Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

# 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)

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

## Sociology Results

The sociology sample consists of randomly sampled faculty at the top 20 sociology PhD-granting departments in the U.S.[34] 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.

[34] Ranking from U.S. News Education

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

**TABLE A3: SOCIOLOGY: PAPERS AND TENURE**

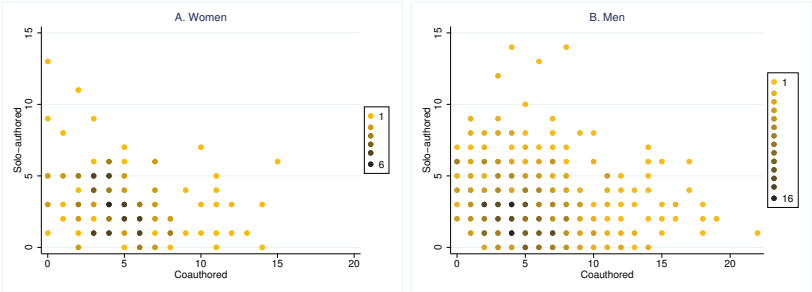
| Dep Var: Tenure | Probit (1) | Probit (2) |
|---|---|---|
| Total first author | 0.050** | |
| | (0.017) | |
| Fem x First Author | 0.026 | |
| | (0.040) | |
| Fraction first author | | 0.403*** |
| | | (0.043) |
| Fem x Frac. First Author | | -0.042 |
| | | (0.172) |
| Solo papers | 0.008 | 0.000 |
| | (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)

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

# Appendix B Additional Figures

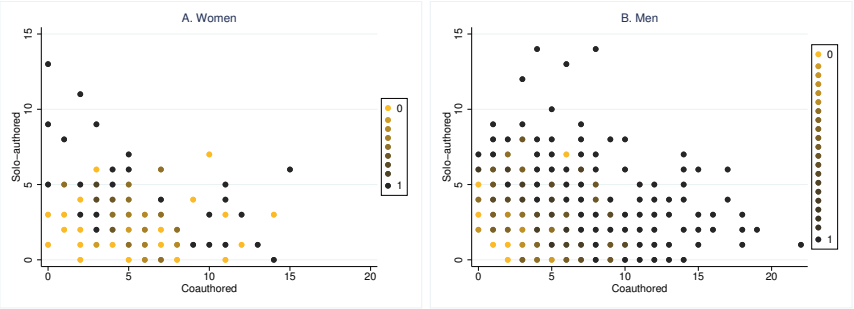
## FIGURE B1: DISTRIBUTION OF PAPER COMBINATIONS




*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 shows the minimum and maximum number of individuals who have a specific paper combination. Panel A is constructed using the full sample of women and Panel B is constructed using the full sample of men.

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

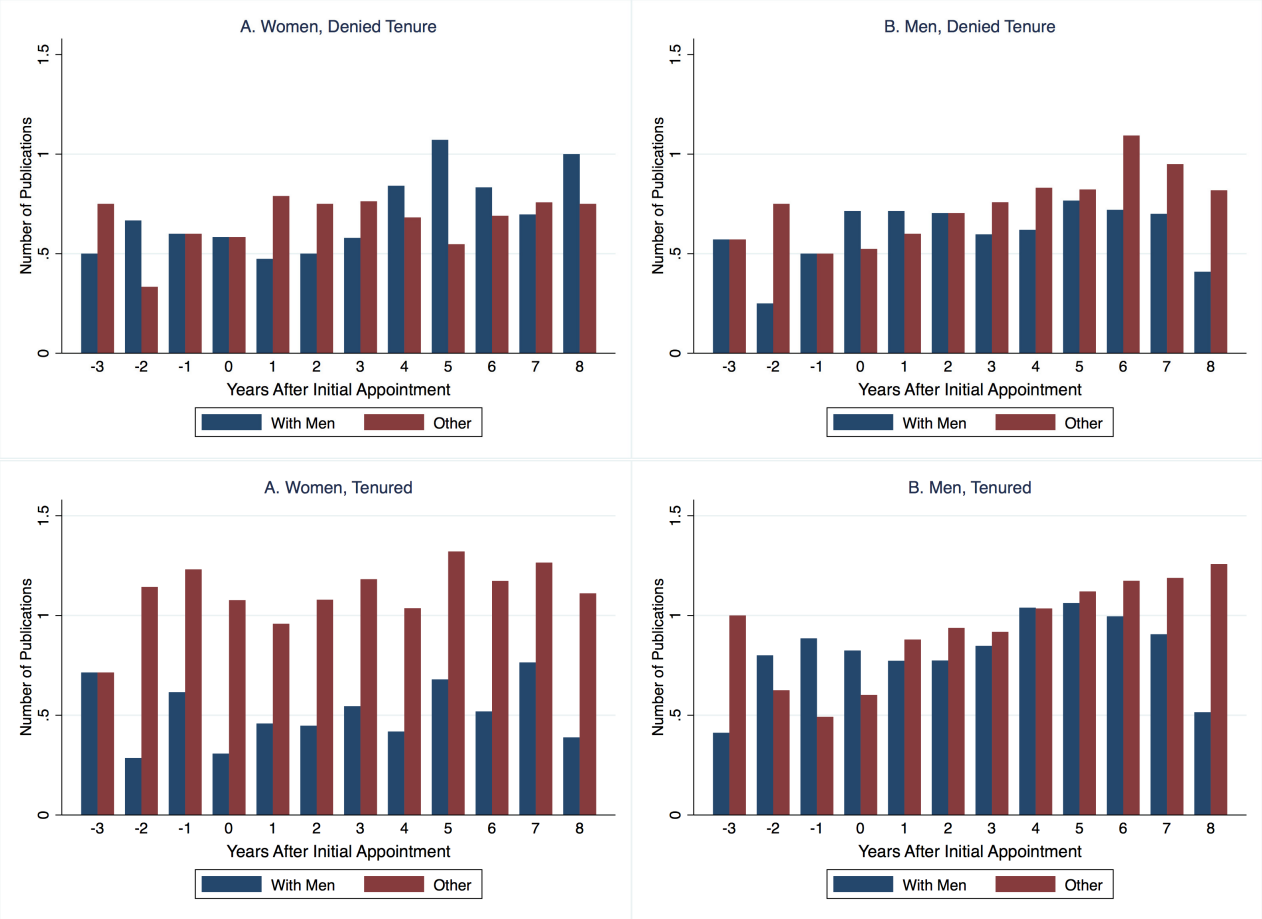
## FIGURE B2: TENURE PROBABILITIES BY PAPER COMBINATIONS




