ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

ECONOMICS

Ciclo 36

Settore Concorsuale: 13/A1 - ECONOMIA POLITICA

Settore Scientifico Disciplinare: SECS-P/01 - ECONOMIA POLITICA

ESSAYS IN GENDER ECONOMICS

Presentata da: Elisa Orlandi

Coordinatore Dottorato

Andrea Mattozzi

Supervisore

Francesca Barigozzi

Esame finale anno 2025

Acknowledgements

I am deeply grateful to Professor Francesca Barigozzi for her academic guidance and personal support; without her I am not sure I could have faced the challenges of this PhD. I could not have asked for a better mentor.

I also thank my co-authors—Marianna Baggio, Ginevra Marandola, Hannah U. Nohlen, Natalia Montinari, Alexia Gaudeul and Andrea Blasco. You taught me how research works.

My gratitude extends to the members of the Economics faculty and my PhD colleagues in Bologna, and to all the members of the Collaborative Centre for Behavioural Insights (JRC), with a special thank to Emanuele Ciriolo for his guidance.

And to F.R., thank you—for everything.

Abstract

This dissertation, titled *Essays in Gender Economics*, brings together four studies, each shedding light on how gender shapes economic behaviour, decision-making and outcomes.

The first essay investigates the effectiveness of Equal Employment Opportunity (EEO) statements in Spain and Germany, using an online experiment. It shows that targeting women with explicit equal opportunity statements encourages them to apply for high-paying managerial roles, without discouraging men.

The second essay uses a pre-registered lab experiment to examine gender differences in *explore-exploit* tasks, focusing on exploration and competition under both gain-only and gain-loss environments. Surprisingly, women exhibit more exploration in gain-only settings, and once individual risk attitudes are considered, the initially lower tendency to compete that they show in gain-only environments aligns with that of men.

The third essay examines employee responses to pay and job discrimination through a large-scale online labour market experiment conducted in Germany and Romania. It finds that discrimination increases complaints, particularly job discrimination, but does not reduce effort. Women under discrimination exert more effort and report stronger emotional distress than men.

Finally, the fourth essay analyses more than five million university syllabi and, using a Monte Carlo benchmark that represents gender-neutral matching, shows that mixed-gender co-teaching teams occur at roughly half the rate expected; teaching team gender composition is also correlated with the novelty, interdisciplinarity, and gender diversity of the readings assigned in courses.

Essays Included in this Dissertation:

1. *"The Impact of Equal Opportunity Statements in Shaping Gender Distribution in Job Applications"*
   Co-authors: Marianna Baggio, Ginevra Marandola, Hannah U. Nohlen

2. *"Gendered Decision Making in Explore-Exploit Tasks"*
   Co-authors: Francesca Barigozzi, Natalia Montinari

3. *"Complaints and Effort under Pay and Job Discrimination: A Labour Market Experiment"*
   Co-authors: Marianna Baggio, Alexia Gaudeul

4. *"Gender in Teaching: Insights from Five Million Syllabi on Collaboration, Interdisciplinarity and Reading Selections"*
   Co-authors: Andrea Blasco

The Impact of Equal Opportunity Statements in Shaping Gender Distribution in Job Applications

Marianna Baggio[*†] Ginevra Marandola[*‡] Hannah U. Nohlen[*] Elisa Orlandi [§]

Abstract

Affirmative action policies remain at the centre of policy debate. Among these, equal employment opportunity (EEO) statements are those whose policy effects are less understood and analysed in the scientific literature. In an incentivised online experiment conducted in Spain and Germany with 3,736 participants (1,860 from Spain and 1,876 from Germany), we investigate the effect of equal opportunity statements on job application behaviour. Specifically, we focus on the choices made by men and women when deciding between two managerial positions with different qualification requirements and, consequently, different pay levels. We manipulated whether EEOs were placed on one or both job types that participants could apply for. Our results indicate that equal opportunity statements targeted towards women successfully increase the likelihood of women applying for higher-paying, high-qualification positions when they are placed on high-qualification jobs only as well as when placed on all job offers. Notably, men are not discouraged by the presence of these EEO statements.

Keywords: Equal Employment Opportunity Statements, Affirmative Action, Gender Differences, Job Application Behaviour, Online Experiment

JEL Codes: C91, J16, J71

[*]Joint Research Centre, European Commission
[†]European Foundation for the Improvement of Living and Working Conditions
[‡]AGCM, Italian Competition Authority
[§]University of Bologna

1 Introduction

Gender disparities in the workplace have been a resistant and long-standing issue. In the European Union, women still earn about 13% less than men on average [European Commission, 2023]. Women frequently encounter numerous barriers to career advancement, including gender biases, stereotypes, occupational segregation, work-life balance challenges, and limited opportunities for networking, mentoring and development - each presenting significant hurdles to their professional progress [Eagly and Carli, 2007]. Given these barriers, and accounting for several behavioural factors further outlined below, women have been found less likely than men to apply for promotions and high-paying jobs [Haegele, 2024, Fluchtmann et al., 2024]. This is also reflected in the persistent underrepresentation of women in top leadership positions. Recent research by the European Institute for Gender Equality (EIGE) highlights that only 23.3% of executives and 8.2% of CEOs in the largest publicly listed companies in the European Union are women. In Germany and Spain, the two countries where we conducted our experiment, women comprise 23.5% and 19.1% of executive positions and only 2.6% and 2.9% of CEOs, respectively [European Institute for Gender Equality, 2023]. Affirmative action policies aim to promote equal opportunities, but there is much discussion about their impact and implementability. Among these policies, equal employment opportunity (EEO) statements in job advertisements are expected to encourage underrepresented groups (e.g. minorities, women) to apply for all kinds of jobs, including those that have more demanding requirements and pay more. The current study aims to understand the effectiveness of EEO statements in promoting gender equality in job applications and identify circumstances where they may have unintended adverse effects.

Workplace gender disparities are influenced by several factors. Next to contextual and societal issues such as gender biases and work-life balance challenges, psychological factors such as confidence, self-assessment, and self-promotion can play a role. Some research suggests that women are less likely to enter competitive environments and tend to shy away from competition [Niederle and Vesterlund, 2007, 2011, Buser et al., 2014]. Women are consistently found to be less confident than men [Möbius et al., 2022, Sarsons and Xu, 2021, Niederle and Vesterlund, 2007], especially in fields that are not stereotypically female [Bordalo et al., 2019]. Women are also less likely to self-promote and provide worse self-evaluations [Exley and Kessler, 2022, Biasi and Sarsons, 2022], which can significantly impact their pursuit of positions with more demanding qualification requirements. Another explanation for gender differences in career advancement is that women have been found to be significantly more risk averse than men (see Croson and Gneezy, 2009, for a comprehensive literature review of gender differences in risk aversion in experimental settings, note that some studies find the results to be context and task-dependent; Filippin and Crosetto, 2016,Filippin, 2022). As a result, even high-performing qualified women may avoid applying to more competitive job positions or

promotions. This is particularly true in environments where qualifications are not easily verifiable [Kanthak and Woon, 2015]. These behavioural factors provide a margin of efficacy for interventions to promote gender equality in the labour market, including affirmative action measures. Furthermore, interventions targeting gender equality should be tailored to the specific needs of the intended beneficiary group while ensuring no detrimental effects on the rest of the population, thus aiming for a Pareto improvement.

Affirmative action measures, such as gender quotas, gender-specific training, and preferential treatment aim at supporting women in accessing more opportunities. Specifically, these measures may help reduce gender differences in employment rates and pay gaps by encouraging women to apply for competitive job positions [Balafoutas and Sutter, 2012, Niederle et al., 2013]. However, they may also have unintended consequences in terms of efficiency, intra-group inequality and intergroup relations [Leibbrandt et al., 2018, Matsa and Miller, 2013]. The implementation of affirmative action measures is, therefore, still at the centre of the economic and political debate (see, for example, Gender Balance on Corporate Boards, European Parliament and the Council of the European Union [2022]).

In this study, we focus on the role of one specific affirmative action measure: Equal Employment Opportunity (EEO) statements. EEO statements are regarded as a promising and relatively low-effort and low-cost affirmative action measure. They are typically included in job advertisements or featured on a company's website, expressing the organisation's commitment to providing non-discriminatory employment opportunities. EEO statements have been largely overlooked in the literature so far. Most previous studies have focused on self-reported measures, such as participants' perceptions of organisational attractiveness (Klysing et al. [2022],Heath et al. [2023]; see also Alahakoon et al. [2024], for a literature review).

For instance, McNab and Johnston [2002] explored the effect of EEOs on perceptions of organisations. They highlighted that EEO statements can make organisations more appealing to women. Similarly, Nater and Sczesny [2016] found that explicitly inviting women to apply for leadership positions increased their perceived fit for the role, which in turn raised their self-reported inclination to apply.

Recent studies, however, have sparked a debate on the impact of EEO statements on the recruitment and retention of underrepresented groups. For example, whereas Flory et al. [2021] suggest that a company explicitly signalling its interest in attracting a diverse workforce encourages minorities to apply, Leibbrandt and List [2018] found that EEO statements may discourage minorities from applying for job opportunities due to stereotype threat and tokenism[1].

[1]A broader body of literature has explored diversity, equity, and inclusion (DEI) materials in recruitment. For example, Walker et al. [2009] examined employee testimonials on recruitment websites, and Walker et al. [2012] analyzed diversity cues, such as employee gender composition, on job seekers' behavior. Ibañez and Riener [2018] studied affirmative action statements describing the use of quotas or preferential treatment in the recruitment process. Delfino [2024] focused on photographs and information about the share of workers who received higher evaluations in the past. However, our focus is specifically on EEO statements, understood as commitments to diversity without promises of specific selection processes or diversity reporting, which we consider especially relevant given how widespread such statements are in job advertisements and recruitment platforms.

One reason for the observed differential effects of EEO statements may be the type of job to which these statements are applied. Flory et al. [2024] found that the effectiveness of diversity signals in job recruitment materials depends on the career stage of the applicants, suggesting that diversity signals may have varying impacts depending on the job level or field.

More generally, the literature has found that gender differences in application behavior are affected by different aspects of job positions. Women are less likely to apply to positions that are stereotyped as masculine [Barbulescu and Bidwell, 2013]. Women are also found to be discouraged by competitive compensation schemes [Samek, 2019, Flory et al., 2015]. In addition, a frequently cited claim, popularized by Sandberg [2015], suggests that women apply for jobs only when they meet 100% of the qualifications, whereas men apply when they meet just 60%. This statement originates from anecdotal evidence in an internal Hewlett-Packard report, as highlighted by Rice [2014], and thus should be taken with caution. However, Abraham et al. [2024] found that qualifications and the language used for qualifications do affect application behavior. In their field experiment, they found thatwomen who applied for the job positions were significantly more skilled than men, and that lowering the required qualifications increased the number of applicants from both sexes while reducing this qualification gap. Similarly, Coffman et al. [2024] found that removing ambiguity from qualification requirements makes qualified women more likely to apply to demanding job positions, reducing the gender gap in application rates. Both studies suggest that fit for the position or quality match might be more important for women than for men. Notably, a recent paper found no gender difference in application intention depending on qualification fit but reported higher psychological hurdles for women; for example, women in their sample strive more for preparedness than men [Salwender and Stahlberg, 2024].

Our study aims to investigate how the effectiveness of EEO statements varies across jobs with different qualification requirements, given their role in shaping application behavior and gender differences in application behavior.

More specifically, we aim to understand the effectiveness of EEO statements in promoting gender equality in job applications and identify circumstances where they may have unintended negative effects for women or men. We investigate the impact of the presence of EEO statements specifically on jobs with lower versus higher qualifications, intended as a combination of prior experience, education and skills. In our study, EEO statements were added to job advertisements for a lower-paying managerial position, a higher-paying managerial position, both, or neither (control), allowing us to assess their impact variability across positions. We collected data in two countries, Germany and Spain, in which the percentage of women in top leadership positions is below the European average. Testing the effect of the presence of EEO statements in these countries allows us to draw conclusions on their effectiveness in contexts where efforts for more gender equality are needed. Controlling for socio-demographic factors and exploring variations across two

countries, our research forms the foundation for comprehending the role of EEO statements in shaping gender-balanced job applications. We seek to discern when EEO statements are effective, when they may fall short, and highlight potential negative effects for both genders.

To summarise, our study offers several unique contributions to the existing literature. First, our study goes beyond typical approaches that study the effectiveness of affirmative action measures using attractiveness ratings of organisations, and instead investigates the decision-making process of job applications in the context of an incentivized experiment. Second, while previous research has predominantly focused on race and ethnicity, our study specifically examines gender differences, addressing a significant gap in the literature. Third, unlike previous studies that are primarily US-centric, our research broadens the context to the European labour markets, providing insights relevant to the EU. Fourth, we investigate the effects of EEO statements on both high-paying and low-paying managerial positions, requiring different qualifications. By focusing on different job types, we add a dimension that could partially explain the conflicting results observed in the literature, suggesting that the effectiveness of EEO statements might depend on the job's qualification requirements and pay levels. Finally, utilising a large, non-student sample from two countries ensures the robustness and generalizability of our findings.

2 Methods

In the current study, we examine how the presence of EEO statements on job ads influences women's and men's job application behaviour. More specifically, we conducted an incentivized online experiment to answer the following questions: a) Do EEO statements explicitly targeting women encourage women to apply to higher-requirement jobs?, b) Does the presence of these EEO statements discourage men from applying to higher-requirement jobs?, and c) What are the consequences of having EEO statements present on either low-paying, low-qualification requirements jobs, or high-paying, high-qualification requirements jobs, or both? The study was approved by the Ethical Review Board of the Joint Research Centre, European Commission (evaluation ID 30506_3_05102021), and preregistered (`https://www.socialscienceregistry.org/trials/8284`). A comprehensive account of all measures collected, including those not analysed in the article, can be found in the annexes.

2.1 Participants

Data collection took place between June 2021 and January 2022, and participants were recruited through a panel provider (Dynata) in two EU member states: Germany and Spain. A total of 3800 participants were recruited for the study, with 1900 participants per country. The decision to on Germany and Spain was

driven by the necessity to conduct a study that holds relevance at the EU level, offering insights into various cultural, policy, and labour market setups, including factors such as gender pay gap, gender participation and segregation in labour markets, and the presence of pay transparency policies, to name but a few. Germany and Spain were chosen due to their differences in gender-related workplace legislation and outcomes as well as due to their size, which facilitated recruiting a large number of participants through an online panel. In Germany, the unadjusted gender pay gap—measured as the percentage difference between the average gross hourly earnings of male and female employees—is 17.7%, which is higher than the EU average of 12.7%. In Spain, the gap is 8.7%, which is lower than the EU average [Eurostat, 2022].

We aimed for nationally representative samples in terms of age ranges and regions while keeping a gender-balanced sample and maintaining coverage of different educational levels. Participants were screened out from the survey if they failed attention checks or took less than 6 minutes or more than 62 minutes to complete the experiment[2]. We considered that participants who completed the survey in such a short time had a higher chance to reply randomly, while those who took too long were likely to engage in unrelated tasks while taking the survey. Additionally, participants whose self-reported gender did not coincide with the gender in the Dynata database were also excluded.

The final sample included 1860 observations for Spain (931 men) and 1876 observations for Germany (935 men). Age ranged from 18 to 88 years, with $M_{age} = 47.45$ years ($SD = 15.57$ years) in Spain, and $M_{age} = 49.04$ years ($SD = 16.69$ years) in Germany. Further details on the sample characteristics can be found in Table S1 (Appendix S1).

The study was created in English and then translated into German and Spanish by professional translators, including back-translation to ensure quality. Additionally, both versions were proofread by native speakers before conducting the study.

Before the experiment, respondents were informed that the study's purpose was to gain a better understanding of job recruitment dynamics. They were reassured about the anonymization of their responses and were given details about the survey's duration. Participants were made aware of the participation fee and were told that the rest of their payment would depend on their performance in the tasks. All participants provided informed consent. They were also informed that they could end their participation in the study at any time and for any reason.

[2]Based on the pilot, the fastest 1% of respondents corresponds to a completion time of under 6.4 minutes. The slowest 1% corresponds to a completion time of 62 minutes.

2.2 Experimental design and procedure

After providing demographic information such as gender, age, education level, employment status, and industry, participants were briefed about the structure and tasks of the experiment (see Appendix S2 for the complete set of instructions and questionnaire as provided to participants). They assumed the role of an *experienced manager* in search of a new job and were required to complete three tasks. Firstly, they were asked to select a resume (*Resume Task*). Next, they had to decide which job position to apply for (*Job Choice Task*). Finally, they were asked to choose a 'managerial statement' to include on their application form. After receiving the hiring decision, whether accepted or rejected, participants were asked several questions about their experience with the decision tasks and their risk propensity.

2.3 Resume Task

In the Resume Task, participants were presented with two fictitious resumes. The two resumes (Appendix S3) were identical in terms of work experience (5 years of experience in managing a team of 8 employees) and education level (Master's degree), differing only in terms of the profile of the attended Master's programme (Business administration vs. Management), names of previous workplaces (CCBI Ltd. vs. Delta Ltd.), type of employees supervised (business experts vs. market analysts) and type of certificate attesting to English proficiency (Cambridge vs. Oxford).

Both resumes presented in the *Resume Task* met the minimum qualification requirements for the two job advertisements presented in the *Job Choice Task*. In other words, regardless of the resume participants selected in the first task, they possessed sufficient qualifications to apply for both types of jobs in the second task. Letting participants choose the resume, rather than assigning it randomly, was solely done to increase participants' attention to the resume and to create a sense of ownership. Additionally, knowing that applicants had chosen their resumes made the choice of recruiters more realistic.

2.4 Job Choice Task

After selecting the resume, participants were asked to choose which of two job positions (Appendix S4) to apply for: a lower-paying job with fewer required qualifications (Position Beta) or a higher-paying job with more stringent qualifications (Position Gamma). This choice served as the main dependent variable in our study. As already stated, both resumes provided to the participants demonstrated qualifications that met the minimum qualification requirements for both the lower-paying and higher-paying positions. (cf. Table 1 for differences in job requirements). Participants based their decision to apply on the job ads only, mimicking

job applications in the real world[3].

Position Gamma offered a wage of 150 tokens, while Position Beta offered a wage of 50 tokens, with a conversion rate of 10 tokens to 0.40 €. Participants would receive the wage listed in the job advertisement only if they were hired; otherwise, they would receive €0. Hence, job applicants who were hired had the opportunity to earn either €2 or €6, in addition to a €3 participation fee. Those who did not secure a job received only the participation fee.

	High-requirements, high-paying position (*Gamma*)	**Low-requirements, low-paying position (*Beta*)**
Wage	150 tokens	50 tokens
Required Experience	Between 5 and 8 years of experience managing a medium size team (8 people or more)	At least 3 years of experience managing a small-medium size team (at least 5 people)
Required Education	PhD preferred, but a Master is accepted	Master
Language	English, a second language is an asset	English

Table 1: Job Choice Task: Wage and Requirements

Important to note that the decision to hire a participant was made by real people fulfilling the role of HR executives in a separate survey (N = 64; 32 per country)[4]. This approach was designed to ensure incentive-compatibility and to simulate real-life hiring processes, avoiding deception.

After the Job Choice Task, participants could select one of two management style statements to send along with their application. Essentially, participants were asked to choose a statement describing their management style. One statement alluded to values more commonly associated with a male stereotype, while the other was more commonly associated with a female stereotype. Participants were informed that this choice would also be reviewed by the HR executive responsible for the hiring decision. The profile, the job vacancy, and the statement describing the management style together represented the participant's job application. The choice of statement was not analysed for this paper. However, for transparency reasons, we want to disclose that any results concerning the managerial statement choice were not significant[5]. Details on the statements' phrasing can be seen in Appendix S5.

[3]When applying for jobs, people are usually unaware of the quality of their competitors, their gender, incentives of HR staff responsible for the hiring, or gender stereotypes of HR staff, similar to the design of our study.

[4]The HR executives in the separate survey had to compare two job applications at a time and could hire only one of the applicants. They were instructed to select the applicant they thought would be most suitable for the job, as if the position was advertised in the real world (see Appendix S2 for the exact wording and the questionnaire as presented to individuals who played the role of HR executives) . This study aims to understand application behaviour and the effectiveness of EEO statements on application behaviour. Therefore, the analysis of HR executives' decisions is not of interest and will not be analysed further.

[5]Analyses can be provided upon request.

2.5 Treatments

Participants were randomly assigned to one of four between-subject conditions, characterised by the presence or absence of Equal Employment Opportunity (EEO) statements in the two job ads participants could apply for (see Table 2). We used the following EEO statement: "Company Beta (Gamma) values a diverse workplace and strongly encourages women to apply." No EEO statement was included in the job adverts presented in the control condition (control). EEO statements were either incorporated only in the job advert of the high-paying position (EEOHigh), only in the advert of the low-paying position (EEOLow), or in both, the advert of the high-paying and the low-paying position (EEOboth).

	High-qualifications, high-paying position	**Low-qualifications, low-paying position**
Control	✗	✗
EEOboth	✓	✓
EEOHigh	✓	✗
EEOLow	✗	✓

Table 2: Experimental conditions: presence of EEO statements

2.6 Post-experimental questionnaire

After choosing the job, participants filled out a post-experimental questionnaire[6].

Before participants were informed of the hiring decision, they ranked the three most relevant reasons for choosing the job out of several provided reasons (Appendix S2). We also asked whether they had noticed the EEO statement. For participants who reported having noticed the EEO statement, we assessed their reaction to the statement in terms of perceived fairness of the hiring process ("It [the EEO statement] made me think my application was treated…" ranging from (0) unfairly to (10) fairly) and confidence in their application ("It [the EEO statement] affected my confidence in my application", ranging from (0) decreased confidence to (10) increased confidence) using sliders.

We also asked whether the statement changed participants' expectation of an equal gender distribution among employees. Agreement with the item "The statement made me expect that there would be an equal gender distribution among employees" was recorded on a 5-point Likert scale (1 = Strongly disagree … 5 = Strongly agree). In a separate item, participants stated how the statement influenced their job choice, selecting one of five ordered categories ("It encouraged me / somewhat encouraged me / neither encouraged

[6]We only report the measures analysed concerning the main dependent variable of the study. Some questions were also included regarding beliefs about the chosen management style, a variable that is excluded from this paper (for the complete questionnaire, see Appendix S2)

nor discouraged me / somewhat discouraged me / discouraged me to choose the higher-paying job").

Finally, we measured whether they believed that the statement influenced the hiring decisions of HR executives who made the hiring decisions in the study ("Do you believe the presence of this statement influenced the choice of HR executives hiring for that job position?") with 5 response options (Appendix S4).

After providing participants with the decision regarding their application, we measured risk aversion ("Please indicate on the scale below how you see yourself: Are you, in general, a person who takes risks or do you try to evade risks?") on a scale ranging from (1) "Not at all prepared to take risk" to (10) "very much prepared to take risk". Finally, we asked participants to indicate which gender they believed the HR executive who took the decision was ("male", "female", "I don't know").

3 Results

3.1 Distribution of Job applications by gender

In the control condition (no EEO), men and women in our sample applied similarly often to the job with higher requirements. Specifically, 51.3% of women applied to the high-paying position compared to 52.1% of men ($\chi^2 = 0.0744$, $p > 0.1$)[7].

Including an EEO statement in the job postings, on the other hand, had a gender-specific effect on applications and the effect was dependent on whether the statement appeared on one or both jobs (see Figure 1).

3.2 Effect on women's job applications

EEO statements on only one of the two jobs encouraged women's applications to the position where the statement was included. This was true for high-paying jobs with higher qualification requirements and low-paying jobs with lower qualification requirements. When the EEO statement was present only in the high-paying job post, over 63% of women applied for that position, an increase of around 12 pp compared to when no statement was present on either job ad ($\chi^2 = 14.713$, $p < 0.001$). In contrast, when the statement appeared only in the low-paying job post, almost 56% of women applied for the low-paying position, an increase of about 7pp compared to the control ($\chi^2 = 4.825$, $p = 0.028$). Symmetrically, the share of applications to the high-paying job posts decreased to 44%. However, when EEO statements were featured in both job posts, women applied more often to high-paying, high-requirement positions: in this case, 60%

[7] In Section 3, we report the unadjusted p-values for all χ^2-tests. As a robustness check, Appendix A includes these p-values adjusted for multiple testing, using three different correction methods: Bonferroni, Holm, and False Discovery Rate (FDR).

of female participants applied to the high-paying job, an increase of around 9 percentage points compared to the control ($\chi^2 = 7.7004$, $p = 0.006$).

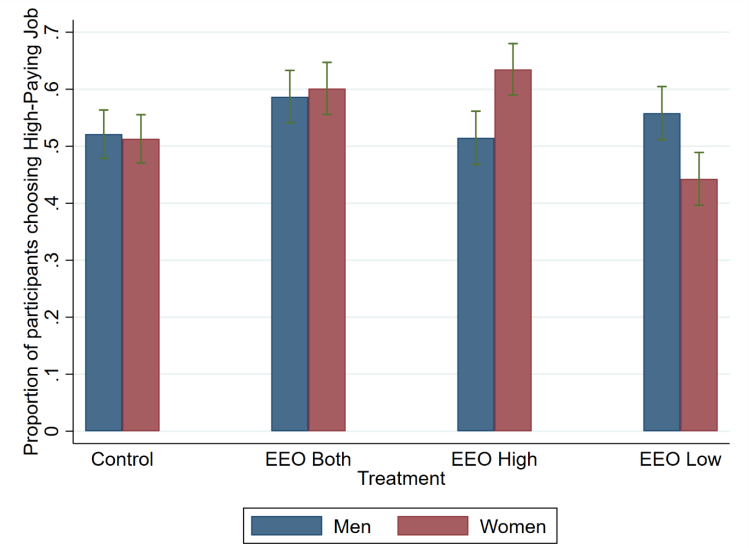


Figure 1: Proportion of participants choosing the High-Paying Job position by treatment and gender

3.3 Differential effects on women and men's job applications

We found that, while having the statement only on the high-requirements job did not discourage men from applying, the strong effect that it had on women made the distribution unbalanced in women's favour, with significantly more women (63.49%) than men (51.47%) applying to the high-requirements job ($\chi^2 = 13.074$, $p < 0.001$).

On the other hand, we found that when the EEO statement was placed only on the low-requirements job, fewer women (44.27%) than men (55.81%) applied to the high-requirements job, resulting in an unbalanced gender distribution in favour of men ($\chi^2 = 11.77$, $p < 0.001$).

Finally, the presence of the statement on both jobs increased the total number of applications to the high-requirements job for both genders, but it did not affect the gender distribution of applicants, as both genders applied more to the high-requirements job with a symmetric increase.

3.4 Regression Analysis: Probability of applying for High-Paying, High-Requirements Job

We conducted two probit regressions, predicting the probability of applying to the high-paying, high-requirements job by EEO presence and gender (Model 1), as well as their interaction term (EEO presence x

gender; Model 2). In both models, we controlled for socio-demographic characteristics (education, age), risk attitude, country, and selected resume (Table 3)[8].

Model 1 results indicate that gender did not predict the likelihood of applying for the high-paying job ($p > 0.1$). We found a significant positive effect on the likelihood of applying for a high-paying job when EEO statements were present in both job posts ($p < 0.001$), and when they were present only in the high-paying job post ($p = 0.005$). Interestingly, the presence of EEO statements only in the low-paying job posts did not have a significant impact ($p > 0.1$).

Model 1 provides context to our analysis. However, the crux of our investigation lies in Model 2, where we specifically examine how the presence of EEO statements impacts men and women differently. In Model 2, we introduce the interaction between gender and EEO statements, and we explore these results more deeply using the marginal effects (Table 4).

Participants' risk attitude positively affects the likelihood of applying to a high-paying job ($p < 0.001$)[9]. Thus, the most risk-embracing participants are more likely to apply for the higher-paying job[10]. Unexpectedly, having selected the CCBI resume also increases this likelihood ($p = 0.037$)[11]. We also found some country effects, in that living in Spain is associated with a higher propensity to apply for the high-paying job compared to Germany ($p < 0.001$). We have conducted a more in-depth analysis of country differences in treatment effects (see Appendix C). This analysis confirms that the effectiveness of EEO statements is country dependent.

Table 3: Equal Employment Opportunity statements effect on Job application

Dependent Variable	**High paying job**	
Model	(1)	(2)
EEOboth	0.205***	0.180*
	(0.0581)	(0.0818)
EEOHigh	0.165**	0.00995
	(0.0580)	(0.0816)
EEOLow	-0.0204	0.124
	(0.0580)	(0.0818)
Female	0.0459	0.0290
	(0.0437)	(0.0786)

[8]Similar probit regression analyses that account for relevant post-experimental questionnaire variables can be found in Appendix B.

[9]On average, women in our sample reported higher risk aversion ($M = 5.975$, $SD = 0.060$) than men did ($M = 6.276$, $SD = 0.057$), $t = 3.635$, $p < 0.001$.

[10]An additional robustness check focusing on the role of risk aversion in job application behavior is presented in Appendix G. While risk aversion affects the probability of applying to the higher-paying job, risk aversion alone does not explain the observed gender differences in job application behavior in response to EEO statements.

[11]Starting from the assumption that we designed the two resumes to be qualitatively similar, the unexpected significance of the coefficient of CCBI is puzzling. Further analysis in Appendix E show that the interactions between CCBI and treatment conditions are not statistically significant. In addition, we only find a significant marginal effect in the control treatment for women; all the other marginal effects are not significant. Thus, this does not undermine the main results and conclusions of this paper.

EEOboth × Female		0.0499
		(0.116)
EEOHigh × Female		0.317**
		(0.116)
EEOLow × Female		-0.284*
		(0.116)
Age	1.82e-05	0.000535
	(0.00180)	(0.00180)
Mid education	0.000548	-0.00156
	(0.0570)	(0.0572)
High education	0.0679	0.0680
	(0.0555)	(0.0557)
Unemployed	-0.0696	-0.0652
	(0.0575)	(0.0577)
Retired	-0.0707	-0.0813
	(0.0673)	(0.0675)
Risk Attitude	0.0964***	0.0961***
	(0.00912)	(0.00915)
CCBI	0.0912*	0.0895*
	(0.0429)	(0.0430)
Spain	0.176***	0.180***
	(0.0486)	(0.0487)
Constant	-0.717***	-0.731***
	(0.120)	(0.124)
N	3736	3736
Pseudo R2	0.0461	0.0508

Note: In the table, statistical significance is denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. The analysis was conducted using a probit regression model with robust standard errors. Standard errors are reported in parentheses. Model 1 only assesses the effect of EEO presence and gender, while Model 2 includes interaction terms between EEO presence and gender. The dependent variable equals 1 if the participant selected the high-paying position (position gamma), and 0 if they selected the low-paying position (position Beta). EEOboth, EEOHigh and EEOLow are treatment dummies, and their coefficients express the effect compared to the control group (Control) where no Equal Employment Opportunity (EEO) statements were present. The 'Female' variable is a gender dummy, equal to 1 if the participant is a woman, and 0 if the participant is a man. 'Risk attitude' is a measure of self-reported risk, with higher values indicating participants who are more risk-loving. 'CCBI' is a dummy variable equal to 1 if the participant chose the CCBI resume in the first task (the resume choice task), and 0 if they chose the Delta resume. 'Spain' is a country dummy equal to 1 for Spanish respondents and 0 for German respondents. The analysis also controls for age, education, and employment status. 'Low education' and 'employed' are the reference levels for education and employment status, respectively.

Table 4 presents the average marginal effects of the different treatments derived from Model 2, broken down by gender. These marginal effects illustrate the change in the predicted probability of an individual applying to the high-paying job for each treatment, relative to the control group (absence of EEO statements). For women, the presence of EEO statements on both job posts increases the probability of applying to the

high-paying job by 8.6 percentage points ($p = 0.005$). The presence of EEO statements solely on the high-paying job increases the likelihood of women applying to that job by 12 percentage points ($p < 0.001$) compared to the control condition. Having an EEO statement only on the low-paying position only marginally affects women's probability of applying to the high-paying position ($p = 0.052$)

For men, the presence of Equal Employment Opportunity (EEO) statements on both job posts increased the probability of applying to the high-paying job by 6.8 percentage points ($p = 0.027$). However, the presence of EEO statements solely on the high-paying job did not significantly alter the likelihood of men applying to the high-paying job ($p > 0.1$). Similarly, when EEO statements were only on the low-paying job, the effect was not significant ($p > 0.1$).

The effect of having EEO statements on both job positions on the probability of applying to the high-paying position does not differ between men and women ($p > 0.1$). On the other hand, the effect of having a statement only on the high-paying job is significantly larger for women than for men ($p = 0.007$). Finally, while, as previously mentioned, having the EEO statements on the low-paying position only does not affect the probability of applying for the high-paying job for either men or women, the differential effect is statistically significant ($p = 0.014$).

Table 4: Average marginal effects of EEO statements by gender

	Women	**Men**	**Difference (Women- Men)**
EEOboth	0.086**	0.068*	0.018
	(0.0308)	(0.0308)	(0.0436)
EEOHigh	0.121***	0.004	0.118**
	(0.0305)	(0.0311)	(0.0436)
EEOLow	-0.061	0.047	-0.108*
	(0.0314)	(0.0314)	(0.0441)

Note: The marginal effects in this table are derived from the probit regression model specified in Model 2, which includes interaction terms between treatments and gender and controls for age, education, risk attitude, resume choice and country effects. These are average marginal effects calculated for a representative man and woman. Standard errors are reported in parentheses. The statistical significance is denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Figure 2 visualises the predicted probability of applying to the high-paying position by treatment conditions and gender.

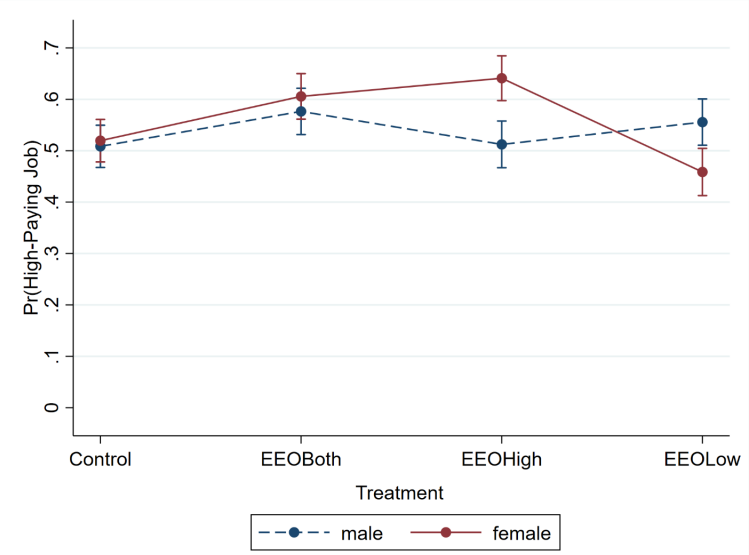


Figure 2: Predicted probability of applying to the high-paying job by gender and treatment condition: Interaction effect between gender and treatment conditions

3.5 Behavioural drivers

At the end of the main task, we included several self-report measures regarding participants' experience during the experiment (see Methods). More specifically, in addition to assessing their risk propensity, we assessed their self-reported perceived fairness of the hiring process, their confidence in their application's success, the gender distribution they expected among employees, and the extent to which the presence of the EEO statement encouraged/discouraged their application to a higher-paying position.

In Figure 3, we report participants' answers to whether they noticed the EEO statements in the three treatments where the statements were effectively present[12].

[12]Robustness check of the main analysis on job application/treatment effects and gender, controlling for whether individuals noticed the statements or not, is reported in Appendix F. The direction of the effects and the significance are in line with our results in the results section.

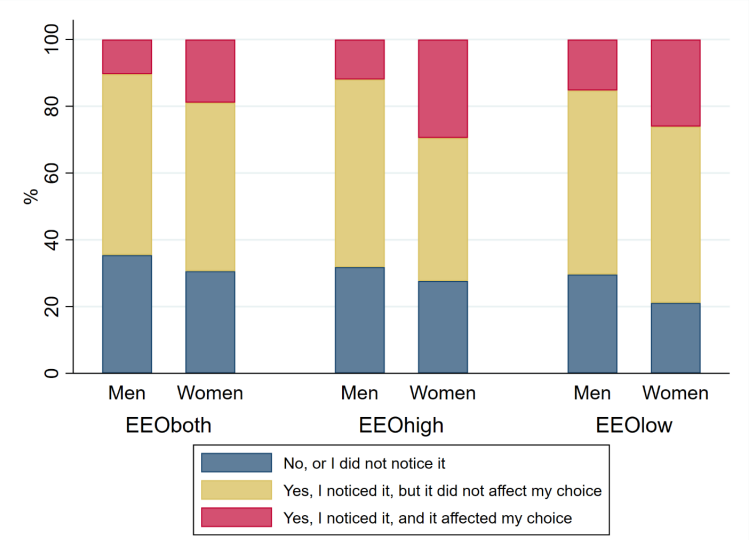


Figure 3: The proportion of participants that reported noticing the statement by gender and treatment condition

In all three treatment conditions, there is a consistent pattern where a higher proportion of female participants noticed the EEO statement and considered it more influential in their application process than male participants did. The differences in the distribution of responses between men and women are significant in all treatments ($\chi^2 = 13.1634$, $p < 0.01$ for EEOBoth; $\chi^2 = 41.9864$, $p < 0.001$ for EEOHigh; and $\chi^2 = 18.9688$, $p < 0.001$ for EEOLow)[13]. This is in line with our results, which show that men are generally unaffected by the presence of Equal Employment Opportunity statements, at least in EEOHigh and EEOLow.

