slightly better (men's mean score is between 1 and 3.99 points larger), and (2) men much better (men's mean score is more than 4 points larger).

Figure E2 shows the distribution of the recruiters' beliefs depending on the task and the gender of the recruiter. The modal belief of male recruiters is that the performance of men and women is about the same in both tasks. Moreover, the remaining answers are more or less evenly distributed among the remaining options, implying that the beliefs of male recruiters are not systematically biased in favor of either male or female candidates. This is confirmed by sign tests evaluating whether the median of the distribution is zero ($p = 0.348$ for the search task and $p = 0.597$ for the vocabulary task). By contrast, the modal answer for female recruiters is that the performance of women is slightly better than that of men, reflecting a slight bias by female recruiters in favor of female candidates (sign tests $p < 0.001$ in both tasks).). Finally, there are no significant differences in the beliefs distributions depending on the task (Fisher's exact tests: $p = 0.283$ for male recruiters and $p = 0.726$ for female recruiters), which confirms that neither task is perceived as more stereotypically male (female) than the other.

FIGURE E2: DISTRIBUTION OF RECRUITERS' BELIEF OF GENDER DIFFERENCES IN PERFORMANCE IN EXPERIMENT II