Does Scarcity of Female Instructors Create Demand for Diversity among Students? Evidence from Observational and Experimental Data*

Patricia Funk† Nagore Iriberri‡ Giulia Savio§

October 25, 2019

Abstract

This paper combines observational and experimental data to investigate whether scarcity of female instructors affects students' preferences for instructor gender. First, we exploit variation in the share of female professors across different faculties at a Swiss university to explore gender patterns in student evaluations. Our differences-in-differences estimates suggest that female students evaluate female professors more favorably (compared to male students and relative to the gender differences in evaluating male professors) but *only* in faculties with a relatively low share of female professors (Economics and Informatics, but not Communication Science). To shed light on scarcity of female professors as a potential channel for the gender gaps in student preferences, we design an incentivized instructor-choice experiment on MTurk. We experimentally vary the gender balancedness of the instructor pool and let subjects choose one additional instructor among one male and one female. Only female subjects are more likely to choose the female instructor when the pool of instructors is male-dominated, suggesting that female students appreciate a more balanced instructor pool if female professors are scarce.

Keywords: instructor-choice experiment, gender bias, teaching evaluations

*We would like to thank Università della Svizzera italiana for providing anonymized teaching evaluations. We are indebted to Alejandro Martínez Marquina for his invaluable help with the implementation of the MTurk experiment. We also thank Muriel Niederle for her very helpful advice on the experimental design. Further thanks for helpful comments goes to Stefano DellaVigna, Anne Boring, Pedro Rey-Biel and seminar participants at Università della Svizzera italiana, Universitat Pompeu Fabra, University of Zürich, University of Milan and Erasmus University Rotterdam. Finally, we thank Silvia Martinez and Ludovica Marcotti for excellent research assistance. Patricia Funk acknowledges financial support from the Swiss National Science Foundation (grant no. 178887). Nagore Iriberri acknowledges financial support from Ministerio de Economía y Competición (ECO 2015-66027-P), Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (IT1367-19). All remaining errors are our own.

†Università della Svizzera italiana. E-mail: patricia.funk@usi.ch.

‡University of the Basque Country UPV/EHU, IKERBASQUE, Basque Foundation for Science. E-mail: nagore.iriberri@gmail.com.

§Università della Svizzera italiana E-mail: giulia.savio@usi.ch.

1 Introduction

The scarcity of female economists has recently attracted considerable attention (Bayer and Rouse, 2016; Chari, Anusha and Paul Goldsmith-Pinkham, 2017). According to the most recent survey by the American Economic Association, 23.5 percent of tenured and tenure-track faculty in economics are women. As such, gender diversity in academia in economics is as poor as in the male-dominated tech industry, where 30 percent of the Silicon Valley workforce is female. Even worse, among full professors in economics, the share of females is often less than 15 percent (Lundberg and Stearns, 2019).

As forcefully argued by Bayer and Rouse (2016), the low share of females in the economic profession may have negative consequences for research. First, May, McGarvey and Whaples (2014) have shown that female and male economists hold different views on economic policies (see also the article "Women in Economics" in the Economist, 2018). As such, research conducted by mostly male economists may not be representative of a more gender-balanced researcher pool and may miss topics relevant for society as a whole. Second, experimental studies have shown that diverse teams tend to be more productive (Apesteguia, Azmat and Iriberri, 2012; Hoogendoorn, Oosterbeek, and van Praag, 2013). Having a more gender-unequal faculty may therefore negatively affect academic output.

In contrast to the literature that notes the potential negative effects of male-dominated faculties on economic research, little is known about potential negative effects on students. However, as a lack of diversity affects the type of research topics studied and taught to the students, this factor may directly channel into female and male students' interest in economics. Moreover, teaching styles may vary with instructor gender and affect student satisfaction of either, or both sexes. In sum, if students value diversity in the instructor pool, a low share of female instructors may make them more valuable in the students' eyes. This taste for females could be driven by all students (general taste for diversity), or among certain subgroups of students in particular. Concerning the latter channel, research in social psychology suggests that an individual's distinctive trait in relation to other people in the environment is more salient if this trait is a numerical minority ("numerical distinctiveness theory", see McGuire and Padawer-Singer, 1976; McGuire and McGuire, 1981). As such, when female professors are scarce, gender may become particularly salient to female students, which may affect their preferences for female (as opposed to male) instructors.

To obtain empirical evidence on this last conjecture (gender-specific taste for diversity), we analyze around 27,000 teaching evaluations from three faculties (Communication, Economics, and Computer Science[1]) of the Università della Svizzera italiana. In all three faculties, women are under-represented in the instructor pool, but the degree of under-representation varies considerably across the faculties. While in Computer Science, only 17% of instructors are female, this share goes up to 23% in Economics, and even up to 33% in Communication Science. As such, the share of female instructors in Communication Science is double the size of Computer Science.

If scarcity makes female professors particularly valuable to female students, we may expect this fact to be reflected in course evaluations. Indeed, we find that female professors are evaluated relatively more positively by female students but only in male-dominated faculties.[2] This relative female preference for female professors in male-dominated faculties cannot be explained by cross-faculty differences in student response rates to the evaluation survey. In fact, response rates are well above 90 percent for all three faculties, as filling out the evaluations is necessary to have access to the course grades. This absence of survey response bias makes our setting unique. Moreover, student selection into the different faculties is unlikely to account for these differences, as surveys among the students reveal similar gender gaps on gender attitudes across the three faculties. This leaves us with a few plausible explanations for the observed correlations: First, our proposed scarcity channel (where the effect of professor scarcity may be amplified by student scarcity). Second, the fact that female professors may teach courses that are more appealing to female students; or they are simply able to motivate female students particularly in male-dominated environments. Third, the

[1]Computer Science is called Informatics at this university.

[2]Specifically, we run regressions with course fixed effects and identify differences-in-differences in the evaluation of female versus male students for courses taught by female professors relative to courses taught by male professors.

selection process of female professors may be different in male-dominated faculties, and forth, the quantitativeness of the subject may matter. In male-dominated faculties, subjects are more technical, and female professors may be particularly valued by female students, if for example, they explain in a more intuitive way.

Due to the difficulty of determining the precise mechanism with observational data, we directly test for the presence of (a potentially gender-specific) taste for instructor diversity in an experimental setting. With this aim, we design a deception-free, incentivized instructor-choice experiment on MTurk, where subjects have to select an instructor who will give them advice on how to solve a given task. The choice set consists of two instructors with comparable qualifications and experience but different gender. Before getting to this choice (which is our key outcome of interest), subjects are told that there is a pool of six instructors, all of which give advice. To test whether scarcity of females affects the choice of the additional instructor (male or female), we experimentally vary the "stock" of six instructors. In the balanced treatment, the subject is presented a stock of three female and three male instructors, whereas in the unbalanced treatment, the subject has a stock of six male instructors.

The main interest of the experiment is analyzing whether the choice of the additional instructor (male or female) depends on the gender balance of the existing instructor pool. To rule out that the order of presenting the two instructors (or details of their profile) affects the subjects' choice, we randomize subjects into permutations that vary according to task type, the order of presenting the two candidates, and the values of the two characteristics attached to the candidates (all details are provided in Section 3). To ensure that subjects take the experiment seriously, we use a variable remuneration that increases with the correct answers in the tasks (next to a fixed show-up fee).