*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).

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

## 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.

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

# 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

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

# 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

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

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.

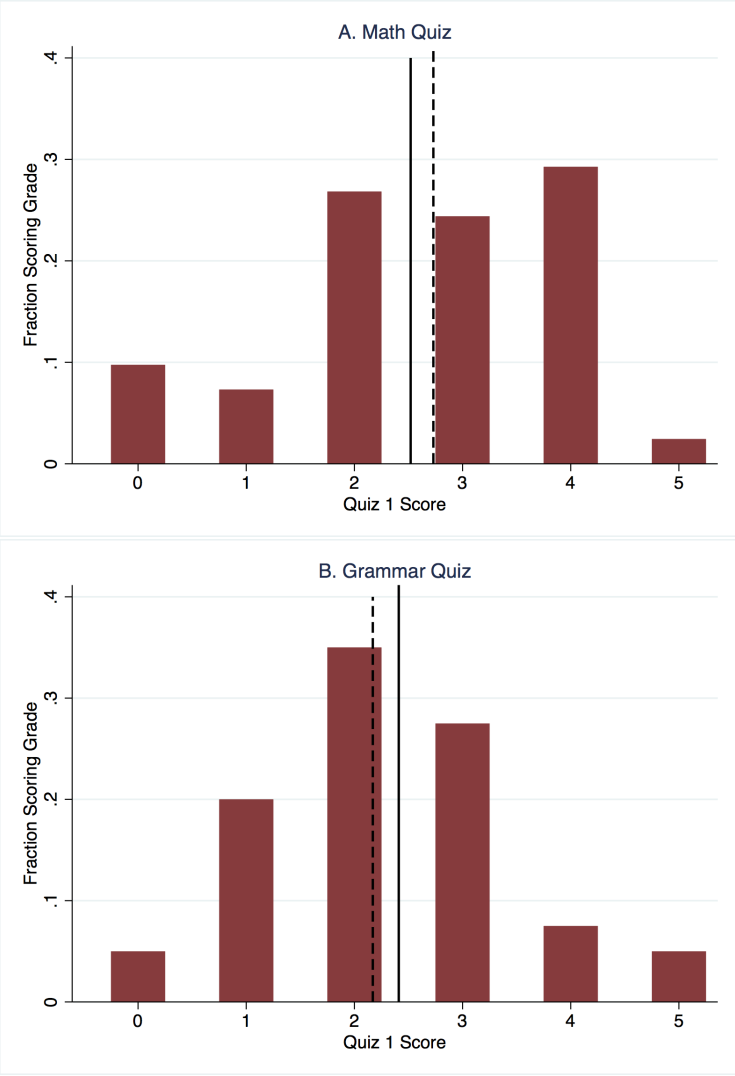
Based on the team's performance, please predict each individual's score on Quiz 2. You can view each quiz by clicking on the link below.

**Histograms**

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

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

## 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.

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

# 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.[35] 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 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).

[35]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.

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

## 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

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

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.[36] 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 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

[36] 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.

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.

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

**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 4 points larger), (-1) women slightly better (women's mean score is between 1 and 3.99

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

**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)

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.

Supplemental Material (not copyedited or formatted) for: Heather Sarsons, Klarita Gerxhani, Ernesto Reuben, Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." Journal of Political Economy 129(1). DOI: https://doi.org/10.1086/711401.

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