Additional questions about fairness, confidence, and expected workforce gender distribution were asked only to individuals who noticed the presence of the statement in the experimental conditions ($N = 1875$).

We tested for gender differences in the distribution of responses regarding fairness perception, confidence in the application and expected gender distribution[14] in the workplace, for each treatment condition and by the type of job the participant applied for using the Mann-Whitney test. We found no gender differences in terms of confidence and gender distribution expectations. However, we did find significant gender differences in the distribution of perceived fairness for participants in the EEOlow treatment who applied to the low-paying position, as well as for those in EEOhigh who applied to the high-paying job. In neither case did men and women differ in terms of median, but women who applied for the low-paying job in the EEOlow treatment reported slightly higher mean fairness than the men in the same group ($z = -2.301$, $p = 0.021$). At the same time, women who applied for the high-paying job in the EEOhigh treatment reported a higher

[13]P-values corrected for multiple testing are provided in Appendix D.

[14]Fairness was assessed on a scale from 0 (unfair) to 10 (fair). Confidence was self-reported on a scale where 0 represents decreased confidence and 10 signifies increased confidence. For expected gender distribution participants expressed their level of agreement with the statement, "It made me expect that there would be an equal gender distribution among employees," on a 5-point scale, ranging from Strongly Disagree (1) to Strongly Agree (5).

perceived fairness than their male counterparts ($z = -1.967$, $p = 0.049$). However, none of these results is robust to multiple error correction (see Appendix D to observe all the tests performed and the multiple testing correction).

To those participants who not only noticed the statement but also reported that the statement influenced their decision ($N = 496$), we also asked whether the statement encouraged or discouraged their application on a scale from 1 to 5, where 1 is encouraged and 5 is discouraged. In the EEOHigh treatment, we find that there is a significant difference in responses between men and women ($z = 3.464$, $p < 0.001$). For women, the median is 2, which means that they are encouraged to apply to the high-paying job, while for men the median is 3, which corresponds to being neither encouraged nor discouraged from applying to the high-paying job. The distribution of responses between men and women is also statistically different in the EEOLow treatment ($z = -3.675$, $p < 0.001$). Even though the median for both genders is 3, meaning no reported discouragement/encouragement effect, the average is slightly lower for women. No gender difference is found in EEOBoth[15].

In Figure 4, the focus shifts to participants' perceptions regarding the gender of the HR executive with whom they were paired during the study. There is a significant difference in the perception of the HR executive's gender based on the participant's gender, both for those who got the job ($\chi^2 = 164.0735$, $p < 0.001$) and for those who did not ($\chi^2 = 35.7294$, $p < 0.001$). The perception of the HR executive's gender also varied significantly depending on whether the participants were hired, and this is true for both men ($\chi^2 = 16.0070$, $p < 0.001$) and women ($\chi^2 = 81.7663$, $p < 0.001$). As observed in the figure, among the women who were hired, nearly half of them (48.83%)[16] perceived the HR executive to be female, compared to 39.76% of women who were not hired. The results seem to suggest that women believe there might be a gender bias in the hiring process and that their hiring is more likely in the presence of a female HR executive. The picture for men is less clear, at least in terms of probable interpretation. From the graph, we notice that men who got hired are more likely to report not knowing the gender of the HR executive, with 43.44% indicating 'I don't know', compared to those who didn't get hired, who reported 'I don't know' at a rate of 34.35%.

[15]In Appendix E we discuss the significance of the name of the resume in more detail and report the results of the robustness check.

[16]P-values corrected for multiple testing are provided in Appendix D.

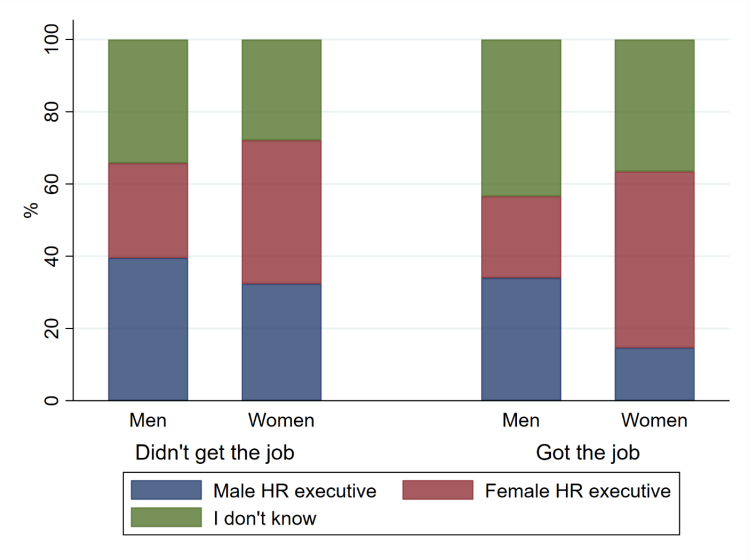


Figure 4: Participants Expectation of HR Executive Gender Matching

4 Discussion

We conducted an online experiment in two countries to investigate the potential of EEO statements to effectively influence job applications and, consequently, their potential to reduce the gender employment gap, particularly in high-paying, high-qualification requirements positions. The study revealed a gender-specific impact of including EEO statements in job postings, and the effect of these statements was contingent on whether the statement appeared on one or both job postings. We found that when EEO statements were placed on higher-paying, high-requirements positions, EEO statements encouraged women to apply to the better-paid position. This was also the case when the EEO statement was placed on both the lower- and higher-paying positions. Overall, the presence of EEO statements did not discourage men from applying. Interestingly, the presence of EEO statements on all positions encouraged both men and women to apply to the high-requirements and better-paid position, increasing the overall number of applications without affecting the gender distribution. A significant increase in the number of applications is also observed when the statement is placed exclusively on the high-paying position, but in this latter case, the gender distribution is affected in favour of women.

This paper contributes to the ongoing scientific discourse regarding the value of incorporating EEO statements and other explicit encouragements for underrepresented groups to apply for job postings [McNab and Johnston, 2002, Leibbrandt and List, 2018, Flory et al., 2021]. We focused on statements explicitly targeting women, setting our work apart from studies that address broader claims about diversity and equality in the workplace. Our study goes beyond typical approaches of attractiveness ratings of organizations and

investigates the decision-making process of job applications in the context of an incentivized experiment, using a non-student sample. This allows us to conclude the impact of EEO statements on actual choices to apply for one job over another. Methodologically, a notable strength of our research is its implementation in two countries using a large sample designed to approach representativeness in terms of age and gender. This approach results in stronger evidence and better generalizability of our findings. Even though, the difference in gross income between men and women are significantly different in Germany and Spain, with the difference in Germany being above and in Spain below the EU average, the percentage of women in top leadership positions in both countries is below the European average Applying actions that support women to apply for these positions may thus be needed. Our findings suggest that including EEO statements can be a powerful tool to encourage women to apply for job postings in these countries. The results also suggest that men do not feel discouraged to apply even if EEOs are specifically targeted at women.

Based on the premise that EEO statements should target qualified women who might otherwise hesitate to apply to high-requirements, competitive positions, two scenarios emerge as optimal: placing EEO statements exclusively on high-paying, high-requirements job postings, or including them on all positions. In both cases, we observe a significant and positive impact on women, while men are not dissuaded from applying to the high-paying position in the first case and are even encouraged to do so in the latter. Our observations indicate that when the EEO statement is included in both types of job postings, both women and men show an equal propensity to apply for the high-requirements, higher-paying positions, thus not impacting the gender ratio in applications. In a competitive labour market, where employers vie for the best candidates, EEO statements can serve as an effective tool to attract diverse job applicants. This may incentivize high-paying employers to use such statements as a signal of their commitment to diversity. This strategy would be particularly appealing because including EEO statements on high-paying job postings (in high-paying or on both high- and low-paying job postings) increases the overall number of applications received.

Based on our results, we can make the following policy recommendations: first, encourage organisations to include EEO statements in high-paying job advertisements where women are under-represented. This approach would increase the number of female applicants without deterring male candidates. Second, promote the inclusion of the statement across all jobs postings. This strategy increases overall application numbers while maintaining a balanced distribution of applicants. Furthermore, raising awareness on the use and placement of EEO statements may empower organisations to work towards promoting gender diversity, creating a more equitable workforce, and in turn closing the employment and pay gap.

While this study provides valuable insights into the potential of EEO statements, several limitations should be acknowledged. First, in favour of experimental control, we simplified the job application process by using standardised fictitious CVs and asking participants to select one of two jobs to apply to. Even though

this provided us with more experimental control, this approach may not fully capture the complexities of real-world employment practices. Linked to this, it is important to note that, differently from Fluchtmann et al. [2024] and Lochner and Merkl [2022] but in line with Barbulescu and Bidwell [2013], we found no significant gender difference in application decisions within the control treatment. This suggests that, in our sample, men did not apply to the high-requirements job more often than women. A possible explanation could be related to the use of a fictitious resume, which may have mitigated the effect of (over-)confidence, thus reducing the gender differences in application behaviour that are commonly reported in the literature. In addition, recent studies have highlighted that gender differences can be context and task-dependent [Filippin and Crosetto, 2016, Filippin, 2022]. Notably, a similar result, meaning no gender difference in application intention while using a fictitious resume, is found in Salwender and Stahlberg [2024]. Not finding any gender differences in the control condition may thus not be inconsistent but reflect the specific conditions and context of our study.

Second, the effectiveness of EEO statements may vary across countries. We found that the effects of EEO statements were more pronounced in Spain than in Germany. This effect was partly explained by socio-demographic differences in our sample, in that our German sample included a higher number of retired individuals and EEO statements did not show any effect in this group. Based on our data, we cannot draw any final conclusions regarding the reasons for country differences of the effect of EEO statements on job applications. However, differences may also be based on societal, cultural and institutional differences in both countries. As gender equality seems somewhat higher in Spain than in Germany according to EIGE's Gender Equality Index, with Spain scoring 76.4 and Germany 70.8 (EU overall: 70.2), there may be, for example, more susceptibility to affirmative action measures in Spain than in Germany[17]. Notably however, for people in the workforce in both countries, women were encouraged, and men were not discouraged by the presence of EEO statements on high-paying positions, highlighting the robustness of this effect.

Future studies could further investigate the nuances of different wordings of EEO statements, which may influence their effectiveness. A logical next step would be to explore the impact of EEO statements in real-world job application scenarios, using actual job postings and real applicants. Additionally, research should be expanded to investigate the effectiveness of EEO statements in different job contexts. The impact of such statements may vary across industries and job roles, and exploring these variations can inform tailored diversity and inclusion strategies for different sectors.

[17]The Gender Equality Index assigns scores from 1 to 100, with 100 indicating full gender equality between women and men.

Acknowledgments

This study was conducted with the internal funding of the Joint Research Centre (JRC) of the European Commission.

References

Lisa Abraham, Johannes Hallermeier, and Alison Stein. Words matter: Experimental evidence from job applications. *Journal of Economic Behavior & Organization*, 225:348–391, 2024.

Thilini Alahakoon, Amanda Beatson, Byron Keating, Frank Mathmann, Gary Mortimer, and Asha Worsteling. Diversity, equity and inclusion statements in recruitment materials: A systematic review and research agenda. *Australasian Marketing Journal*, 2024. 14413582241255680.

Loukas Balafoutas and Matthias Sutter. Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science*, 335:579–582, 2012.

Roxana Barbulescu and Matthew Bidwell. Do women choose different jobs from men? mechanisms of application segregation in the market for managerial workers. *Organization Science*, 24(3):737–756, 2013.

Barbara Biasi and Heather Sarsons. Flexible wages, bargaining, and the gender gap. *The Quarterly Journal of Economics*, 137(1):215–266, 2022.

Pedro Bordalo, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. Beliefs about gender. *American Economic Review*, 109(3):739–773, 2019.

Thomas Buser, Muriel Niederle, and Hessel Oosterbeek. Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447, 2014.

Katherine B. Coffman, Michelle R. Collis, and Leena Kulkarni. Whether to apply. *Management Science*, 70 (7):4649–4669, 2024.

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

Annalisa Delfino. Breaking gender barriers: Experimental evidence on men in pink-collar jobs. *American Economic Review*, 114(6):1816–1853, 2024.

Alice H. Eagly and Linda L. Carli. *Through the Labyrinth: The Truth About How Women Become Leaders*. Harvard Business School Press, 2007.

European Commission. Gender Pay Gap in the EU remains at 13% on Equal Pay Day. Statement, Nov 14, 2023, Brussels (STATEMENT/23/5692), 2023. URL https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT/23/5692. [Online; accessed 15 November 2023].

European Institute for Gender Equality. Gender statistics database – largest listed companies: Ceos, executives and non-executives, 2023. Retrieved December 14, 2023.

European Parliament and the Council of the European Union. Directive (EU) 2022/2381 of the European Parliament and of the Council of 23 November 2022 on improving the gender balance among directors of listed companies and related measures. OJ L 315, 7.12.2022, p. 44–58, 2022.

Eurostat. Gender pay gap statistics. European Commission, 2022. Retrieved November 24, 2024, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Gender_pay_gap_statistics.

Christine L. Exley and Judd B. Kessler. The gender gap in self-promotion. *The Quarterly Journal of Economics*, 137(3):1345–1381, 2022.

Antonio Filippin. Gender differences in risk attitudes. *IZA World of Labor*, 2022.

Antonio Filippin and Paolo Crosetto. A reconsideration of gender differences in risk attitudes. *Management Science*, 62(11):3138–3160, 2016.

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

Jeffrey A. Flory, Andreas Leibbrandt, Caroline Rott, and Olga Stoddard. Increasing workplace diversity: Evidence from a recruiting experiment at a fortune 500 company. *Journal of Human Resources*, 56(1): 73–92, 2021.

Jeffrey A. Flory, Andreas Leibbrandt, Caroline Rott, and Olga Stoddard. Leader signals and "growth mindset": A natural field experiment in attracting minorities to high-profile positions. *Management Science*, 70(8):4953–4973, 2024.

Jonas Fluchtmann, Anita M Glenny, Nikolaj A Harmon, and Jonas Maibom. The gender application gap: Do men and women apply for the same jobs? *American Economic Journal: Economic Policy*, 16(2): 182–219, 2024.

Isabelle Haegele. The broken rung: Gender and the leadership gap. *arXiv preprint arXiv:2404.07750*, 2024.

Alexandra J. Heath, Mikael Carlsson, and Jens Agerström. What adds to job ads? the impact of equality and diversity information on organizational attraction in minority and majority ethnic groups. *Journal of Occupational and Organizational Psychology*, 96(4):872–896, 2023.

Marcela Ibañez and Gerhard Riener. Sorting through affirmative action: Three field experiments in colombia. *Journal of Labor Economics*, 36(2):437–478, 2018.

Kristin Kanthak and Jonathan Woon. Women don't run? election aversion and candidate entry. *American Journal of Political Science*, 59(3):595–612, 2015.

Amanda Klysing, Emma A. Renström, Marie Gustafsson-Sendén, and Anna Lindqvist. Gender diversity in recruitment: Influence of gender trouble on applicant attraction and evaluation. *Journal of Applied Social Psychology*, 52(8):781–802, 2022.

Andreas Leibbrandt and John A. List. Do equal employment opportunity statements backfire? evidence from a natural field experiment on job-entry decisions. Technical Report 25035, National Bureau of Economic Research, 2018.

Andreas Leibbrandt, Lian Chaw Wang, and Charmaine Foo. Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women. *Management Science*, 64(8):3501–3516, 2018.

Benjamin Lochner and Christian Merkl. Gender-specific application behavior, matching, and the residual gender earnings gap. Technical Report 22/2022, IAB-Discussion Paper, 2022.

David A. Matsa and Amalia R. Miller. A female style in corporate leadership? evidence from quotas. *American Economic Journal: Applied Economics*, 5(3):136–169, 2013.

Sophie M. McNab and Lucy Johnston. The impact of equal employment opportunity statements in job advertisements on applicants' perceptions of organizations. *Australian Journal of Psychology*, 54(2):105–109, 2002.

Markus M. Möbius, Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. Managing self-confidence: Theory and experimental evidence. *Management Science*, 68(11):7793–7817, 2022.

Christa Nater and Sabine Sczesny. Affirmative action policies in job advertisements for leadership positions: How they affect women's and men's inclination to apply. *European Journal of Social Psychology*, 46(7): 891–902, 2016.

Muriel Niederle and Lise Vesterlund. Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101, 2007.

Muriel Niederle and Lise Vesterlund. Gender and competition. *Annual Review of Economics*, 3(1):601–630, 2011.

Muriel Niederle, Carmit Segal, and Lise Vesterlund. How costly is diversity? affirmative action in light of gender differences in competitiveness. *Management Science*, 59(1):1–16, 2013.

Cheryl Rice. How mckinsey's story became sheryl sandberg's statistic-and why it didn't deserve to. Huffington Post, 2014.

Mira Salwender and Dagmar Stahlberg. Do women only apply when they are 100% qualified, whereas men already apply when they are 60% qualified? *European Journal of Social Psychology*, 2024.

Anya Samek. Gender differences in job entry decisions: A university-wide field experiment. *Management Science*, 65(7):3272–3281, 2019.

Sheryl Sandberg. *Lean In: Women, Work, and the Will to Lead*. Knopf, 2015.

Heather Sarsons and Guo Xu. Confidence men? evidence on confidence and gender among top economists. In *AEA Papers and Proceedings*, volume 111, pages 65–68. American Economic Association, 2021.

H. Jack Walker, Hubert S. Feild, William F. Giles, Achilles A. Armenakis, and Jeremy B. Bernerth. Displaying employee testimonials on recruitment web sites: effects of communication media, employee race, and job seeker race on organizational attraction and information credibility. *Journal of Applied Psychology*, 94(5):1354, 2009.

H. Jack Walker, Hubert S. Feild, Jeremy B. Bernerth, and J. Bret Becton. Diversity cues on recruitment websites: Investigating the effects on job seekers' information processing. *Journal of Applied Psychology*, 97(1):214, 2012.

Appendix S1

Table S1: Sample Size and Demographic characteristics by Country

	Country		
	Spain (N=1860)	**Germany (N=1876)**	**Total**
Treatment Condition			
Control	540	541	1,081
	29.03%	28.84%	28.93%
EEOBoth	441	446	887
	23.71%	23.77%	23.74%
EEOHigh	439	445	884
	23.60%	23.72%	23.66%
EEOLow	440	444	884
	23.66%	23.67%	23.66%
Gender			
Men	931	935	1,866
	50.05%	49.84%	49.95%
Women	929	941	1,870
	49.95%	50.16%	50.05%
Age	47.57	49.04	48.31
	(15.57)	(16.69)	(16.15)
Education Level[1]			
Low education	225	892	1,117
	12.10%	47.55%	29.90%
Middle education	673	479	1,152
	36.18%	25.53%	30.84%
High education	962	505	1,467
	51.72%	26.92%	39.27%
Employment status			
Employed	1,108	1,059	2,167
	59.57%	56.45%	58.00%
Unemployed	407	316	723
	21.88%	16.84%	19.35%
Retired	345	501	846
	18.55%	26.71%	22.64%

[1]Low education corresponds to ISCED 0-2, namely Grund-/Hauptschulabschluss and Mittlerer Schulabschluss (Realschulabschluss/Mittlere Reife oder Vergleichbares) in Germany, and Educación Primaria and Educación Secundaria (ESO, o equivalente) in Spain. Medium education corresponds to ISCED 3-4, namely Höherer Schulabschluss (Abitur, Fachhochschulreife oder Vergleichbares) in Germany, and Educación Secundaria (Selectividad, FP, o equivalente) in Spain. High education corresponds to ISCED 5-8, namely Erster akademischer Studienabschluss, Universität, and Postgraduiertenabschluss, Universität (z. B. MSc, Promotion) in Germany, and Universidad (diplomatura/licenciatura/grado) and Estudios de Posgrado (p. ej. Máster, Doctorado) in Spain.

Appendix S2

S2.1 Job applicant survey

Section 1

Welcome! This study is being conducted by The Behaviouralist on behalf of the Joint Research Centre.

The purpose of this study is to better understand job recruitment dynamics.

The survey should take around 15 – 20 minutes to complete. All answers will be anonymised so you will not be identified. By clicking the button below, you acknowledge that your participation in the study is voluntary and that you are aware that you can end your participation in the study at any time and for any reason. If you decide to withdraw from this survey, we will delete your answers. Click here for more information on how we will use your data.

You will, in addition to receiving your regular survey completion compensation of 3 €, receive a variable payment for completing the tasks in the study. The variable payment depends on the decision that you make in this study and on the decisions of other participants. You will be compensated only if you complete the survey in full. The variable payment will be distributed within the next 2 weeks.

It's important that you pay attention to the questions and that you complete the survey in one go. To ensure that the data is of high quality, we have included a set of attention checks and we will be tracking the time you spend completing the survey. Participants that complete the survey either too fast or too slow or that fail to respond to the attention checks correctly will be excluded. Excluded participants will not receive the variable payment.

Remember, participants who complete the survey too fast or too slow or who do not respond correctly to the attention checks will be excluded. Excluded participants will not receive the variable payment. Please take your time, don't rush.

Section 2

Before starting with the study we would like to ask you some socio-demographic questions.

Section 3

What year were you born?

Section 4

What is the highest level of education that you have achieved?

- Primary school
- Secondary school (GCSE, O-level, AS level, or equivalent)
- Secondary school (A-level, BTEC, or equivalent)
- Undergraduate degree, university
- Postgraduate degree, university (e.g. MSc, PhD)

Section 5

What is your gender?

- Male
- Female
- Other

Thank you for your time spent taking this survey. We have reached the quota for this profile.

Section 6

What is your current employment status?

- Employed full time (36 or more hours per week)
- Employed part-time (up to 35 hours per week)
- Unemployed and currently looking for work
- Unemployed and not currently looking for work
- Furloughed
- Student
- Retired
- Homemaker
- Self-employed
- Unable to work

Section 7

What sector do you work in?

- Construction
- Transport
- Industry
- Agriculture
- Public administration and defence

- Financial services
- Other services
- Commerce and hospitality
- Education
- Health
- Other

Section 8

Which of the following best describes your role in your occupation?

- Upper Management
- Middle Management
- Junior Management
- Administrative Staff
- Support Staff
- Trained professional
- Researcher
- Skilled Labourer
- Consultant
- Temporary Employee
- Trainee
- Other

Section 9

Thank you for your answers.

The study is about to begin.

Section 10

This study requires your complete, undistracted attention.

Please remove any distractions that could interfere during the study (e.g. turn off email notifications, chat windows).

We ask that you please read these instructions carefully.

You will only be able to proceed with the study if you understand the instructions.

You can earn tokens during the study, which we will automatically convert to euros at the following rate:

10 tokens = 0.40 €

Respondents who complete this survey will, on average, earn tokens worth 4 €. The maximum amount that you can earn is 6.80 €.

You will know your payoff immediately at the end of the study.

Section 11

In this survey, you will be asked to undertake some tasks.

Participants to this survey can be assigned to one of two roles:

- Job applicant who is looking for a new job as a manager
- HR Executive who evaluates job applications

There are other real people undertaking each of these roles. For example, if you are assigned to the role of a job applicant, you will be matched with another participant that has been assigned the role of an HR executive.

Click 'Next' to know which role you have been assigned to.

Section 12

Your role: Job applicant

You are a job applicant and you are looking for a new job as a manager.

We will ask you to put together a job application. You will have to choose:

1. A resume
2. A job advertisement
3. A statement describing your management style

These 3 choices constitute your job application.

A fictitious name based on your gender will appear on your application. This will ensure that your privacy is protected.

Section 13

Your job application will be evaluated by an HR Executive who decides whether you get the job or not.

The HR Executive will compare two job applications (yours and another) and can only hire one applicant. The HR executive is instructed to select the applicant they think would be most suitable for the job if it was advertised in the real world.

The amount that you can earn for this task depends on the job advertisement that you choose and on the decision the HR executive makes (i.e. whether the HR executive decides to hire you or not).

Section 14

Before choosing which job advertisement to apply for, we ask you to select a resume.

The resume will be shown to the HR Executive who reviews your job application, along with the particular job advertisement and management style you will choose. A fictitious name will automatically appear on the resume.

You will be asked to select a job advertisement and a statement describing your management style later in the survey.

Section 15

Before applying to a job advertisement, you have to choose a resume. You can choose between the two resumes below.

Once you've chosen a resume, you will be informed about the requirements needed for two different job vacancies.

Resume A

- Manager at CCBI Ltd. for 5 years
- Managed a team of 8 business experts
- Master in Business Administration
- Proficient English (Oxford University – level B2)
- Project management certificate level B

Resume B

- Manager at Delta Ltd. for 5 years
- Managed a team of 8 market analysts
- Master in Management
- Proficient English (Cambridge University – level B2)
- Project management certificate level B

Section 16

Your job application consisting of your resume (which you have already chosen), the job advertisement, a statement describing your management style and a fictitious name are shown to the HR Executive.

The HR Executive decides whether to hire you or not. The HR Executive receives two applications and can hire only one applicant.

If you get the job, you receive the wage displayed in the job advertisement. If you do not get the job, you receive 0 tokens.

Section 17

Before choosing your job, we will ask you to respond to 3 questions to make sure you understand the study.

Please, be reassured that giving wrong answers will not jeopardize your participation in the study. Mistakes are normal and accepted, and you will receive explanations for each question.

Section 18

Question 1: How many tokens do you earn if you do not get the job?

- 50 tokens
- 150 tokens
- 0 tokens

Correct!

If you do not get the job, you will earn 0 tokens.

Please press 'Next' to continue.

Incorrect

If you do not get the job, you will earn 0 tokens.

Please select the correct answer to continue.

Section 19

Question 2: The HR Executive decides whether or not to grant the job to you or to another job applicant.

- Yes
- No

Correct!

The HR Executive decides between you and the other job applicant.

Please press 'Next' to continue.

Incorrect

The HR Executive decides between you and the other job applicant.

Please select the correct answer to continue.

Section 20

Question 3: How much do you earn if you get the job?

- The HR Executive will determine how much I will earn
- 100 tokens
- I will earn the wage displayed in the job advertisement

Correct!

If you get the job, you receive the wage displayed in the job advertisement.

Please press 'Next' to continue.

Incorrect

If you get the job, you receive the wage displayed in the job advertisement.

Please select the correct answer to continue.

Section 21

Your resume.

Please select the job you wish to apply for:

Manager position at Gamma Ltd. (150 tokens)

- Ideally, you have between 5 to 8 years of experience, with previous experience in managing medium-size teams (8 people or more).
- Preferably, you have a PhD, but a relevant master is accepted.
- We expect good knowledge of English, a second language would be an asset.
- A project manager certificate with level B is needed.
- The average wage for this type of job is 150 tokens.
- Gamma Ltd. values a diverse workplace and strongly encourages women to apply

Manager position at Beta Ltd. (50 tokens)

- Ideally, you have at least 3 years of experience, with previous experience in managing small-medium size teams (at least 5 people).
- You have a relevant master.
- We expect good knowledge of English.
- A project manager certificate with level B is needed.
- The average wage for this type of job is 50 tokens.
- Beta Ltd. values a diverse workplace and strongly encourages women to apply

Pop-up: *Do you confirm that you want to select job Gamma/Beta?*

Section 22

You can now choose a statement that best describes your management style. You can choose between two statements:

Remember: The HR Executive will be shown the statement that you choose, together with the resume, and will determine whether you are suitable for the job (or whether they prefer to grant the job to the other

applicant).

Management style A

I am a mature, dynamic and result-oriented manager. I make decisions easily and maintain calm in stressful situations. I am straightforward and uncomplicated in maintaining relations.

Management style B

I am a dedicated, respectful and persistent manager. I am responsible for the management decisions, but I find it important to involve my staff and other relevant persons in order to reach our joint objectives.

Pop-up: *Do you confirm that you want to select management style A/B?*

Section 23

To ensure that this study remains anonymous, a fictitious name will appear on your resume. This is based on the gender you selected at the beginning of the study.

This is your fictitious name:

- Julie Smith
- Sally Brown
- Mark Taylor
- Robert Johns

Section 24

Before knowing the outcome of your application, we would like to ask you some questions.

Section 25

Which job did you apply to?

- Beta, advertised wage 50 tokens
- Gamma, advertised wage 150 tokens

Section 26 - Question asked only to individuals who chose Job Gamma (High-paying Job)

You applied for the job that is highlighted below:

Beta/Gamma

Please explain why you applied to that role?

Among the options listed below, select those that have most influenced your decision.

- I was overqualified for the other job (Beta)
- I was not worried about the prospect of earning 0 tokens, so I took the risk

- It was the job with the highest wage
- I thought not many others would apply for this job

Section 27 - Question asked only to individuals who chose Job Beta (Low-paying Job)

You applied for the job that is highlighted below:

Beta/Gamma

Please explain why you applied to that role?

Among the options listed below, select those that have most influenced your decision.

- I was underqualified for the other job (Gamma)
- With my profile it would have been unprofessional to apply to the other job
- With my profile it would have been presumptuous to apply to the other job
- I did not want to risk to not get hired and earn 0 tokens
- I thought not many others would apply for this job

You stated that you applied for the job Gamma/Beta for the reasons listed below.

Please rank these options in order of importance. The option ranked 1 being the most important one.

Section 28

We will now show you the two statements that you could send along with your application.

[Management style statement A]

In your opinion, this statement has been selected more often by:

You will earn 10 tokens if your guess is correct.

- Men
- Women
- The same

[Management style statement B]

In your opinion, this statement has been selected more often by:

You will earn 10 tokens if your guess is correct.

- Men
- Women

- The same

Section 29

Was this statement included in at least one of the two jobs displayed to you?

- No, or I did not notice it
- Yes, I noticed it, but it did not affect my choice.
- Yes, I noticed it, and it affected my choice

Section 30

What was your reaction to the statement?

Please complete the sentences below by moving the slider towards the options that apply to you by the extent you see fit.

- It made me think that my application was treated:
    - [unfairly, fairly]
- It affected my confidence in my application
    - [decreased confidence, increased confidence]

Please indicate the extent to which you agree with the following statement:

[Strongly disagree, Disagree, Neither agree or disagree, Agree, Strongly agree]

- It made me expect that there would be an equal gender distribution among employees.

Section 31

How did the statement affect your choice?

- It encouraged me to choose the higher paying job
- It somewhat encouraged me to choose the higher paying job
- It neither encouraged nor discouraged me to choose the higher paying job
- It somewhat discouraged me to choose the higher paying job
- It discouraged me to choose the higher paying job

Section 32

Do you believe the presence of this statement influenced the choices of HR executives?

- Yes, it made HR executives more likely to hire women, regardless of the quality of the application
- Yes, it made HR executives more likely to hire men, regardless of the quality of the application
- Yes, it made HR executives more likely to hire women when applications were similar

- Yes, it made HR executives more likely to hire men when applications were similar
- No, I do not think HR executives have been influenced by the statement

Section 33

HR executive has made the following decision:

You did not get the job you applied for

You got the job you applied for!

Section 34

You completed the main activity.

We will now ask you some additional questions about your experience in the study.

Section 35

"Gamma/Beta Ltd. values a diverse workplace and strongly encourages women to apply."

In your opinion, what would happen if this type of statement was introduced in job advertisements for leadership positions in real life?

[Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree]

- It would make women more likely to apply
- It would make men more likely to apply
- It would lead to a more equal number of female and male leaders
- It would increase the diversity among leaders
- It would make the company a more desirable employer
- I would be more likely to want to work at such a company
- It would generate conflict if a woman is hired
- It would generate conflict if a man is hired
- It would have positive effects on productivity
- It would have no effect because women who want to apply for a leadership position would do so anyways
- It would have no effect because these statements have no consequences

Section 36

Please indicate on the scale below how you see yourself: Are you, in general, a person who takes risks or do you try to evade risks?

- Not at all prepared to take risk
- Very much prepared to take risk

Section 37

Do you think the HR executive that you were matched with was:

- Men
- Female
- I don't know

Section 38

Did you get the job you applied for?

- Yes
- No

Section 39.a

Thank you for taking part in this study!

In addition to receiving your regular survey completion compensation of 3€, you will receive a variable payment.

These are your earnings:

${e://Field/earnings_tokens} tokens

Conversion rate:

10 tokens = 0.40 €

The variable payment will be distributed within the next 2 weeks.

Section 39.b

Thank you for taking part in this study!

You will receive a compensation of 3€ for completing this survey.

a. To ensure that the data is of high quality, we tracked the time you spent completing the survey. Since you completed the survey too fast (under 6 minutes), you will not be eligible to receive the additional variable payment.

b. To ensure that the data is of high quality, we tracked the time you spent completing the survey. Since you completed the survey too slowly (over 62 minutes), you will not be eligible to receive the additional variable payment.

c. To ensure that the data is of high quality, we included a set of attention checks. Since you failed the attention checks, you will not be eligible to receive the additional variable payment.

d. To ensure that the data is of high quality, we included a set of attention checks and tracked the time you spend completing the survey. Since you failed the attention checks and completed the survey too fast (under 6 minutes), you will not be eligible to receive the additional variable payment.

e. To ensure that the data is of high quality, we included a set of attention checks and tracked the time you spend completing the survey. Since you failed the attention checks and completed the survey too slowly (over 62 minutes), you will not be eligible to receive the additional variable payment.

S2.2 HR Executive survey

Section 1

Welcome! This study is being conducted by The Behaviouralist on behalf of the Joint Research Centre.

The purpose of this study is to better understand job recruitment dynamics.

The survey should take around 10 - 15 minutes to complete. All answers will be anonymised so you will not be able to be identified. By clicking the button below, you acknowledge that your participation in the study is voluntary and that you are aware that you can end your participation in the study at any time and for any reason. If you decide to withdraw from this survey, we will delete your answers. Click here for more information on how we will use your data.

You will receive a survey completion compensation of 2 €. You will be compensated only if you complete the survey in full.

It's important that you pay attention to the questions and that you complete the survey in one go. To ensure that the data is of high quality, we will be tracking the time you spend completing the survey.

Section 2

This study requires your complete, undistracted attention.

Please remove any distractions that could interfere during the study (e.g. turn off email notifications, chat windows)

We ask that you please read these instructions carefully.

You will only be able to proceed with the study if you understand the instructions.

You will earn 2 € for completing the study.

Your responses are valuable. Please make your decisions as you would in real life. The results of this study will be used to inform policy-making.

Section 3

In this survey, you will be asked to undertake some tasks.

Participants to this survey can be assigned to one of two roles:

- Job applicant who is an experienced manager looking for a new job
- HR Executive who evaluates job applications and hires an experienced manager

There are other real people undertaking each of these roles.

Click 'Next' to know which role you have been assigned to.

Section 4

Your role: HR Executive

You are an HR Executive and you must hire a new experienced manager for a position that has opened.

You will be presented with two job applicants, and you must decide which applicant to hire.

The job applications include:

1. A resume
2. A statement describing the applicant's management style

Section 5

Before deciding who you want to hire, we would like to ask you two questions to make sure you understand the instructions.

Section 6

Question 1: Your role is?

- Job applicant
- HR executive

Correct!

You are an HR executive.

Please press 'Next' to continue.

Incorrect

You are an HR executive.

Please select the correct answer to continue.

Section 7

Question 2: Your task is to?

- Hire one of two applicants
- Assign wages

Correct!

As an HR executive, your task is to hire one of two applicants. You will not have to assign wages to applicants.

Please press 'Next' to continue.

Incorrect

As an HR executive, your task is to hire one of two applicants. You will not have to assign wages to applicants.

Please select the correct answer to continue.

Section 8

You must hire a new experienced manager for a position that has opened.

Each applicant has submitted a job application containing:

1. Their resume (they chose between two resumes)
2. A statement describing their management style (they chose between two statements)

The job applicant you decide to hire will receive a wage of X tokens.

The job applicant that you do not hire will receive 0 tokens.

Tokens are converted in euros at the following rate: 10 tokens = 0.40 €.

Your decision will affect the real earnings of participants playing the role of applicants in the survey.

Section 9

As an HR executive, you work for a company called Gamma/Beta Ltd. You have received two applications and you have to choose who to hire.

The job advertisement is shown below: X

Choose the job applicant you want to hire by clicking on their job application.

Section 10

You chose X.

She will receive the wage of X tokens, as displayed in the job advertisement.

He will receive the wage of X tokens, as displayed in the job advertisement.

The other applicant, X, will receive 0 tokens.

Section 11

You completed the main activity.

We will now ask you some additional questions about your experience in the study and some socio-demographic questions.

Section 13

Was this statement included in the job advertisement for which you received the applications?

- No
- Yes

Section 14

Did the statement affect your choice?

- Yes, it made me hire a woman, regardless of the quality of the applications
- Yes, it made me hire a man, regardless of the quality of the applications
- Yes, it made me hire a woman when the two applications were similar
- Yes, it made me hire a man when the two applications were similar
- No, I was not influenced by the statement

Section 15

Thank you for taking part in this study!