Our main findings are the following. First, scarcity in the stock of female instructors positively affects the probability of having a female chosen as the additional instructor. On average, the female instructor is 11 percentage points more likely to be selected if the stock of instructors is gender-unbalanced. Second, female and male subjects react differently to scarcity in the instructor pool. If female instructors are scarce, female subjects are 12.3 percentage points more likely than male subjects to select the female instructor (this difference is highly statistically significant).[3] Moreover, in contrast to female subjects, male subjects do not react to scarcity of female instructors in a statistically significant way.

While the experimental setting mimics the case of underrepresentation of females (as present in STEM and economics), two plausible mechanisms can explain the results. First, female subjects prefer female instructors when female instructors are scarce. Second, female subjects have a general preference for diversity (independent of whether scarcity refers to their own gender). To investigate the plausibility of the second channel, we compare instructor-choices of the gender-balanced treatment with a new unbalanced treatment, where all instructors are female. We find that females also value diversity when male instructors are scarce but to a lesser extent than when female instructors are scarce. By contrast, men value diversity *only* if the scarcity is related to their own gender.

Our study contributes to three main strands of the literature. The first is on gender and hiring decisions. As studies relying on observational data are problematic because of unobserved quality differences between candidates, the most convincing studies exploit experimental variation. An early study by Steinpreis, Anders and Ritzke (1999) studied a hypothetical hiring decision among psychology faculty, where the gender of the candidate was experimentally varied. The main finding was that both male and female faculty were less favorable towards the female candidate. More recently, Williams and Ceci (2015) conducted a similar hypothetical hiring experiment among faculty in biology, engineering, economics, and psychology. Surprisingly, the results show a consistent preference for females, with the exception of male economists, who were found to be gender-neutral. Our main contribution to these papers is to rigorously test for scarcity effects in a setting where the hiring decision is incentivized. Our results support the view that, especially in settings where women are scarce, female candidates have an edge.

[3] When decomposing the tasks according to task type (mathematics versus English task), we find larger effects for mathematics, but the differences are not statistically different from the English task.

Second, we find that on average, female subjects prefer female candidates, especially when female candidates are scarce. This result complements previous research on ingroup favoritism and outgroup bias (see Tajfel et al., 1971; Chen and Li, 2009; Chen and Chen, 2011; and Chen et al., 2014; and Coffman, Exley and Niederle, 2018). We add to this strand of literature a connection between the strength of this ingroup preference and the scarcity of the ingroup. As becomes apparent from our study, ingroup preferences may become amplified when the ingroup gets relatively smaller.

Third, our paper relates to literature that documents gender differences in student evaluations. While female students in Economics appear to be more critical than males when evaluating male professors, the same does not hold when evaluating female professors (Boring, A., 2017; Mengel, Sauermann, and Zölitz, 2019). Our study complements this literature by providing across-discipline evidence. While we replicate previous results with our sample, we also document that gender differences in instructor evaluations are completely absent in more gender-balanced faculties (Communication), whereas they are aggravated in even more unbalanced faculties (Computer Science).

In summary, our two pieces of evidence (experimental and observational) indicate that gender-related preferences emerge differently in different contexts. When females are scarce, they become more valuable, particularly among the subgroup of female decision makers. As such, increasing the share of females in male-dominated faculties (e.g., STEM disciplines) may increase student satisfaction and act as a pull-factor for future female students.

The remainder of this article is structured as follows. Section 2 presents the results based on observational data. Section 3 shows the evidence from experimental data, and Section 4 concludes.

2 Observational Evidence from Teaching Evaluations

2.1 Setting and Data

We obtained student evaluation data from the Università della Svizzera italiana (USI) for three different faculties (Communication, Economics, and Computer Science). As shown in Figure 1 and the summary statistics (Appendix Table A.1), a stark variation exists in the presence of female professors across faculties: the share of female instructors is lowest in Computer Science (17 percent), followed by Economics (23 percent) and Communication (33 percent). We collected teaching evaluations for all courses taught by the three faculties for the consecutive academic years of 2015-2016 and 2016-2017.[4]

The academic year is organized into two semesters, where students take approximately 7 classes per semester. Teaching evaluations are done online after the students have taken the courses and completed the exams but before they know their actual grade. As filling out the teaching evaluations is necessary to access the grades, the response rate is close to 100% (see the variable "Dummy students not reporting Teaching Evaluations Score" in Appendix Table A.1). The teaching evaluation questionnaire consists of 10 questions. We focus on the question that represents the summary evaluation of the course: "Please express your overall satisfaction with this course".

Our database contains 26,996 teaching evaluations from 1,910 different students for 847 different courses taught by 318 different instructors. As shown in Appendix Table 1, approximately one-half of the students are female, although the gender composition varies significantly across disciplines. The average student is 24 years old, between 50% and 66% are doing their Bachelor degree, while the rest are at the Masters level, and Italian and Swiss nationalities are roughly equally represented. Regarding the courses, more than one-half of the courses are compulsory. The proportion of quantitative

[4]In the academic year 2017-2018, a new evaluation system was introduced, so the newer data were no longer comparable.

courses varies substantially across disciplines: 90% are taught in Computer Science, 51% in Economics, and only 14% in Communication. The average class size is also smaller in Computer Science than in the other faculties, as is the overall number of students enrolled. Regarding the instructors, the large majority of instructors are lecturers, followed by full professors. We measure their research productivity through citations (received from the database "Publish or Perish"). Finally, with respect to course-student characteristics, only a minority of students is repeating a course and, as mentioned previously, a very small minority does not complete the teaching evaluation (5%). Students earn an average grade of 7.5 out of 10, where the highest grades are in Communication, followed by Economics and Computer Science. Overall, the students show a high satisfaction level, with an average of 7.2 out of 10, which does not differ across disciplines. In conclusion, the most significant differences across disciplines are observed in the gender composition of both, the faculty as well as the student body, and in the proportion of quantitative courses.

2.2 Regression Equation and Results

We test for gender gaps in evaluations with the following regression, which we estimate separately for each faculty:

$$TE_{spc} = \alpha + \beta * F_s + \gamma * F_p + \delta * F_s * F_p + T_s + X_{sc} + J_p + \rho_c + \epsilon_{scp} \qquad (1)$$

The dependent variable TE_{spc} is the teaching evaluation score (ranging from 1 to 10) given by student s to professor p teaching course c. F_s is a dummy variable, taking a value of one for female students, F_p is a dummy variable taking a value of one for female professors, and $F_s * F_p$ refers to the interaction between F_s and F_p. The terms T_s, J_p and X_{sc} denote student, professor and student-course covariates (see Appendix Table A.1 for an overview of the control variables). δ is the main coefficient of interest, as it measures the differences-in-differences in the evaluation of female versus male students for courses taught by female professors relative to courses taught by male professors. In other words, $\delta > 0$ would suggest that females have a relative preference for female professors (which we loosely call "same-sex preferences"). Moreover, if scarcity of female instructors is the main driver of same-sex preferences, we would expect δ to be higher for more male-dominated faculties (Economics and Computer Science).

The results are presented in Table 1. Note that we always include course fixed effects, meaning that we compare evaluations for the *same course*. Baseline estimates with course-year fixed effects are presented in columns 1, 4 and 7. In columns 2, 5, and 8, we add student fixed effects. Last, we add professor-course fixed effects that vary by year to account for the fact that some courses are co-taught (see columns 3, 6, and 9).

First, we note that male students tend to give lower evaluations to female than to male professors (estimated γ), with the largest differences observed in Computer Science. Female students, by contrast, show no differences in how they evaluate male and female professors on average ($\gamma+\delta$). The differences-in-differences (δ) in how female students (relative to males) evaluate female professors (compared to male professors) depend on the faculty. We find evidence of same-gender preferences in Economics and Computer Science but not in Communication. Note that although there are fewer students in Computer Science, the estimated same-gender preference (coefficient δ) is almost doubled compared to that observed in Economics. Estimates of δ suggest that female professors receive 0.2 points more when they are evaluated by female students (compared to when they are evaluated by male students, and relative to the gender gap in evaluating male professors) in Economics and receive almost 0.4 points more in Computer Science. The results are quite stable across specifications. The fact that female students give relatively more generous evaluations to female professors when the faculty is male-dominated may support the view that scarcity matters for student evaluations. While scarcity of female professors seems a very plausible mechanism, scarcity of female students may also play a role (possibly amplifying the effect of scarcity on the professor side).

One caveat could be that students self-select into different disciplines and might exhibit different attitudes towards gender. This characteristic would be troublesome for us if gender *gaps* in attitudes differ across faculties. To shed light on this issue, we designed and administered an 11 question survey for first-year undergraduate students. We interviewed students from three classes during the first semester, one from each faculty. Although the questionnaire was advertised as a survey on students' labor market attitudes and aspirations, five questions (out of eleven) were related to gender stereotypes. In these five questions, students were asked whether they agree or disagree with the following statements: "When jobs are scarce men should have more right to a job than women" (question "Jobs Scarce"), "Having a job is the best way to gain independence for a women" (question "Having a Job"), "When a mother works for pay, children suffer" (question "Mother Works"), "A university education is more important for a boy than for a girl" (question "University Education"), and finally "On the whole, men make better business executives than women" (question "Men better executives"). As shown in Appendix Table A.2, few gender gaps exist within the faculties. Only for the question "Men are better business executives than women" do we observe that male (compared to female) students enrolled in Economics are significantly more likely to agree with this statement. In sum, we find little support for differences in gender-equality values driving same-sex preferences across disciplines.

To strengthen our suggestive evidence that instructor-scarcity shapes preferences for instructor-gender, we conduct an instructor-choice experiment on M-Turk. The experiment has the advantage that we can isolate the effect of variation in instructor-gender on subjects' hiring preferences (female versus male instructors), and keep constant other factors that could have confounded the results from observational data (e.g. student gender composition).

3 Field Experiment on MTurk

3.1 Design and Data

We design an incentivized and deception-free instructor-choice experiment on MTurk. Subjects choose an instructor to give them advice on how to solve a given task under time pressure. Subjects are randomized into two types of tasks: mathematical multiplications ("math task") or spelling certain English words correctly ("English task"). At the beginning of the experiment, students are given information on the payoff structure (payoff depends on task performance) and the type of task (math or English, without giving any further details). All subjects are informed that they will receive 1 dollar for their participation plus 40 cents for each correct answer. Most importantly for this experiment, subjects are told that six instructors (selected by us) will give them tips on how to solve the tasks and that they can choose *one additional instructor*. In the end, they will have to choose among one male and one female instructor with comparable qualifications and experience. The key feature of this experiment is that we experimentally vary the "stock" of six instructors. In the balanced treatment, the student has a stock of three female and three male instructors, whereas in the unbalanced treatment, the student has a stock of six male instructors. The only information given to the students is the instructor's name and the fact that he/she is a graduate student (see Figure 2 for a screenshot).[5] Concerning the additional instructor to choose, the student gets information on the name (Margaret or Richard), the GPA (3.5 or 3.6 out of 4), and the accumulated hours as a teaching assistant (29 or 31). The main interest of the experiment is analyzing whether the choice of Margaret (as opposed to Richard) as an additional instructor depends on the treatment.

We design 16 permutations, 8 for the math task and 8 for the English task. For each task type, 4 permutations have a balanced instructor pool and 4 permutations have

[5] In the balanced treatment, the subjects are told that they have six instructors "Jim", "Mary", "John", "Patricia", "Robert" and "Linda", all graduate students. In the unbalanced treatment, the subjects are told that they have six instructors "Jim", "Kevin", "John", "William", "Robert" and "David", all graduate students. The actual tips are obtained from real graduate students who were shown the task and were asked to describe the task in written form.

an unbalanced instructor pool. These 4 permutations differ in the order of instructor presentation (Margaret first or second) and characteristics (Margaret with a higher GPA but fewer accumulated hours as TA or Margaret with a lower GPA but more accumulated hours as TA). The goal was to obtain roughly 100 subjects for each permutation, for a total of 1,600 subjects. We managed to collect 1,955 observations. However, we removed all subjects who tried to run the experiment twice and those who appeared to be doing the experiment together with a second person.[6] This left us with a subject pool of 1,478: summary statistics are reported in Appendix Table A3. As shown in Panel A, randomization of CV-characteristics (GPA and hours of experience as TA) across the two candidates' profiles worked well, as the likelihood that Margaret comes first or that Margaret has a higher GPA is always approximately 50%. Most importantly, as evident from Panel B, all demographic covariates are balanced across treatments. In contrast to the teaching evaluation setting, the share of male and female subjects is stable across balanced and unbalanced instructor pools. In Panel C, we report subjects' behavior during the experiment. As expected, the main endogenous variable of interest (instructor choice) differs across treatments. Margaret is chosen more frequently when the treatment is "Unbalanced" (when female instructors are scarce). Regarding the duration of the task, the number of times instructor advice was sought, or performance in terms of correct answers, we do not see any differences across treatments.

While the summary statistics indicate that the subjects receive advice slightly more than 4 times on average, it is also interesting to look at *which* type of advice the subjects seek. In Appendix Figure A1, we document the percentage of subjects who click on a specific advice, starting with advice from the instructor to the farthest left of the instructor pool (Tip 1, referring to the advice from Jim), followed by advice from the second-leftmost instructor (Tip 2, referring to advice from Kevin in the unbalanced treatment and Mary in the balanced treatment), etc. The advice number 7 (Tip 7) is advice from the instructor chosen by the subject (Margaret or Richard). As shown in

[6] That is, two subjects running the experiment with the same IP address.

Appendix Figure 1, a spike is observed for Tip 7 (for both male and female subjects), meaning that advice is most frequently sought from the subject-selected instructor.

Of 1,478 subjects, only 267 did not look at the advice of their chosen instructor. We present the main results for the 1,009 subjects who actually looked at the advice of their selected instructor. Arguably, these subjects took the instructor-choice decision most seriously, as they did (and likely planned to) look at the instructor's advice. However, we will also document the robustness to alternative data samples.