Appendix S3

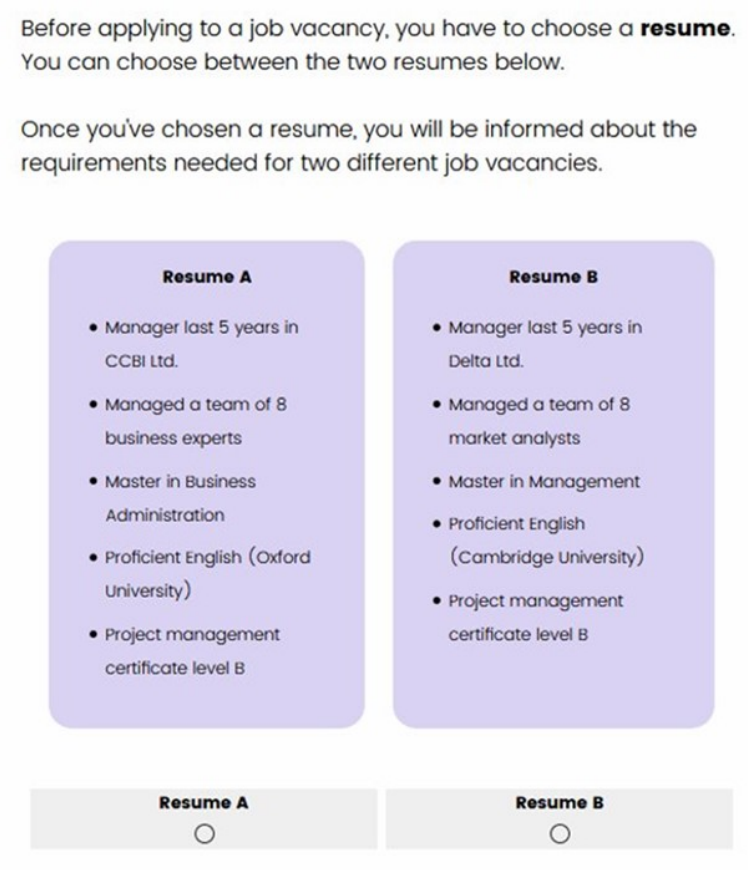


Figure S3.1: Resume task

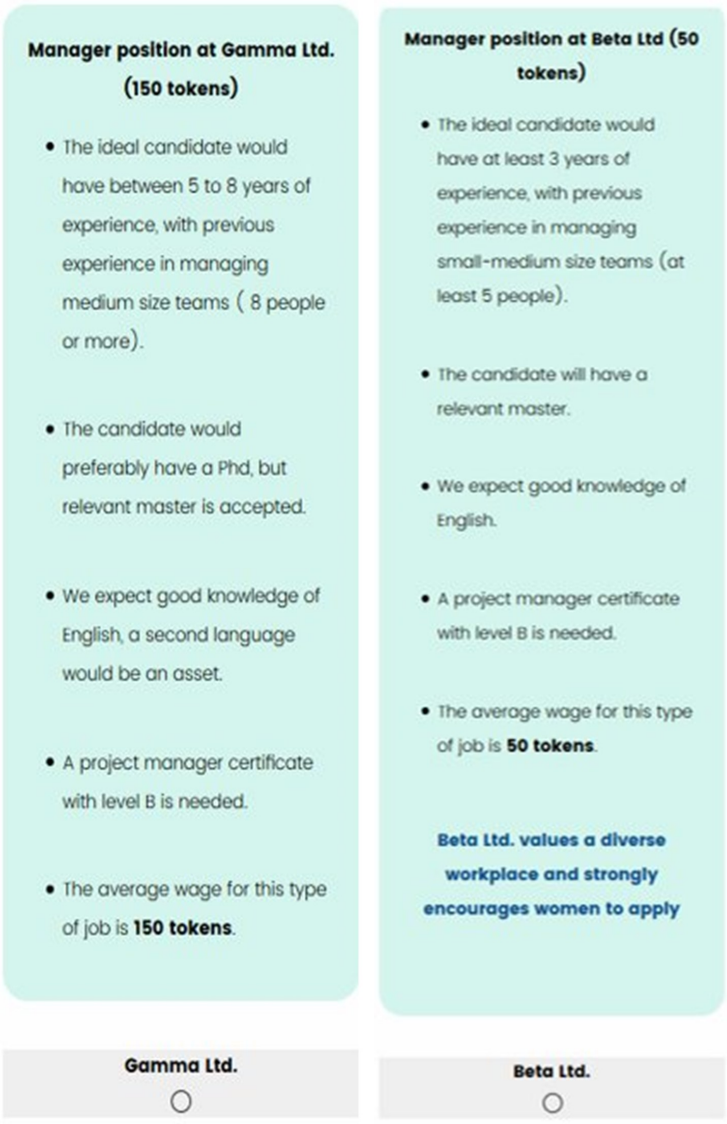


Figure S3.2: Job choice

Appendix S4

Figure S4.1: **Job choice in each treatment:** Control (T0)

Figure S4.2: **Job choice in each treatment:** Both jobs (T1) - EEOBoth

Figure S4.3: **Job choice in each treatment:** High paying job (T2) - EEOHigh

Figure S4.4: **Job choice in each treatment:** Low paying job (T3) - EEOLow

Appendix S5

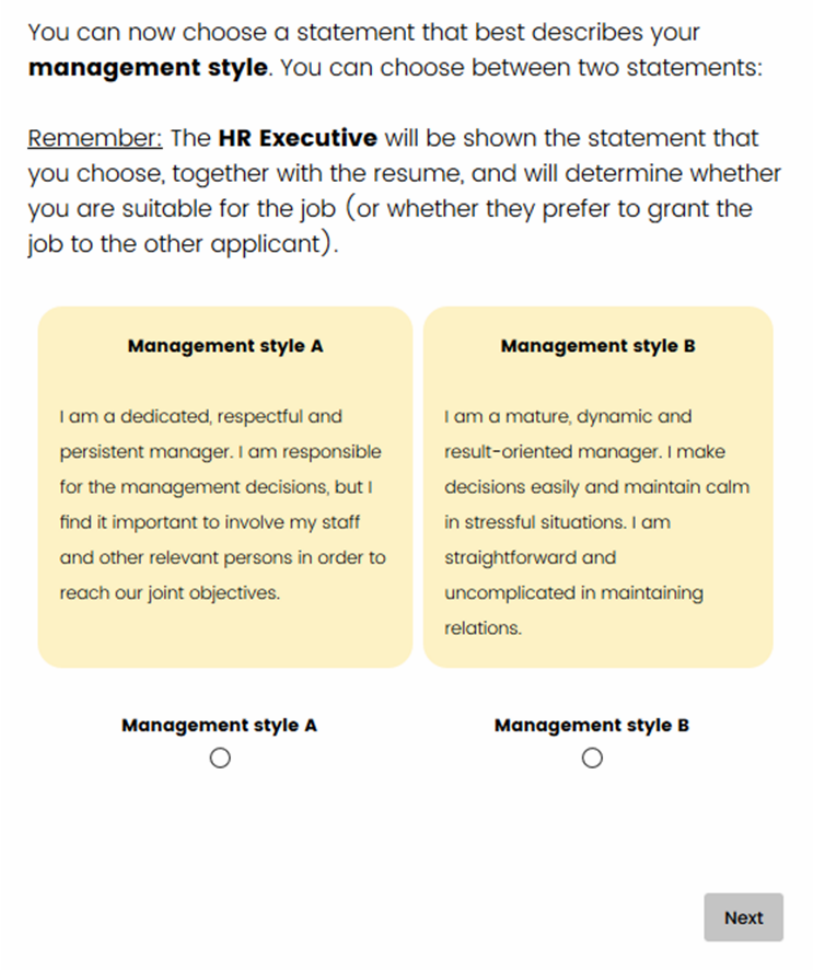


Figure S5.1: **Management style choice**

Appendix A

Table A1: Multiple Testing p-values Adjustments: Bonferroni, FDR, and Holm Correction Methods

Tests	χ^2	**Unadjusted p-values**	**Bonferroni Adjusted p-values**	**FDR q-values**	**Holm Adjusted p-values**
Gender Diff in Control	.0744491	.7849658	1	.8721842	1
Gender Diff in EEOboth	.191852	.6613797	1	.8267246	1
Gender Diff in EEOHigh	13.0737	.0002995	.0029947	.0014974	.0026952
Gender Diff in EEOLow	11.76977	.000602	.00602	.0020067	.004816
Control vs EEOboth - Women	7.700354	.005521	.05521	.0138025	.038647
Control vs EEOboth - Men	4.242783	.0394175	.394175	.0656958	.1970875
Control vs EEOHigh - Women	14.71291	.0001252	.0012519	.0012519	.0012519
Control vs EEOHigh - Men	.0422921	.8370639	1	.8721842	1
Control vs EEOLow - Women	4.824867	.028052	.2805203	.0561041	.1683122
Control vs EEOLow - Men	1.322433	.2501558	1	.3573655	1

The table reports p-values adjusted for multiple testing, taking into account a total of 10 performed χ^2 tests, according to Bonferroni, Holm, and FDR Benjamin-Hochberg correction mechanisms.

The Bonferroni method tends to be excessively conservative when a large number of tests are performed, hence methods like Holm's and Benjamini-Hochberg can offer a better balance in terms of power. Given these considerations, our focus will be on the Holm and FDR methods (using the Benjamini-Hochberg approach) moving forward.

We note that the difference in the proportion of male applicants between Control and EEOboth, as well as the difference in the proportion of female applicants between Control and EEOLow, which were initially identified as significant using unadjusted p-values, do not remain significant when adjusting for multiple testing.

All the other tests that were found significant using unadjusted p-values - specifically, the gender differences in EEOHigh, the gender difference in EEOLow, the difference in the proportion of female applicants applying to the high-paying job between Control and EEOboth, and the difference in the proportion of female applicants applying to the high-paying job between Control and EEOHigh - remain robust when subjected to both Holm and FDR correction mechanisms.

In terms of interpretation, the major drawback regards the loss of significance in the difference in the proportion of women applying to the high-paying job between Control and EEOLow.

This result is confirmed by the probit regression in subsection 3.2, where it is evident that, while con-

trolling for socio-demographic characteristics, employment situation, resume choice, and country, having a statement on a low-paying job only does not significantly decrease the number of women applying to the low-paying job.

Appendix B

Table B1: Robustness check using post-experimental variables

Control	**Fairness**			**Confidence**			**Gender Equality**		
Gender	**Pooled**	**Women**	**Men**	**Pooled**	**Women**	**Men**	**Pooled**	**Women**	**Men**
EEOHigh	-0.638*	-0.537	-0.672	-0.841**	-0.425	-1.077**	-0.572	0.0614	-1.260**
	(0.323)	(0.481)	(0.439)	(0.275)	(0.396)	(0.386)	(0.305)	(0.435)	(0.434)
EEOLow	0.922**	0.845	0.902*	0.686**	0.325	1.021**	0.366	0.606	0.263
	(0.309)	(0.448)	(0.422)	(0.265)	(0.374)	(0.376)	(0.290)	(0.410)	(0.412)
Female	0.0715			0.0557			0.0733		
	(0.0638)			(0.0639)			(0.0627)		
Age	-0.000437	0.00172	-0.00101	-0.000107	0.00185	-3.33e-05	0.000552	0.00269	0.000529
	(0.00262)	(0.00349)	(0.00411)	(0.00261)	(0.00347)	(0.00410)	(0.00260)	(0.00346)	(0.00410)
Mid education	-0.0292	-0.0631	-0.0126	-0.0372	-0.0948	0.0156	-0.0273	-0.0747	-0.000989
	(0.0822)	(0.116)	(0.119)	(0.0829)	(0.116)	(0.120)	(0.0815)	(0.115)	(0.119)
High education	0.0926	0.130	0.0585	0.103	0.122	0.0969	0.111	0.150	0.0728
	(0.0807)	(0.117)	(0.114)	(0.0808)	(0.116)	(0.114)	(0.0799)	(0.115)	(0.114)
Unemployed	-0.0966	-0.0773	-0.134	-0.0839	-0.0734	-0.0972	-0.0791	-0.0585	-0.107
	(0.0813)	(0.101)	(0.146)	(0.0814)	(0.100)	(0.145)	(0.0810)	(0.101)	(0.146)
Retired	-0.155	-0.319*	-0.0770	-0.161	-0.318*	-0.103	-0.174	-0.360*	-0.105
	(0.0981)	(0.155)	(0.132)	(0.0979)	(0.154)	(0.133)	(0.0971)	(0.153)	(0.132)
Risk Attitude	0.0897***	0.0913***	0.0899***	0.0870***	0.0916***	0.0832***	0.0933***	0.0970***	0.0894***
	(0.0131)	(0.0185)	(0.0194)	(0.0134)	(0.0186)	(0.0199)	(0.0129)	(0.0182)	(0.0191)
CCBI	0.0638	-0.0269	0.162	0.0594	-0.0305	0.158	0.0640	-0.0316	0.159
	(0.0617)	(0.0859)	(0.0903)	(0.0619)	(0.0855)	(0.0910)	(0.0615)	(0.0855)	(0.0903)
Spain	0.160*	0.0793	0.242*	0.162*	0.101	0.234*	0.165*	0.0848	0.258*
	(0.0708)	(0.105)	(0.0978)	(0.0710)	(0.104)	(0.0987)	(0.0715)	(0.105)	(0.100)
Fair	0.0318	0.0469	0.0152						
	(0.0314)	(0.0447)	(0.0438)						
EEOHigh × Fair	0.0992*	0.119	0.0714						
	(0.0415)	(0.0616)	(0.0566)						
TEEOLow × Fair	-0.160***	-0.167**	-0.138*						
	(0.0400)	(0.0577)	(0.0549)						
Confidence				0.0271	0.0140	0.0402			
				(0.0282)	(0.0395)	(0.0406)			
EEOHigh × Confidence				0.141***	0.116*	0.142**			
				(0.0382)	(0.0543)	(0.0543)			
EEOLow × Confidence				-0.139***	-0.106*	-0.166**			
				(0.0367)	(0.0511)	(0.0529)			
Equal Gender							0.0105	0.104	-0.0752
							(0.0567)	(0.0813)	(0.0797)
EEOHigh × Equal Gender							0.179*	0.0854	0.290**
							(0.0778)	(0.112)	(0.110)
EEOLow × Equal Gender							-0.168*	-0.273*	-0.0938
							(0.0746)	(0.107)	(0.105)
Constant	-0.635*	-0.705	-0.563	-0.576*	-0.454	-0.751*	-0.513	-0.819*	-0.249
	(0.284)	(0.392)	(0.400)	(0.250)	(0.337)	(0.367)	(0.266)	(0.359)	(0.385)
N	1875	978	897	1875	978	897	1875	978	897
Pseudo R2	0.0737	0.113	0.0602	0.0818	0.108	0.0813	0.0636	0.103	0.0586

Note: In the table, statistical significance is denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. The analysis was conducted using a probit regression model with robust standard errors. Standard errors are reported in parentheses. Model 1 only assesses the effect of EEO presence and gender, while Model 2 includes interaction terms between EEO presence and gender. The dependent variable equals 1 if the participant selected the high-paying position (position gamma), and 0 if they selected the low-paying position (position Beta). EEOboth, EEOHigh and EEOLow are treatment dummies, and their coefficients express the effect compared to the control group (Control) where no Equal Employment Opportunity (EEO) statements were present. The 'Female' variable is a gender dummy, equal to 1 if the participant is a woman, and 0 if the participant is a man. 'Risk attitude' is a measure of self-reported risk, with higher values indicating participants who are more risk-loving. 'CCBI' is a dummy variable equal to 1 if the participant chose the CCBI resume in the first task (the resume choice task), and 0 if they chose the Delta resume. 'Spain' is a country dummy equal to 1 for Spanish respondents and 0 for German respondents. The analysis also controls for age, edu cation, and employment status. 'Low education' and 'employed' are the reference levels for education and employment status, respectively. Fairness was assessed on a scale from 0 (unfair) to 10 (fair). Confidence was self-reported on a scale where 0 represents decreased confidence and 10 signifies increased confidence. For expected gender distribution participants expressed their level of agreement with the statement, "It made me expect that there would be an equal gender distribution among employees," on a 5-point scale, ranging from Strongly Disagree (1) to Strongly Agree (5), Confidence, fairness and gender equality are treated as continuous variables.

Appendix C

In the results of subsection 3.2, we found a significant dummy for Spain in the probit regressions for both Model 1 and Model 2. To delve deeper into this finding, we run separate probit regression analyses for Germany and Spain, as outlined in Table C1. As for the pooled sample, the aim is to predict the likelihood of individuals applying for the high-requirements, higher-paying job based on gender, EEO presence (Model 1), and their interaction term (Model 2), while controlling for variables such as age, education, employment condition, risk attitudes, and the selected resume.

In the German sample, results did not indicate a significant impact of any of the treatment conditions on the likelihood of applying for a high-paying job, regardless of the presence of Equal Employment Opportunity (EEO) statements. This result persists even when controlling for the interaction effects between treatments and gender in Model 2, suggesting that in Germany, EEO statements did not significantly influence job application behaviour for either gender.

For the Spanish sample, the results in Model 1 are aligned with those observed in the pooled sample (Table C1). Specifically, the presence of EEO statements in both job positions, or solely in the high-paying job position, significantly increases the probability of applying for a high-paying job ($p < 0.001$ and $p = 0.005$, respectively). In the interaction model (2), the coefficient for the interaction between the high-paying job treatment condition and gender is positive and significant ($p = 0.002$), whereas the coefficient for the interaction between EEOLow and female is negative and significant ($p = 0.016$). The marginal treatment effects by gender are not reported for brevity, but the main conclusions when considering only Spain perfectly

mirror the pooled results.

Table C1: Equal Employment Opportunity Statements effect on Job application by country

Dependent variable	**High paying job**			
Country	**Germany**		**Spain**	
Model	**(1)**	**(2)**	**(1)**	**(2)**
EEOboth	0.108	0.101	0.312***	0.263*
	(0.0815)	(0.116)	(0.0844)	(0.119)
EEOHigh	0.101	0.0359	0.233**	-0.0153
	(0.0812)	(0.115)	(0.0835)	(0.116)
EEOLow	-0.00968	0.0829	-0.0319	0.174
	(0.0819)	(0.116)	(0.0819)	(0.118)
Female	0.110	0.121	-0.0125	-0.0579
	(0.0611)	(0.111)	(0.0634)	(0.112)
EEOboth × Female		0.0137		0.0989
		(0.163)		(0.169)
EEOHigh × Female		0.129		0.521**
		(0.163)		(0.168)
EEOLow × Female		-0.183		-0.396*
		(0.164)		(0.165)
Age	-0.00264	-0.00243	0.00316	0.00401
	(0.00253)	(0.00254)	(0.00259)	(0.00261)
Mid education	-0.0419	-0.0456	-0.0217	-0.0204
	(0.0739)	(0.0741)	(0.0999)	(0.101)
High education	0.131	0.133	0.0140	0.0146
	(0.0716)	(0.0717)	(0.0972)	(0.0982)
Unemployed	-0.0633	-0.0635	-0.0671	-0.0529
	(0.0854)	(0.0854)	(0.0784)	(0.0796)
Retired	-0.0810	-0.0869	-0.0641	-0.0757
	(0.0906)	(0.0908)	(0.103)	(0.103)
Risk Attitude	0.0894***	0.0893***	0.106***	0.106***
	(0.0118)	(0.0118)	(0.0146)	(0.0147)
CCBI	-0.0548	-0.0561	0.259***	0.258***
	(0.0594)	(0.0595)	(0.0621)	(0.0624)
Constant	-0.473**	-0.485**	-0.835***	-0.850***
	(0.165)	(0.171)	(0.198)	(0.205)
N	1876	1876	1860	1860
Pseudo R2	0.0326	0.0339	0.0407	0.0518

Note: In the table, statistical significance is denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. The analysis was conducted using a probit regression model with robust standard errors. Standard errors are reported in parentheses. Model 1 only assesses the effect of control for EEO presence and gender, while Model 2 includes interaction terms between EEO presence treatments and gender. The dependent variable equals 1 if the participant selected the high paying position (position Gamma), and 0 if they selected the low paying position (position Beta). EEOboth, EEOHigh and EEOLow are treatment dummies, and their coefficients express the effect with respect to the control group (Control) where no Equal Employment Opportunity (EEO) statements were present.
The 'Female' variable is a gender dummy, equal to 1 if the participant is a woman, and 0 if the participant is a man. 'Risk attitude' is a measure of self-reported risk, with higher values indicating participants who are more risk-lover. 'CCBI' is a dummy variable equal to 1 if the participant chose the CCBI resume in the first task (the resume choice task), and 0 if they chose the Delta resume. The analysis also controls for age, education, and employment status. 'Low education' and 'employed' are the reference levels for education and employment status and thus coded 0.

Of note is that the German subsample contains a higher proportion of retirees compared to its Spanish counterpart ($\chi^2 = 35.4801$, $p < 0.001$); overall, our sample comprises 26.71% of retirees in Germany and 18.55% in Spain (see Table C2 to see the distribution of retirees by treatment condition). Thus, in Table C3 we run a few regressions to assess whether treatment effects diverge between retirees and those in the labour force (comprising the unemployed and employed). It emerges that there are no significant treatment effects for retirees in either model—those excluding (Model 1) or including gender interactions (Model 2). Conversely, for the subsample including only the employed and unemployed people, treatment effects are in line with the pooled sample probit regressions (Table 3 in subsection 3.1). With the only exception of a significant and positive coefficient in Model 2 for EEOLow ($p = 0.005$) that we did not found in the main result section. However the coefficient of the Spain dummy is still significant ($p = 0.044$).

Table C2: Employment Condition by treatment and country

Country	**Germany**					**Spain**				
Treatment	$\mathbf{T_{Control}}$	$\mathbf{T_{EEOBoth}}$	$\mathbf{T_{EEOHigh}}$	$\mathbf{T_{EEOLow}}$	**Total**	$\mathbf{T_{Control}}$	$\mathbf{T_{EEOBoth}}$	$\mathbf{T_{EEOHigh}}$	$\mathbf{T_{EEOLow}}$	**Total**
Employed	327	260	241	231	1059	310	271	268	259	1108
	60.44	58.30	54.16	52.03	56.45	57.41	61.45	61.05	58.86	59.57
Unemployed	82	66	85	83	316	122	83	85	117	407
	15.16	14.80	19.10	18.69	16.84	22.59	18.82	19.36	26.59	21.88
Retired	132	120	119	130	501	108	87	86	64	345
	24.40	26.91	26.74	29.28	26.71	20.00	19.73	19.59	14.55	18.55
Total	541	446	445	444	1876	540	441	439	440	1860

Note: First row has absolute frequencies and second row has column percentages

Table C3: Equal Employment Opportunity Statements effect on Job application by Employment Status

Dependent Variable	**High Paying Job**			
	Model (1)		Model (2)	
Employment Status	Employed + Unemployed	Retired	Employed + Unemployed	Retired
EEOboth	0.251***	0.0619	0.236*	0.0558
	(0.0662)	(0.122)	(0.0978)	(0.151)
EEOHigh	0.281***	-0.233	0.111	-0.232
	(0.0662)	(0.122)	(0.0977)	(0.150)
EEOLow	0.0485	-0.242	0.274**	-0.197
	(0.0657)	(0.125)	(0.0977)	(0.152)
Female	0.0498	0.00476	0.0701	0.0306
	(0.0487)	(0.0960)	(0.0887)	(0.177)
EEOboth × Female			0.0280	0.0177
			(0.133)	(0.258)
EEOHigh × Female			0.314*	-0.00124
			(0.133)	(0.261)
EEOLow × Female			-0.402**	-0.133
			(0.132)	(0.267)
Age	0.00136	-0.0115	0.00202	-0.0116
	(0.00183)	(0.00693)	(0.00185)	(0.00696)
Mid education	-0.00516	0.0426	-0.00805	0.0418
	(0.0650)	(0.123)	(0.0653)	(0.124)
High education	0.0949	0.0406	0.0967	0.0415
	(0.0644)	(0.109)	(0.0647)	(0.109)
Spain	0.111*	0.365***	0.111*	0.365***
	(0.0551)	(0.103)	(0.0553)	(0.103)
Risk Attitude	0.104***	0.0779***	0.105***	0.0772***
	(0.0105)	(0.0186)	(0.0105)	(0.0187)
CCBI	0.0767	0.154	0.0764	0.152
	(0.0488)	(0.0905)	(0.0490)	(0.0906)
Constant	-0.863***	0.188	-0.907***	0.191
	(0.128)	(0.485)	(0.134)	(0.487)
N	2890	846	2890	846
Pseudo R2	0.0465	0.0590	0.0531	0.0594

Note: In the table, statistical significance is denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. The analysis was conducted using a probit regression model with robust standard errors. Standard errors are reported in parentheses. Model 1 only assesses the effect of control for EEO presence and gender, while Model 2 includes interaction terms between EEO presence treatments and gender. The dependent variable equals 1 if the participant selected the high paying position (position Gamma), and 0 if they selected the low paying position (position Beta). EEOboth, EEOHigh and EEOLow are treatment dummies, and their coefficients express the effect with respect to the control group (Control) where no Equal Employment Opportunity (EEO) statements were present.

The 'Female' variable is a gender dummy, equal to 1 if the participant is a woman, and 0 if the participant is a man. 'Risk attitude' is a measure of self-reported risk, with higher values indicating participants who are more risk-lover. 'CCBI' is a dummy variable equal to 1 if the participant chose the CCBI resume in the first task (the resume choice task), and 0 if they chose the Delta resume. The analysis also controls for age, education, and employment status. 'Low education' and 'employed' are the reference levels for education and employment status and thus coded 0. Sapin is a dummy variable equal to 1 if the respondent is from Spain, 0 if they are from Germany.

To further investigate country differences, we run probit regressions by employment status and country, as detailed in Table C4. First, it's important to note that not all treatment effects observed in the main pooled sample regression (Table 3 in subsection 3.1) are significant in the subgroup regressions, including those for the employed/unemployed subgroups. This could be due to insufficient power.

However, it's noteworthy that in Model 1, for people in the German labour force, the coefficients for EEOhigh and EEOboth are statistically significant ($p = 0.011$ and $p = 0.038$ respectively). For the same subgroup of participants, in Model 2, the interaction term with gender for EEOlow also shows significance ($p = 0.046$). And for women the treatment effect of EEOHigh is positive and significant ($z = 0.2817$, $p = 0.03$). These results align with our conclusions in the main results section.

Moreover, as observed in Table C3, for the labour force in Germany, we see a significant and positive coefficient in Model 2 for EEOLow ($p = 0.005$), a result we did not find in the main results section.

So, in general, the effectiveness of EEO statements seems to be country-dependent. When considering only people in the workforce (employed and unemployed individuals), we do find some treatment effects in Germany; however, we believe that more research on the country differences is needed, and we think it might be an interesting focus for future research development.

Table C4: Equal Equal Employment Opportunity Statements effect on Job application by Employment Status and Country

Dependent Variable	**High Paying Job**							
Country	**Germany**				**Spain**			
Employment Status	Employed + Unemployed		Retired		Employed + Unemployed		Retired	
Model	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
EEOboth	0.197*	0.219	-0.145	-0.147	0.304**	0.235	0.352	0.321
	(0.0947)	(0.140)	(0.161)	(0.204)	(0.0939)	(0.140)	(0.194)	(0.231)
EEOHigh	0.241*	0.190	-0.321*	-0.298	0.327***	0.0333	-0.0915	-0.109
	(0.0947)	(0.139)	(0.161)	(0.204)	(0.0938)	(0.138)	(0.186)	(0.217)
EEOLow	0.119	0.332*	-0.355*	-0.388	-0.0145	0.204	-0.0794	0.108
	(0.0959)	(0.142)	(0.159)	(0.205)	(0.0898)	(0.137)	(0.203)	(0.237)
Female	0.139	0.219	-0.0115	-0.0201	-0.0317	-0.0820	-0.000832	0.0972
	(0.0711)	(0.128)	(0.121)	(0.230)	(0.0679)	(0.125)	(0.159)	(0.272)
EEOboth × Female		-0.0437		0.00498		0.125		0.0839
		(0.191)		(0.333)		(0.189)		(0.431)
EEOHigh × Female		0.0921	-0.0641		0.546**		0.0832	
		(0.191)	(0.336)		(0.189)		(0.418)	
EEOLow × Female		-0.385*	0.0875		-0.385*		-0.804	
		(0.193)	(0.329)		(0.182)		(0.480)	
Age	0.000609	0.000982	-0.0210*	-0.0210*	0.00302	0.00384	0.0153	0.0171
	(0.00261)	(0.00262)	(0.00829)	(0.00834)	(0.00262)	(0.00264)	(0.0130)	(0.0131)
Mid education	0.00476	-0.00309	-0.184	-0.181	-0.118	-0.116	0.321	0.333
	(0.0834)	(0.0837)	(0.166)	(0.166)	(0.116)	(0.117)	(0.210)	(0.212)
High education	0.187*	0.192*	0.0597	0.0605	-0.0319	-0.0341	0.136	0.141
	(0.0847)	(0.0849)	(0.137)	(0.138)	(0.112)	(0.114)	(0.193)	(0.193)
Risk Attitude	0.105***	0.106***	0.0660**	0.0662**	0.105***	0.106***	0.105**	0.0991**

	(0.0140)	(0.0140)	(0.0225)	(0.0226)	(0.0162)	(0.0163)	(0.0342)	(0.0348)
CCBI	-0.143*	-0.142*	0.184	0.187	0.295***	0.295***	0.111	0.111
	(0.0695)	(0.0696)	(0.116)	(0.116)	(0.0688)	(0.0692)	(0.148)	(0.148)
Constant	-0.788***	-0.851***	1.006	1.005	-0.814***	-0.826***	-1.716	-1.833
	(0.179)	(0.188)	(0.574)	(0.575)	(0.210)	(0.220)	(0.952)	(0.958)
N	1375	1375	501	501	1515	1515	345	345
Pseudo R2	0.0419	0.0452	0.0349	0.0352	0.0443	0.0557	0.0478	0.0567

Appendix D

Table D1: Median and Mean perceived fairness by gender, treatment condition and job choice

	Perceived Fairness					
	Low-Paying Job			**High-Paying Job**		
	Women	**Men**	**Mann-Whitney U test**	**Women**	**Men**	**Mann-Whitney U test**
EEOBoth	8 [7.434]	8 [7.464]	$z = 0.294$ $p = 0.7691$	8 [7.731]	8 [7.707]	$z = 0.207$ $p = 0.8359$
EEOHigh	7 [7.082]	7 [6.985]	$z = -0.176$ $p = 0.8602$	8 [8.175]	8 [7.752]	$z = -1.967$ $p = 0.0492$
EEOLow	8 [7.849]	8 [7.422]	$z = -2.301$ $p = 0.0214$	7 [7.191]	8 [6.868]	$z = -1.105$ $p = 0.2692$

Note: The median and the mean (in parentheses) of the perceived fairness of the application process, reported by gender, treatment, and job chosen by the participants, are provided. Fairness was assessed on a scale from 0 (unfair) to 10 (fair). Tests corrected for multiple comparisons are reported in Appendix E.

Table D2: Median and Mean reported confidence by gender, treatment condition and job choice

	Reported Confidence					
	Low-Paying Job			**High-Paying Job**		
	Women	**Men**	**Mann-Whitney U test**	**Women**	**Men**	**Mann-Whitney U test**
EEOBoth	7 [7]	7 [6.795]	$z = -0.771$ $p = 0.4405$	7 [7.296]	8 [7.167]	$z = -0.551$ $p = 0.5814$
EEOHigh	6 [6.388]	6 [5.810]	$z = -1.894$ $p = 0.0583$	8 [7.491]	8 [7.394]	$z = -0.312$ $p = 0.7554$
EEOLow	8 [7.226]	7 [6.9333]	$z = -1.621$ $p = 0.1049$	6 [6.618]	6 [6.270]	$z = -1.315$ $p = 0.1885$

Note: The median and the mean (in parentheses) of the reported confidence in the application, listed by gender, treatment, and job chosen by the participants, are provided. Confidence was self-reported on a scale where 0 represents decreased confidence and 10 signifies increased confidence. Tests corrected for multiple comparisons are reported in Appendix E.

Table D3: Median and Mean of expected gender distribution by gender, treatment condition and job choice

	Expected gender distribution					
	Low-Paying Job			**High-Paying Job**		
	Women	**Men**	**Mann-Whitney U test**	**Women**	**Men**	**Mann-Whitney U test**
EEOBoth	4 [3.680]	4 [3.830]	z=1.336 p=0.1817	4 [3.876]	4 [3.787]	z=-0.414 p=0.6791
EEOHigh	4 [3.6]	4 [3.657]	z=0.667 p=0.5050	4 [3.949]	4 [4.055]	z=1.035 p=0.3009
EEOLow	4 [3.784]	4 [3.844]	z=0.476 p=0.6339	4 [3.612]	4 [3.707]	z=1.245 p=0.2132

Note: The median and the mean (in parentheses) of the expected gender distribution on the workplace, reported by gender, treatment, and job chosen by the participants, are provided. Participants expressed their level of agreement with the statement, "It made me expect that there would be an equal gender distribution among employees," on a 5-point scale, ranging from Strongly Disagree (1) to Strongly Agree (5). Tests corrected for multiple comparisons are reported in Appendix E.

Table D4: Median and Mean of encouragement/discouragement effect of EEO statement by gender and treatment condition

	Encouragement/discouragement effect		
	Women	**Men**	**Mann-Whitney U test**
EEOBoth	2 [2]	2 [2.3696]	$z = 1.529$ $p = 0.1262$
EEOHigh	2 [1.9308]	3 [2.6415]	$z = 3.464$ $p = 0.0005$
EEOLow	3 [2.9052]	3 [2.3433]	$z = -3.675$ $p = 0.0002$

Note: The median and the mean (in parentheses) of the reported Encouragement/discouragement effect by gender and treatment are provided. Participants indicated whether the EEO statemnt encouraged or discouraged their application to the higher paying job position on a 5-point scale, where 1 is encouraged and 5 is discouraged.

Table D5: Statement noticed or not: Multiple Testing p-values Adjustments (Bonferroni, FDR, and Holm Correction Methods)

Tests	χ^2	**Unadjusted p-values**	**Bonferroni Adjusted p-values**	**FDR q-values**	**Holm Adjusted p-values**
Gender Difference in Noticing EEO in EEOBoth	13.16343	.0013855	.0041564	.0013855	.0013855
Gender Difference in Noticing EEO in EEOHigh	41.9864	7.63e-10	2.29e-09	2.29e-09	2.29e-09
Gender Difference in Noticing EEO in EEOLow	18.96881	.000076	.0002281	.000114	.0001521

Table D6: Fairness, Confidence and Expected gender equality: Multiple Testing p-values Adjustments (Bonferroni, FDR, and Holm Correction Methods)

Tests	**z**	**Unadjusted p-values**	**Bonferroni Adjusted p-values**	**FDR q-values**	**Holm Adjusted p-values**
Gender Difference in Fairness in EEOBoth Low-paying Job	.2935761	.7690819	1	.8652171	1
Gender Difference in Fairness in EEOBoth High-paying Job	.2071535	.83589	1	.88506	1
Gender Difference in Fairness in EEO-High Low-paying Job	-.17609	.8602232	1	.8602232	1
Gender Difference in Fairness in EEO-High High-paying Job	-1.966755	.0492115	.885807	.4429035	.8365955
Gender Difference in Fairness in EE-OLow Low-paying Job	-2.300508	.0214194	.38555	.38555	.38555
Gender Difference in Fairness in EE-OLow High-paying Job	-1.104792	.2692498	1	.6058121	1
Gender Difference in Confidence in EEOBoth Low-paying Job	-.7713641	.4404911	1	.7928841	1
Gender Difference in Confidence in EEOBoth High-paying Job	-.551268	.58145	1	.872175	1
Gender Difference in Confidence in EEOHigh Low-paying Job	-1.893618	.0582757	1	.349654	.9324107
Gender Difference in Confidence in EEOHigh High-paying Job	-.3115901	.7553521	1	.9064225	1
Gender Difference in Confidence in EEOLow Low-paying Job	-1.621417	.1049283	1	.4721774	1
Gender Difference in Confidence in EEOLow High-paying Job	-1.315095	.1884778	1	.5654335	1
Gender Difference in Equality in EEOBoth Low-paying Job	1.335559	.1816933	1	.6540959	1
Gender Difference in Equality in EEOBoth High-paying Job	-.4136963	.6790965	1	.8731241	1
Gender Difference in Equality in EEOHigh Low-paying Job	.6665875	.5050356	1	.826422	1
Gender Difference in Equality in EEOHigh High-paying Job	1.034601	.3008552	1	.6017105	1
Gender Difference in Equality in EE-OLow Low-paying Job	.4762498	.6338965	1	.8777028	1
Gender Difference in Equality in EE-OLow High-paying Job	1.244786	.2132104	1	.5482552	1

Table D7: Encouragement/discouragement effect: Multiple Testing p-values Adjustments (Bonferroni, FDR, and Holm Correction Methods

Tests	**z**	**Unadjusted p-values**	**Bonferroni Adjusted p-values**	**FDR q-values**	**Holm Adjusted p-values**
Gender Difference in encouragement/discouragement effect in EEOBoth	1.529416	.1261613	.3784838	.1261613	.1261613
Gender Difference in encouragement/discouragement effect in EEOHigh	3.464037	.0005321	.0015964	.0007982	.0010643
Gender Difference in encouragement/discouragement effect in EEOLow	-3.674772	.0002381	.0007142	.0007142	.0007142

Table D8: HR Executive expected gender: Multiple Testing p-values Adjustments (Bonferroni, FDR, and Holm Correction Methods)

Tests	χ^2	**Unadjusted p-values**	**Bonferroni Adjusted p-values**	**FDR q-values**	**Holm Adjusted p-values**
Gender Difference in expectation about HR's gender for participants that didn't get the job	35.72944	1.74e-08	6.97e-08	2.32e-08	3.49e-08
Gender Difference in expectation about HR's gender for participants that got the job	164.0735	2.35e-36	9.42e-36	9.42e-36	9.42e-36
Difference in expectation about HR's gender between men that got the job and men who didn't	16.00704	.0003343	.0013371	.0003343	.0003343
Difference in expectation about HR's gender between women that got the job and women who didn't	81.7663	1.76e-18	7.03e-18	3.51e-18	5.27e-18

While gender differences in noticing the equal opportunity statement, encouragement / discouragement effect and the differences in HR's gender expectation (including both gender differences and differences between those who got the job and those who didn't) remain significant even after correcting for multiple testing, the same cannot be said about the perceived fairness considerations.