3.2 Regression Equations and Results

We run two regression equations:

$$Margaret_i = \alpha + \beta * Unbalanced_i + \eta * Math_i + \theta * MargFirst_i + \psi * MargTA_i + \iota * X_i + \epsilon_i \quad (2)$$

$$\begin{aligned} Margaret_i = \alpha + \beta * Unbalanced_i + \gamma * Female_i * Unbalanced_i + \eta * Math_i \\ + \theta * MargFirst_i + \psi * MargTA_i + \iota * X_i + \epsilon_i \quad (3) \end{aligned}$$

The dependent variable $Margaret_i$ is a dummy equal to one if subject i chooses the female candidate (Margaret) over the male candidate (Richard). $Female_i$ is a dummy equal to one if the subject is female. $Unbalanced_i$ is a dummy equal to one if subject i is exposed to a pool of six male instructors. The variables $Math_i$, $MargFirst_i$ and $MargTA_i$ control for the experimental permutation: $Math_i$ is a dummy equal to one for the math task; $MargFirst_i$ is a dummy equal to one if - in the instructor choice step - the name of the female candidate (Margaret) comes before the name of the male candidate (Richard); and $MargTA_i$ is a dummy variable taking a value of one if the female candidate (Margaret) is more experienced as a teaching assistant than is the male candidate (Richard). Finally, X_i is a vector of all individual covariates listed in Appendix Table A.3.

The main coefficient of interest in equation 2 is β. A positive β suggests that female instructors are in more demand when scarce. Equation 3 adds an interaction term $Female \cdot Balanced$, which enables testing for whether the potential effect of (female) scarcity on demand for females is gender-specific.

Our main experimental results are shown in Table 2. Columns 1-2 show the results of regression equation 2. Clearly, being exposed to a pool of male instructors increases the demand for the female instructor. The probability of choosing Margaret increases by 11 percentage points if a subject is exposed to the gender-unbalanced instructor pool. Adding controls (column 2) hardly affects the estimated coefficient of the treatment "Unbalanced", as would be the case in successful randomization. Column 3 displays the results for regression equation 2, indicating that the stronger preference for female instructors in the treatment "Unbalanced" is *entirely* driven by female subjects. Males are *not* more likely to choose Margaret if the teacher pool is unbalanced (see the estimated β).[7]

Clearly, the likely mechanism behind these results is that females (but not males) value the diversity brought in by a female instructor if the pool of instructors is all male. Female students may expect a different type of advice from female instructors that would help them to answer the questions correctly and earn more money. In this instrumental view, females select the female instructor because they would like to receive the advice. Alternatively, the decision could be entirely subconscious, where females select the female instructor (when the instructor pool is all male), even though they have no expectations in terms of the advice they would receive from the female instructor. To test for this possibility, we conduct a placebo analysis. We restrict the sample to those subjects who did not check any single advice (remember - before seeing the task!). These are likely subjects whose strategy is to earn the participation fee but have no intention of exerting any additional effort to answer the questions correctly. As shown in Table 3 columns 1 and 2, there is no effect of the treatment "Unbalanced" (nor the interaction of "Unbalanced" with subject gender) on the probability that Margaret is chosen. As such, in the sample of subjects who exert very little effort in

[7]Note that we also estimated models with triple interaction terms to see whether effects differ between task type (English or math). Since the estimated coefficient before the triple interaction $Unbalanced \cdot Female \cdot Math$ is statistically insignificant, we report results for the two tasks combined.

the experiment, scarcity of females in the instructor pool does not affect the probability that the female instructor is chosen.

Most interesting to us are the subjects who exert at least some minimal effort to correctly answer the questions. We presented the results for subjects who looked at the advice of the chosen instructor in Table 2. As additional evidence, we now report the results for different subject samples, varying in the number of advice seen. As shown in Table 3, columns 3-8, gender-specific preference for diversity get larger in samples where subjects ask for more advice and, arguably, take the task more seriously. Last, we focus on those subjects who could answer the last survey question: "How many women were in the initial instructor pool of six instructors?" Again, the effect is strong (in fact the strongest) among those subjects who appeared to pay close attention to the experiment.

As such, we provided evidence that female subjects value female instructors when female instructors are scarce. While we were most interested in the setting lacking female diversity, we were curious whether females would also value diversity in a setting where males are scarce. We therefore ran an additional treatment, where the scarce group is male (i.e., the unbalanced treatment is all female). As shown in Appendix Table A.5, females also value diversity in this scenario but to a smaller extent (compare the estimated $(\beta + \gamma)$ with those in Table 2). By contrast, males value diversity *only* when male instructors are underrepresented.[8]

4 Conclusions

Female underrepresentation in science (especially STEM faculties) is a topic of heated debate. While numerous articles explore potential causes (e.g., stereotypes (Reuben, 2014), family and career incompatibilities (Goldin, 2014), or publishing hurdles (Card et al., 2019)), little is known about the consequences of a lack of academic diversity. Since diverse teams are often found to be more productive, a reasonable conjecture is that academic diversity would favor research output (Bayer and Rouse, 2016). What

[8]The estimated β is large and highly significant in Appendix Table A.5 but not in Table 2.

about teaching quality? We first document that the *perceived* teaching quality of the same course may differ across student gender. Importantly, gender differences in preferences are amplified when faculties are male-dominated. While gender gaps in Communication are negligible, female students are (relatively) more satisfied with female professors (vis-à -vis male professors) in male-dominated faculties such as Economics and Computer Science. Does that mean that if female professors are scarce, female students would prefer to hire more female professors? Our experimental evidence suggests that this might be the case. Female (but not male) subjects show a clear preference for female instructors if females are scarce in the instructor pool. Luckily for the few existing female students in STEM faculties, hiring preferences seem to become more female friendly as long as female candidates are equal to or better than male candidates (Williams and Ceci, 2015a, 2015b).

References

[1] Apesteguia, J., Azmat, G., Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, 58(1), 78-93.

[2] Bayer, A., Rouse, C. E. (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives*, 30(4), 221-42.

[3] Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27-41.

[4] Card, D., DellaVigna S., Funk P., and Iriberri N. (2019). Are Referees and Editors in Economics Gender Neutral? *Quarterly Journal of Economics*, accepted.

[5] Chari, A., and Goldsmith-Pinkham, P. (2017). Gender representation in economics across topics and time: Evidence from the NBER summer institute (No. w23953). *National Bureau of Economic Research.*

[6] Chen, R., and Chen, Y. (2011). The potential of social identity for equilibrium selection. *American Economic Review*, 101(6), 2562-89.

[7] Chen, Y., and Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, 99(1), 431-57.

[8] Chen, Y., Li, S. X., Liu, T. X., and Shih, M. (2014). Which hat to wear? Impact of natural identities on coordination and cooperation. *Games and Economic Behavior*, 84, 58-86.

[9] Coffman, K. B., Exley, C. L., and Niederle, M. (2017). When gender discrimination is not about gender (No. 18-054). *Harvard Business School Working Paper.*

[10] "Women in Economics" (2017). *The Economist.*

[11] Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4), 1091-1119.

[12] Hoogendoorn, S., Oosterbeek, H., and Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science*, 59(7), 1514-1528.

[13] Lundberg, S. and Stearns, J. (2019). Women in Economics: Stalled Progress. *Journal of Economic Perspectives* 33(1), 3-22.

[14] May, A. M., McGarvey, M. G., and Whaples, R. (2014). Are disagreements among male and female economists marginal at best?: A survey of AEA members and their views on economics and economic policy. *Contemporary Economic Policy*, 32(1), 111-132.

[15] McGuire, W. J., and Padawer-Singer, A. (1976). Trait salience in the spontaneous self-concept. *Journal of personality and social psychology*, 33(6), 743.

[16] McGuire, W. J., and McGuire, C. V. (1981). The spontaneous self-concept as affected by personal distinctiveness. *Self-concept: Advances in theory and research*, 147-171.

[17] Mengel, F., Sauermann, J., and Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2), 535-566.

[18] Reuben, E., Sapienza P., and Zingales L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences* 111: 44034408.

[19] Steinpreis, R. E., Anders, K. A., and Ritzke, D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex roles*, 41(7-8), 509-528.

[20] Tajfel, H., Billig, M. G., Bundy, R. P., and Flament, C. (1971). Social categorization and intergroup behaviour. *European journal of social psychology*, 1(2), 149-178.

[21] Williams, W. M., and Ceci, S. J. (2015a). National hiring experiments reveal 2: 1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences*, 112 (17) 5360-5365.

[22] Williams, W. M., and Ceci, S. J. (2015b). Women have substantial advantage in STEM faculty hiring, except when competing against more-accomplished men *Frontiers in Psychology*, 6: 1532.

Figures and Tables

Figure 1: Gender Ratio of Professors by Discipline

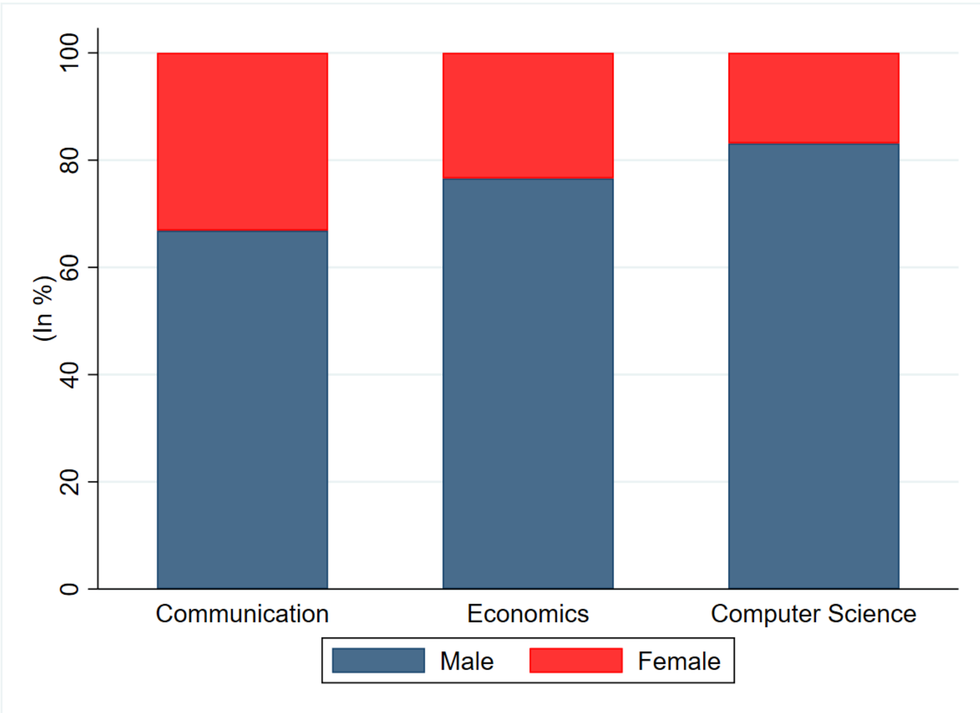
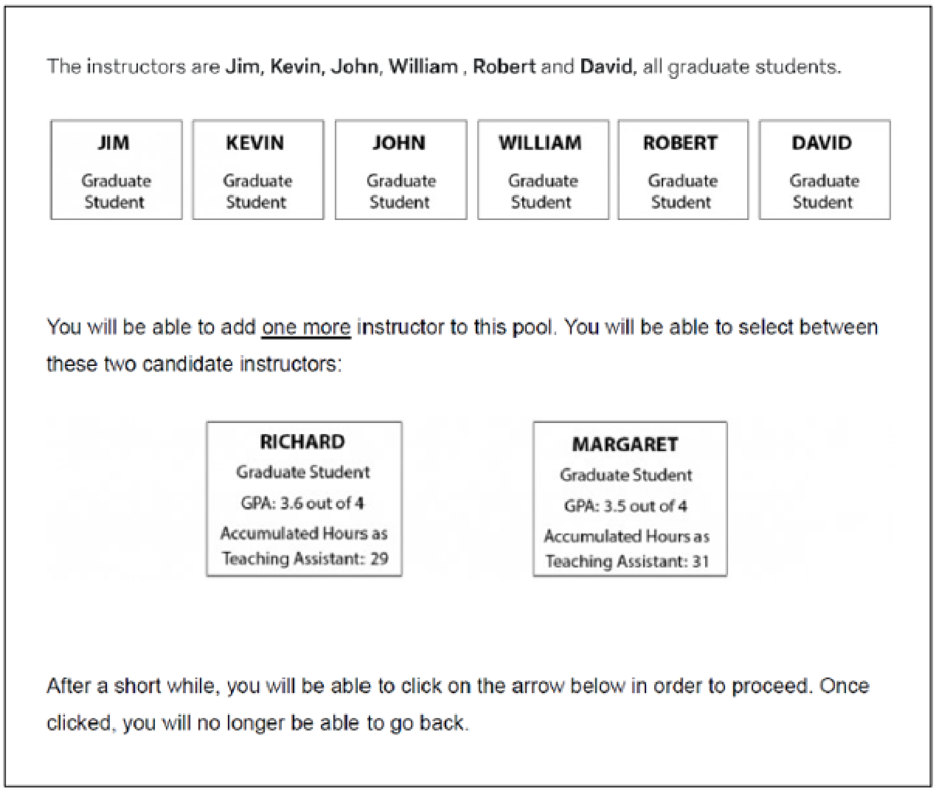