Appendix E

Table E1: Robustness Check: Interactions between CCBI and treatment conditions

Dependent Variable	**High paying job**	
Model	(1)	(2)
EEOboth	0.194*	0.103
	(0.0897)	(0.131)
EEOHigh	0.210*	-0.103
	(0.0885)	(0.130)
EEOLow	0.0834	0.0941
	(0.0918)	(0.136)
CCBI	0.146	0.0386
	(0.0791)	(0.112)
EEOboth × CCBI	0.0201	0.128
	(0.118)	(0.168)
EEOHigh × CCBI	-0.0794	0.191
	(0.117)	(0.167)
EEOLow × CCBI	-0.175	0.0495
	(0.118)	(0.170)
Female	0.0450	-0.0879
	(0.0437)	(0.122)
EEOboth × Female		0.170
		(0.180)
EEOHigh × Female		0.592***
		(0.179)
EEOLow × Female		-0.0119
		(0.184)
CCBI × Female		0.209
		(0.158)
EEOboth × CCBI × Female		-0.206
		(0.237)
EEOHigh × CCBI × Female		-0.495*
		(0.236)
EEOLow × CCBI × Female		-0.489*
		(0.238)
Age	3.25e-05	0.000569
	(0.00180)	(0.00181)
Mid education	-0.000625	-0.00361
	(0.0570)	(0.0573)
High education	0.0658	0.0654
	(0.0555)	(0.0557)
Unemployed	-0.0707	-0.0679
	(0.0575)	(0.0578)
Retired	-0.0719	-0.0832
	(0.0674)	(0.0676)
Risk Attitude	0.0961***	0.0958***
	(0.00912)	(0.00916)
Spain	0.179***	0.182***
	(0.0486)	(0.0488)
Constant	-0.747***	-0.698***
	(0.127)	(0.141)
Observations	3,736	3,736
Pseudo R2	0.0467	0.0530

Note: In the table, statistical significance is denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. The analysis was conducted using a probit regression model with robust standard errors. Standard errors are reported in parentheses. Model 1 only assesses the effect of EEO presence interacted with CCBI, while Model 2 includes interaction terms between EEO presence, gender and CCBI. The dependent variable equals 1 if the participant selected the high-paying position (position gamma), and 0 if they selected the low-paying position (position Beta). EEOboth, EEOHigh and EEOLow are treatment dummies, and their coefficients express the effect compared to the control group (Control) where no Equal Employment Opportunity (EEO) statements were present. The 'Female' variable is a gender dummy, equal to 1 if the participant is a woman, and 0 if the participant is a man. 'Risk attitude' is a measure of self-reported risk, with higher values indicating participants who are more risk-loving. 'CCBI' is a dummy variable equal to 1 if the participant chose the CCBI resume in the first task (the resume choice task), and 0 if they chose the Delta resume. 'Spain' is a country dummy equal to 1 for Spanish respondents and 0 for German respondents. The analysis also controls for age, edu cation, and employment status. 'Low education' and 'employed' are the reference levels for education and employment status, respectively.

Starting from the assumption that we designed the two resumes to be qualitatively similar, the unexpected significance of the coefficient of CCBI is puzzling. However, as we can see in Model 1 of Table E1, the interactions between CCBI and treatment conditions are not statistically significant. In addition, in Figure E1, we can see the marginal effects estimated from Model 2 of choosing the CCBI resume by gender and treatment condition. Only in the control treatment for women do we have significant marginal effects; all the others are not. Thus, this does not undermine the main results and conclusions of this paper.

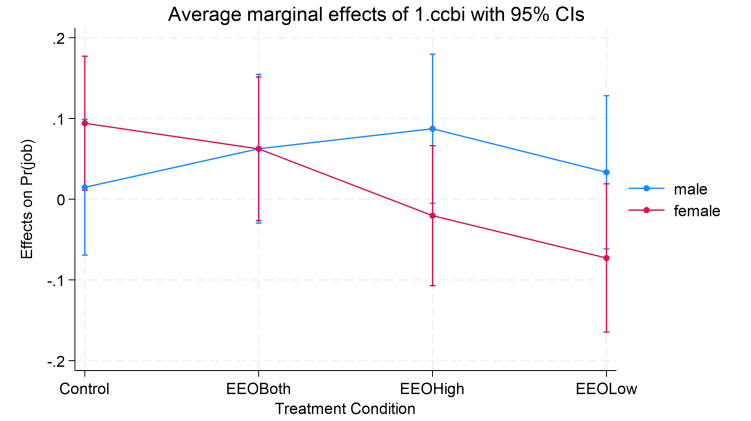


Figure E1: Marginal effects of CCBI

Appendix F

As an additional robustness check, in Table F1 we report the results of the probit regressions (as specified in Model 1 and Model 2 in the methods section) on the subsample of individuals who correctly noticed or correctly did not notice the EEO statements, meaning individuals who reported having noticed the statement when it was present (EEO Both, EEO High, and EEO Low) and who reported not having noticed the statement when it was not present (Control).

Table F1: Robustness Check: Probit regressions on the subsample of people who correctly noticed or did not notice the EEO statements.

Dependent Variable	**High paying job**	
Model	(1)	(2)
EEOboth	0.241***	0.280**
	(0.0700)	(0.1000)
EEOHigh	0.352***	0.132
	(0.0696)	(0.0980)
EEOLow	-0.0220	0.199*
	(0.0678)	(0.0971)
Female	0.0988	0.119
	(0.0523)	(0.0929)
EEOboth × Female		-0.0746
		(0.140)
EEOHigh × Female		0.456**
		(0.140)
EEOLow × Female		-0.413**
		(0.136)
Age	0.00151	0.00221
	(0.00215)	(0.00217)
Mid education	0.00883	0.00241
	(0.0685)	(0.0689)
High education	0.134*	0.137*
	(0.0667)	(0.0672)
Unemployed	-0.0332	-0.0301
	(0.0682)	(0.0689)
Retired	-0.0878	-0.106
	(0.0803)	(0.0806)
Risk Attitude	0.0979***	0.0979***
	(0.0109)	(0.0110)
CCBI	0.0809	0.0782
	(0.0513)	(0.0515)
Spain	0.128*	0.134*
	(0.0583)	(0.0587)
Constant	-0.842***	-0.883***
	(0.145)	(0.151)
Observations	2,654	2,654
Pseudo R2	0.0531	0.0630

Note: In the table, statistical significance is denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. The analysis was conducted using a probit regression model with robust standard errors. Standard errors are reported in parentheses. Model 1 only assesses the effect of EEO presence and gender, while Model 2 includes interaction terms between EEO presence and gender. The dependent variable equals 1 if the participant selected the high-paying position (position gamma), and 0 if they selected the low-paying position (position Beta). EEOboth, EEOHigh and EEOLow are treatment dummies, and their coefficients express the effect compared to the control group (Control) where no Equal Employment Opportunity (EEO) statements were present. The 'Female' variable is a gender dummy, equal to 1 if the participant is a woman, and 0 if the participant is a man. 'Risk attitude' is a measure of self-reported risk, with higher values indicating participants who are more risk-loving. 'CCBI' is a dummy variable equal to 1 if the participant chose the CCBI resume in the first task (the resume choice task), and 0 if they chose the Delta resume. 'Spain' is a country dummy equal to 1 for Spanish respondents and 0 for German respondents. The analysis also controls for age, education, and employment status. 'Low education' and 'employed' are the reference levels for education and employment status, respectively.

Table F2: Marginal effects by gender and treatment conditions on the subsample of people who correctly noticed or did not notice the EEO statements.

	Women	**Men**	**Difference (Women - Men)**
EEOboth	0.077*	0.106**	-0.029
	(0.0366)	(0.0374)	(0.0524)
EEOHigh	0.207***	0.0502	0.158**
	(0.0343)	(0.0372)	(0.0508)
EEOLow	-0.081	0.0756*	-0.156**
	(0.0361)	(0.0367)	(0.0514)

Note: The marginal effects in this table are derived from the probit regression model specified in Model 2 of Table F1, which includes interaction terms between treatments and gender and controls for age, education, risk attitude, resume choice and country effects. These are average marginal effects calculated for a representative man and woman. Standard errors are reported in parentheses. The statistical significance is denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

In Table F2, we report the marginal effects estimated from the probit regression Model 2 in Table F1. As we can see, the direction of the effects and the significance are in line with our results in the results section (see Table 4). The only difference is that in this robustness check, we find that when EEO statements are solely on the low-paying job, men are more than 7 percentage points more likely to apply to the high-paying job, a result we did not find in the main analysis.

Appendix G

In Table G1, we report an additional robustness check focusing on the role of risk aversion in job application behavior. Model 1 examines the interaction between risk attitude and treatment conditions, while Model 2 explores the interaction between risk attitude and gender. Both models include the same control variables used in the main analysis to ensure consistency.

Table G1: Robustness Check: Interactions between Risk Attitude and treatment conditions (1) and Risk Attitude and Female (2)

Dependent Variable	**High paying job**	
Model	(1)	(2)
EEOboth	0.299	0.205***
	(0.158)	(0.0581)
EEOHigh	0.182	0.166**
	(0.155)	(0.0580)
EEOLow	-0.00324	-0.0204
	(0.155)	(0.0580)
Risk Attitude	0.101***	0.105***
	(0.0164)	(0.0129)
EEOboth × Risk Attitude	-0.0153	
	(0.0237)	
EEOHigh × Risk Attitude	-0.00256	
	(0.0234)	
EEOLow × Risk Attitude	-0.00270	
	(0.0235)	
Female	0.0461	0.142
	(0.0437)	(0.113)
Age	1.53e-05	-1.26e-05
	(0.00180)	(0.00180)
Mid education	0.00140	-0.00140
	(0.0570)	(0.0570)
High education	0.0690	0.0659
	(0.0555)	(0.0555)
Unemployed	-0.0710	-0.0687
	(0.0575)	(0.0574)
Retired	-0.0717	-0.0702
	(0.0674)	(0.0674)
CCBI	0.0921*	0.0904*
	(0.0429)	(0.0429)
Spain	0.176***	0.178***
	(0.0486)	(0.0486)
Female × Risk Attitude		-0.0159
		(0.0170)
Constant	-0.747***	-0.767***
	(0.147)	(0.132)
Observations	3,736	3,736
N	3736	3736
Pseudo R2	0.0462	0.0463

Note: In the table, statistical significance is denoted as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. The analysis was conducted using a probit regression model with robust standard errors. Standard errors are reported in parentheses . Model 1 only assesses the effect of EEO presence interacted with Risk Attitude, while Model 2 includes interaction terms between gender and Risk Attitude. The dependent variable equals 1 if the participant selected the high-paying position (position gamma), and 0 if they selected the low-paying position (position Beta). EEOboth, EEOHigh and EEOLow are treatment dummies, and their coefficients express the effect compared to the control group (Control) where no Equal Employment Opportunity (EEO) statements were present. The 'Female' variable is a gender dummy, equal to 1 if the participant is a woman, and 0 if the participant is a man. 'Risk attitude' is a measure of self-reported risk, with higher values indicating participants who are more risk-loving. 'CCBI' is a dummy variable equal to 1 if the participant chose the CCBI resume in the first task (the resume choice task), and 0 if they chose the Delta resume. 'Spain' is a country dummy equal to 1 for Spanish respondents and 0 for German respondents. The analysis also controls for age, education, and employment status. 'Low education' and 'Employed' are the reference levels for education and employment status, respectively.

In a manner consistent with the main results section, we estimated the marginal effects to better interpret the interaction terms. For Model 1, the marginal effects of risk attitude indicate a positive and significant influence on the probability of applying for the high-paying job across all treatment conditions. Specifically, a one-unit increase in risk attitude increases the probability of applying by 3.84 pp in the control group ($p < 0.001$), 3.21 pp in the EEOBoth condition ($p < 0.001$), 3.68 pp in the EEOHigh condition ($p < 0.001$), and 3.75 pp in the EEOLow condition ($p < 0.001$). Tests of differences in these marginal effects between treatment conditions and the comntrol condition show no significant variation (p-values ranging from 0.449 to 0.910), suggesting that risk attitude influences application behavior similarly across all treatments.

For Model 2, the marginal effects show that risk attitude has a comparable positive and significant effect for both genders. For men, a one-unit increase in risk attitude raises the probability of applying by 3.93 pp ($p < 0.001$), while for women, the effect is 3.35 pp ($p < 0.001$). The difference between these effects is -0.58 pp which is not statistically significant ($p = 0.334$). These results indicate that risk attitude influences job application behavior similarly for men and women.

Based on this analysis, we conclude that risk aversion alone does not explain the observed gender differences in job application behavior in response to EEO statements. Other factors, beyond risk aversion, are likely driving these differences in our setting.

Gendered decision making in explore-exploit tasks

Francesca Barigozzi* Natalia Montinari* Elisa Orlandi*

Abstract

Many real-world scenarios involve explore-exploit decisions, balancing the pursuit of better opportunities with securing a certain but potentially suboptimal outcome. Do gendered approaches to these decisions exist? This study investigates gender differences in exploration and competition behaviour through a pre-registered lab experiment with 432 participants (50% female). Specifically, we examine behaviour in the context of the explore-exploit dilemma, both under a piece rate payment scheme and in a competitive tournament setting. Participants completed three computerized tasks: the grain game featuring the explore-exploit dilemma, which included two treatments, one allowing only gains and another incorporating both gains and losses, a risk elicitation task (BRET), and a loss aversion task. These tasks were followed by a questionnaire designed to assess various individual characteristics The results show that, contrary to the initial pre-registered hypotheses, women do not explore less than men; in fact, they explore more in environments where only gains are possible. However, no gender differences emerge in exploration when the environment entails the possibility of losses. Regarding competition, women are initially less likely to choose competitive settings than men in gain-only environments, but this difference disappears once individual characteristics, such as risk and loss aversion, are taken into account. These findings contribute to understanding gendered tendencies in risk-taking, exploration and competition.

Keywords: Explore-exploit tasks, Gender differences, Exploration behaviour, Competitive behaviour

JEL Codes: C91, D81, D03, J16

*University of Bologna,
Email addresses: francesca.barigozzi@unibo.it; natalia.montinari2@unibo.it; elisa.orlandi15@unibo.it.
Acknowledgments: We express our sincere gratitude to the participants of the IDEE meeting at the University of Bologna, the 2022 European ESA Meeting in Bologna, and the seminar attendees at the University of Bath (UK) for their insightful comments and suggestions. We are particularly thankful to Stephanie Heger and Maria Bigoni for their invaluable feedback, and to Alycia Chin, Charley M. Wu, and Azzurra Ruggeri for their generosity in sharing their data and code. Our thanks also go to Riccardo Ghioni for his outstanding research assistance with the experiment code. Any errors that remain are entirely our own responsibility.

1 Introduction

Many significant real-world scenarios involve explore-exploit decisions, characterized by the tension between seeking potentially better opportunities and securing a certain, albeit possibly suboptimal, outcome. The dynamic interplay between exploration and exploitation may describe, for example, the uncertainty and decision-making challenges inherent in the job search process. For instance, a young job seeker encounters the explore-exploit dilemma when deciding how many applications to submit and when to accept a job offer. Initially, exploration dominates as they apply broadly to learn about available opportunities. However, as offers start coming in, they face the critical decision of when to shift to exploitation, accepting an offer while weighing the risks of missing better opportunities against the possibility of losing viable ones.

Do gendered approaches to these decisions exist? Cortés et al. [2023] provide evidence that college-educated women tend to accept jobs significantly earlier than men with the same qualifications. Similarly, Theodossiou and Zangelidis [2009] find that men exhibit higher voluntary job-to-job mobility compared to women. These findings suggest the existence of gendered job search strategies.

Gender roles are shared societal expectations regarding the attributes and behaviours of men and women, or boys and girls, based on their self-identification as male or female [Eagly and Karau, 2002]. These roles are shaped by two types of norms: descriptive norms, which reflect consensual beliefs about what members of a group typically do (e.g., "girls are more caring than boys," often aligning with traditional gender stereotypes), and injunctive norms, which define what members of a group should do (e.g., "boys should be brave") [Cialdini and Trost, 1998]. Research in Developmental Psychology has highlighted how these norms are internalized during early childhood, shaping developmental pathways and behaviours [Fulcher and Coyle, 2011].

Gender norms become further internalized through early play and childhood interactions. Boys are not only encouraged to be brave and adventurous, while girls are encouraged to be well-behaved, but they are also frequently steered toward different types of play. Boys are guided toward action-oriented or movement-based games, as well as those involving experimentation and construction, including electronic games. In contrast, girls are often surrounded by dolls and other toys that reflect the societal roles expected of them in the future; see, among others, Davis and Hines [2020], Kung [2022], and Harbin [2023]. Later in life, these early experiences further shape societal dynamics and behaviours ([Francis, 2010]), including decision-making processes such as the explore-exploit dilemma.

This study employs a lab experiment to examine whether gendered patterns emerge in explore-exploit tasks. For this purpose, we use and adapt the Grain Game designed by Chin et al. [2023] to study how

the potential to experience losses during a search affects individuals' willingness to explore. In the game, participants face the decision whether to *explore* new opportunities, *exploit* current gains, or *retreat* to safer options. Participants face two treatments, one in which only gains are possible and another in which both gains and losses are possible.[1,2] However, we depart from Chin et al. [2023] because we introduce competition. Willingness to compete is an important individual characteristic that, together with risk aversion, self-confidence, and loss aversion, may shape decisions in explore-exploit tasks, especially in job-seeking-related environments. To study the impact of the competitive pressure on these decisions, we nest the Grain Game into the framework of Niederle and Vesterlund [2007]. Hence, participants play the Grain Game three times, each under a distinct payment scheme: (1) a piece rate payment, (2) a tournament format within groups of four, and (3) a choice between piece rate payment or participating in the tournament again.

Niederle and Vesterlund [2007] examine willingness to compete within a stereotyped context, such as a math-related task, where women are often perceived as weaker performers. In the Grain Game, instead, the task is gender-neutral: participants play the role of farmers tasked with maximizing their harvests by planting seeds in a linear field. To further ensure gender neutrality, each player controls an on-screen character that matches their own gender. We deliberately chose an abstract and gender-neutral task to isolate individuals' attitudes toward exploration, independent of environmental influences, creating a setting where men and women should feel equally comfortable.

Our results provide insights into gender differences in exploratory and competitive behaviours. However, they challenge some of our initial hypotheses. Contrary to expectations, in the *Gain* treatment, women explore significantly more than men, particularly during the Tournament and Choice stages. However, when potential negative payoffs are introduced in the *Gain & Loss* treatment, no gender differences in exploration are observed. Random-effects multinomial logit estimates confirm that, in *Gain* environments, being female increases the likelihood of exploring rather than exploiting, but this effect disappears in the *Gain & Loss* treatment.

Regarding competition, we find that in the *Gain* treatment, women are less likely than men to select the competitive scheme (21% vs. 36%). This gap disappears in the *Gain & Loss* treatment (27% vs. 28%). Further analysis reveals that this gender gap in the *Gain* setting is fully explained by individual differences in confidence, risk preferences, and related traits, offering no support for our third hypothesis. Similarly,

[1]The two treatments are relevant because participants are likely to be motivated to avoid losses, which increases exploration when they are incurring losses but decreases exploration when they face the potential for losses. The authors predict and find this result.

[2]A possible interpretation of the two treatments in our motivating example of job search strategies is as follows. In the Only Gains environment, players face no risk of negative outcomes, mirroring scenarios where job seekers can experiment without fear of setbacks. Exploration involves pursuing new opportunities with lower pay but growth potential, exploitation reflects staying in stable roles with modest gains, and retreat signifies neutral actions like reskilling. In the Gains and Losses environment, setbacks such as unemployment or financial strain are possible, making exploration riskier, akin to applying for challenging positions. Exploitation represents safer, stagnant choices, while retreat entails reskilling, with potential opportunity costs.

we find no evidence that women in STEM fields are more competitive than men, as the interaction between gender and STEM major has no significant effect.

Finally, our results highlight the role of individual characteristics in shaping exploration and competition decisions. Participants who are more risk-averse or loss-averse tend to explore less and are less inclined to compete. However, we would like to stress that women do not shy away from competition in the *Gain and Losses* treatment, where men and women exhibit very similar behaviours across all stages. Notably, the gap in willingness to compete is entirely eliminated in this treatment.

Although we did not initially formulate a specific assumption on payoffs, our findings reveal slight, consistently male-favoring gaps in payoffs across the stages and treatments of the game, with the gender gap becoming statistically significant only in the Choice stage of the *Gain* treatment. Focusing on this specific stage of the *Gain* treatment, our results show that not only do women not explore less than men, but they explore more while earning less. This suggests that women may actually *overexplore*. This interpretation is supported by a comparison with a benchmark derived from reinforcement learning models.

Section 7 discusses how insights from our lab experiment relate to real-world dynamics in early career decision-making. While evidence suggests that women accept jobs earlier than men, potentially reflecting less exploration, our findings show that women tend to overexplore in the *Gain* treatment, resulting in lower payoffs. On a positive note, in the *Gains and Losses* treatment, all gender disparities vanish, with men and women behaving and competing equally, suggesting that high-stakes, riskier frameworks might mitigate women's overexploration and reluctance to compete.[3] We elaborate on policy implications based on these intriguing results in the concluding section.

The remainder of the paper is organized as follows. Section 2 reviews the related literature, providing context for our study, it also outlines and motivates our hypotheses. Section 3 details the experimental design, while Section 4 describes the data used in the analysis. Section 5 presents our main findings, focusing on measures of exploration, willingness to participate in the tournament in the Choice stage, and participants' payoffs. Section 6 offers an overview of the results, followed by Section 7, which explores an application to job seekers' behaviours. Section 8 provides concluding remarks.

2 Literature Review

Our paper contributes to two key strands of literature. The first examines the dynamics of the explore-exploit dilemma using experimental methods, focusing on how individuals navigate trade-offs between exploration

[3]This pattern resonates with the "gender-equality paradox" in STEM fields, where one possible explanation is that the high-stakes nature of entering STEM careers drives women in less gender-equal countries toward these fields, while in wealthier, gender-equal societies, women may pursue careers more aligned with personal preferences; see Stoet and Geary [2018].

and exploitation. The second investigates the determinants of gender differences in competitiveness, building on the seminal work of Niederle and Vesterlund [2007]. With respect to the first strand, we analyze a specific version of the explore-exploit dilemma, focusing on gender differences, which have not been studied in this context before. Regarding the second strand, we introduce a novel variation in the task used to study competitive behaviour, providing fresh insights into this well-researched area.

The exploration–exploitation dilemma, the trade-off between seeking new options and capitalizing on known ones, has been widely investigated across diverse areas, including animal foraging behaviour [Mehlhorn et al., 2015], computational models of reinforcement learning [Yogeswaran and Ponnambalam, 2012], procurement under incomplete information [Azoulay-Schwartz et al., 2004], organizational learning and performance [March, 1991, Gupta et al., 2006], and the neural mechanisms involved in decision-making [Daw et al., 2006, Blanchard and Gershman, 2018], just to name a few. For broader overviews, see also Mehlhorn et al. [2015], Berger-Tal et al. [2014].

In psychological, behavioural, and cognitive sciences, the exploration–exploitation dilemma has also been extensively analyzed through multi-armed bandit (MAB) frameworks, which model how individuals learn to allocate choices among multiple options with initially uncertain payoffs [e.g., Berry and Fristedt, 1985, Steyvers et al., 2009, Speekenbrink and Konstantinidis, 2015, Reverdy et al., 2014, Wu et al., 2018, Schulz et al., 2020, 2019, Meder et al., 2021]. Notably, a number of these papers focus on children, modelling how younger participants develop and adjust their exploration strategies [Meder et al., 2021, Wilson et al., 2021, Schulz et al., 2019].

The Grain Game developed by Chin et al. [2023], which we adapt in this paper, shares some similarities with Wu et al. [2018] in that participants explore a space where rewards are locally correlated, it is unique in that exploration is restricted to one step at a time. This constraint more closely mimics real-world settings in which exploring vast spaces is neither free nor costless, reflecting the practical limitations people face when gathering information.

While several papers have utilized lab experiments to examine individual behaviours in explore-exploit tasks, to the best of our knowledge, this study is the first to specifically investigate gender differences within this framework.

Considering now more in details the second strand of the literature we contribute to. Niederle and Vesterlund [2007] established a foundational result in experimental economics, showing that men are more likely than women to self-select into competitive environments. This finding has been widely replicated [Niederle and Vesterlund, 2011, Azmat and Petrongolo, 2014, Buser et al., 2014], yet subsequent research has demonstrated that the magnitude and even the presence of this gender gap depend on various contextual factors. For example, the nature of the task influences competitiveness, with men often being more competitive

in stereotypically male-dominated tasks, while gender-neutral or female-oriented tasks reduce or eliminate the gap [Dreber et al., 2014, Große and Riener, 2010, Shurchkov, 2012]. Cultural and societal differences also play a significant role, as cross-cultural studies have shown that gender gaps in competitiveness vary according to levels of gender equality and prevailing norms [Gneezy et al., 2009, Booth and Nolen, 2012]. Furthermore, differences in the conceptual frameworks used, such as variations in task difficulty or how competition is framed, highlight the contextual dependency of these differences [Buser et al., 2017]. This body of evidence highlight the importance of incorporating task-specific, cultural, and methodological dimensions into the study of gender and competitiveness. In this study, we contribute to this literature by employing the explore-exploit dilemma as a novel task to investigate gender differences in competitive behaviour, shifting the focus to decision-making under uncertainty in a gender-neutral context.

2.1 Hypothesis

Following the existing literature, we formulate the following hypotheses to address gendered differences in exploration and competition behaviours. These hypotheses were pre-registered prior to data collection to ensure transparency and credibility in our research design.[4] The first hypothesis addresses gender differences in behaviour within the context of the explore-exploit dilemma in absence of strategic uncertainty due to competition.

Hypothesis 1 *Women explore less than men, particularly in decision-making contexts that entail the possibility of facing losses.*

Gender differences in exploration behaviours may stem from early socialization processes and societal expectations regarding risk-taking. Research in developmental psychology suggests that gender norms, internalized from an early age, influence individuals' behaviours and decision-making processes in adulthood [Eagly and Karau, 2002, Fulcher and Coyle, 2011]. Women are often socialized to be more risk-averse and cautious, which may lead to lower exploratory behaviour in uncertain environments [Croson and Gneezy, 2009]. This tendency is further exacerbated in the presence of potential losses, as women exhibit greater loss aversion than men, inhibiting exploration in such contexts [Eckel and Grossman, 2008, Dohmen et al., 2011]. Additionally, women may adopt a more conservative approach due to early exposure to gendered play patterns that discourage risk-taking and adventurous decision-making [Kung, 2022, Harbin, 2023]. Women are also found to have higher levels of anxiety then men about spacial navigation [Lawton, 1994].

[4]The link for the pre-registration is available at the following OSF webpage: `https://osf.io/24eah/?view_only=1a9950422ed54210a9da08ea2caa5654`

Our second and third hypotheses address gender differences in competitive environments, examining both behaviour within the explore-exploit dilemma (with a focus on stage 2) and the decision to compete (with a focus on stage 3).

Hypothesis 2 *Controlling for individual characteristics, risk aversion, and loss aversion, individuals (both men and women) who choose to compete in the Choice stage engage in greater exploratory behaviour.*[5]

Hypothesis 3 *Controlling for individual characteristics, risk aversion, and loss aversion, women are less likely to choose competition compared to men.*

Competition has been shown to increase exploratory behaviour across genders, as individuals seek to outperform their peers and maximize their outcomes [Gill and Prowse, 2012, Mago et al., 2016]. Competitive settings often act as a motivational force, driving individuals to adopt strategies that involve greater exploration and risk-taking to gain a competitive edge [Gërxhani et al., 2023]. However, significant gender differences persist in competitive environments, with women being less likely to self-select into competition [Niederle and Vesterlund, 2011].

This reluctance is often attributed to lower self-confidence, internalized social norms that discourage competitive behaviour in women, and perceived stereotypes about gender roles [Eagly and Karau, 2002, Francis, 2010]. Women may also experience a higher sensitivity to social penalties or a fear of underperformance in competitive contexts [Buser et al., 2017]. Moreover, studies highlight that men are more likely to engage in voluntary job-to-job mobility, reflecting a greater willingness to compete and explore new opportunities [Theodossiou and Zangelidis, 2009, Cortés et al., 2023].

Together, these findings highlight the complex interplay of individual traits, societal norms, and environmental factors in shaping exploratory and competitive behaviours across genders, making it crucial to account for these elements when analyzing decision-making in competitive settings.

Our fourth hypothesis is based on the attitude toward competitiveness of women attending STEM fields:

Hypothesis 4 *Women attending STEM fields of study are more competitive than men.*

[5]The pre-registered version of Hypothesis 2 stated: "*Controlling for individual characteristics, risk aversion and loss aversion, in competitive settings individuals (both men and women) explore more.*" While the two formulations are related, the version presented in the paper narrows the empirical focus to the Choice stage, where participants self-select into a competitive or non-competitive payment scheme. This allows us to test whether those who actively choose to compete also explore more—capturing the endogenous relationship between willingness to compete and exploratory behavior. We acknowledge, however, that this design does not allow for a within-subject comparison of behavior across incentive schemes (e.g., Tournament vs. Piece Rate), which would have been more directly aligned with the original formulation. This analytical choice reflects a practical constraint of the experimental design (i.e., fixed order and lack of random assignment in Choice, which is common in the experiments on competitiveness, see, e.g. Niederle and Vesterlund [2007]), and we have updated the hypothesis wording to match the operationalization in the data.

The under-representation of women in STEM fields has led to increased scrutiny of competitiveness in male-dominated environments. Some studies suggest that women pursuing STEM careers tend to adopt more competitive behaviours as they navigate traditionally male-dominated spaces [Buser and Yuan, 2019, Cárdenas et al., 2012]. Moreover, the "gender-equality paradox" suggests that in countries with fewer gender-equal opportunities, women are more likely to pursue challenging STEM careers as a means of economic advancement [Stoet and Geary, 2018]. Women in these fields may develop resilience and adaptability, making them more competitive than their male counterparts in similar environments.

Finally, our last hypothesis refer to the impact of individual characteristics:

Hypothesis 5 *Individual risk preferences, loss aversion, cognitive abilities, and self-confidence affect both the likelihood to explore and the likelihood to compete.*

A wide range of literature highlights the significant role of personality traits and economic preferences in shaping decision-making under uncertainty and competition [Camerer and Hogarth, 1999, Dohmen et al., 2010]. Higher risk aversion and loss aversion have been associated with a lower tendency to explore and compete, whereas cognitive abilities and self-confidence have a positive influence. Gender norms further intersect with these individual traits, influencing behaviours such as over-exploration and reluctance to engage in competition in specific environments [Chin et al., 2023]. The tendency to explore and compete is shaped not only by inherent traits but also by the external environment, which can encourage or inhibit certain behaviours.

3 Experimental Design

This study investigates gender differences in exploration and competition behaviours through a controlled incentivised laboratory experiment conducted at the BLESS (Bologna Laboratory for Experiments in Social Sciences)[6]. The experiment was pre-registered on OSF (https://osf.io/6kxsp) and received approval from the Bioethics Committee of the University of Bologna.

The participants took part in three computerised games presented in fixed order: the Grain Game [Chin et al., 2023], a Loss Aversion Task [Gächter et al., 2022] and the Bomb Risk Elicitation Task (BRET; Crosetto and Filippin 2013). These tasks were followed by a post-experimental questionnaire designed to collect additional data on socio-demographic characteristics, personality traits [Gosling et al., 2003],

[6]The experiment was conducted in Italian, an English version of the experimental instructions are reproduced in Appendix H.

competitiveness and cognitive abilities [Frederick, 2005]. See Figure 1 for a blueprint of the experimental design.

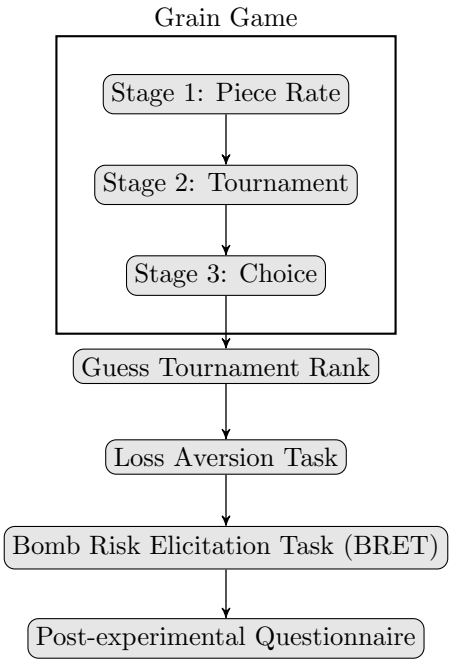


Figure 1: Experimental Design. *Note*: Participants first completed the Grain Game (composed of three stages: Piece rate, Tournament, and Choice), guessed their tournament rank, followed by the Loss Aversion Task, the BRET, and finally the post-experimental questionnaire.

3.1 The Grain Game

The Grain Game, adapted from Chin et al. [2023], placed participants in the role of farmers tasked with maximising their harvests by planting seeds in a linear field of 70 planting slots. During the course of the game, participants planted a total of 70 seeds, but made 69 decisions. This is because in the first round, participants automatically planted in position 1, the leftmost slot in the field, without making a choice. From the second round onwards, they could decide whether to plant in the same slot as the previous round (exploit), in the slot immediately to the right (explore) or in the slot immediately to the left (retreat).[7] In each round, a seed had to be planted. De facto constraining the movement of participants to one step at a time at most. An illustration of the game interface, showing the planting field and participant options, is provided in Figure 2. More screenshots can be found in Appendix A.

Payoffs in each slot were calculated as the sum of a fixed value derived from one of six predefined payoff sequences and random noise. The sequences were generated as random walks with errors drawn from a

[7]It follows that the furthest slot a participant could reach, given the 70 seeds and 69 decisions, is the planting slot in position 70.

Figure 2: Illustration of the Grain Game. *Note*: During the play, participants could see the total amount of accumulated points in the top left of the screen and the remaining number of seeds or rounds to play in the top right. They could also reread the instructions at any time by clicking on the question mark icon on the screen.

normal distribution with mean 0 and standard deviation 2.[8] The noise followed a symmetric distribution with probabilities assigned as follows: 0 points with a probability of 0.56, ± 1 point with a probability of 0.15 for each, and ± 2 points with a probability of 0.07 for each. Thus, the payoffs are locally correlated.

As in Chin et al. [2023], the experiment included two treatments. In the Gain treatment, participants encountered only positive payoffs, with fixed values ranging from 1 to 25 points. In the Gain & Loss treatment, participants faced potential losses as payoffs ranged from -10 to 15 points. To account for the possibility of negative payoffs, participants in the Gain & Loss treatment received an initial endowment of 700 points.[9]

Differently than Chin et al. [2023], we used three payoff sequences for each treatment. The sequences in the Gain & Loss treatment were identical to those in the Gain treatment, with a -10 offset applied to all values. All sequences rewarded exploration in the sense that the global maximum of every payoff sequence was placed in the second half of the field. Participants played the Grain Game three times, with each stage using a different compensation scheme, following the experimental framework of Niederle and Vesterlund [2007]. Therefore, the first part of the Niederle and Vesterlund [2007], where participants are paid piece rate represents the replication of Chin et al. [2023] with the only difference that we vary the sequences of

[8]The sequences can be found in Appendix B.

[9]The average payoff in the Gain treatment was 15 points and 5 points in the Gain & Loss treatment. The 700-point endowment ensured equal expected payoffs across the two treatments.

payoffs encountered by participants within treatments while in Chin et al. [2023] participants all face the same sequence. At the end of the game, only one of the three stages was randomly selected for payment and participants were informed of this rule before the experiment began.