Figure 2: Experimental Treatments

Panel A: Treatment Unbalanced



Panel B: Treatment Balanced

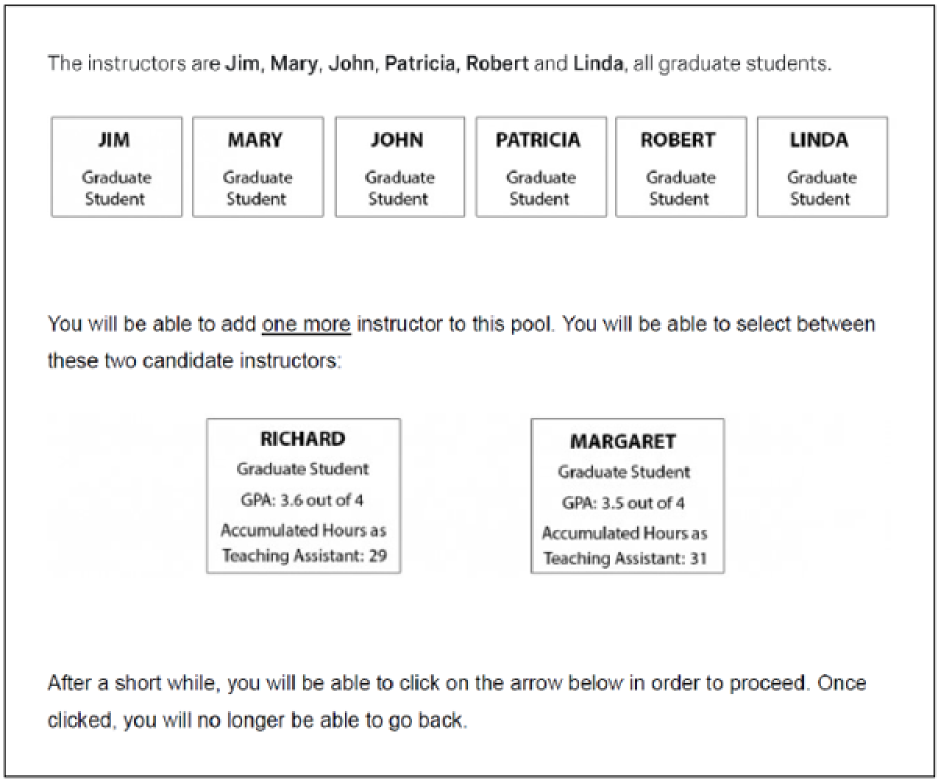


Table 1: Gender Gaps in Teaching Evaluations, by Discipline

Disciplines	Communication			Economics			Computer Science		
Column	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female_s (β)	-0.046			-0.266***			0.234*		
	(0.055)			(0.048)			(0.132)		
Female_p (γ)	-0.180	-0.251*		-0.075	-0.065		-0.326***	-0.343***	
	(0.152)	(0.151)		(0.124)	(0.129)		(0.025)	(0.038)	
$\text{Female}_S \times \text{Female}_P$ (δ)	0.019	0.110	0.106	0.270**	0.225**	0.232***	0.414	0.444*	0.442*
	(0.089)	(0.082)	(0.073)	(0.111)	(0.096)	(0.085)	(0.257)	(0.250)	(0.256)
Constant	5.424***	5.346***	5.535***	7.081***	7.407***	7.564***	1.583	5.992***	6.527***
	(0.380)	(0.366)	(0.369)	(0.419)	(0.375)	(0.419)	(1.801)	(0.540)	(0.479)
$\gamma + \delta$	-0.161	-0.141		0.195	0.16		0.088	0.10	
	(0.163)	(0.165)		(0.127)	(1.333)		(0.236)	(0.235)	
Course-Year FE	YES	YES	NO	YES	YES	NO	YES	YES	NO
Student FE	NO	YES	YES	NO	YES	YES	NO	YES	YES
Professor-Course-Year FE	NO	NO	YES	NO	NO	YES	NO	NO	YES
Student-Course Control	YES	YES	YES	YES	YES	YES	YES	YES	YES
Student Control	YES	NO	NO	YES	NO	NO	YES	NO	NO
R-squared	0.221	0.472	0.480	0.187	0.490	0.494	0.381	0.568	0.569
N	10799	10820	10891	11250	11266	11266	2402	2410	2410

Notes. The dependent variable is the teaching evaluation score received by instructor i for course j. Evaluations in courses with less than six students are excluded from the analysis. Columns 1,4,7 include Course-Year fixed effects, Columns 2,5,8 include Course-Year fixed effects and Student fixed effects, and Columns 3,6,9 include Professor-Course-Year fixed effects and Student fixed effects. Standard errors, clustered at course-year level, are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Choice of Female Instructor when Female Instructors are Scarce

	(1)	(2)	(3)
Unbalanced (β)	0.116***	0.111***	0.049
	(0.029)	(0.029)	(0.041)
Female X Unbalanced (γ)			0.125**
			(0.053)
Math Task	0.006	0.009	0.006
	(0.025)	(0.024)	(0.024)
Margaret First	-0.055**	-0.059**	-0.059**
	(0.026)	(0.026)	(0.025)
Margaret TA	-0.022	-0.023	-0.021
	(0.038)	(0.039)	(0.038)
Female		0.050**	-0.014
		(0.023)	(0.036)
Age		0.004***	0.004***
		(0.001)	(0.001)
White		0.009	0.010
		(0.042)	(0.042)
College Degree		0.021	0.016
		(0.047)	(0.046)
Post-graduate Degree		0.013	0.005
		(0.051)	(0.050)
Constant	0.670***	0.494***	0.531***
	(0.038)	(0.069)	(0.067)
$\beta + \gamma$			0.174***
			(0.035)
R-squared	0.019	0.032	0.036
N	1009	1009	1009

Notes. The dependent variable is a dummy equal to one if Margaret is chosen. Treatment "Unbalanced" is a dummy equal to one if the subject is exposed to a pool of six male instructors, and zero if he/she is exposed to a gender balanced pool of instructors. All included subjects checked the advice by the chosen instructor. Robust standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Choice of Female Instructor when Female Instructors are Scarce: Robustness Checks

	Zero advices		One advice		Two advices		Three advices		Guessed right	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Unbalanced	-0.039	0.040	0.081***	0.057	0.098***	0.058	0.090***	0.028	0.110***	0.039
	(0.082)	(0.096)	(0.023)	(0.039)	(0.024)	(0.041)	(0.024)	(0.041)	(0.038)	(0.055)
Female X Unbalanced		-0.174		0.050		0.085		0.127**		0.128*
		(0.146)		(0.059)		(0.060)		(0.060)		(0.075)
Math Treatment	-0.048	-0.049	0.016	0.015	0.010	0.007	0.005	0.002	0.010	0.007
	(0.083)	(0.083)	(0.022)	(0.022)	(0.025)	(0.025)	(0.024)	(0.024)	(0.032)	(0.032)
Margaret First	0.062	0.061	-0.050**	-0.050**	-0.049*	-0.050*	-0.058**	-0.058**	-0.023	-0.026
	(0.074)	(0.074)	(0.022)	(0.022)	(0.026)	(0.025)	(0.028)	(0.028)	(0.037)	(0.037)
Margaret TA	0.002	0.016	-0.056*	-0.056*	-0.063*	-0.061*	-0.057*	-0.055	-0.083**	-0.079*
	(0.076)	(0.076)	(0.033)	(0.032)	(0.032)	(0.032)	(0.033)	(0.033)	(0.041)	(0.041)
Female	0.065	0.160**	0.067***	0.042	0.057**	0.016	0.057**	-0.005	0.025	-0.029
	(0.079)	(0.071)	(0.023)	(0.034)	(0.023)	(0.037)	(0.024)	(0.041)	(0.038)	(0.039)
Age	0.003	0.002	0.003***	0.003***	0.002**	0.002**	0.002*	0.002*	0.004***	0.004***
	(0.003)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
White	-0.089	-0.081	0.033	0.033	0.042	0.042	0.029	0.030	0.075*	0.073
	(0.100)	(0.099)	(0.042)	(0.042)	(0.048)	(0.048)	(0.043)	(0.043)	(0.044)	(0.045)
College Degree	0.113	0.096	0.051	0.049	0.053	0.050	0.034	0.028	-0.014	-0.022
	(0.140)	(0.136)	(0.050)	(0.050)	(0.050)	(0.049)	(0.050)	(0.049)	(0.066)	(0.064)
Post-graduate Degree	0.060	0.037	0.033	0.031	0.046	0.042	0.036	0.029	0.024	0.010
	(0.153)	(0.146)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.068)	(0.065)
Constant	0.474***	0.458***	0.459***	0.472***	0.480***	0.505***	0.523***	0.562***	0.482***	0.524***
	(0.148)	(0.151)	(0.071)	(0.071)	(0.071)	(0.067)	(0.073)	(0.071)	(0.092)	(0.085)
R-squared	0.030	0.037	0.029	0.030	0.028	0.030	0.025	0.030	0.042	0.047
N	202	202	1276	1276	1077	1077	1005	1005	645	645