In the first stage, participants earned 1 cent for each point they scored if this stage was selected for payment. In the second stage, participants were divided into gender-balanced groups of four, with two men and two women in each group. In this stage, only the participant with the highest score in the group earned 4 cents for each point if the stage was selected for payment, while the other three participants received nothing. If there was a tie, the winner was selected randomly among those with the highest scores.

In the third stage, participants could choose between the piece-rate scheme or the tournament scheme before starting the game. This choice is a proxy for competitiveness. If this stage was selected for payment, those who chose the piece-rate scheme earned 1 cent for each point as in the first stage. Those who chose the tournament scheme earned 4 cents for each point only if they outscored the top scorer from their group in the second stage (Tournament); otherwise they received no payment. If their score tied with the prior top scorer, the winner was selected randomly.

Each participant played all three payoff sequences within their assigned treatment. The order of sequences was randomised across participants at the session level, to mitigate potential order effects.

At the end of the Grain Game, participants were asked to guess their ranking in the Tournament stage. Correct guesses were incentivized as in Niederle and Vesterlund [2007].

3.2 Loss aversion task and BRET

After completing the Grain Game, in the second part of the experiment, participants were asked to perform two additional tasks: a loss aversion task and a task to measure their risk attitudes. Before starting, participants were informed that their performance in only one of these two tasks would be randomly selected for payment of part 2.

To assess participants' loss aversion, we implemented the task developed by Gächter et al. [2022]. Participants were presented with six small-stakes lotteries. For each lottery, they decided whether to participate in a coin toss. In all six lotteries, participants could win €6 if the coin landed on tails, but they risked losing an increasing amount of money (from €2 in the first lottery to €7 in the sixth) if the coin landed on heads. If participants opted not to participate, they received €0. If the loss aversion task was selected for payment, one of the six lotteries was randomly chosen to determine their final earnings.

The Bomb Risk Elicitation Task (BRET), adapted from Crosetto and Filippin [2013], was used to measure participants' risk attitudes in a controlled and incentivised setting. Participants were presented with a grid

of 100 cells (10 x 10) and asked to decide how many cells to colour red. Each red cell represented a potential gain of 5 cent. After their decision, the computer randomly selected one cell. If the chosen cell was grey, participants earned 5 cent for each cell they chose to colour red. However, if the chosen cell was red, they earned nothing. If the BRET task was selected for payment, their earnings were calculated based on the outcome of the task.

For an illustration of the Loss Aversion Task and BRET interfaces, see Appendix C.

3.3 Experimental Payments

At the end of the experiment, participants' final payment depended on the payoff from one randomly selected stage of the Grain Game (Piece Rate, Tournament, or Choice), on the payoff from one randomly selected additional task (Loss Aversion or BRET), plus a fixed 5 Euro show-up fee. Detailed payoff rules for each stage and task are summarized in Table 1. Formally:

$$\begin{aligned}\text{Final Payment} = & \underbrace{(\text{Payment from one of the three Grain Game stages})}_{\text{randomly chosen}} \\ & + \underbrace{(\text{Payment from one of the two additional tasks})}_{\text{randomly chosen}} \\ & + \underbrace{5\,\text{Euros}}_{\text{show-up fee}} .\end{aligned}$$

Table 1: Overview of the Experimental Compensation Scheme

Part	**Description**
Part 1: Grain Game	**Stage 1 (Piece Rate)**: 1 cent per point. **Stage 2 (Tournament)**: Groups of 4 (2 men, 2 women); highest score earns 4 cents/point (ties random). **Stage 3 (Choice)**: Each participant chooses either (i) to not compete and get Piece Rate payment (1 cent/point), or (ii) to compete and get Tournament payment (4 cents/point if beating the top scorer from Stage 2, else 0).
Part 2: Additional Tasks	**Loss Aversion Task**: One of six coin-toss lotteries randomly determines payoffs. **BRET**: Earn 5 cents per chosen red cell unless the random draw is red; then earnings are 0.
Show-up Fee	A fixed 5-Euro payment, independent of performance.

4 Data

Our pre-registered plan aimed to collect data from 432 participants (50% female). To achieve this, we invited 432 individuals to take part in the study, which was conducted at the Bologna Laboratory for Experiments in Social Sciences (BLESS) between June 2022 and March 2023. Recruitment was carried

out through the ORSEE system [Greiner, 2015], ensuring random assignment to sessions and a balanced gender composition across experimental groups. Each session accommodated up to 32 participants, who were randomly assigned to gender-balanced groups of four (two men and two women). However, due to a software problem encountered in the Grain Game, data from 13 participants were lost, resulting in a final sample of 419 participants (206 women and 213 men) for the analysis presented in this paper.

Table 2 summarises the distribution of participants across treatments and genders.

Table 2: Participants by Treatment and Gender

	Gain	**Gain & Loss**	**Overall**
Men	106	107	213
Women	104	102	206
Overall	210	209	419

While the main sample consists of 419 participants, data for the loss aversion task and the BRET are unavailable for the 62 participants from the first two sessions due to technical issues. As a result, depending on the model specification used in the paper, the sample size might be smaller. In addition, in the post-experimental questionnaire, some individuals chose to skip certain questions, so there is missing data from that source as well. However, the sample size used in each regression is always specified in the Results section, and a table with descriptive statistics for all relevant variables is provided in Appendix D.

Furthermore, when constructing the λ_{loss} variable, our measure of loss aversion based on the loss aversion task, we exclude participants whose decisions are non-monotonic (following the procedure in Gächter et al. [2022]). This exclusion generates additional missing values for those participants.

5 Results

In this section we present our results organized as follow: Section 5.1 focuses on the behaviour in the explore-exploit dilemma under piece rate. Section 5.2 analyse the willingness to compete and section 5.3 discusses the differences in earnings.

5.1 Exploration

5.1.1 Descriptive Evidence

We begin by examining descriptive evidence on exploration behaviour and whether it differs by gender and treatment. We use three different measures to study gender gaps in exploration: (i) the average percentage of "explore" choices; (ii) the farthest position ("*Maximum position*") that participants reach at any point in

the game, and (iii) the final position where they plant their last seed ("*Final position*").[10]

Figure 3 shows the average percentage of *Explore*, *Exploit*, and *Retreat* decisions across each stage of the Grain Game, disaggregated by gender (men/women) and by treatment condition (Gain vs. Gain & Loss).

Contrary to our pre-registered hypothesis 1, women do not explore less than men. Instead, we find evidence suggesting that women, in some specific settings, explore more than men, particularly in the Tournament and Choice stages of the Grain Game.

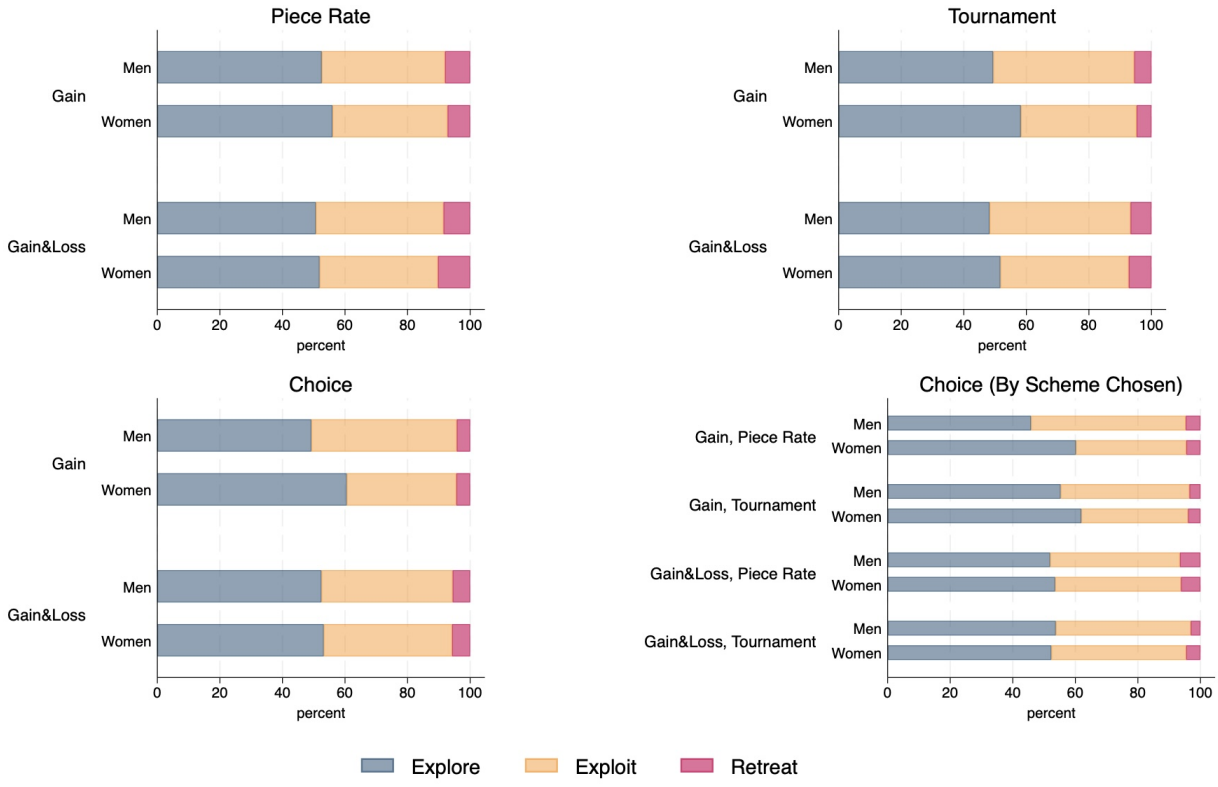


Figure 3: Explore, Exploit, Retreat % Across Stages, Gender and Treatment. *Note*: This figure displays the proportion of exploration, exploitation and retreat choices made by participants in each stage of the experiment, disaggregated by gender (men/women) and by treatment condition (Gain vs. Gain & Loss).

In the Tournament stage, women explore in 55.0% of decisions (SD = 24.9%), compared to 48.8% for men (SD = 25.7%). This difference is statistically significant ($t = -2.50$, $p = 0.0130$). Similarly, in the final choice stage, women explore in 56.9% of decisions (SD = 26.7%), compared to 50.9% for men (SD = 27.9%), and this difference is also statistically significant ($t = -2.28$, $p = 0.0234$).

When analysing by treatment, the gender difference is driven by the Gain treatment. In this setting, women explore 58.3% of decisions in the Tournament stage (compared to 49.4% for men; $t = -2.47$, $p =$

[10]Because participants can retreat after planting in a new slot, *maximum position* and *final position* do not necessarily coincide.

0.0143) and 60.6% of decisions in the final choice stage (compared to 49.2% for men; $t = -3.02$, $p = 0.0028$). In the Gain & Loss treatment, gender differences in exploration are not statistically significant.

The other two measures, "*Maximum position*" and "*Final position*", confirm that in the Tournament and Choice stages under the Gain treatment, women tend to explore more than men; specifically, they reach a higher maximum position and conclude the game farther to the right. By contrast, in the Gain & Loss treatment, no significant gender difference emerges. Detailed results on those two additional measures and the percentage of explore are provided in Appendix E.

5.1.2 Regression Analysis

To assess whether women explore more or less than men controlling for individual characteristics, we focus on the *Choice* stage of the Grain Game.[11] This stage is particularly interesting because it allows us to verify if and how competition is associated with the players' choices. In this stage, as in every stage, each participant plants a total of 70 seeds and therefore makes 69 sequential decisions over time.

We define $Move_{it}$ as a categorical variable indicating the participant's i decision at round t, where $Move \in \{\text{Explore}, \text{Exploit}, \text{Retreat}\}$. To analyze these decisions, we estimate three random-effects multinomial logistic regressions with *Exploit* as the baseline outcome. Our general model can be written as:

$$\begin{aligned} \ln\!\left(\frac{\Pr(Move_{it}=m)}{\Pr(Move_{it}=Exploit)}\right) = \beta_{0,m} \;+\; \beta_{1,m}\, Female_i \;+\; \beta_{2,m}\, Gain \;\&\; Loss_i \\ +\; \beta_{3,m}\,[Female_i \times Gain \;\&\; Loss_i] \;+\; \beta_{4,m}\, LagPayoff_{it-1} \\ +\; \beta_{5,m}\, Threshold_{it-1} \\ +\; \beta_{6,m}\, Compete_i \;+\; \boldsymbol{\gamma}_m' \mathbf{X}_i \;+\; u_{i,m}, \end{aligned} \tag{1}$$

where $m \in \{Retreat, Explore\}$, $Female_i$ is a dummy taking value 1 for women and 0 otherwise, and $Gain \;\&\; Loss_i$ indicates whether participant i is assigned to the Gain & Loss treatment (= 1) or not (= 0). $LagPayoff_{it-1}$ is the payoff from the previous planting round, adjusted by adding 10 points in the Gain & Loss treatment to ensure comparability with the non-negative payoffs in the Gain setting. $Threshold_{it-1}$ is a dummy set to 1 if the adjusted payoff in the previous round was below 10 (i.e., a negative payoff in the Gain & Loss environment), and 0 otherwise. $Compete_i$ indicates whether participant i chose the competitive payment scheme in this final stage of the Grain game. Finally, $u_{i,m}$ denotes participant-level random intercepts, and $\mathbf{X}_i$ is a vector of additional participant-level controls (e.g., demographics, risk attitudes, personality traits) which differ across model specifications.[12]

[11] In Appendix F, we report the regression results for the *Piece Rate* and *Tournament* stages.

[12] As a robustness check, we also estimated these regressions while controlling for the payoff sequences that participants ultimately observed. For brevity, we do not include those additional results here, but they are available upon request. The interpretation of the hypothesis remains unchanged.

We estimate three versions of Equation (1) that differ in the participant-level controls included in $\mathbf{X}_i$. Model A is the baseline and includes only the core regressors shown in Equation (1) ($\mathbf{X}_i$ is empty). Model B augments the baseline by adding socio-demographic characteristics (age, nationality, field of study and GPA), personality and cognitive measures (TIPI, CRT), piece-rate payoff, competitiveness, and self-reported risk attitudes. Finally, Model C replaces the self-reported risk attitude with the incentivized measure from the BRET task. It also includes the loss-aversion parameter λ_{loss} from the loss-aversion task.[13] Table 3 presents the results from these three random-effects multinomial logit regressions.

Table 3: Random Effects Multinomial Logistic Results for *Move*

Dependent Variable	*Move* (Baseline = *Exploit*)					
	Model A		**Model B**		**Model C**	
	Retreat	Explore	Retreat	Explore	Retreat	Explore
Female	1.165	2.096*	1.199	2.330*	1.131	2.305
	(0.360)	(0.782)	(0.409)	(0.869)	(0.485)	(1.097)
Gain & Loss	0.996	1.171	1.147	1.190	0.730	1.323
	(0.340)	(0.430)	(0.424)	(0.443)	(0.297)	(0.564)
Female × (Gain & Loss)	1.250	0.432	1.070	0.353*	1.444	0.269*
	(0.583)	(0.226)	(0.520)	(0.181)	(0.801)	(0.161)
Lag Payoff	0.722***	0.705***	0.723***	0.704***	0.729***	0.707***
	(0.022)	(0.021)	(0.022)	(0.021)	(0.024)	(0.024)
Threshold	0.265***	0.565***	0.276***	0.579***	0.277***	0.587***
	(0.055)	(0.074)	(0.058)	(0.078)	(0.064)	(0.094)
Piece Rate Payoff	–	–	0.9999	1.0039***	0.9999	1.0040***
			(0.0008)	(0.0009)	(0.0010)	(0.0012)
Risk (self-reported)	–	–	1.091	1.281**	–	–
			(0.084)	(0.097)		
Risk (Bret)	–	–	–	–	1.015	1.021*
					(0.010)	(0.011)
λ_{loss}	–	–	–	–	0.518**	0.545**
					(0.106)	(0.125)
Competitiveness (self-reported)	–	–	0.966	1.074	0.936	1.077
			(0.056)	(0.070)	(0.063)	(0.082)
CRT	–	–	1.227	0.977	1.295	1.023
			(0.152)	(0.128)	(0.181)	(0.160)
Extraversion	–	–	1.036	0.967	0.984	0.989
			(0.079)	(0.083)	(0.088)	(0.105)
Agreeableness	–	–	1.553***	1.331*	1.480**	1.444*

(Continued on next page)

[13] Because some participants did not complete the BRET and the loss-aversion task and because non-monotonic (inconsistent) responses in the loss-aversion task were dropped following Gächter et al. [2022], Model C is estimated on a slightly smaller sample.

	Model A		Model B		Model C	
	Retreat	Explore	Retreat	Explore	Retreat	Explore
			(0.187)	(0.168)	(0.202)	(0.209)
Conscientiousness	–	–	1.028	0.928	0.989	1.000
			(0.095)	(0.094)	(0.105)	(0.119)
Emotional Stability	–	–	0.903	0.800*	1.042	0.883
			(0.081)	(0.082)	(0.099)	(0.099)
Openness	–	–	0.837	0.826	0.896	0.953
			(0.107)	(0.108)	(0.128)	(0.150)
GPA	–	–	0.950	0.921	0.875	0.957
			(0.148)	(0.144)	(0.152)	(0.170)
Born in Italy	–	–	0.263*	0.229*	0.192*	0.136**
			(0.152)	(0.134)	(0.135)	(0.106)
Age	–	–	0.987	1.016	0.988	1.041
			(0.033)	(0.042)	(0.047)	(0.055)
STEM	–	–	0.726	0.782	0.910	0.993
			(0.196)	(0.215)	(0.281)	(0.317)
Compete	0.627	1.383	0.600	0.999	0.449*	0.804
	(0.165)	(0.395)	(0.167)	(0.284)	(0.148)	(0.274)
Constant	4.513***	222.366***	5.823	16.359	21.436	5.098
	(2.134)	(114.842)	(10.166)	(31.442)	(42.361)	(11.517)
Model Statistics						
Observations	28,492		27,676		21,080	
Number of Groups	419		407		310	
Wald χ^2	193.85		242.75		216.16	
Log pseudolikelihood	−16139.104		−15671.713		−11755.302	

Notes: All entries are Relative Risk Ratios (RRRs) from random-effects multinomial logit regressions. Robust standard errors (clustered at participant level) are shown in parentheses. The dependent variable is *Move*, and the baseline (omitted) outcome is *Exploit*. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. While most variables are self-explanatory, note that λ_{loss} is our measure of loss aversion, where higher values indicate higher loss aversion, and *CRT* is the Cognitive Reflection Test score, where a higher value indicates more reflective (rather than intuitive) thinking.

The estimates from the random-effects multinomial logistic models indicate that in the gain treatment women are significantly more likely than men to explore (rather than exploit) in the final (Choice) stage of the Grain Game. In Models A and B of Table 3, the coefficient on *Female* in the "Explore vs. Exploit" comparison exceeds 2.0 and is statistically significant at the 5% level. In Model C, which relies on a smaller sample, the coefficient remains above 2.0 but is only marginally significant ($p \approx 0.08$). An RRR above 2.0 implies that, for women, the relative probability of choosing to *Explore* (compared to *Exploit*) is more than twice the relative probability for men.

By contrast, the interaction term *Female*×*Gain & Loss* is below 1.0 in all three models and is signifi-

cantly different from 1.0 in Models B and C. This pattern suggests that the higher propensity of women to explore, relative to men, is noticeably decreased when negative payoffs become possible. Furthermore, linear combination tests confirm that under *Gain & Loss* the net female effect fully vanishes in every specification, implying no statistically discernible difference in exploratory behaviour between women and men in the Gain & Loss tretament.[14] Hypothesis 1 is thus not supported.

Across all models, a larger payoff in the previous round strongly reduces participants' tendency to deviate from exploitation in the subsequent round, and if the adjusted payoff from the prior round falls below a certain threshold (i.e., is negative under the Gain & Loss treatment), individuals become even less inclined to explore.

Notably, choosing the competitive scheme (*Compete*) does not appear to increase exploration in any of the three model specifications. In Model C, *Compete* significantly reduces "Retreat vs. Exploit" (RRR ≈ 0.45, $p < 0.05$), indicating that those who select the competitive scheme are less likely to retreat (as opposed to exploit) than those in the noncompetitive setting. However, this effect does not translate into a higher relative probability of exploring. Thus, we find no evidence that individuals who self-select into competition systematically explore more, which does not support Hypothesis 2.

Turning to other variables, risk attitudes (both self-reported and measured via the BRET) consistently raise the relative probability of exploring rather than exploiting. In Model B, self-reported risk is strongly associated with "Explore vs. Exploit" (RRR ≈ 1.28, $p < 0.001$), and in Model C the incentivized risk measure (*RiskBret*) is smaller but still significant (RRR ≈ 1.02, $p < 0.05$). By contrast, loss aversion (λ_{loss}) substantially reduces the RRR for both "Retreat vs. Exploit" and "Explore vs. Exploit" (RRRs of about 0.52–0.55), suggesting that more loss-averse participants are less likely to deviate from exploiting their current slot. Among the remaining controls, *Agreeableness* increases the likelihood of leaving an exploited slot (whether via retreat or exploration), whereas being born in Italy lowers it (RRRs < 1). *Competitiveness*, *CRT*, *STEM*, *GPA*, and *Extraversion* show no consistent effects. Hence, Hypothesis 5 is supported, in that risk preferences, loss aversion, some personality traits, and certain demographic characteristics significantly influence exploration decisions.

[14] Specifically, we test whether the sum of the main *Female* coefficient and the *Female*×*Gain & Loss* interaction coefficient is zero in each model. For Model A, $\chi^2(1) = 0.07$ and $p = 0.7843$; for Model B, $\chi^2(1) = 0.23$ and $p = 0.6283$; and for Model C, $\chi^2(1) = 1.24$ and $p = 0.2659$. Thus, in all cases we fail to reject the null, indicating that the net female effect under *Gain & Loss* is not significantly different from zero.

5.2 Competition

5.2.1 Descriptive Evidence

We now turn to Hypothesis 3, which posits that, controlling for individual characteristics, women are less likely than men to enter competition, which in our setting means we expect women to be less likely to choose the competitive payment scheme in the final (Choice) stage of the Grain Game. In this stage, participants had to decide between receiving a piece-rate compensation of 1 cent per point or challenging the highest scorer from the previous (Tournament) stage, with a higher payoff of 4 cents per point awarded only if they outscored the tournament winner.

Looking at the descriptive statistics, in the Gain Treatment women are significantly less likely to compete compared to men (21% vs 36% for men, p=0.0183, $\chi^2 = 5.5548$), while no gender difference is found in the Gain & Loss Treatment(27% for women vs. 28% for men p 1, $\chi^2 = 0.009$). See Figure 4.

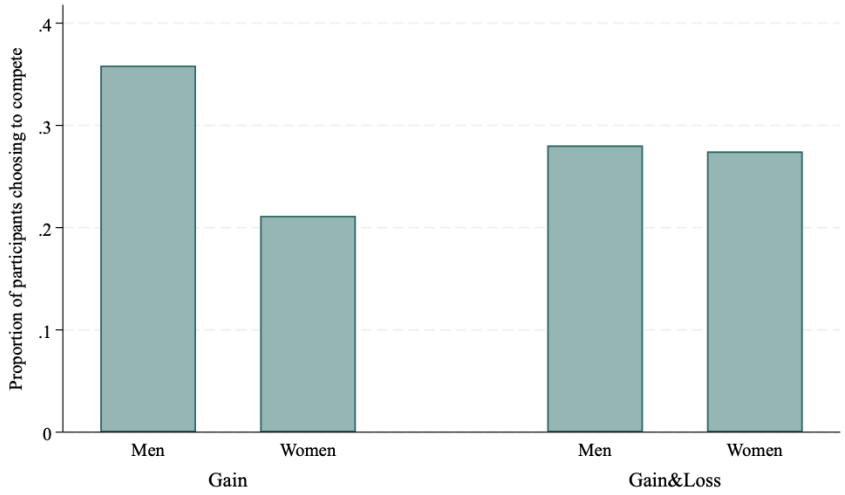


Figure 4: Decision to compete by Gender and Treatment. *Note*: This figure displays the proportion of participants who chose to compete in the Choice stage of the Grain Game disaggregated by gender (men/women) and by treatment condition (Gain vs. Gain & Loss).

5.2.2 Regression Analysis

To test Hypothesis 3, we estimate four logistic regressions of the binary decision to select the competitive payment scheme ($Compete = 1$) versus the piece-rate one ($Compete = 0$). Formally, let $\Pr(Compete_i = 1)$ denote the probability that participant i chooses to compete in the Choice stage. Our model is:

$$\ln\left(\frac{\Pr(Compete_i=1)}{\Pr(Compete_i=0)}\right) = \beta_0 \;+\; \beta_1\, Female_i \;+\; \beta_2\, Gain\&Loss_i \\ +\; \beta_3\left(Female_i \times Gain\&Loss_i\right) \;+\; \beta_4\, PieceRatePayoff_i \\ +\; \boldsymbol{\gamma}'\mathbf{X}_i \;+\; \varepsilon_i, \tag{2}$$

where $Female_i$ is a dummy variable equal to 1 for women and 0 for men, $Gain\&Loss_i$ indicates whether participant i is assigned to the Gain&Loss treatment (= 1) or not (= 0). $PieceRatePayoff_i$ denotes participant i's payoff in the Piece Rate stage (a proxy for ability), $\mathbf{X}_i$ is a vector of additional controls, and ε_i is the error term. We estimate four versions of Equation (2) that differ in the participant-level controls included in $\mathbf{X}_i$. In Model 1, we use the parsimonious specification, with no additional participant-level controls. Model 2 adds self-assessed relative rank from the Tournament stage (a proxy for perceived ability). Model 3 introduces self-reported risk attitudes, a self-reported measure of competitiveness, cognitive reflection (CRT), personality traits (TIPI), and demographic controls (including a STEM-major indicator). Finally, Model 4 replaces self-reported risk with the incentivized Bomb Risk Elicitation Task (BRET) measure and incorporates the loss-aversion parameter λ_{loss}.[15] Table 4 presents the results from all four models.[16]

Table 4: Logistic Regression Results on the Decision to Compete

Dependent Variable	*Compete* = 1 if Tournament, 0 if Piece Rate			
Model:	(1)	(2)	(3)	(4)
Panel A: AME, ×100				
Female	-7.28	-4.95	4.68	3.61
	(4.33)	(4.38)	(4.75)	(5.22)
Gain & Loss	-2.11	-1.03	-0.28	-0.55
	(4.33)	(4.27)	(4.23)	(4.78)
Piece rate payoff	0.03*	0.03	0.02	0.02
	(0.01)	(0.02)	(0.01)	(0.01)
Rank Belief	–	-10.13**	-9.80**	-11.78**
		(3.07)	(3.03)	(3.40)
Risk (self-reported)	–	–	2.88*	–
			(1.26)	
Competitiveness (self-reported)	–	–	3.49**	3.43**
			(1.01)	(1.11)
Risk (Bret)	–	–	–	0.31*
				(0.15)
λ_{loss}	–	–	–	-12.58**
				(4.44)
CRT	–	–	1.50	0.88
			(2.12)	(2.43)
Extraversion	–	–	0.87	2.44
			(1.49)	(1.63)
Agreeableness	–	–	2.42	4.27
			(2.24)	(2.48)
Conscientiousness	–	–	-0.52	0.10
			(1.66)	(1.79)
Emotional Stability	–	–	1.46	2.18
			(1.74)	(2.01)
Openness	–	–	-0.78	0.28
			(2.25)	(2.48)
GPA	–	–	-6.64**	-7.36*
			(2.52)	(2.89)
Born in Italy	–	–	6.98	6.03

(Continued on next page)

[15] Following Gächter et al. [2022], we exclude non-monotonic responses in the loss-aversion task. Because some participants did not complete these tasks or provided non-monotonic responses, the sample size is smaller in Model 4 than in the other specifications

[16] As a robustness check, we also ran regressions controlling for both piece-rate performance and tournament performance, and found no difference in the interpretation of the results.

			(7.47)	(7.65)
Age	–	–	0.75	0.94
			(0.70)	(0.91)
STEM	–	–	3.52	4.60
			(4.42)	(4.90)
Observations	419	419	407	310
Pseudo R^2	0.0210	0.0480	0.1169	0.1693
Panel B: Female AME (conditional on Gain vs. Gain&Loss), $\times 100$				
AME(Female | Gain only)	-14.55*	-11.07	-1.68	0.35
	(6.23)	(6.23)	(6.26)	(7.73)
AME(Female | Gain & Loss)	-0.19	1.16	11.17	5.94
	(6.04)	(6.04)	(6.42)	(6.36)

Note: $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$. Panel A reports average marginal effects (AME), multiplied by 100, with robust standard errors (also $\times 100$) in parentheses. Panel B shows the AME of being *Female* conditional on (*Gain*) vs. (*Gain & Loss*), again multiplied by 100. While most variables are self-explanatory, note that λ_{loss} is our measure of loss aversion, where higher values indicate higher loss aversion, and *CRT* is the Cognitive Reflection Test score, where a higher value indicates more reflective (rather than intuitive) thinking.

As we can see in Panel A of Table 4 being female has no significant effect on the probability of competing in any specification. Panel B of Table 4 shows the average marginal effects (AMEs) of being female under the two treatments: Gain and Gain & Loss, separately. Focusing on Model (1), we see that in the Gain treatment, women are on average 14.6 percentage points less likely ($p < 0.05$) to choose to compete than men. By contrast, in Gain & Loss treatment, the AME of being female is statistically insignificant. Once we include participants' self-assessed rank (Model 2) and additional controls for risk preferences, competitiveness, cognitive reflection, personality traits, and demographics (Models 3 and 4), the gender gap in "Gain" disappears, dropping from -14.6 to -1.7 percentage points or smaller and no longer statistically significant. In other words, the initial female disadvantage in choosing competition under gains can be fully explained by differences in confidence, risk attitudes, personality traits, competitiveness and demographic characteristics.

Turning to the other covariates, *Rank Belief* has a significant and negative effect on the decision to compete in every specification where it is included.[17] Unsurprisingly, both *risk attitude* measures (self-reported risk (Model 3) and the BRET measure (Model 4)) also affect the decision to compete, with more risk-averse individuals being less likely to chose the competitive payment scheme. Similarly, *loss aversion* reduces the probability of competing, whereas *competitiveness* increases it. Finally, an increase in *GPA* is associated with a decrease in the probability of competing. This finding also supports the competition aspect of Hypothesis 5, as loss aversion, risk preferences, and self-confidence (measured by rank belief) all affect the likelihood of entering competition.

Finally, in order to test whether women attending STEM fields of study are more competitive than men, we run two additional specifications based on Equation (2). Specifically, we extend Model 3 and Model 4 by adding an interaction term between *STEM* (an indicator for studying a STEM field) and *Female* (let's rename thes models Model 3a and Model 4a). These augmented models control for the same covariates as

[17] Recall that rank ranges from 1 (winner) to 4 (last place), so a one-point increase in believed rank (i.e. believing one performed worse) decreases the probability of competing by about 10–11 percentage points.

before (e.g.., risk aversion, competitiveness, GPA, and personality traits).[18].

Table 5: Female STEM vs. Men: Predicted Probability (%) of Choosing Competition (Models 3a and 4a)

	Model 3a (N=407)	Model 4a (N=310)
Panel A: Predicted Probabilities (%)		
Male Non-STEM	23.3	23.5
	(3.8)	(4.6)
Male STEM	29.1	28.7
	(4.3)	(4.8)
Female Non-STEM	30.4	27.8
	(4.3)	(4.9)
Female STEM	30.4	31.4
	(5.3)	(5.1)
Panel B: Key Pairwise Differences (p.p.)		
Female STEM − Male STEM	+1.3	+2.7
	(6.9)	(7.2)
	[p=0.85]	[p=0.71]
Female STEM − Male Non-STEM	+7.1	+7.9
	(6.6)	(6.8)
	[p=0.28]	[p=0.25]

Note: Panel A shows the predicted probability (%) of choosing the competitive payment scheme for each subgroup. Panel B shows the corresponding pairwise differences in percentage points (p.p.). Standard errors are in parentheses; p-values in brackets. Both Model 3a and Model 4a include the same covariates as Models 3 and 4, respectively, with the addition of an interaction between *Female* and *STEM*.

The results in Table 5 indicate that, once these covariates are accounted for, female STEM participants do not differ significantly from men (whether in STEM or not) in their probability of entering competition. All pairwise comparisons of marginal effects yield p-values above the conventional thresholds for statistical significance. Hence, we do not find evidence that women in STEM fields are more likely to choose the competitive payment scheme than men.

5.3 Participants' performance

5.3.1 Payoffs and maximum positions reached by treatment and gender

Table 6 reports average payoffs and maximum positions reached by gender and treatment for each of the three stages. While there are no significant differences between men and women in the Piece Rate and Tournament stages as for the payoffs, a marginally significant gap arises in the Choice stage, with men

[18]The regressions results are reported in Appendix G

earning more points than women overall. At the treatment level, this marginal significance is present only in the Gain treatment and not in Gain&Loss. Notably, men's payoffs also exceed women's in earlier stages and both treatments, but these differences are never statistically significant.

Turning to the maximum position reached by participants, gender differences are statistically significant in the Gain treatment during both the Tournament and Choice stages. Women tend to reach positions farther from their initial ones compared to men. Consistently, across all stages and Treatments—except for the Gain & Loss Treatment in the piece-rate stage—women systematically reach positions farther from their initial ones than men. However, these differences are never statistically significant in the Gain & Loss Treatment.

To sum up, women tend to earn fewer points than men and reach positions farther from their initial ones in the Gain Treatment. These results suggest that women underperform relative to men due to a tendency to over-explore in the Gain Treatment, whereas no gender differences in overall performance are observed in the Gain & Loss Treatment.

Since a closed-form solution to the exploit-explore dilemma participants face in our game does not exist, we employed a reinforcement learning model to derive a benchmark outcome. This benchmark allows us to further evaluate participants' performance by comparing it to their output in the piece-rate stage of the game, where no strategic interaction occurs among participants.

Table 6: Payoff and Maximum Position, by Gender, Treatment and Stage (Two-Sided p-values)

	Payoff			**Max Position**		
	Men (SE)	Women (SE)	p-value	Men (SE)	Women (SE)	p-value
Piece Rate						
Gain	836.64 (17.60)	828.57 (15.15)	0.7288	32.16 (1.59)	35.03 (1.71)	0.221
Gain & Loss	880.05 (13.03)	867.69 (13.72)	0.5141	30.58 (1.56)	30.23 (1.65)	0.876
Overall	858.45 (11.01)	847.94 (10.30)	0.4867	31.37 (1.11)	32.65 (1.20)	0.432
Tournament						
Gain	920.44 (13.78)	894.82 (13.16)	0.1803	31.73 (1.81)	38.09 (1.72)	0.012
Gain & Loss	912.95 (16.20)	896.09 (14.54)	0.4408	29.98 (1.73)	31.78 (1.73)	0.462
Overall	916.68 (10.62)	895.45 (9.77)	0.1426	30.85 (1.25)	34.97 (1.24)	0.020
Choice						
Gain	961.89 (11.10)	929.83 (14.73)	0.0829	32.37 (1.92)	39.67 (1.84)	0.007
Gain & Loss	947.35 (13.32)	929.43 (12.41)	0.3273	33.31 (1.86)	34.01 (1.90)	0.792
Overall	954.58 (8.67)	929.63 (9.62)	0.0544	32.84 (1.33)	36.87 (1.33)	0.033

Notes: Reported values are the means of payoff and max position (with standard errors in parentheses) for "Gain", "Gain & Loss" and "Overall" (pooled). Each row compares Men vs. Women. All p-values are two-sided from two-sample t-tests.

5.3.2 Reinforcement-learning benchmark in the piece-rate stage

We trained two reinforcement learning models, Q Learning and SARSA, to build a numerical benchmark for the first stage of the Grain Game. Each model plays the role of the farmer who learns through trial and error to maximise total points. At every slot, the farmer can *retreat* by moving one step left, *exploit* by planting in the same slot, or *explore* by moving one step right. The payoff in a slot equals the deterministic value of the underlying sequence plus a zero-mean random fluctuation, mirroring de facto the experimental conditions for the piece-rate stage (see Section 3).

The two models are trained on a grid of 1 728 parameter combinations. Each combination runs for 10 000 learning episodes followed by 5 000 test episodes. The full algorithmic details, grid search procedure, additional observations as well as the top performing configurations appear in Appendix I. We use the resulting outcomes as reference points to evaluate human performance, both in terms of final payoff and the maximum position achieved. We keep the configuration that attains the highest average payoff for each of the six payoff sequences. Recall that the first three sequences refer to the Gain Treatment, while the others refer to the Gain & Loss Treatment.[19]

Table 7 reports these six benchmarks both in terms of final payoff and maximum position achieved (second and sixth columns, respectively). Those outcomes are compared to the average experimental payoffs and the average maximum position achieved for male and female participants in each sequence of the piece-rate stage.[20] Participants fall short of the benchmark in all sequences except sequence 4, where they exceed it by about 350 points.[21] Since the algorithm's performance is unreliable in this sequence, we refrain from commenting on sequence 4.