Notes. The dependent variable is a dummy equal to one if Margaret is chosen. Treatment "Unbalanced" is a dummy equal to one if the subject is exposed to a pool of six male instructors, and zero if he/she is exposed to a gender balanced pool of instructors. In columns from 1 to 10, we report results of MTurk experiments for different samples of subjets, namely those who did not check any advice (1-2), those who checked at least one advice (3-4), those who checked at least two advices (5-6), those who checked at least 3 advices (7-8), and those who guessed corretly how many female instructors were in the pool (9-10). Robust standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A Online Appendix

A.1 Description of MTurk Experiment

The experiment is structured in seven steps, which are listed below. In every step, subjects are shown a screen window. In the first four steps, subjects are free to choose when to move forward by clicking on the arrow in the lower right corner of the screen. Once the subjects click on the arrow, they move to the next step and cannot go back. We made this rule clear by warning subjects with this sentence at the bottom of the screen window in step 1 to step 4: “After a short while, you will be able to click on the arrow below in order to proceed. Once clicked, you will no longer be able to go back.”

Step 1. All the subjects are given the following information:[9]

- They will have to solve simple math/language tasks (10 questions) under time pressure.
- They will be paid based on performance (40 cents for each correct answer).
- They will all receive $1 for their participation.
- Before the test, they can read tips on how to solve the tasks written by different instructors.

Step 2. Two different lists of 6 instructors are shown to subjects. They are not given any information other than the instructors’ first names and qualification as “graduate student” (see Figure 2, panel A and B, upper part).

- Treatment subjects are exposed to a pool of 6 male instructors.
- Control subjects are exposed to a pool of 3 female and 3 male instructors.

Step 3. Subjects are asked to choose one additional instructor; they can choose between one female and one male candidate (see Figure 2, panel A and B, lower part).

- The two candidates are Margaret (female candidate) and Richard (male candidate).
- The two candidates have the same educational background: they are both enrolled in a PhD.
- Subjects are given some additional information about the two candidates: GPA and hours of experience as TA.

Step 4. Subjects may read as many tips as they want. They do not have any time limit in this stage.

[9] Subjects randomly assigned to the math task visualized precisely the following message: “Thank you for your participation in this study. You will receive 1 dollar for your participation, that is, if you complete the study. We estimate it will not take more than 15-20 minutes. We will ask you to perform a MATH task and we will pay you according to how well you do the task. In particular, we will ask you 10 questions with limited time to respond, and we will pay you 40 cents per correct answer. If you answer correctly all the 10 questions you will receive 4 dollars in addition to the 1 dollar for your participation. Before you do the task, you will be able to read explanations on the task, and you will receive tips on how to get the correct answer for the MATH questions quickly. You will have 10 seconds to answer each question. In the next screen you will find the pool of instructors, all of whom will explain the task and give you tips on how to solve the task correctly under limited time. After a short while, you will able to click on the arrow below in order to proceed. Once clicked, you will no longer be able to go back.”. Subjects randomly assigned to the english task visualized the same message, with the only difference that the word MATH was replaced by the word ENGLISH.

Step 5. Whenever they feel ready, subjects can proceed with the exercise solving part. They have 10 seconds for each question.

- If subjects are randomized into the math task, they have to solve 10 multiplications of the number 11 with a two or more digit number.
- If subjects are randomized into the language task, they have to spell 10 English words correctly.

Step 6. Subjects are asked to give some personal information (age, gender, education).

Step 7. At the end, subjects are asked to answer the question "In the pool of six instructors how many women were there?". Options were in a range from zero to three.

Figure A.1: Percentage of Subjects Choosing Each Advice

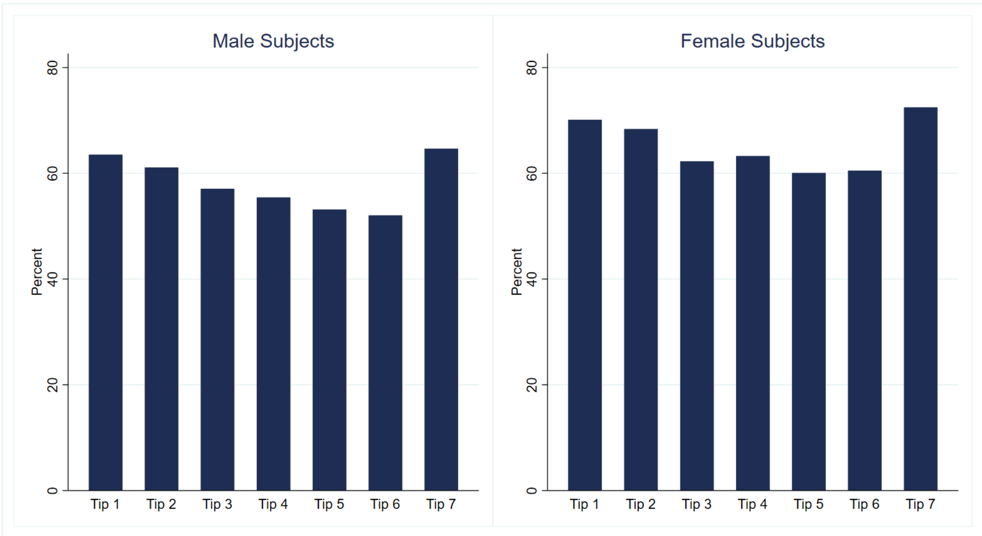


Table A.1: Descriptive Statistics of Teaching Evaluations

	Comm.	Econ.	Comp. Sc.	Δ(E,CO) *P*-value	Δ(E,CS) *P*-value
Column	(1)	(2)	(3)	(4)	(5)
Panel A: Students Characteristics					
No. of Students	770	922	218	-	-
Dummy Female Student	0.69	0.43	0.12	0.00	0.00
Dummy Swiss Students	0.44	0.34	0.35	0.00	0.66
Dummy Italian Students	0.41	0.49	0.30	0.00	0.00
Dummy Other Nationalities	0.15	0.17	0.34	0.25	0.00
Dummy Bachelor Students	0.59	0.48	0.66	0.00	0.00
Student Age	24.56	23.89	24.41	0.00	0.03
Panel B: Course Characteristics					
No. of Courses	430	420	191	-	-
Dummy Compulsory Courses	0.60	0.45	0.71	0.00	0.00
Dummy Quantitative Courses	0.14	0.51	0.90	0.00	0.00
Class Size	34.30	39.36	24.61	0.05	0.00
Panel C: Instructor Characteristics					
No. of Instructors	181	171	89	-	-
Dummy Female Instructors	0.33	0.23	0.17	0.04	0.22
Dummy Full Professors	0.28	0.32	0.36	0.42	0.54
Dummy Associate Professors	0.13	0.17	0.18	0.33	0.84
Dummy Assistant Professors	0.07	0.10	0.10	0.26	0.96
Dummy Lecturers	0.52	0.40	0.34	0.02	0.30
Publish or Perish Citations	87.23	131.68	1225.48	0.11	0.00
Panel D: Student-Course Characteristics					
No. of Teaching evaluations (TE)	11,768	12,435	2,793	-	-
Dummy Students repeating courses	0.02	0.04	0.06	0.00	0.00
Dummy Students not reporting TE-Score	0.06	0.07	0.04	0.00	0.00
Student Grade	7.92	7.51	7.51	0.00	0.84
TE-Score: Overall satisfaction with the course	7.21	7.22	7.28	0.82	0.42

Notes. Table reports summary statistics related to students (Panel A), courses offered (Panel B), professors (Panel C), and students-course characteristics (Panel D) for the academic years 2015 to 2017. In each panel, we report sample numerosity in the first row. For each variable, we report the mean of the variable by faculty (Columns 1-3). In Column 4 we report the P-value of the difference between the mean values of Economics and Communication. In Column 5 we reports the P-value of the difference between the mean values of Economics and Computer Science.