When comparing human performance to the benchmark, we find that—excluding sequence 4—the benchmark payoff consistently exceeds that of the participants. Regarding the maximum position reached, we observe statistically significant deviations from the benchmark: over-exploration in sequences 3 and 6, with women tending to over-explore more than men, and under-exploration in sequence 5.

Across sequences, the difference in payoffs between male and female participants never exceeds approximately 40 points, and none of these differences is statistically significant.

Table 7: Piece-rate performance relative to reinforcement-learning (RL) benchmarks

	Payoff			**Maximum position**		

[19] See Appendix B for a description of the payoff sequences.

[20] Note that the information presented in Table 7 differs from that in Table 6, as the former reports performance by stage and treatment, while the latter reports performance by sequence.

[21] Both algorithms learn poorly in sequence 4, so the positive gap there reflects a weakness of the benchmark rather than exceptional human performance.

Seq	RL bench.	Gap: Overall	Gap: Men	Gap: Women	RL bench.	Gap: Overall	Gap: Men	Gap: Women
1	808.90	-62.43**	-61.92	-62.97*	35	-2.67	-3.38	-1.90
		(22.10) [87]	(34.07) [45]	(28.04) [42]		(1.82) [87]	(2.66) [45]	(2.49) [42]
2	933.68	-85.13***	-74.53**	-95.35***	29	2.93	2.07	3.75
		(14.58) [55]	(21.21) [27]	(20.23) [28]		(2.36) [55]	(3.19) [27]	(3.51) [28]
3	966.64	-36.61***	-29.20**	-44.02**	25	11.51***	8.74**	14.29***
		(8.37) [68]	(9.94) [34]	(13.49) [34]		(1.99) [68]	(2.51) [34]	(3.05) [34]
4	509.22	351.99***	362.63***	341.35***	43	-7.64***	-8.30**	-6.97*
		(24.23) [66]	(34.34) [33]	(34.61) [33]		(1.85) [66]	(2.42) [33]	(2.84) [33]
5	944.08	-100.67***	-83.81***	-120.08***	41	-14.08***	-12.71***	-15.67***
		(12.03) [71]	(16.79) [38]	(16.85) [33]		(1.94) [71]	(2.72) [38]	(2.78) [33]
6	977.57	-61.64***	-69.13***	-54.16***	15	14.31***	14.22***	14.39***
		(9.21) [72]	(13.75) [36]	(12.31) [36]		(1.96) [72]	(2.86) [36]	(2.72) [36]

Note. The RL benchmark payoff is the return generated by the best performing reinforcement learning algorithm described in Appendix I. The maximum-position benchmark is the furthest position reached by the same algorithm in each sequence. Each gap equals the experimental value minus the corresponding RL benchmark. Stars indicate two-sided t-test significance: * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$. Standard errors are in parentheses and observation counts in brackets.

6 Discussion

The findings challenge our initial expectations regarding gender differences in exploration (Hypotheses 1). Contrary to the notion that women explore less, we observe that, in the *Gain* treatment, women explore significantly more often than men, especially in the Tournament and Choice stages. Yet, when the possibility of negative payoffs is introduced (*Gain* & *Loss*), no gender difference arises. Random-effects multinomial logit estimates confirm that in *Gain* environments, being female substantially increases the likelihood of exploring rather than exploiting, whereas under *Gain* & *Loss* the net gender effect disappears. Hence Hypothesis 1 is supported. Furthermore, the data provide no evidence that competition, once risk and loss aversion are taken into account, spurs more exploration (Hypothesis 2).

Regarding entry into competition, our descriptive statistics show that in the *Gain* treatment women are less likely to choose the competitive scheme than men (21% vs. 36%). By contrast, there is no gap in the *Gain* & *Loss* treatment (27% vs. 28%). Once we control for risk preferences, loss aversion, self-assessed rank (i.e. confidence in own performance), and other characteristics, the gender gap in the *Gain* setting vanishes, indicating that differences in confidence, risk attitudes, and other individual traits can fully explain the raw difference in women's propensity to compete. This finding does not support the idea that women remain less likely to compete once individual-level heterogeneity is taken into account (Hypothesis 3). We also do not detect any evidence that women in STEM fields are more competitive than men (Hypothesis 5), since adding

an interaction between being female and studying a STEM major yields no significant effect. The absence of a gender gap in competition contrasts with the results documented in Niederle and Vesterlund [2007] and subsequent studies, such as Niederle and Vesterlund [2011] and Buser et al. [2014]. A possible explanation for this divergence lies in the specific task employed in our experiment, which differs substantially from the addition tasks used in Niederle and Vesterlund [2007] and the majority of studies replicating that paradigm. Our task was designed to be more abstract and gender-neutral, as suggested by Shurchkov [2012] and Gneezy et al. [2009], to minimize stereotype-related effects and ensure equal comfort across genders. This design choice may have mitigated the gendered behaviours typically observed in tasks perceived as stereotypically male-oriented.

Finally, we find substantial support for the influence of risk preferences, loss aversion, and cognitive or personality traits on both exploration and competition decisions (Hypothesis 6). Participants who are more risk-averse or more loss-averse generally explore less and show a lower inclination to enter competition.

Although we did not initially formulate a specific assumption regarding this aspect, our findings reveal slight differences in payoffs between women and men across the stages and treatments of the game, consistently favouring men. However, the gender gap in payoffs reaches statistical significance at the 10% level only in the Choice stage of the *Gain* treatment. Examining the maximum position reached by participants, we observe that women systematically reach positions farther from their initial ones than men, except in the Gain & Loss Treatment during the piece-rate stage. Since women tend to earn fewer points than men and reach positions farther from their initial ones in the Gain Treatment, this underperformance may be attributed to a tendency to over-explore in that treatment. Comparison with the benchmark performance derived from reinforcement learning models supports these observations.

Overall, our analysis suggests that gender differences in exploratory and competitive behaviour are primarily driven by differences in beliefs and preferences, rather than inherent gender traits. Interestingly, all gender differences disappear in the *Gain and Losses* treatment, where men and women exhibit similar behaviours at all stages. In this treatment, the effect acts in opposite directions for men and women: it reduces men's willingness to compete while increasing women's willingness to compete compared to the *Gain Only* treatment. This dynamic completely eliminates the gender gap in the willingness to compete.

7 An application: jobseekers' behaviours

In most countries, girls achieve higher academic performance than boys during their school years [Del Pero and Bytchkova, 2013]. Later in life, women attain higher levels of education than men [Bertrand, 2020] and,

in many countries, surpass men in college achievement.[22] Despite these educational advantages, women are less likely to be employed than men, and when employed, they earn less [Bertrand, 2020].

To explain the early gender gaps emerging in the labor market, even among young workers with comparable academic skills and backgrounds, recent research has highlighted gendered preferences for job attributes. Women are more likely to seek family-friendly occupations with shorter commuting times (e.g., Le Barbanchon et al., 2021; Fluchtmann et al., 2024). Additionally, studies have emphasized gender differences in personality traits, with women generally being more risk-averse, less willing to compete, and less confident compared to men (e.g., Vesterlund, 1997; Flinn et al., 2020). Among high-skilled young graduates, Cortés et al. [2023] find a clear gender difference in the timing of job offer acceptance, with women accepting jobs substantially earlier than men. The authors show that gendered job search strategies are driven by differences in risk preferences and overconfidence. Women, being more risk-averse, tend to have lower reservation wages, begin their job search earlier, and accept offers sooner. Conversely, men exhibit greater optimism regarding potential job offers, maintain higher reservation wages, and delay acceptance. The empirical evidence provided by Cortés et al. [2023] is further supported by findings from a laboratory experiment designed to test the previous intuition. In their setup, jobseekers are uncertain about their skills and choose their reservation wages over multiple rounds.

Although jobseekers' attitudes toward *exploration* are critical for understanding early career decision-making, this aspect has not been previously examined. While stylized and abstract, our lab experiment may provide valuable insights into job seekers' behaviours, complementing the findings of Cortés et al. [2023]. Furthermore, incorporating competition allows us to explore whether gender differences in another key individual characteristic, e.g. willingness to compete, play a role in job search behaviours.

Empirical evidence indicating that women accept jobs significantly earlier than men could suggest that women tend to engage less in exploration during an explore-exploit task. In the *Gain* treatment, we do observe a gender gap in payoffs that disadvantages women, potentially reflecting broader gender disparities in labor market outcomes. However, contrary to our initial expectations, we also find that women tend to overexplore.

Several factors could explain this result. It is possible that our experiment, being abstract and stylized, fails to capture the complexities of the labor market setting. Alternatively, crucial real-world factors not accounted for in the lab experiment, such as social norms influencing job market behaviour (e.g., societal expectations regarding men's and women's roles), may play a significant role. Additionally, factors related to the employer side of the labor market could shape job seekers' behaviours in ways not reflected in our

[22] See, for example, Conger and Long [2010] for the USA; Piazzalunga [2018] and Bovini et al. [2024] for Italy; Verbree et al. [2023] for the Netherlands; and Carroll [2023] for the UK.

experimental design.

On a positive note, our findings reveal that in the *Gains and Losses* treatment, all gender disparities disappear, with men and women exhibiting equal behaviour and performance in the explore-exploit task. Notably, the gap in willingness to compete is entirely eliminated for women from all academic backgrounds, including STEM and humanities. This result is particularly intriguing, as it suggests that a riskier, high-stake framework involving both gains and losses curtails women's tendency to overexplore and mitigates their reluctance to compete. This pattern echoes the "gender-equality paradox" in STEM fields documented by Stoet and Geary [2018], who explore the relationship between gender equality and women's representation in STEM disciplines. They show that countries with higher levels of gender equality have lower percentages of female STEM graduates. As a possible explanation, the authors propose that in wealthier, gender-equal countries, women are more likely to pursue careers aligned with personal preferences. Conversely, in poorer, less gender-equal countries, economic necessity drives women toward STEM fields due to their higher earning potential and career opportunities.

In the next section, we discuss policy implications inspired by the finding that environments with greater risks and high stakes reduce gender gaps. Proposed policies include career counselling programs emphasizing the high-stakes nature of early career decisions, mentorship initiatives to guide young women in navigating high-reward paths, and public awareness campaigns highlighting the long-term impact of initial job choices.

8 Conclusion

Our study explores gendered decision-making behaviours in the context of explore-exploit tasks, a topic that has been neglected in the extensive literature on gender gaps in economic behaviours and outcomes. Developmental psychology highlights how internalized gender norms and early childhood experiences shape individuals' preferences and attitudes, forming the foundation of their decision-making processes. From a young age, girls are taught to be "well-behaved," while boys are encouraged to be "brave," reinforcing gendered expectations that influence behaviour later in life (Fulcher and Coyle [2011] and Francis [2010], among many others). Building on these insights, we expected that women might explore less than men and specifically underexplore. Contrary to this intuition at the basis of our pre-registered hypothesis, our findings reveal that in gain-only environments, women engage in exploration more frequently. However, this heightened exploration does not translate into better outcomes. Women earn less despite their greater exploratory efforts, pointing to overexploration: they explore too much to fully exploit known alternatives but not enough to discover and capitalize on unknown prospects.

Interestingly, when the possibility of losses is introduced, all observed gender differences in exploration

and competition behaviours disappear, underscoring the critical role of environmental framing in shaping behavioural patterns. These results suggest that gender disparities in decision-making often reflect differences in beliefs and preferences rather than inherent behavioural differences.

Future research should investigate whether alternative contexts or additional social and psychological factors amplify or mitigate these dynamics. Exploring such factors could provide a deeper understanding of the mechanisms behind gendered behaviours in high-stakes environments.

Our most striking result is the complete elimination of gender gaps—in both behaviours and outcomes—in riskier, high-stakes environments with gains and losses. Such settings appear to reduce women's tendency to overexplore while fostering equal participation in competition across genders. This outcome offers valuable insights for policy interventions aimed at promoting gender equality in competitive domains.

For instance, consider a very important explore-exploit decision-making setting, e.g. the job search strategies of young applicants entering the labor market for the first time. A large empirical evidence documents gender gaps in labor market outcomes (among many others, Bertrand [2020], Le Barbanchon et al. [2021]). Other papers emphasize that those gaps emerge from the very beginning: despite outperforming men academically, women face disadvantages in employment rates, salaries, and the quality of job positions from their initial entry into the workforce (Piazzalunga [2018], Bovini et al. [2024], Cortés et al. [2023]). Inefficient search strategies among women may partially explain this phenomenon.Our findings suggest that gender gaps in search strategies could be mitigated by emphasizing the high-stakes nature of early labor market entry to job applicants.

This could be achieved through targeted interventions in schools and universities, such as career counselling programs that focus on the high-stakes implications of initial career choices, guiding young women in selecting roles and industries strategically. These programs could provide detailed insights into salary trajectories, job security, and career progression, emphasizing how these decisions shape long-term outcomes. Mentorship programs pairing women with experienced professionals could highlight the risks and rewards of different career paths, offering personalized advice to navigate high-stakes early decisions effectively. Public awareness campaigns could further stress that initial career choices carry significant long-term consequences, aligning with the high-stakes framework of gains and losses. By equipping women with the tools and knowledge to approach these critical decisions, such initiatives could improve early career choices.

References

Ghazala Azmat and Barbara Petrongolo. Gender and the labor market: What have we learned from field and lab experiments? *Labour economics*, 30:32–40, 2014.

Rina Azoulay-Schwartz, Sarit Kraus, and Jonathan Wilkenfeld. Exploitation vs. exploration: choosing a supplier in an environment of incomplete information. *Decision support systems*, 38(1):1–18, 2004.

Oded Berger-Tal, Jonathan Nathan, Ehud Meron, and David Saltz. The exploration-exploitation dilemma: a multidisciplinary framework. *PloS one*, 9(4):e95693, 2014.

Donald A Berry and Bert Fristedt. Bandit problems: sequential allocation of experiments (monographs on statistics and applied probability). *London: Chapman and Hall*, 5(71-87):7–7, 1985.

Marianne Bertrand. Gender in the twenty-first century. In *AEA Papers and proceedings*, volume 110, pages 1–24. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2020.

Tommy C Blanchard and Samuel J Gershman. Pure correlates of exploration and exploitation in the human brain. *Cognitive, Affective, & Behavioral Neuroscience*, 18:117–126, 2018.

Alison Booth and Patrick Nolen. Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81(2):542–555, 2012.

Giulia Bovini, Marta De Philippis, and Lucia Rizzica. The origins of the gender pay gap: education and job characteristics. Technical report, SSRN Working Paper No. 5021036, 2024.

Thomas Buser and Huaiping Yuan. Do women give up competing more easily? evidence from the lab and the dutch math olympiad. *American Economic Journal: Applied Economics*, 11(3):225–252, 2019.

Thomas Buser, Muriel Niederle, and Hessel Oosterbeek. Gender, competitiveness, and career choices. *The quarterly journal of economics*, 129(3):1409–1447, 2014.

Thomas Buser, Noemi Peter, and Stefan C Wolter. Gender, competitiveness, and study choices in high school: Evidence from switzerland. *American economic review*, 107(5):125–130, 2017.

Colin F Camerer and Robin M Hogarth. The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of risk and uncertainty*, 19:7–42, 1999.

Juan-Camilo Cárdenas, Anna Dreber, Emma Von Essen, and Eva Ranehill. Gender differences in competitiveness and risk taking: Comparing children in colombia and sweden. *Journal of Economic Behavior & Organization*, 83(1):11–23, 2012.

Matthew Carroll. Sex gaps in education in england: Research report. Technical report, Cambridge University Press & Assessment, 2023. Available via ERIC.

Alycia Chin, David Hagmann, and George Loewenstein. Fear and promise of the unknown: How losses discourage and promote exploration. *Journal of Behavioral Decision Making*, 36(3):e2309, 2023.

Robert B. Cialdini and Melanie R. Trost. Social influence: Social norms, conformity, and compliance. In Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey, editors, *The Handbook of Social Psychology*, volume 2, pages 151–192. McGraw-Hill, 4th edition, 1998.

Dylan Conger and Mark C Long. Why are men falling behind? gender gaps in college performance and persistence. *The Annals of the American Academy of Political and Social Science*, 627(1):184–214, 2010.

Patricia Cortés, Jessica Pan, Laura Pilossoph, Ernesto Reuben, and Basit Zafar. Gender differences in job search and the earnings gap: Evidence from the field and lab. *The Quarterly Journal of Economics*, 138 (4):2069–2126, 2023.

Paolo Crosetto and Antonio Filippin. The "bomb" risk elicitation task. *Journal of Risk and Uncertainty*,

47:31–65, 2013.

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

Jac TM Davis and Melissa Hines. How large are gender differences in toy preferences? a systematic review and meta-analysis of toy preference research. *Archives of Sexual Behavior*, 49(2):373–394, 2020.

Nathaniel D Daw, John P O'doherty, Peter Dayan, Ben Seymour, and Raymond J Dolan. Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095):876–879, 2006.

Angelica Salvi Del Pero and Alexandra Bytchkova. A bird's eye view of gender differences in education in oecd countries. Technical Report 149, OECD Publishing, 2013.

Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238–1260, 2010.

Thomas Dohmen, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G Wagner. Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, 9(3):522–550, 2011.

Anna Dreber, Emma Von Essen, and Eva Ranehill. Gender and competition in adolescence: task matters. *Experimental Economics*, 17:154–172, 2014.

Alice H Eagly and Steven J Karau. Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3):573–598, 2002.

Catherine C Eckel and Philip J Grossman. Differences in the economic decisions of men and women: Experimental evidence. *Handbook of experimental economics results*, 1:509–519, 2008.

Christopher Flinn, Petra Todd, and Weilong Zhang. Personality traits, job search and the gender wage gap. Technical report, Faculty of Economics, University of Cambridge, 2020. URL `https://www.repository.cam.ac.uk/handle/1810/314752`.

Jonas Fluchtmann, Anita M Glenny, Nikolaj A Harmon, and Jonas Maibom. The gender application gap: Do men and women apply for the same jobs? *American Economic Journal: Economic Policy*, 16(2): 182–219, 2024.

Becky Francis. Gender, toys and learning. *Oxford Review of Education*, 36(3):325–344, 2010.

Shane Frederick. Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42, 2005.

Megan Fulcher and Emily F Coyle. Breadwinner and caregiver: A cross-sectional analysis of children's and emerging adults' visions of their future family roles. *British Journal of Developmental Psychology*, 29(2): 330–346, 2011.

Simon Gächter, Eric J Johnson, and Andreas Herrmann. Individual-level loss aversion in riskless and risky choices. *Theory and Decision*, 92(3):599–624, 2022.

Klarita Gërxhani, Jordi Brandts, and Arthur Schram. Competition and gender inequality: A comprehensive analysis of effects and mechanisms. *American Journal of Sociology*, 129(3):715–752, 2023.

David Gill and Victoria Prowse. A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1):469–503, 2012.

Uri Gneezy, Kenneth L Leonard, and John A List. Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5):1637–1664, 2009.

Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. A very brief measure of the big-five personality

domains. *Journal of Research in Personality*, 37(6):504–528, 2003.

Ben Greiner. Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125, 2015.

Niels Daniel Große and Gerhard Riener. Explaining gender differences in competitiveness: Gender-task stereotypes. Technical report, Jena Economic Research Papers, 2010.

Anil K Gupta, Ken G Smith, and Christina E Shalley. The interplay between exploration and exploitation. *Academy of management journal*, 49(4):693–706, 2006.

S Julie Harbin. Gender differences in rough and tumble play behaviors. *International Journal of Undergraduate Research and Creative Activities*, 8(1):4, 2023.

Karson TF Kung. Recalled childhood gender-related play behaviour and current gender-related occupational interests in university students: Examining the mediating roles of gender compatibility, goal endorsement, and occupational stereotype flexibility. *Frontiers in Psychology*, 13:927–998, 2022.

Carol A Lawton. Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex roles*, 30:765–779, 1994.

Thomas Le Barbanchon, Roland Rathelot, and Alexandra Roulet. Gender differences in job search: Trading off commute against wage. *The Quarterly Journal of Economics*, 136(1):381–426, 2021.

Shakun D Mago, Anya C Samak, and Roman M Sheremeta. Facing your opponents: Social identification and information feedback in contests. *Journal of Conflict Resolution*, 60(3):459–481, 2016.

James G March. Exploration and exploitation in organizational learning. *Organization science*, 2(1):71–87, 1991.

Björn Meder, Charley M Wu, Eric Schulz, and Azzurra Ruggeri. Development of directed and random exploration in children. *Developmental science*, 24(4):e13095, 2021.

Katja Mehlhorn, Ben R. Newell, Peter M. Todd, Michael D. Lee, Kate Morgan, Victoria A. Braithwaite, Daniel Hausmann, Klaus Fiedler, and Cleotilde Gonzalez. Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, 2(3):191–215, 2015.

Muriel Niederle and Lise Vesterlund. Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101, 2007.

Muriel Niederle and Lise Vesterlund. Gender and competition. *Annu. Rev. Econ.*, 3(1):601–630, 2011.

Daniela Piazzalunga. The gender wage gap among college graduates in italy. *Italian Economic Journal*, 4 (1):33–90, 2018.

Paul B Reverdy, Vaibhav Srivastava, and Naomi Ehrich Leonard. Modeling human decision making in generalized gaussian multiarmed bandits. *Proceedings of the IEEE*, 102(4):544–571, 2014.

Eric Schulz, Charley M Wu, Azzurra Ruggeri, and Björn Meder. Searching for rewards like a child means less generalization and more directed exploration. *Psychological science*, 30(11):1561–1572, 2019.

Eric Schulz, Nicholas T Franklin, and Samuel J Gershman. Finding structure in multi-armed bandits. *Cognitive psychology*, 119:101261, 2020.

Olga Shurchkov. Under pressure: gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, 10(5):1189–1213, 2012.

Maarten Speekenbrink and Emmanouil Konstantinidis. Uncertainty and exploration in a restless bandit problem. *Topics in cognitive science*, 7(2):351–367, 2015.

Mark Steyvers, Michael D Lee, and Eric-Jan Wagenmakers. A bayesian analysis of human decision-making

on bandit problems. *Journal of mathematical psychology*, 53(3):168–179, 2009.

Gijsbert Stoet and David C Geary. The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological science*, 29(4):581–593, 2018.

Ioannis Theodossiou and Alexandros Zangelidis. Should i stay or should i go? the effect of gender, education and unemployment on labour market transitions. *Labour Economics*, 16(5):566–577, 2009.

Anne-Roos Verbree, Lisette Hornstra, Lientje Maas, and Leoniek Wijngaards-de Meij. Conscientiousness as a predictor of the gender gap in academic achievement. *Research in Higher Education*, 64(3):451–472, 2023.

Lise Vesterlund. The effects of risk aversion on job matching: Can differences in risk aversion explain the wage gap? Unpublished manuscript, Iowa State University, 1997.

Robert C Wilson, Elizabeth Bonawitz, Vincent D Costa, and R Becket Ebitz. Balancing exploration and exploitation with information and randomization. *Current Opinion in Behavioral Sciences*, 38:49–56, 2021.

Charley M Wu, Eric Schulz, Maarten Speekenbrink, Jonathan D Nelson, and Björn Meder. Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, 2(12):915–924, 2018.

Mohan Yogeswaran and SG Ponnambalam. Reinforcement learning: Exploration–exploitation dilemma in multi-agent foraging task. *Opsearch*, 49:223–236, 2012.

Appendix A

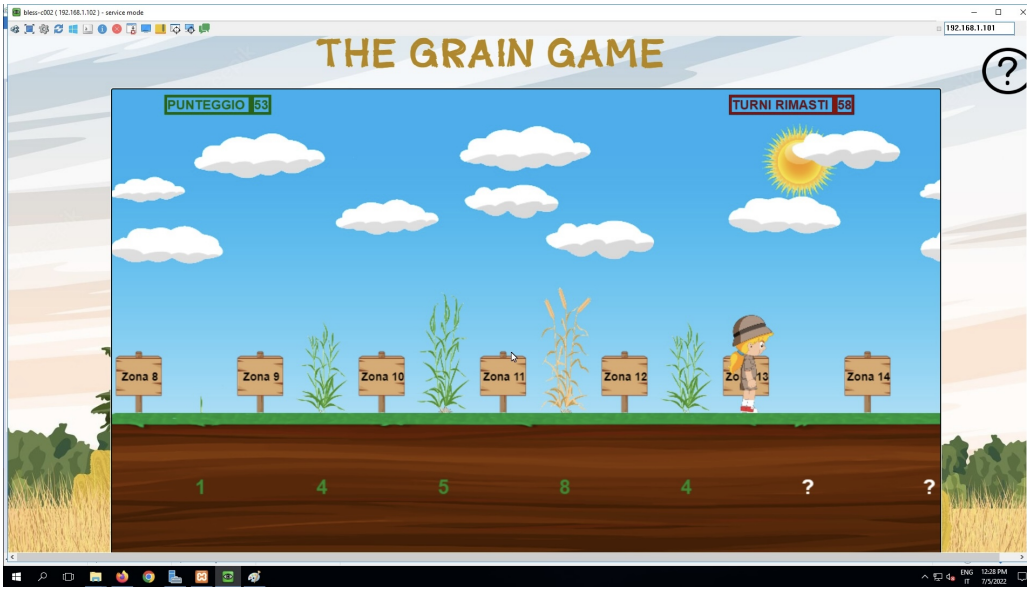


Figure A.1: Grain Game interface under the *Gain* treatment.

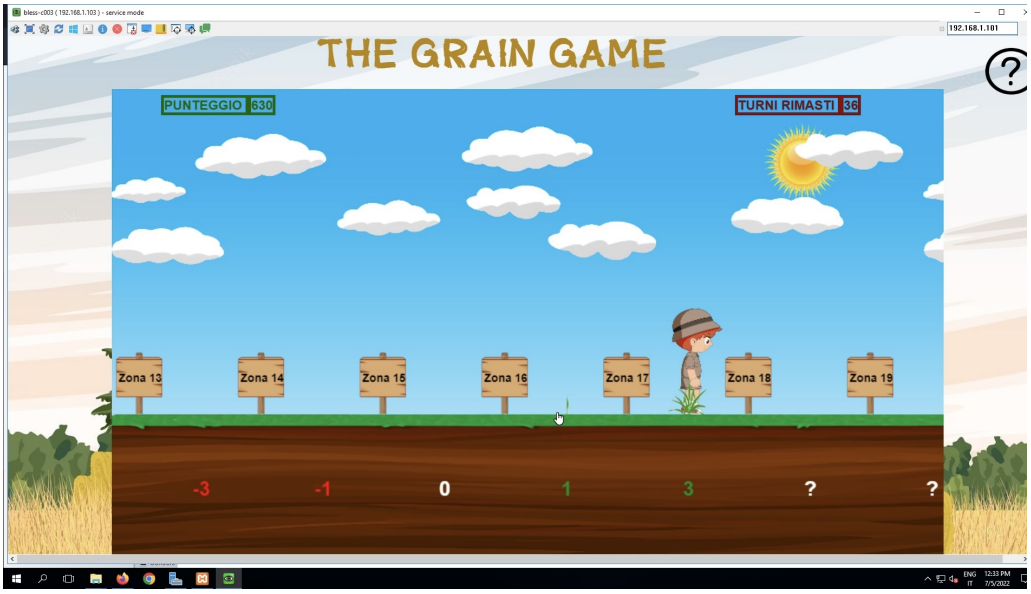


Figure A.2: Grain Game interface under the *Gain & Loss* treatment.

Appendix B

The payoff sequences used in the experiment were generated using a random walk process, with errors drawn from a normal distribution $\mathcal{N}(0, 2)$. These sequences were designed to reward exploration by ensuring that the global maximum of every sequence was placed in the second half of the field.

The sequences for the Gain treatment are shown below:

- **S1**: {7, 7, 7, 5, 0, 3, 2, 1, 4, 5, 8, 4, 7, 9, 10, 10, 12, 15, 18, 17, 17, 16, 12, 12, 11, 12, 13, 13, 11, 13, 15, 17, 15, 17, 17, 18, 16, 18, 17, 18, 18, 19, 22, 25, 23, 18, 20, 19, 20, 23, 24, 23, 23, 21, 24, 25, 25, 24, 23, 23, 23, 20, 19, 22, 21, 20, 17, 15, 17, 17}
- **S2**: {11, 8, 5, 8, 9, 11, 12, 13, 14, 11, 16, 15, 11, 10, 6, 7, 7, 10, 12, 13, 13, 11, 13, 13, 12, 11, 12, 10, 12, 12, 15, 14, 14, 13, 16, 13, 13, 15, 14, 16, 16, 13, 16, 15, 16, 20, 20, 21, 24, 23, 23, 23, 24, 25, 24, 22, 21, 18, 17, 17, 18, 18, 17, 19, 19, 18, 20, 17, 15, 19}
- **S3**: {7, 10, 14, 14, 15, 12, 13, 14, 15, 11, 12, 11, 12, 11, 12, 9, 10, 8, 9, 13, 13, 12, 15, 17, 13, 13, 16, 16, 13, 15, 17, 14, 16, 16, 17, 14, 14, 15, 15, 15, 18, 16, 11, 12, 13, 13, 11, 12, 10, 11, 12, 13, 14, 13, 15, 20, 22, 23, 25, 24, 21, 20, 21, 23, 20, 19, 19, 17, 14, 11}

The payoff sequences for the Gain & Loss treatment are derived by subtracting 10 points from the respective Gain sequences. For example:

- **S4 (Gain & Loss sequence derived from S1)**: {-3, -3, -3, -5, -10, -7, -8, -9, -6, -5, -2, -6, -3, -1, 0, 0, 2, 5, 8, 7, 7, 6, 2, 2, 1, 2, 3, 3, 1, 3, 5, 7, 5, 7, 7, 8, 6, 8, 7, 8, 8, 9, 12, 15, 13, 8, 10, 9, 10, 13, 14, 13, 13, 11, 14, 15, 15, 14, 13, 13, 13, 10, 9, 12, 11, 10, 7, 5, 7, 7}

Figure B.1 below provides a visual representation of the three payoff sequences for the Gain treatment, with a horizontal line at $Payoff = 10$. The threshold marks the transition point to negative payoffs in the Gain & Loss treatment, where a -10 offset was applied.

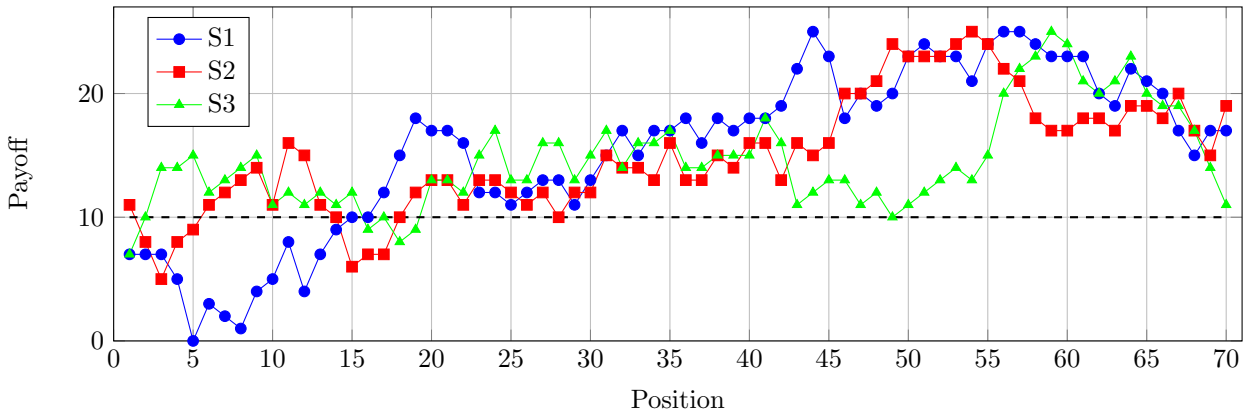


Figure B.1: Payoff sequences for the Gain treatment with a threshold line at $Payoff = 10$. *Note*: The threshold indicates the transition to negative payoffs in the Gain & Loss treatment, where a -10 offset was applied to all values.

Appendix C

This appendix provides illustrations of the interfaces used for the Loss Aversion Task (Figure C1) and the Bomb Risk Elicitation Task (BRET) (Figure C2).

Figure C1: Screen for the Loss Aversion Task. *Note*: Participants decided whether to accept or reject monetary gambles with potential gains and losses.

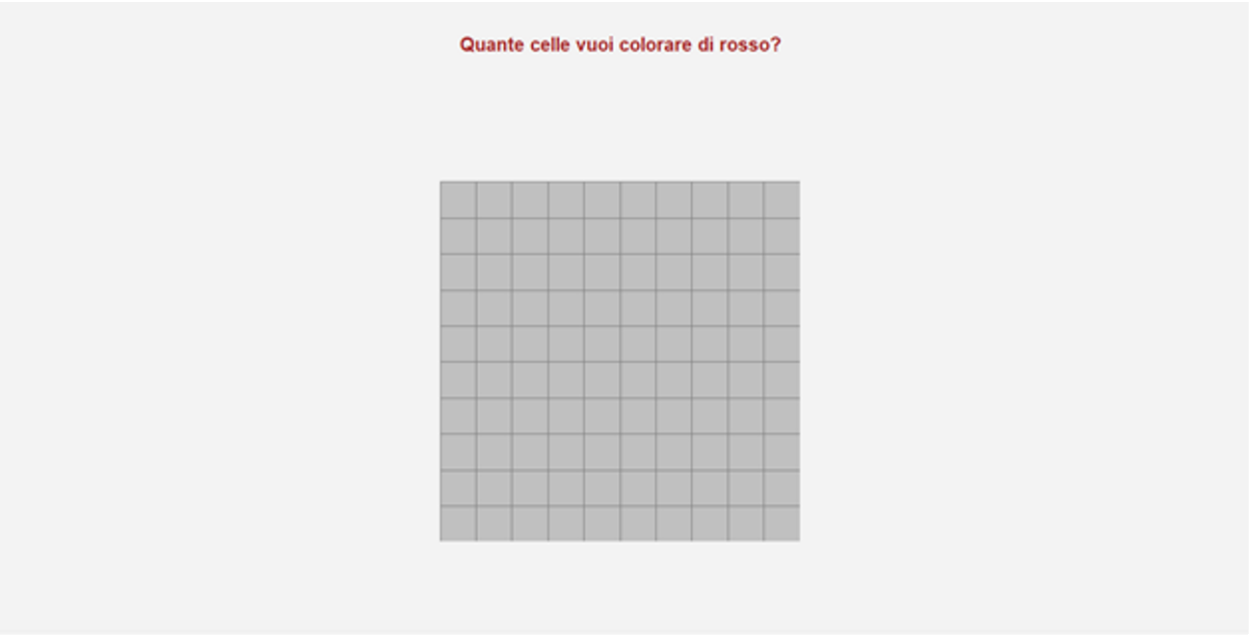


Figure C2: Screen for the Bomb Risk Elicitation Task (BRET). *Note*: Participants could adjust the number of red cells by selecting cells with their mouse and confirm their choice by clicking "Confirm".

Appendix D

In Table D1, we present the summary statistics for the main variables used in our analysis. The *Cognitive Reflection Test (CRT)* is measured following Frederick [2005], where a higher score indicates a greater tendency for reflective thinking rather than relying on intuitive responses. We define and construct the *TIPI* (Ten-Item Personality Inventory) as in Gosling et al. [2003], which captures concise measures of the Big Five personality traits. The measure of *Loss Aversion* (λ_{loss}) follows Gächter et al. [2022], where higher values indicate stronger aversion to losses; any non-monotonic decisions in this task are coded as missing observations, consistent with Gächter et al. [2022]. Finally, our measures of risk attitudes (*Risk (Bret)* and *Risk (Self-Reported)*) are such that higher values reflect lower levels of risk aversion.

Table D1: Descriptive statistics by Gender

	Overall (N=419)	**Men (N=213)**	**Women (N=206)**	**p-value**
Age				0.003
Mean (SD)	24.25 (3.33)	24.73 (3.68)	23.76 (2.85)	
Median [Min, Max]	24.00 [18.00, 52.00]	24.00 [19.00, 52.00]	24.00 [18.00, 41.00]	
Missing	3	3	0	
CRT				< 0.001
Mean (SD)	1.64 (1.10)	1.97 (1.02)	1.30 (1.09)	
Median [Min, Max]	2.00 [0.00, 3.00]	2.00 [0.00, 3.00]	1.00 [0.00, 3.00]	
Missing	3	1	2	
GPA				0.2
Mean (SD)	3.93 (0.87)	3.87 (0.91)	3.98 (0.82)	
Median [Min, Max]	4.00 [1.00, 5.00]	4.00 [1.00, 5.00]	4.00 [2.00, 5.00]	
Missing	3	2	1	
Risk (Self-Reported)				0.6
Mean (SD)	5.39 (1.83)	5.44 (1.94)	5.35 (1.73)	
Median [Min, Max]	5.00 [0.00, 10.00]	5.00 [0.00, 10.00]	5.00 [1.00, 10.00]	
Missing	1	1	0	
Risk (Bret)				0.2
Mean (SD)	46 (17)	45 (17)	48 (17)	
Median [Min, Max]	49 [1, 100]	46 [1, 100]	50 [1, 100]	
Missing	62	31	31	
Agreeableness				0.033
Mean (SD)	5.01 (1.11)	4.89 (1.12)	5.13 (1.10)	
Median [Min, Max]	5.00 [1.50, 7.00]	5.00 [2.00, 7.00]	5.00 [1.50, 7.00]	
Missing	1	1	0	
Extraversion				0.031
Mean (SD)	3.90 (1.54)	3.74 (1.47)	4.07 (1.60)	
Median [Min, Max]	4.00 [1.00, 7.00]	3.50 [1.00, 7.00]	4.00 [1.00, 7.00]	
Missing	1	1	0	
Conscientiousness				0.066
Mean (SD)	5.07 (1.33)	4.96 (1.29)	5.19 (1.35)	
Median [Min, Max]	5.50 [1.00, 7.00]	5.00 [1.50, 7.00]	5.50 [1.00, 7.00]	
Missing	1	1	0	
Emotional Stability				< 0.001
Mean (SD)	4.05 (1.38)	4.38 (1.37)	3.70 (1.31)	
Median [Min, Max]	4.00 [1.00, 7.00]	4.50 [1.00, 7.00]	3.50 [1.00, 7.00]	
Missing	1	1	0	
Openness				< 0.001
Mean (SD)	4.90 (1.15)	4.67 (1.06)	5.14 (1.19)	
Median [Min, Max]	5.00 [1.50, 7.00]	4.50 [2.00, 7.00]	5.00 [1.50, 7.00]	
Missing	1	1	0	
Competitiveness (self-reported)				< 0.001
Mean (SD)	5.89 (2.40)	6.48 (2.27)	5.29 (2.39)	
Median [Min, Max]	6.00 [0.00, 10.00]	7.00 [0.00, 10.00]	5.00 [0.00, 10.00]	
Missing	1	1	0	
λ_{loss}				0.2
Mean (SD)	1.77 (0.65)	1.81 (0.71)	1.72 (0.57)	
Median [Min, Max]	1.50 [0.86, 4.00]	2.00 [0.86, 4.00]	1.50 [0.86, 4.00]	
Missing	100	46	54	
Rank Belief				0.006
Mean (SD)	2.08 (0.81)	1.97 (0.86)	2.19 (0.74)	
Median [Min, Max]	2.00 [0.00, 4.00]	2.00 [0.00, 4.00]	2.00 [1.00, 4.00]	
Born in Italy	389 (93%)	198 (93%)	191 (93%)	0.8
Missing	1	1	0	
STEM	159 (39%)	95 (46%)	64 (31%)	0.003
Missing	7	5	2	

Continued on next page

Notes: p-values in the last column reflect a Men vs. Women comparison. They come from Two Sample t-tests for continuous variables or Pearson's Chi-squared tests for categorical variables.

Appendix E

Table E1: Final and Max Position, by Gender, Treatment and Stage

	Final Position			**Max Position**		
	Men (SE)	Women (SE)	p-value (stars)	Men (SE)	Women (SE)	p-value (stars)
Piece Rate						
Gain	30.74 (1.69)	33.67 (1.82)	0.238	32.16 (1.59)	35.03 (1.71)	0.221
Gain & Loss	29.10 (1.63)	28.71 (1.74)	0.868	30.58 (1.56)	30.23 (1.65)	0.876
Overall	29.92 (1.17)	31.21 (1.27)	0.452	31.37 (1.11)	32.65 (1.20)	0.432
Tournament						
Gain	30.32 (1.90)	36.93 (1.81)	0.013	31.73 (1.81)	38.09 (1.72)	0.012
Gain & Loss	28.67 (1.79)	30.68 (1.80)	0.431	29.98 (1.73)	31.78 (1.73)	0.462
Overall	29.49 (1.30)	33.83 (1.29)	0.019	30.85 (1.25)	34.97 (1.24)	0.020
Choice						
Gain	31.03 (2.01)	38.75 (1.92)	0.006	32.37 (1.92)	39.67 (1.84)	0.007
Gain & Loss	32.34 (1.92)	32.75 (2.00)	0.883	33.31 (1.86)	34.01 (1.90)	0.792
Overall	31.69 (1.39)	35.78 (1.40)	0.039	32.84 (1.33)	36.87 (1.33)	0.033

Notes: Reported values are means of Final Position and Max Position, with standard errors in parentheses. All p-values come from two-sample t-tests comparing Men vs. Women within each row. "Overall" pools Gain and Gain & Loss within each stage.

Table E2: Explore Percent, by Gender, Treatment and Stage

	Explore Percent		
	Men (SE)	**Women** (SE)	**p-value**
Piece Rate			
Gain	0.526 (0.021)	0.560 (0.024)	0.284
Gain & Loss	0.507 (0.022)	0.519 (0.022)	0.697
Overall	0.516 (0.015)	0.540 (0.016)	0.293
Tournament			
Gain	0.494 (0.026)	0.583 (0.025)	0.014
Gain & Loss	0.483 (0.024)	0.517 (0.024)	0.311
Overall	0.488 (0.018)	0.550 (0.017)	0.013
Choice			
Gain	0.492 (0.028)	0.606 (0.025)	0.003
Gain & Loss	0.525 (0.026)	0.532 (0.027)	0.842
Overall	0.509 (0.019)	0.569 (0.019)	0.023

Notes: Reported values are the mean fraction of rounds in which participants explored, with standard errors in parentheses. All p-values come from two-sample t-tests comparing Men vs. Women in each row. "Overall" pools Gain and Gain & Loss within each stage.

Appendix F

Table F1: Random Effects Multinomial Logistic Results for *Move* (Piece Rate Stage)

Dependent Variable	*Move* (Baseline = *Exploit*)					
	Model A1		**Model B1**		**Model C1**	
	Retreat	Explore	Retreat	Explore	Retreat	Explore
Female	0.949	1.162	0.969	1.339	1.060	1.559
	(0.242)	(0.253)	(0.268)	(0.310)	(0.371)	(0.432)
Gain & Loss	1.099	0.936	1.079	0.970	1.135	1.155
	(0.291)	(0.200)	(0.303)	(0.212)	(0.371)	(0.288)
Female × (Gain & Loss)	1.511	0.968	1.547	0.977	1.463	0.847
	(0.543)	(0.298)	(0.567)	(0.304)	(0.652)	(0.311)
Lag Payoff	0.846***	0.853***	0.845***	0.850***	0.826***	0.819***
	(0.013)	(0.015)	(0.014)	(0.015)	(0.017)	(0.018)
Threshold	0.470***	0.604***	0.470***	0.589***	0.544***	0.622***
	(0.060)	(0.064)	(0.061)	(0.063)	(0.075)	(0.076)
Risk (self-reported)	–	–	0.966	1.030	–	–
			(0.057)	(0.049)		
Risk (Bret)	–	–	–	–	1.006	1.006
					(0.007)	(0.006)
λ_{loss}	–	–	–	–	0.645*	0.646**
					(0.112)	(0.089)
Competitiveness (self-reported)	–	–	0.946	1.037	0.936	1.031
			(0.042)	(0.039)	(0.047)	(0.042)
CRT	–	–	1.170	1.086	1.357**	1.261*
			(0.109)	(0.088)	(0.146)	(0.123)
Extraversion	–	–	1.085	0.974	1.025	0.922
			(0.070)	(0.051)	(0.078)	(0.055)
Agreeableness	–	–	1.114	1.173*	1.073	1.184*
			(0.107)	(0.088)	(0.120)	(0.100)
Conscientiousness	–	–	1.069	0.856*	1.051	0.827**
			(0.080)	(0.053)	(0.091)	(0.060)
Emotional Stability	–	–	0.972	0.939	1.068	0.985
			(0.070)	(0.059)	(0.087)	(0.069)
Openness	–	–	0.949	0.922	1.040	0.947
			(0.089)	(0.068)	(0.112)	(0.075)
GPA	–	–	0.836	0.916	0.810	0.950
			(0.094)	(0.086)	(0.103)	(0.094)
Born in Italy	–	–	0.804	0.491	0.511	0.482
			(0.348)	(0.199)	(0.225)	(0.213)
Age	–	–	1.035	1.052	1.080	1.105***
			(0.032)	(0.037)	(0.042)	(0.032)
STEM	–	–	0.926	1.080	0.935	1.144
			(0.185)	(0.180)	(0.222)	(0.214)
Constant	1.168	13.093***	0.739	12.540*	0.438	6.791
	(0.320)	(3.387)	(1.004)	(15.518)	(0.707)	(9.140)
Model Statistics						
Observations	28,492		27,676		21,080	
Number of Groups	419		407		310	
Wald χ^2	136.51		188.22		182.36	
Log pseudolikelihood	−21695.549		−21062.913		−15947.562	

Notes: All entries are Relative Risk Ratios (RRRs) from random-effects multinomial logit regressions. Robust standard errors (clustered at participant level) are shown in parentheses. The dependent variable is *Move*, and the baseline (omitted) outcome is *Exploit*. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table F2: Random Effects Multinomial Logistic Results for *Move* (Tournament Stage)

Dependent Variable	*Move* (Baseline = *Exploit*)					
	Model A2		**Model B2**		**Model C2**	
	Retreat	Explore	Retreat	Explore	Retreat	Explore
Female	1.218	1.530	1.309	1.936*	1.083	2.228*
	(0.383)	(0.473)	(0.450)	(0.608)	(0.454)	(0.881)
Gain & Loss	1.338	0.933	1.454	0.874	1.093	1.030
	(0.419)	(0.283)	(0.487)	(0.258)	(0.410)	(0.357)
Female × (Gain & Loss)	1.018	0.732	0.954	0.666	1.044	0.491
	(0.451)	(0.313)	(0.433)	(0.273)	(0.555)	(0.241)
Lag Payoff	0.771***	0.763***	0.768***	0.759***	0.760***	0.758***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.020)	(0.021)
Threshold	0.397***	0.670***	0.399***	0.681**	0.302***	0.683*
	(0.065)	(0.078)	(0.066)	(0.081)	(0.060)	(0.105)
Piece Rate Payoff	–	–	1.000	1.004***	1.001	1.004***
			(0.001)	(0.001)	(0.001)	(0.001)
Risk (Self-Reported)	–	–	0.911	1.120	–	–
			(0.061)	(0.066)		
Risk (Bret)	–	–	–	–	1.007	1.007
					(0.008)	(0.009)
λ_{loss}	–	–	–	–	0.551**	0.457***
					(0.101)	(0.079)
Competitiveness (self-reported)	–	–	0.946	1.024	0.909	0.987
			(0.049)	(0.053)	(0.054)	(0.059)
CRT	–	–	1.209	1.072	1.237	1.157
			(0.136)	(0.118)	(0.161)	(0.149)
Extraversion	–	–	0.982	0.977	0.948	0.953
			(0.076)	(0.067)	(0.085)	(0.076)
Agreeableness	–	–	1.320*	1.243*	1.320*	1.300*
			(0.149)	(0.132)	(0.166)	(0.155)
Conscientiousness	–	–	0.938	0.949	0.932	0.983
			(0.082)	(0.076)	(0.094)	(0.093)
Emotional Stability	–	–	0.913	0.925	0.914	0.961
			(0.075)	(0.079)	(0.085)	(0.092)
Openness	–	–	0.807	0.835	0.837	0.919
			(0.098)	(0.088)	(0.113)	(0.108)
GPA	–	–	0.989	0.922	1.020	0.983
			(0.137)	(0.123)	(0.161)	(0.152)
Born in Italy	–	–	0.225**	0.204**	0.183**	0.157**
			(0.121)	(0.106)	(0.113)	(0.105)
Age	–	–	1.023	1.043	1.024	1.069
			(0.036)	(0.038)	(0.046)	(0.050)
STEM	–	–	0.708	1.122	0.815	1.508
			(0.183)	(0.258)	(0.242)	(0.402)
Constant	1.661	58.353***	9.047	2.578	13.922	2.870
	(0.620)	(22.630)	(15.714)	(4.403)	(27.148)	(5.708)
Observations	28,492		27,676		21,080	
Number of Groups	419		407		310	
Wald χ^2	179.50		252.06		199.75	
Log pseudolikelihood	−18233.956		−17616.035		−13283.393	

Notes: All entries are Relative Risk Ratios (RRRs) from random-effects multinomial logit regressions. Robust standard errors (clustered at participant level) are shown in parentheses. The dependent variable is *Move*, and the baseline (omitted) outcome is *Exploit*. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table F3: Random Effects Multinomial Logistic Results for *Move* (Tournament Stage with Compete control)

Dependent Variable	*Move* (Baseline = *Exploit*)					
	Model A3		**Model B3**		**Model C3**	
	Retreat	Explore	Retreat	Explore	Retreat	Explore
Female	1.140	1.593	1.267	1.932*	1.047	2.217*
	(0.367)	(0.495)	(0.442)	(0.607)	(0.450)	(0.877)
Gain & Loss	1.281	0.954	1.377	0.871	1.010	1.023
	(0.403)	(0.291)	(0.463)	(0.260)	(0.380)	(0.355)
Female × (Gain & Loss)	1.086	0.705	1.037	0.669	1.137	0.496
	(0.485)	(0.303)	(0.475)	(0.277)	(0.611)	(0.244)
Lag Payoff	0.771***	0.763***	0.768***	0.759***	0.760***	0.758***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.020)	(0.021)
Threshold	0.397***	0.670**	0.399***	0.682**	0.302***	0.684*
	(0.065)	(0.078)	(0.066)	(0.081)	(0.060)	(0.105)
Compete (Tournament)	0.653	1.311	0.640	0.974	0.414**	0.893
	(0.166)	(0.296)	(0.172)	(0.220)	(0.138)	(0.250)
Piece Rate Payoff	–	–	1.000	1.004***	1.001	1.004***
			(0.001)	(0.001)	(0.001)	(0.001)
Risk (Self-Reported)	–	–	0.923	1.121	–	–
			(0.062)	(0.067)		
Risk (Bret)	–	–	–	–	1.010	1.007
					(0.008)	(0.009)
λ_{loss}	–	–	–	–	0.492***	0.451***
					(0.092)	(0.080)
Competitiveness (self-reported)	–	–	0.962	1.025	0.939	0.992
			(0.052)	(0.055)	(0.058)	(0.061)
CRT	–	–	1.217	1.072	1.248	1.159
			(0.137)	(0.118)	(0.164)	(0.150)
Extraversion	–	–	0.989	0.977	0.974	0.956
			(0.076)	(0.067)	(0.087)	(0.076)
Agreeableness	–	–	1.334*	1.243*	1.371*	1.306*
			(0.152)	(0.133)	(0.174)	(0.156)
Conscientiousness	–	–	0.937	0.949	0.923	0.983
			(0.081)	(0.076)	(0.092)	(0.092)
Emotional Stability	–	–	0.916	0.925	0.932	0.963
			(0.075)	(0.079)	(0.086)	(0.091)
Openness	–	–	0.800	0.835	0.828	0.920
			(0.097)	(0.088)	(0.112)	(0.108)
GPA	–	–	0.958	0.920	0.954	0.974
			(0.135)	(0.124)	(0.155)	(0.150)
Born in Italy	–	–	0.226**	0.205**	0.182**	0.158**
			(0.123)	(0.106)	(0.113)	(0.105)
Age	–	–	1.028	1.043	1.038	1.071
			(0.037)	(0.038)	(0.047)	(0.050)
STEM	–	–	0.722	1.122	0.857	1.509
			(0.187)	(0.258)	(0.256)	(0.402)
Constant	1.936	52.860***	7.494	2.535	9.764	2.690
	(0.751)	(20.931)	(13.025)	(4.353)	(18.887)	(5.374)
Observations	28,492		27,676		21,080	
Number of Groups	419		407		310	
Wald χ^2	185.44		256.03		214.48	
Log pseudolikelihood	−18232.316		−17615.063		−13280.770	

Notes: All entries are Relative Risk Ratios (RRRs) from random-effects multinomial logit regressions. Robust standard errors (clustered at participant level) are shown in parentheses. The dependent variable is *Move*, and the baseline (omitted) outcome is *Exploit*. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Appendix G

To test whether women attending STEM fields of study are more likely to choose the competitive payment scheme than men, we augment Equation (2) by adding the interaction term $\beta_5\big(Female_i \times STEM_i\big)$. Because $STEM_i$ is already included in the vector of controls $\mathbf{X}_i$, the only modification to the original model is the inclusion of this interaction. Formally, the augmented specification is:

$$\begin{aligned}\ln\Big(\tfrac{\Pr(Compete_i=1)}{\Pr(Compete_i=0)}\Big) = \beta_0 \;+\; \beta_1\, Female_i \;+\; \beta_2\, Gain\&Loss_i \\ +\; \beta_3\left(Female_i \times Gain\&Loss_i\right) \;+\; \beta_4\, PieceRatePayoff_i \\ +\; \beta_5\left(Female_i \times STEM_i\right) \;+ \boldsymbol{\gamma}'\mathbf{X}_i \;+\; \varepsilon_i,\end{aligned} \tag{G.1}$$

where $\mathbf{X}_i$ includes the same additional controls as in Models 3 and 4.

In Table G1, we report the average marginal effects (AMEs) of the regressions estimated from Equation (G.1).

Table G1: Logistic Regression Results on the Decision to Compete

Dependent Variable	*Compete* = 1 if Tournament, 0 if Piece Rate	
Model:	(3a)	(4a)
Female	4.70	3.64
	(4.74)	(5.23)
Gain & Loss	-0.30	-0.58
	(4.23)	(4.80)
Piece rate payoff	0.02	0.01
	(0.01)	(0.02)
Rank Belief	-9.97**	-11.88**
	(3.06)	(3.51)
Risk (self-reported)	2.82*	–
	(1.27)	
Risk (Bret)	–	0.31*
		(0.14)
λ_{loss}	–	-12.58**
		(4.45)
Competitiveness (self-reported)	3.48**	3.43**
	(1.02)	(1.11)
CRT	1.65	0.91
	(2.12)	(2.45)
Extraversion	0.89	2.46
	(1.49)	(1.62)
Agreeableness	2.36	4.25
	(2.24)	(2.48)
Conscientiousness	-0.46	0.12
	(1.65)	(1.80)
Emotional Stability	1.55	2.19
	(1.73)	(2.00)
Openness	-0.86	0.21
	(2.26)	(2.53)
GPA	-6.66**	-7.40*
	(2.52)	(2.90)
Born in Italy	6.91	6.12
	(7.40)	(7.61)
Age	0.72	0.93
	(0.71)	(0.92)
STEM	3.25	4.53
	(4.39)	(4.86)
Observations	407	310
Pseudo R^2	0.1169	0.1693

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All entries are average marginal effects (AME), multiplied by 100, with robust standard errors (also $\times 100$) in parentheses.

Appendix H

This appendix is organized as follows. In Section H.1 we give the full English translation of the instructions exactly as participants saw them on-screen; in Section H.2 we show the concise help popup summary that appeared when they clicked the "?" icon.

H.1 Instructions shown to participants (English translation)

INTRODUCTION

You are participating in a study financed by the University of Bologna. During this study, you will earn a sum of money determined by the rules described in the following pages. Payment will be made via PayPal in a confidential manner in the days following the completion of the study.
Today's study will last approximately 1 hour. The study is composed of 2 parts. Your final earnings will consist of the earnings obtained in each of the two parts plus 5 Euros as a participation fee.
The rules we will follow to determine your earnings are different in each part, and you will receive the instructions for each part in sequence.
It is forbidden to communicate with other participants during the study. If you have any questions, raise your hand and an assistant will come to your workstation to answer in private.

INSTRUCTIONS PART 1

The first part is composed of 3 stages. At the end of Part 1, the computer will make a random draw and select one of the 3 stages that will be used to calculate your earnings in Part 1.

PART 1 – STAGE 1: INSTRUCTIONS

In the first stage, you will act as a farmer who has purchased a 70-hectare plot of land for growing wheat. Unfortunately, you do not know which parts of the land are more fertile (i.e., those capable of producing more wheat). You have 70 wheat seeds available, and you must decide where to plant them, one after another.

Each planted seed will generate points, and your goal is to obtain the highest number of points possible.

For each seed, you have two choices. You can plant a wheat seed in a hectare of land where you have never planted anything before (indicated by "?") or in a hectare where you have already planted. Note that if you decide to plant the seed in a hectare where you have already sown, you may receive a number of points slightly different from those you received in the previous turn.

At the end of each turn, you will receive points based on how much wheat has grown in the selected hectare. In each turn, you can obtain between -10 and 15 points per hectare. Since it is possible to obtain fewer than zero points, we will give you an initial endowment of 700 points.

On your screen, you will see an image like the one reproduced in Figure 1. At the top, you will see the total number of accumulated points and the number of seeds left to plant. You can also consult the instructions by clicking on the "?".

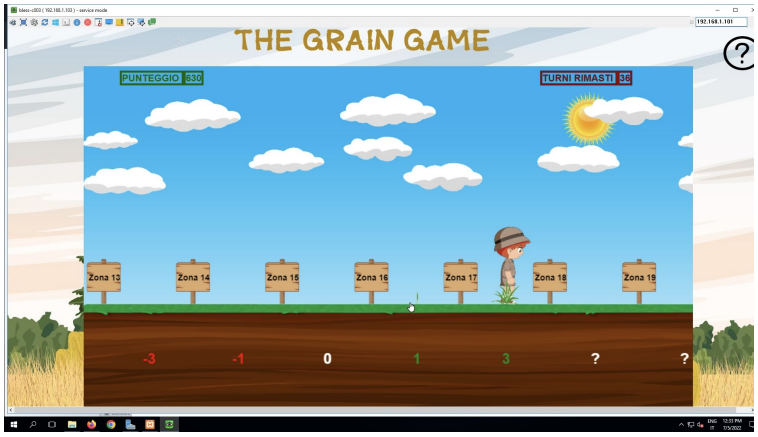


Figure 1

HOW TO PLANT THE SEEDS

1. To move right or left, use the arrow keys on your keyboard. Movements are possible one step at a time, both to the right and to the left.

2. To plant a seed, press the space bar.

3. You can plant each wheat seed in a hectare where you have already planted before, or you can move to a hectare to the right or to the left. The hectares where you can decide to plant are highlighted with a "?", while those where you have already planted are colored green (if you obtained a positive result) or red (if you obtained a negative result).

4. You can view the total number of points you have accumulated in the green box at the top left, and the number of seeds you still have to plant (i.e., remaining turns) in the red box at the top right.

5. You can use the "A" and "D" keys to scroll the field view

Remember: you have 70 wheat seeds available. The first seed has already been planted for you, and you can see it on your left. You must decide how to plant the remaining 69. Each planted seed will generate points. In each turn, you can obtain between -10 and 15 points per hectare. Since it is possible to obtain fewer than zero points, we will give you an initial endowment of 700 points.

EARNINGS FOR STAGE 1

At the end of Stage 1, we will sum all the points you have obtained. If Stage 1 is randomly drawn for payment in Part 1, you will receive 1 cent of a euro for every point obtained. We will call this payment system "piece rate."

Before starting, you will see a screen like the one reproduced in Figure 2, and you must select "NUOVA PARTITA" ("New Game").

Figure 2

After that, a screen like the one shown in Figure 3 will appear, and you will have to indicate your gender, enter the identification code you find at your workstation, and choose the level. For Stage 1, you must choose the first level and then click "NUOVA PARTITA" ("New Game").

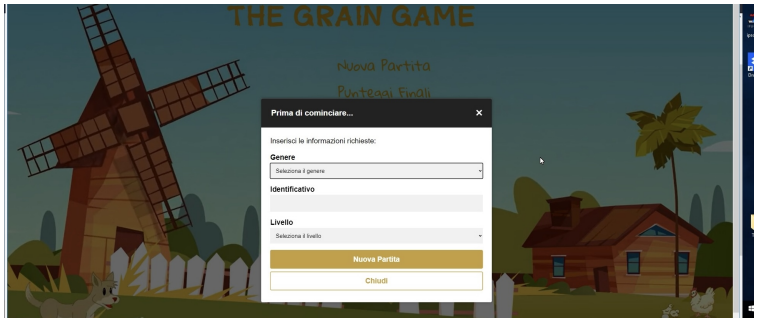


Figure 3

WHAT HAPPENS NOW?

We will give you a few minutes to answer the control questions at your workstation. If you have any questions, please raise your hand; an assistant will come to your workstation to answer. After answering any questions, Stage 1 will begin.

PART 1 – STAGE 2: INSTRUCTIONS

As in the first stage, you will act as a farmer who has purchased a 70-hectare plot of land to grow wheat, and you will have 70 wheat seeds available to plant one after another. Note that the fertility of the various hectares will be different from what you encountered in Stage 1, but you can still obtain between -10 and 15 points per hectare, and you will receive a new initial endowment of 700 points.

As in Stage 1, each planted seed will generate points, and your goal is to obtain the highest number of points possible.

For each seed, you have two choices. You can plant a wheat seed in a hectare of land where you have never planted anything before (indicated by "?") or in a hectare of land where you have already planted. If you decide to plant the seed in a hectare where you have already sown, you may receive a number of points slightly different from those you received in the previous turn. Each planted seed will generate points, and in each turn you can obtain between -10 and 15 points per hectare.

EARNINGS FOR STAGE 2

Payment in this stage depends on your performance relative to the performance of a group of 4 participants. Groups are formed randomly at the beginning of this stage, and each participant remains in the same group for the entire duration of this stage and the next stage.

GROUP ALLOCATION

Each group is composed of 4 participants, 2 men and 2 women. The identity of the group members is never revealed, neither during nor after the study, so all decisions remain anonymous. In Stage 2, your earnings depend on the number of points you obtain compared to the other three people in your group. The person in your group who obtains the highest number of points is the winner of the tournament.

The winner receives 4 cents of a euro for every point obtained, while the other participants do not receive any payment. In the event of a tie, the final ranking is determined by a random draw. We will call this payment system a "tournament."

At the end of Stage 2, you will know the total number of points you obtained, but you will not know whether you were among the tournament winners until the study is completed. If Stage 2 is randomly drawn for payment, your earnings depend on whether you were the tournament winner or not.

As in Stage 1, you will have to indicate your gender, enter the identification code you find at your workstation, and choose the level. For Stage 2, you must choose the second level and then click "NUOVA PARTITA" ("New game").

WHAT HAPPENS NOW?

We will give you a few minutes to answer the control questions at your workstation. If you have any questions, please raise your hand; an assistant will come to your workstation to answer. After addressing any questions, Stage 1 will begin.

PART 1 – STAGE 3: INSTRUCTIONS

STAGE 3: ACTIVITY AND EARNINGS

As in the first and second stages, you will act as a farmer who has purchased a 70-hectare plot of land to grow wheat, and you will have 70 wheat seeds available to plant one after another. Note that the fertility of the land's various hectares will be different from what you experienced in Stage 1 and Stage 2, but each planted seed will generate between -10 and 15 points per hectare. As in the previous stages, you will have a new initial endowment of 700 points.

In Stage 3, you yourself will choose the payment method you prefer: you can choose piece rate payment (as in Stage 1) or tournament payment (as in Stage 2).
Summarizing:

• Earnings in Stage 3 if you choose the Piece Rate scheme: you receive 1 cent for each point accumulated.

• Earnings in Stage 3 if you choose the Tournament scheme: If in Stage 3 you obtain more points than the person who won the tournament in Stage 2, then you will receive 4 cents for each point (that is, 4 times the piece-rate payment); otherwise, you will receive no payment. If you obtain exactly the same number of points as the Stage 2 tournament winner, a random draw will select either you or the other winner. If in Stage 2 you were the winner, then you have to surpass your own previous score.

The group composition (2 men and 2 women) is the same as in Stage 2. If you choose the Tournament payment scheme, you will not be informed about the outcome of the tournament until the end of the study. You will see a screen like the one in Figure 4 below, asking whether you want to choose piece rate payment or tournament payment for your performance in Stage 3.

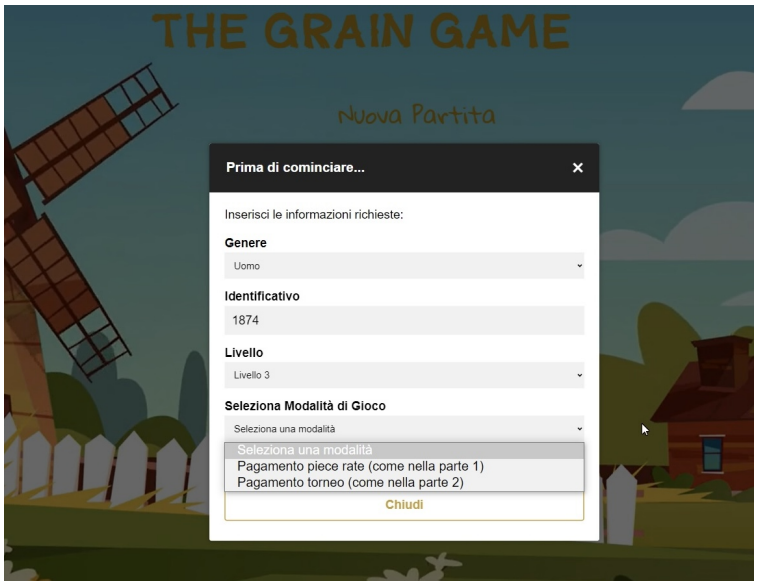


Figure 4

If Stage 3 is randomly drawn for final payment, your earnings depend on which payment choice you made and on your performance in this activity.

WHAT HAPPENS NOW?

We will give you a few minutes to answer the control questions at your workstation. If you have any questions, please raise your hand; an assistant will come to your workstation to answer. After addressing any questions, Stage 1 will begin.

INSTRUCTIONS FOR PART 2

Part 2 is composed of 2 stages. At the end of Part 2, the computer will make a random draw and select one of these 2 stages to calculate your earnings in Part 2.

PART 2 – STAGE 1: INSTRUCTIONS

In this stage, you will be asked to make some decisions. The decisions you make affect only your earnings and have no consequences for other participants. On your computer screen, you will see a table like the one reproduced in Figure 5 below. The 6 rows correspond to 6 decisions.

Figure 5

For each possible choice, there are different losses associated with a "heads" outcome of a coin toss. If you decide to toss the coin, you will win if the outcome is tails, according to the amounts shown in Figure 5. If you decide not to toss the coin, you will earn zero.
For example, in decision 4, if you decide not to toss the coin, you get 0. If you decide to toss the coin, you receive 6 if the result is tails; you get -5 if the result is heads.
If the result you obtain in this stage is negative, we will subtract the losses from your participation fee and from the other earnings obtained in Part 1.
To determine your earnings in this stage, one of the six rows will be randomly selected by the computer. If for the randomly selected row you indicated you want to toss the coin, the coin toss will be made by the computer, and you will learn your earnings associated with that outcome.

PART 2 – STAGE 2: INSTRUCTIONS
In this stage, you will be asked to make only one decision. The decision you make affects only your earnings and has no consequences for other participants. On your computer screen, you will see a table of 100 cells, as shown in Figure 6 below.

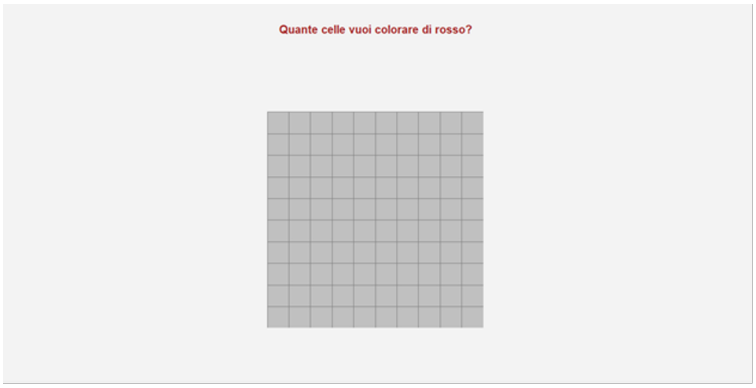


Figure 6

Your task is to decide how many cells you want to colour red.
You will receive 10 cents of a euro for each red cell. However, this earning is only potential because, after you have decided how many cells to colour red, the computer will randomly select one cell (out of the 100 in the table) for each participant.
If a gray cell is selected, you earn 10 cents of a euro for every cell you chose to colour red. If a red cell is selected, you earn 0 cents of a euro.
To choose how many cells to colour red, you must select them with your mouse. A message will appear showing the exact number of cells you have decided to colour red. If you are sure of your choice, press "Conferma" (Confirm); otherwise, press "Cancella" (Cancel) and make another choice.

H.2 Help Icon ("?") Instruction Recap (English Translation)

At any point during the grain-planting task, participants could click the "?" icon in the top bar to review a concise summary of the rules. Below is the English translation of that help popup.

Figure 7: Instructions recap for the Grain Game (Gain Treatment) as presented after clicking on Help Icon ("?")

Welcome! In this game, you will take on the role of a farmer who has purchased a 70-hectare plot of land to grow wheat. Unfortunately, you do not know which parts of the land are the most fertile (i.e., those capable of producing more wheat). You have 69 wheat seeds at your disposal and must decide where to plant them (the first one has already been planted at the start of the field). Each seed planted will generate points, and your goal is to earn as many points as possible. Once you have planted the final seed, we will add up the total number of points you have scored and give you one cent of a euro for every point. For each seed, you can move one hectare to the left or right of where you last planted, or remain in the same area. At the end of each turn, you can earn between 0 and 25 points for each hectare.[23]

Controls: - Use the arrow keys (← →) to move from one area of the field to another. - Press the space bar to plant a seed in the area you are currently in. - Use the A and D keys to scroll through the field view.

Appendix I

This appendix provides an overview of the reinforcement learning algorithm benchmark and the range of hyperparameters adopted.

Setup. We treat the Grain Game as a finite, episodic decision problem:

- The state variable $s \in \{0, \ldots, 69\}$ is the index of the current planting slot.
- In each decision round t, the agent chooses an action $a_t \in \{\text{Retreat, Exploit, Explore}\}$. At $s = 0$, the available actions are limited to {Exploit, Explore}.

[23] "Each hectare" is to be understood as a hectare in which a seed has been planted.

- An episode ends after $T = 69$ moves. The algorithm plays one of the six sequences, receiving the corresponding points in each planting slot plus the same zero-mean fluctuation as in the lab (see Section 3 and Appendix B).
- In the Gain&Loss setting, we add 700 points to the final cumulative reward to replicate the experimental endowment.

Action-selection rules. We implement two exploration mechanisms:

1. **ε-greedy.** Let $Q_t(s, a)$ denote the current estimate of the value of taking action a in state s at time t. With probability $1 - \varepsilon$ the agent selects the greedy action $\arg\max_a Q_t(s_t, a)$; with probability ε it chooses one admissible action at random:

$$a_t = \begin{cases} \arg\max_a Q_t(s_t, a), & \text{prob. } 1 - \varepsilon, \\ \text{uniform random action}, & \text{prob. } \varepsilon. \end{cases} \tag{I.1}$$

2. **ε-soft-max.** For every admissible action compute the *soft-max probability*

$$P_\tau(a) = \frac{\exp\big(Q_t(s_t, a)/\tau\big)}{\sum_{a'} \exp\big(Q_t(s_t, a')/\tau\big)}, \qquad \tau > 0 \text{ (temperature)}. \tag{I.2}$$

The same ε split is then applied, but in the exploration branch the agent takes the action with the *largest* $P_\tau(a)$ (no random draw):

$$a_t = \begin{cases} \arg\max_a Q_t(s_t, a), & \text{prob. } 1 - \varepsilon, \\ \arg\max_a P_\tau(a), & \text{prob. } \varepsilon. \end{cases} \tag{I.3}$$

Small τ values make $P_\tau(a)$ concentrate on the best actions; large values flatten the distribution.

Value update. After executing a_t in state s_t the agent observes the reward r_t and the next state s_{t+1}. The *temporal–difference error* δ_t depends on the learning algorithm:

$$\delta_t^{\text{Q-Learning}} = r_t + \gamma \max_{a_{t+1}} Q_t(s_{t+1}, a_{t+1}) \; - \; Q_t(s_t, a_t), \qquad \text{(off-policy)} \tag{I.4}$$

$$\delta_t^{\text{SARSA}} = r_t + \gamma\, Q_t(s_{t+1}, a_{t+1}) \; - \; Q_t(s_t, a_t), \qquad \text{(on-policy)} \tag{I.5}$$

With SARSA, $\gamma \in [0, 1]$ is the discount factor and a_{t+1} is the action actually taken at the next step. With QL, $\gamma \in [0, 1]$ will be randomized by the algorithm.

The estimate for the *current state–action pair* (s_t, a_t) is then updated as

$$Q_{t+1}(s_t, a_t) \; = \; Q_t(s_t, a_t) + \alpha_t\, \delta_t,$$

where α_t is the learning rate.

The exploration rate ε and the learning rate α are either kept constant or set to decay as $\varepsilon_t = \alpha_t = 1/\sqrt{t+1}$.

Training and evaluation. For every combination of $(\varepsilon, \alpha, \gamma, \tau)$, exploration rule ($\varepsilon$-greedy or ε-soft-max) and update algorithm (Q-Learning or SARSA) the agent is trained for 10 000 episodes and tested for a further 5 000 episodes with learning disabled. The procedure is repeated independently for each of the six payoff sequences.