Table A.2: Descriptive Statistics of Gender-Related Questions in the Survey

	All	Male	Female	N_M	N_F	$\Delta(M,F)$ P-value
	(1)	(2)	(3)	(4)	(5)	(6)
University education						
Communication	0.017	0	0.023	16	43	0.546
Economics	0.074	0.102	0.028	59	35	0.195
Computer Science	0	0	0	21	8	-
Having a job						
Communication	0.146	0.2	0.129	10	31	0.592
Economics	0.285	0.325	0.222	43	27	0.358
Computer Science	0.444	0.5	0.33	12	6	0.531
Mother works						
Communication	0.017	0.062	0	16	43	0.102
Economics	0.086	0.086	0.086	58	35	0.993
Computer Science	0.310	0.380	0.125	21	8	0.195
Men better executives						
Communication	0.137	0.25	0.095	16	42	0.131
Economics	0.276	0.39	0.086	59	35	0.0012
Computer Science	0	0	0	21	8	-
Jobs scarce						
Communication	0.018	0	0.025	15	40	0.545
Economics	0.106	0.095	0.121	42	33	0.722
Computer Science	0	0	0	18	8	-

Notes: The table shows gender differences in attitudes toward gender stereotypes within the faculties of Communication, Economics and Computer Science. We focus on the following survey questions: "A university education is more important for a boy than for a girl" , "Having a job is the best way to gain independence for a women", "When a mother works for pay, children suffer", "On the whole, men make better business executives than women", and "When jobs are scarce men should have more right to a job than women". We report the share of students enrolled in a given faculty who agree with the statement, in total (column 1) and by gender (columns 2-3), the number of students filling out the survey by gender (columns 4-5), and the P-value of a two-sample t-test for differences between female and male answers (column 6).

Table A.3: Summary Statistics of MTurk Experiment: Balanced versus Unbalanced (Female Scarce)

Group	Balanced			Unbalanced			(B-U)
	No.Obs	Mean	Std.Dev	No. Obs	Mean	Std.Dev	P-value
Panel A: Permutation variables							
Math Task	743	0.47	0.50	735	0.47	0.50	0.886
Margaret First	743	0.49	0.50	735	0.50	0.50	0.756
Margaret TA	743	0.49	0.50	735	0.52	0.50	0.404
Panel B: Sociodemographic variables							
Female	743	0.45	0.50	735	0.48	0.50	0.18
Age	743	35.78	11.32	735	36.36	11.41	0.33
White	743	0.77	0.42	735	0.76	0.43	0.76
College degree	743	0.60	0.49	735	0.61	0.49	0.59
Post-graduate degree	743	0.30	0.46	735	0.31	0.46	0.75
Panel C: Subjects' performance							
Margaret chosen	743	0.63	0.48	735	0.69	0.46	0.013
Duration	743	819.90	352.43	735	840.87	502.71	0.353
No. of advices	743	4.35	2.70	735	4.25	2.77	0.472
No. of correct answers	743	7.03	3.48	735	6.97	3.36	0.732

Notes. The group "Balanced" includes all subjects exposed to a gender balanced pool of instructors, while the group "Unbalanced" includes all subjects exposed to a pool of six male instructors. For each variable of interest, we report the number of observations, mean and standard deviation. The last column reports P-values of a t-test between variables in control and treatment group.

Table A.4: Summary Statistics of MTurk Experiment: Balanced versus Unbalanced (Male Scarce)

Group	Balanced			Unbalanced			(B-U)
	No.Obs	Mean	Std.Dev	No. Obs	Mean	Std.Dev	P-value
Panel A: Permutation variables							
Math Task	743	0.47	0.50	699	0.47	0.50	0.76
Margaret First	743	0.49	0.50	699	0.50	0.50	0.83
Margaret TA	743	0.49	0.50	699	0.50	0.50	0.80
Panel B: Sociodemographic variables							
Female	743	0.45	0.50	699	0.42	0.49	0.29
Age	743	35.78	11.32	699	34.67	10.32	0.051
White	743	0.77	0.42	699	0.75	0.44	0.28
College degree	743	0.60	0.49	699	0.70	0.46	0.00
Post-graduate degree	743	0.30	0.46	699	0.20	0.40	0.00
Panel C: Subjects' performance							
Richard chosen	743	0.37	0.48	699	0.52	0.50	0.00
Duration	743	819.90	352.43	699	827.60	354.50	0.68
No. of advices	743	4.35	2.70	699	4.69	2.58	0.01
No. of correct answers	743	7.03	3.47	699	7.20	3.23	0.34

Notes. The group "Balanced" includes all subjects exposed to a gender balanced pool of instructors, while the group "Unbalanced" includes all subjects exposed to a pool of six female instructors. For each variable of interest, we report the number of observations, mean and standard deviation. The last column reports P-values of a t-test between variables in control and treatment group.

Table A.5: Choice of Male Instructor when Male Instructors are Scarce

	(1)	(2)	(3)
Unbalanced (β)	0.136***	0.131***	0.161***
	(0.035)	(0.034)	(0.041)
Female X Unbalanced (γ)			-0.064
			(0.053)
Math Task	-0.005	-0.012	-0.012
	(0.028)	(0.028)	(0.028)
Margaret First	0.087**	0.089***	0.087***
	(0.033)	(0.033)	(0.032)
Margaret TA	-0.016	-0.013	-0.014
	(0.038)	(0.039)	(0.039)
Female		-0.020	0.012
		(0.028)	(0.036)
Age		-0.002	-0.002
		(0.001)	(0.001)
White		0.027	0.025
		(0.033)	(0.033)
College Degree		0.082*	0.085*
		(0.046)	(0.046)
Post-graduate Degree		0.026	0.028
		(0.046)	(0.046)
Constant	0.334***	0.338***	0.320***
	(0.044)	(0.069)	(0.070)
$\beta + \gamma$			0.096**
			(0.045)
R-squared	0.029	0.037	0.038
N	994	994	994

Notes. The dependent variable is a dummy equal to one if Richard is chosen. Unbalanced is a dummy equal to one if the subject is exposed to a pool of six female instructors, and zero if he/she is exposed to a gender balanced pool of instructors. All included subjects checked the advice by the chosen instructor. Robust standard errors are reported in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.