Hyper-parameter grid. The search ranges are

$$\varepsilon \in \{0.5,\ 1,\ \text{decay}\}, \quad \tau \in \{0.5,\ 1,\ 5\}, \quad \alpha \in \{0.2,\ 0.5,\ 0.8,\ \text{decay}\}, \quad \gamma \in \{0.1,\ 0.5,\ 0.9\}.$$

Crossing these values with the two exploration rules, the two updated algorithms and the six payoff sequences yield 1 728 distinct configurations. Table I1 lists the three best configurations for each payoff sequence. The full output and code are available upon request.

Table I1: Top 3 simulation results by payoff sequence

Payoff Sequence	Algorithm	Exploration Strategy	ε	γ	α	Avg. Payoff	Maximum Position
1	Q-Learning	ϵ-greedy	0.5	0.9	0.2	808.9	35
	Q-Learning	ϵ-greedy	0.5	0.9	0.5	808.3	32
	SARSA	ϵ-greedy	0.5	0.9	0.2	791.0	46
2	Q-Learning	ϵ-greedy	0.5	0.9	0.5	933.7	29
	Q-Learning	ϵ-greedy	0.5	0.9	0.2	932.9	21
	Q-Learning	ϵ-greedy	0.5	0.9	0.8	932.5	30
3	Q-Learning	ϵ-greedy	0.5	0.9	decay	966.6	25
	SARSA	ϵ-greedy	0.5	0.5	0.5	965.9	24
	Q-Learning	ϵ-greedy	0.5	0.5	0.2	965.7	16
4	Q-Learning	ϵ-greedy	0.5	0.9	decay	509.2	43
	Q-Learning	ϵ-greedy	0.5	0.9	0.5	501.1	24
	Q-Learning	ϵ-greedy	0.5	0.9	0.8	499.6	25
5	Q-Learning	ϵ-greedy	0.5	0.9	0.2	944.1	41
	Q-Learning	ϵ-greedy	0.5	0.9	0.8	943.8	20
	Q-Learning	ϵ-greedy	0.5	0.9	0.5	943.4	19
6	Q-Learning	ϵ-greedy	0.5	0.9	0.5	977.6	15
	Q-Learning	ϵ-greedy	0.5	0.9	decay	977.1	15
	Q-Learning	ϵ-greedy	0.5	0.5	0.5	976.2	17

Notes: Each model was trained for 10 000 episodes and tested for 5 000. For payoff sequences 4–6 (Gain & Loss treatment) the values include the 700-point endowment. "Maximum Position" is the furthest slot reached.

Appendix J

For the Choice stage, Table J1 and Table J2 report an additional specification in which we include the interaction between the *Gain & Loss* treatment and the threshold dummy (*Threshold*= 1 if the payoff in the previous round was negative). The regression is estimated separately for men and women.

To evaluate whether crossing the threshold changes the likelihood of *Explore* (versus *Exploit*) once losses are possible, we test, within each gender, the linear restriction

$$\beta_{\text{Threshold}} + \beta_{\text{Threshold}\times\text{Gain\&Loss}} = 0.$$

For men, the null hypothesis is rejected in every specification (Model A: $\chi^2(1) = 17.28$, $p < 0.001$; Model B: $\chi^2(1) = 16.87$, $p < 0.001$; Model C: $\chi^2(1) = 10.74$, $p < 0.01$), implying that a negative payoff in the preceding round reduces the relative probability of exploring (versus exploiting).

For women, the same joint test is never significant (Model A: $\chi^2(1) = 2.09$, $p = 0.148$; Model B: $\chi^2(1) = 2.10$, $p = 0.148$; Model C: $\chi^2(1) = 0.93$, $p = 0.335$).

Hence, under the *Gain & Loss* being below the threshold in previous turn implies a decrease in exploration for men (relative to exploitation), while women's exploration behaviour remains unchanged. For men the result is in line with what is found by Chin et al. [2023].

Table J1: Random Effects Multinomial Logistic Results for *Move* – Women Only (Choice stage)

Dependent Variable	*Move* (Baseline = *Exploit*)					
	Model A_W		**Model B_W**		**Model C_W**	
	Retreat	Explore	Retreat	Explore	Retreat	Explore
Gain & Loss	1.326	0.543	1.198	0.467*	0.896	0.356*
	(0.449)	(0.187)	(0.435)	(0.159)	(0.382)	(0.145)
Threshold	0.343**	0.598*	0.364*	0.668	0.186***	0.436**

	(0.131)	(0.137)	(0.148)	(0.157)	(0.085)	(0.132)
Gain & Loss × Threshold	0.805	1.139	0.761	1.017	1.920	1.740
	(0.380)	(0.415)	(0.377)	(0.373)	(0.969)	(0.743)
Lag Payoff	0.780***	0.780***	0.782***	0.780***	0.780***	0.768***
	(0.027)	(0.027)	(0.028)	(0.027)	(0.031)	(0.031)
Compete (Tournament)	0.515	1.131	0.560	0.848	0.788	0.840
	(0.194)	(0.410)	(0.228)	(0.309)	(0.393)	(0.374)
Piece Rate Payoff	–	–	1.000	1.003**	1.001	1.003*
			(0.001)	(0.001)	(0.002)	(0.001)
Risk (Bret)	–	–	–	–	1.007	1.001
					(0.011)	(0.012)
Risk (Self-Reported)	–	–	1.130	1.482***	–	–
			(0.114)	(0.153)		
Competitiveness	–	–	0.788**	0.994	0.742**	1.020
			(0.060)	(0.075)	(0.064)	(0.095)
CRT	–	–	1.487*	1.165	1.594*	1.228
			(0.233)	(0.194)	(0.290)	(0.237)
Extraversion	–	–	0.944	0.924	0.949	1.040
			(0.095)	(0.095)	(0.111)	(0.134)
Agreeableness	–	–	1.347*	1.241	1.326	1.532*
			(0.202)	(0.195)	(0.240)	(0.296)
Conscientiousness	–	–	0.994	0.886	1.095	0.923
			(0.096)	(0.106)	(0.123)	(0.133)
Emotional Stability	–	–	0.874	0.763*	0.944	0.830
			(0.105)	(0.089)	(0.130)	(0.121)
Openness	–	–	0.770	0.749	0.884	0.881
			(0.123)	(0.117)	(0.158)	(0.176)
GPA	–	–	1.229	1.289	1.085	1.185
			(0.282)	(0.257)	(0.292)	(0.297)
Born in Italy	–	–	0.874	0.639	0.954	0.316
			(0.444)	(0.451)	(0.615)	(0.267)
Age	–	–	1.040	1.049	1.029	1.075
			(0.058)	(0.070)	(0.072)	(0.084)
STEM	–	–	0.601	0.838	0.664	1.148
			(0.230)	(0.287)	(0.289)	(0.471)
λ_{loss}	–	–	–	–	0.763	0.424*
					(0.238)	(0.148)
Constant	1.998	99.670***	0.771	1.565	0.762	3.434
	(1.047)	(55.910)	(1.722)	(4.245)	(1.967)	(11.970)
Observations	14,008		13,668		10,132	
Groups	206		201		149	
Wald χ^2	85.67		139.76		115.93	
Log PL	−8513.96		−8289.73		−6103.07	

Table J2: Random Effects Multinomial Logistic Results for *Move* – Men Only (Choice stage)

Dependent Variable	*Move* (Baseline = *Exploit*)					
	Model $\mathbf{A}_M$		**Model $\mathbf{B}_M$**		**Model $\mathbf{C}_M$**	
	Retreat	Explore	Retreat	Explore	Retreat	Explore
Gain & Loss	0.961	1.292	1.199	1.332	0.619	1.547
	(0.350)	(0.589)	(0.480)	(0.627)	(0.281)	(0.807)
Threshold	0.292*	0.659	0.312*	0.670	0.327	0.939
	(0.142)	(0.234)	(0.152)	(0.241)	(0.187)	(0.407)
Gain & Loss × Threshold	0.605	0.520	0.584	0.497	0.734	0.407
	(0.361)	(0.241)	(0.353)	(0.236)	(0.505)	(0.225)
Lag Payoff	0.637***	0.610***	0.640***	0.610***	0.657***	0.633***
	(0.034)	(0.033)	(0.034)	(0.033)	(0.039)	(0.039)
Compete (Tournament)	0.794	1.820	0.525	1.042	0.259**	0.754
	(0.296)	(0.852)	(0.211)	(0.479)	(0.120)	(0.374)
Piece Rate Payoff	–	–	0.999	1.005**	0.999	1.005**
			(0.001)	(0.001)	(0.001)	(0.002)
Risk (Self-Reported)	–	–	1.046	1.109	–	–

			(0.119)	(0.125)		
Risk (Bret)	–	–	–	–	1.043*	1.055*
					(0.018)	(0.018)
λ_{loss}	–	–	–	–	0.493**	0.682
					(0.134)	(0.208)
Competitiveness	–	–	1.252*	1.141	1.227	1.122
			(0.115)	(0.124)	(0.129)	(0.133)
CRT	–	–	1.031	0.786	0.982	0.722
			(0.182)	(0.167)	(0.205)	(0.193)
Extraversion	–	–	1.110	0.976	1.012	0.946
			(0.128)	(0.137)	(0.137)	(0.157)
Agreeableness	–	–	1.751**	1.379	1.696*	1.363
			(0.325)	(0.296)	(0.365)	(0.320)
Conscientiousness	–	–	1.193	1.105	1.040	1.302
			(0.205)	(0.198)	(0.199)	(0.262)
Emotional Stability	–	–	0.960	0.846	1.135	0.867
			(0.127)	(0.148)	(0.160)	(0.147)
Openness	–	–	0.893	1.018	0.809	1.060
			(0.178)	(0.234)	(0.181)	(0.270)
GPA	–	–	0.750	0.595*	0.889	0.758
			(0.162)	(0.145)	(0.219)	(0.200)
Born in Italy	–	–	0.047**	0.063**	0.005***	0.008**
			(0.047)	(0.066)	(0.007)	(0.013)
Age	–	–	0.958	0.985	0.951	0.987
			(0.039)	(0.054)	(0.061)	(0.073)
STEM	–	–	0.933	0.740	1.257	0.872
			(0.348)	(0.338)	(0.541)	(0.445)
Constant	19.257***	1940.643***	68.240	581.495*	795.754*	89.917
	(14.832)	(1704.032)	(171.513)	(1712.102)	(2203.712)	(297.266)
Observations	14,484		14,008		10,948	
Groups	213		206		161	
Wald χ^2	137.69		203.03		208.08	
Log PL	−7418.08		−7167.44		−5502.25	

Notes: Entries are Relative Risk Ratios (RRRs); robust s.e. in parentheses, clustered at participant level. Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Complaints and effort under pay and job discrimination

A labour market experiment

Marianna Baggio*† Alexia Gaudeul* Elisa Orlandi ‡

Abstract

The authors conducted an online incentivized job market experiment with 3,279 participants in two European countries (Germany and Romania). The experiment simulated the employer/employee relationship, presenting participants with different pay and job assignments that involved discrimination in terms of average pay for performing a real effort task. Discrimination was either based on differences in pay for the same job (pay discrimination) or differences in assignments to jobs with varying pay levels (job discrimination). We also varied the gender of the discriminated individuals; this is unusual, as studies on pay transparency usually focus on women and on the effects of pay transparency versus the absence of transparency.

The study measured participants' reactions to discrimination, which included complaining and asking for a pay raise or job promotion, and/or reducing productivity. Results indicate that participants in the discriminated group complained more frequently than participants in the privileged group. Overall, discrimination did not affect effort levels. However, discriminated individuals reported lower fairness and satisfaction with job allocations and wages, with wage satisfaction being notably lower under pay discrimination than under job discrimination.

We found that participants complained more under job discrimination than under pay discrimination, with this result driven by the high number of ungrounded complaints from non-discriminated individuals.

Gender differences in complaint rates were not significant, but women demonstrated higher effort levels under discrimination and reported more negative emotions than men. These gender differences (and, in some cases, the lack thereof) will be discussed in more detail in the discussion section.

Keywords: Gender Differences, Labour market experiment, Pay discrimination, Job discrimination

JEL Codes: C91, J16, J31, J71

*Joint Research Centre, European Commission
†European Foundation for the Improvement of Living and Working Conditions
‡University of Bologna

Executive summary

Policy context

The European Union has been committed to the principle of equal pay for equal work since the 1957 Treaty of Rome. Despite decades of efforts, a significant gender pay gap persists. To tackle this issue, the EU introduced the Pay Transparency Directive (Directive (EU) 2023/970) in June 2023, with a transposition deadline of June 2026. This directive aims to enhance pay transparency and address existing complexities, such as unclear pay systems and difficulties in defining work of equal value.

Key conclusions

Our study shows that, under pay transparency, employees who face discrimination are more likely to file complaints compared to those who do not face discrimination, without any negative impact on their effort, performance, time spent on tasks, or motivation to perform the task. Providing transparent information about average pay and job allocations by gender helps employees identify and challenge discrimination. This is the case even when the discrimination is more subtle than getting lower pay for the same job. Indeed, transparency is equally effective in helping to challenge discrimination that translates in reduced chances to obtain the better paid job positions. However, it's important to present pay and job information carefully, and help employees interpret it, so as to prevent unwarranted complaints from employees who are not discriminated against. This is because they may fail to recognize their privileged status and mistakenly complain. This is especially the case when challenging allocation to better paid job positions. These unfounded complaints can create administrative challenges, as they require more resources to validate their legitimacy.

Main findings

— **Complaint Behavior**: The study found that employees are more likely to complain when they experience job discrimination compared to pay discrimination. Discriminated employees are more inclined to complain than non-discriminated employees. However, non-discriminated employees also tend to complain about job assignments if they do not receive the higher-paying job even when they are not discriminated against. This suggests that some complaints may be driven more by dissatisfaction with outcomes like salary or job role rather than a clear recognition of discrimination.
— **Effort and Performance**: The study found that neither type of discrimination—whether pay or job discrimination—led to reduced effort among discriminated employees. While these employees reported lower perceived fairness in job allocation and less satisfaction with their wages, these feelings did not negatively impact their effort, performance, or the time they dedicated to tasks.
— **Gender Differences**: The study did not find significant gender differences in the likelihood of filing complaints among discriminated employees. However, women who faced discrimination tended to put in more effort, spend more time on tasks, and complete more work than men. Additionally, these women were less satisfied with their wage allocations and reported more negative emotions compared to their male counterparts.

Quick guide

This study investigates the impact of pay transparency on employee behaviour in the context of gender-based discrimination using an incentivized performance-based online experiment. Key terms include pay

discrimination (where wages differ within the same occupation) and job discrimination (disparities in job assignments). The experiment involved a simulated work environment where participants' reactions to discrimination were measured.

1 Introduction

The principle of equal pay for equal work, particularly between men and women, has been a cornerstone of European Union policy since the 1957 Treaty of Rome. Despite decades of legislation, recommendations, and initiatives aimed at promoting gender equality in the workplace, the effective implementation of this principle remains a significant challenge. The persistent gender pay gap and slowing progress in recent years [England et al., 2020, Eurostat., 2024, Goldin, 2014] have prompted renewed efforts to address this issue. In response to these challenges, the European Union introduced the Pay Transparency Directive[1] in June 2023, with a transposition deadline[2] of June 2026 for Member States. This directive aims to enhance pay transparency and address the complexities identified in previous evaluations, such as unclear pay systems and difficulties in establishing what constitutes work of equal value.

Pay transparency, while seemingly a straightforward policy instrument, is in fact a complex concept as its implementations involves various aspects of employment, including labor law, company remuneration policies, collective agreements, and individual behavioral responses. This research focuses on individual behavioural responses, as pay transparency measures are effective only if people react to them. The Pay Transparency Directive encourages the involvement of equality bodies, organizations, associations, and workers' representatives to help workers who wish, as stated in the Directive, "to seek to exercise their right to equal pay." However, this is possible only if discriminated workers actually file complaints. It is also crucial to consider that complaints made for false reasons—such as those stemming from individuals who believe they have been discriminated against but have not actually been, or from individuals who weaponize complaints for personal gain—can have negative consequences. These include the costs associated with evaluating unfounded claims and the potential for undermining the overall effectiveness of pay transparency measures.

Following this reasoning, our research specifically examines how individuals react to pay transparency in terms of complaint behaviour, effort levels, satisfaction, enjoyment, and emotions towards work tasks in a context where discrimination is present.

Using an incentivized performance-based online experiment, we simulate a work environment to explore two types of discrimination: pay discrimination, where wages differ within the same occupation, and job discrimination, involving disparities in job assignments between different groups. In the context of the European principle of "Equal Pay for Equal Work," pay discrimination can be viewed as unequal pay for the same job, while job discrimination represents unequal job assignments for work of equal value. Both types of discrimination violate this principle, as they lead to unequal treatment in either compensation or job allocation for comparable work. In our experiment, discrimination is based on gender, with participants assigned roles according to their reported gender identity.

By examining these factors, our study aims to contribute to the broader understanding of how pay transparency policies might influence workplace dynamics, particularly in contexts where discrimination persists. This research is crucial for informing future policy decisions and organizational practices aimed at promoting equality and fairness in the workplace.

[1](Directive (EU) 2023/970 of the European Parliament and of the Council of 10 May 2023 to Strengthen the Application of the Principle of Equal Pay for Equal Work or Work of Equal Value between Men and Women through Pay Transparency and Enforcement Mechanisms (Text with EEA Relevance), 2023)

[2]In EU legislation, a *transposition deadline* is the date by which every Member State must incorporate the directive's provisions into its own national law.

1.1 Scope and Relevant Policies

The gender pay gap (GPG) in the European Union (EU) was 12.7% in 2022, with substantial variation between countries, ranging from -0.7% in Luxembourg to 21.3% in Estonia [Eurostat., 2024]. This disparity indicates that women, on average, earn 0.87€for every 1€earned by men, requiring them to work an additional 1.5 months per year to match men's earnings. Despite continuous efforts, the gap has only narrowed by 3.7 percentage points over the last decade, indicating slow progress (Figure 1).

In addition, gender segregation by occupation and sector remains evident. According to recent statistics from the European Institute of Gender Equality in 2024, only 23.7% of executives and 7.8% of CEOs in the largest publicly listed companies in the European Union are women [EIGE, 2024]. More broadly, in the EU-27 job market, only one in every three managers is a woman [Hurley et al., 2021]. Women are also more often found working in lower-paying sectors, such as education and health, and frequently hold part-time positions [Eurofound, 2021].

Figure 1: Unadjusted gender Pay gap for the economy as a whole, 2010 – 2022. *Source: Eurostat online data code sdg_05_20*

Empirical research divides the gender pay gap (GPG) into two components: the "explained gap" and the "unexplained gap." The explained gap results from measurable differences in education, experience, industry, roles, and hours worked. The unexplained gap, on the other hand, encompasses fewer measurable factors, including caregiving responsibilities, family choices, gender preferences, and gender discrimination. Over the last four decades, the explained gap has been decreasing, often referred to as the "great convergence" [Blau and Kahn, 2006, 2017, England et al., 2020, Goldin, 2014, Olivetti and Petrongolo, 2016]. This decrease is largely attributed to the narrowing gender gap in education and experience, with women catching up in terms of human capital [Blau and Kahn, 2017, Goldin, 2014]. Despite these gains, a persistent unexplained gap remains, growing in relative importance as the explained gap decreases [Blau and Kahn, 2006, 2017, Goldin, 2014, Redmond and Mcguinness, 2019]. The unexplained gap, potentially linked to gender discrimination, thus remains a crucial focus in both academic and policy discussions.

Recent literature has focused on the role of pay transparency legislation as a policy instrument to narrow the GPG. The underlying hypothesis is that a lack of information favours discrimination, as information asymmetry hinders workers' awareness of discriminatory practices and their ability to take action [Baker et al., 2023, Eisenberg, 2011, Gulyas et al., 2023, Trotter et al., 2017]. The results on the effectiveness of pay transparency legislation in reducing the GPG are mixed. Some papers find a decrease in the gender pay gap following the introduction of pay transparency laws [Baker et al., 2023, Obloj and Zenger, 2022], although some scholars highlight that this reduction is due to a decrease in wage growth for male employees rather than an increase in female wage growth [Bennedsen et al., 2022]. Other studies find no impact of pay transparency legislation [Böheim and Gust, 2021, Gulyas et al., 2023]. For a comprehensive review of the literature, see [Bennedsen et al., 2023, Cullen, 2024].

In June 2023, the Pay Transparency Directive[3] entered into force to address the unexplained gender pay gap in the European Union. This directive aims to enhance pay transparency and provide more information on gender-based pay differences for equivalent work. Key measures include better access to justice for pay discrimination victims and ensuring pay transparency for workers and employers. The directive mandates that job seekers be informed about starting salaries or pay ranges, prohibits employers from asking about candidates' pay history, and allows employees to request average pay levels by gender for similar roles. Companies must report pay levels annually if they have over 250 employees and every three years for those with between 150 and 250 employees. Corrective action is required if an unjustified pay gap greater than 5% is found. Compensation for gender pay discrimination includes back pay, bonuses, and payments in kind, with the burden of proof on the employer.

1.2 Literature Review and Hypotheses

Our analysis encompasses three interconnected layers: (1) comparing individuals who experience discrimination versus those who do not, (2) examining the differential outcomes of pay discrimination versus job discrimination, and (3) investigating gender-specific reactions to these forms of discrimination and their interactions. And all the hypotheses we decided to test fall under one of these three categories.

The behavioral effects of pay transparency, which correspond to our first layer of analysis, have already been the subject of several studies, revealing both positive and negative impacts. According to the fair wage-effort hypothesis, confirmed by the literature on gift-giving, workers exert effort proportional to the fairness of the wage they receive [Akerlof and Yellen, 1990, Breza et al., 2018, Clark et al., 2010, Gächter and Thöni, 2010]. Studies on relative wages suggest that workers who perceive their wages as lower than their peers tend to reduce effort [Clark et al., 2010], work fewer hours [Gagnon et al., 2020], and quit more often [Dube et al., 2019]. Conversely, higher-wage workers may show increased effort [Charness et al., 2016] or, if they previously experienced higher income rankings, might reduce effort levels regardless of current income and rank [Clark et al., 2010]. Cullen and Perez-Truglia [2022] found that the impact of wage transparency varies depending on whether the wage gap is with peers or managers; a gap with managers motivates harder work, while a gap with peers leads to demotivation.

A recent study by Baggio and Marandola [2023] highlights the behavioral effects of pay transparency, investigating its impact on employees' performance, effort provision, and actions to correct pay disparities. Their findings indicate that pay transparency does not affect average performance but reduces effort provision for individuals with below-average wages. Additionally, pay transparency increases potentially justified requests to correct pay disparities while decreasing unjustified requests.

Following this line of literature, we formulated the following hypothesis:

Hypothesis 1. *Complaints will be more frequent, and effort will be lowered more, in the group that is discriminated against than in the group that is not discriminated against (everything else equal, i.e., given the same job and/or same pay).*

Moreover, Grasser et al. [2023] highlight that for underpaid individuals, pay transparency negatively affects perceived fairness and thus lowers motivation. Conversely, they found that pay transparency does not

[3](Directive (EU) 2023/970 of the European Parliament and of the Council of 10 May 2023 to Strengthen the Application of the Principle of Equal Pay for Equal Work or Work of Equal Value between Men and Women through Pay Transparency and Enforcement Mechanisms (Text with EEA Relevance), 2023)

significantly impact individuals who are paid more than they deserve according to their performance or those who are paid in alignment with their performance.

Building on these findings, and the previous ones, we propose our second hypothesis:

Hypothesis 2. *Satisfaction with wage and job allocation will be lower among the discriminated group. Similarly, willingness to exert effort and experienced enjoyment of effort will be lower among the discriminated group.*

While these hypotheses do not differ much from the existing literature, we believe the real novelty of this research lies in expanding the analysis to include the interplay between pay transparency and different forms of discrimination, as well as gender-specific responses to these factors. As explained in the introduction, we focus on two dimensions of discrimination: pay discrimination, where the discriminated group is paid less for the same job, and job discrimination, where there are disparities in job allocations.

This type of investigation is important in light of the still existing occupational and sectoral gender segregation in the European labor market (see figure presented in subsection 1.1). Additionally, evidence from the literature shows that men have higher promotion rates than women [Blau and Devaro, 2007], and in general, women face barriers advancing in the hierarchy of the labor market, dealing with what the literature describes as a glass ceiling [Bertrand, 2018, Blau and Kahn, 2017, Ransom and Oaxaca, 2005].

An additional reason to investigate the behavioral impacts of pay transparency in different discrimination contexts is connected with how easily understandable the information about pay structures/job allocations (and thus potential discrimination) in a firm is. While pay discrimination is relatively easy to spot and recognize once individuals are provided with data on wages by gender, data on job discrimination are not necessarily as readable and comprehensible at first glance. This is because job discrimination encompasses more complex reasoning about abilities, competency, and meritocracy in general. It follows that, from a behavioral point of view, the impact of getting information on discriminatory practices that take place in the workplace on workers' efforts, satisfaction, complaints, and so on might depend on how easy the discrimination is to recognize and comprehend.
Considering the above, we hypothesize that:

Hypothesis 3. *Pay discrimination is more likely to give rise to complaints, and lowers effort more, among the discriminated, than job discrimination does.*

Hypothesis 4. *Satisfaction with wage among the discriminated is lower if wage differs within occupation (pay discrimination), and satisfaction with occupation is lower if job assignments differ across groups (job discrimination).*

Concerning our last layer of analysis, we believe that gender might be a relevant factor, as the literature has shown differences between men and women in retaliation [Dehdari et al., 2019], risk aversion [Croson and Gneezy, 2009]; although some studies find it to be context-dependent, Filippin and Crosetto 2016), self-evaluation [Exley and Kessler, 2022], willingness to compete [Niederle and Vesterlund, 2007, 2011] and

confidence [Exley and Nielsen, 2024, Möbius et al., 2022]. All factors that might affect how individuals react to discrimination and pay transparency.

Additionally, there is a gender gap in self-reporting skills [Murciano-Goroff, 2022] and self-promotion [Exley and Kessler, 2022], as well as what the literature has often described as the probability of "speaking up," with women being less likely to intervene or participate than men [Coffman, 2014]. Recent research has also focused on gender differences in negotiations, finding women to be both less likely to negotiate and less successful at it[4] [Babcock and Leschevar, 2003, Hernandez-Arenaz and Iriberri, 2019, Säve-Söderbergh, 2019]. For a comprehensive review of the gender gap in negotiation, see Recalde and Vesterlund [2023] and Recalde and Vesterlund [2020].

Women are also found to ask for lower salaries, referred to as the gender ask gap [Roussille, 2024, Säve-Söderbergh, 2019] and in general have lower salary expectations [Biasi and Sarsons, 2022]. Furthermore, they are less likely to apply for promotions or ask for promotions [Bosquet et al., 2019, Haegele, 2024].

In addition, the literature shows that stereotypes and gender roles also affect behaviour. Stereotype threats[5] negatively impact women's performance, as found in studies on gaming [Kaye and Pennington, 2016, Vermeulen et al., 2016]. Similarly, both descriptive and normative stereotypes have been shown to have negative effects on women's career outcomes [Heilman, 2012, Heilman et al., 2024]. Gender norms are also significant, with social penalties associated with not conforming to these norms. For example, women are penalized more than men when initiating negotiations [Bowles et al., 2007], assertive women face societal backlash [Amanatullah and Tinsley, 2013], self-promoting women are considered less socially appealing than similar men [Rudman, 1998], power-seeking intentions negatively affect women but not men [Okimoto and Brescoll, 2010], and there are also gender-specific norms in risk-taking behaviour [Grimm, 2019]. Furthermore, research shows that women take potential backlash into account, being as assertive as men in negotiations only when negotiating on behalf of someone else [Amanatullah and Morris, 2010], exhibiting less self-promotion out of fear of backlash [Moss-Racusin and Rudman, 2010] and avoiding talking too much or speaking up to avoid social penalties [Brescoll, 2011].

Given all this evidence, we expect women to be less likely to complain in the face of discrimination and to ask for a different job allocation or different wage in our experimental context. We formalize this in Hypothesis 5.

Hypothesis 5. *Women are less likely than men to ask for a pay rise or a change in job allocation in response to discrimination, and they are comparatively more likely than men to reduce their effort intensity in response to discrimination.*

To check if participants are able to correctly assess whether the discrimination stems from an unfair pay structure or rather from an unfair job allocation, we propose the following hypothesis:

Hypothesis 6. *Assignment to a worse job is more likely to be ascribed to discrimination by the discriminated group if it is job discrimination rather than pay discrimination. Lower pay is more likely*

[4]Leibbrandt and List [2015] found that when the possibility of negotiation is explicit, there are no gender differences in negotiation outcomes, suggesting that context and clarity play significant roles in mitigating gender disparities in negotiation behaviour.

[5]The term stereotype threat refers to a situation in which the existence of (negative) stereotypes about a social group may affect the performance of individuals in that group [Nguyen and Ryan, 2008, Spencer et al., 2016, Steele and Aronson, 1995].

to be ascribed to discrimination by the discriminated group if it is pay discrimination rather than job discrimination.

Hypothesis 7 can be seen as linked with Hypothesis 5. Given what has already been said about women asking for lower salaries, negotiating less, and being less likely to report high performance [Biasi and Sarsons, 2022], in addition to women paying themselves less than their male counterparts for the same amount of fixed work [Breza et al., 2018], we decided to test whether this means that they are also less likely to recognize unfair wages and job allocations as discrimination. We thus formulate Hypothesis 7 as:

Hypothesis 7. *Women are less likely than men to ascribe their assignment to job A (especially under job discrimination), or their lower pay (especially under pay discrimination), to discrimination.*

The last hypothesis is strictly related to Hypotheses 1 and 2, as it focuses on the impact of discrimination on discriminated versus non-discriminated groups. It relates to the literature on the morale effects of discrimination, which found that in the presence of pay disparity, workers report being more unhappy [Breza et al., 2018] and perceive lower fairness [Grasser et al., 2023].

Hypothesis 8. *Groups that are discriminated experience higher levels of negative emotions and lower levels of positive emotions than groups that are not discriminated. This difference is larger for women than for men.*

All hypotheses tested in this report have been preregistered in OSF (`https://osf.io/y3ch4`). The data and analytic code are available on the project website at `https://osf.io/ea2sp/`.

2 Experimental Setup

In the following section, we will describe the experimental design, treatments, methodology, recruitment procedures, and reasoning. This study was approved by the JRC's Ethical Review Committee (Protocol Number 30506). All participants provided informed consent before taking part in the experiment.

We conducted an incentivized performance based online experiment in two European countries: Germany and Romania. The experiment was conducted online to accommodate a large number of participants, eliminate the need for travel amidst COVID-19 restrictions, reduce operational costs, and easily allowed matching people from different locations. The quality of the data was guaranteed through several quality assurance measures: panel provider checks such as ratings, feedback, and updates; attention and quality checks to evaluate participant comprehension and exclude submissions of low quality, such as those completed in implausibly short durations; and technical checks to ensure unique IP addresses and non-reusable links.

Both the recruitment and the experiment were run by DevStat. Recruitment was conducted through email invitations sent to a random sample of participants from Schlesinger's online panels. The selection criteria included employed individuals aged between 18 and 65 who, during the reference week, performed work for pay, profit, or family gain. The sampling strategy ensured that the sample was representative of the employed population in terms of age, education level, and region (NUTS1) for each combination of country

and treatment, based on the latest Eurostat data from the 2020 Labour Force Survey (LFS). The experiment was run on Devstat's virtual server.

In Appendix A, we provide some screenshots of the experiment we ran to give an idea of the participants' experience in our experiment (specifically, these screenshots are for a male participant in a context with job discrimination against men).

2.1 Experimental blueprint

Participants were informed that the experiment included two different roles: employers and employees. See Figure 2 for the blueprint of the experimental design.

Participants selected to play the role of employees were required to complete both a preliminary task and a main task, based on the string task by Isen and Reeve 2005. This involves identifying strings of 6 letters in alphabetical order, meaning that participants had to inspect each string and decide whether the six letters in the string are in alphabetical order or not (as read from left to right). The task was chosen because it is a real effort task and is not gender-stereotypical; additionally, available data from a previous study [Baggio and Marandola, 2023] supported the design of the current study.

Participants in the role of employees (employees for short henceforth) were informed that employers[6] would assess their performance on the preliminary task and their gender to determine their wages and allocate them to one of two different job positions: Job A (for worse performers) or Job B (for better performers). Employees were explicitly informed that the main task for both positions would be exactly the same.

Figure 2: Employees Experiment Blueprint

In the preliminary task, employees were required to classify 120 strings within 5 minutes. At the end of the task, participants were asked to guess how they performed (meaning the number of strings they correctly classified), how other people performed, and which gender performed better in the task. All measures of self-confidence in the experiment will be named from here forward as Confidence1, Confidence2 and Confidence3.[7]

Participants were then asked to decide which piece of information about job, wage, and gender distribution in the firm they wished to receive. They could choose to see either the average pay by job position and the percentage of people per job position, or the average pay by gender and the percentage of men and women in the firm. Regardless of their choice, both pieces of information were shown to them, but the information they

[6] A small group of participants were actually assigned to the employer role in a separate, auxiliary experiment, thus avoiding any form of deception.

[7] The questions related to the measures of self-confidence were as follows: Confidence1 asked participants, "Out of the strings you did, how many strings do you think you classified correctly?" with responses recorded as an integer between 0 and 120. Confidence2 asked, "On average, how well do you think other candidates performed?" with responses on a scale from 1 to 4: 1 = Better than me, 2 = Slightly better than me, 3 = Slightly worse than me, 4 = Worse than me. Confidence3 asked, "On average, which gender do you think performed best in this task?" with responses on a scale from 1 to 4: 1 = Your gender better than Opposite gender, 2 = Your gender slightly better than Opposite gender, 3 = Your gender slightly worse than Opposite gender, 4 = Your gender worse than Opposite gender.

chose was shown first. [8] Additionally, after the first two aforementioned pieces of information, employees were presented with all wage structure information on a single page. All these pieces of information depend on the treatment participants were assigned to. See Figure 3 and Figure 4 for visual representations of the information presented to participants.

Then, employees were asked to guess which job they were allocated to (variable belief) and then were informed which job they actually got assigned to. Additionally, they were questioned about the fairness of the job obtained and satisfaction with the wage obtained, as well as fairness concerning other people's job allocation and satisfaction with other people's wages (variables that will be named Own fairness, Own satisfaction, Other fairness, and Other satisfaction).

They were also asked to express their emotions concerning the job they were allocated to, their salary, and the general gender distribution of the firm (negative feeling variable). Before the main task, they were also asked whether they feel close to same-sex employees (closeness variable). For the exact phrasing of the relevant questions and to check how the questions were aggregated and variables transformed, see Appendix B.

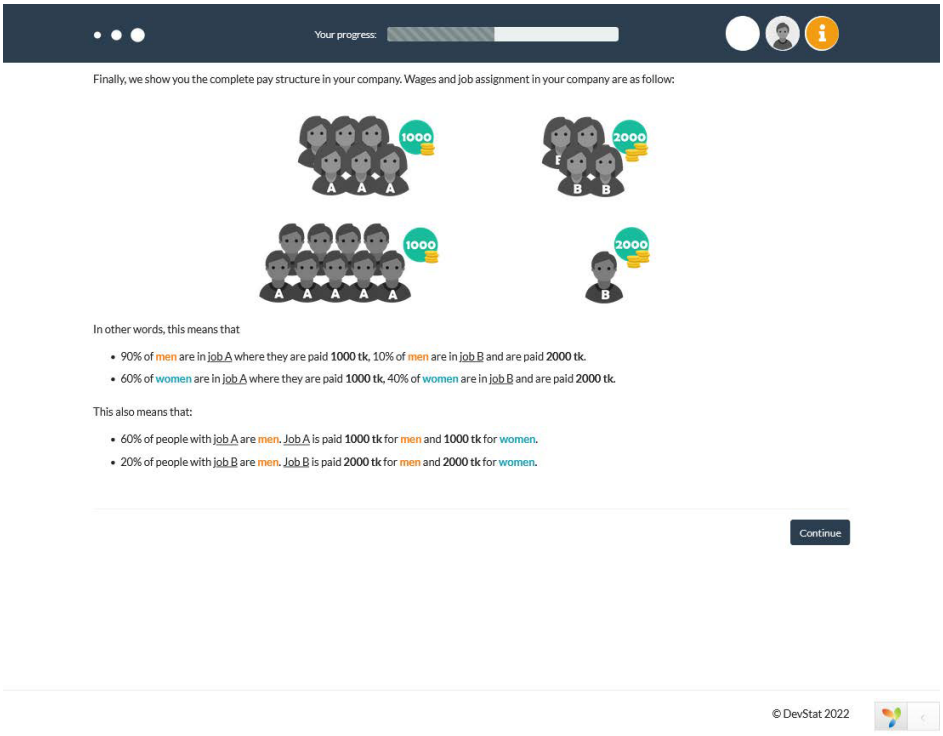


Figure 3: Job Discrimination Information Presentation. *Source: DevStat*

[8] In hindsight, a clearer prompt would have been "Which information would you like to see *first*?" so that participants knew both pieces would appear. Because the choice only determined display order and all information was always revealed, we regard this as a harmless ordering preference rather than substantive deception, and we believe it did not affect the experimental outcomes.

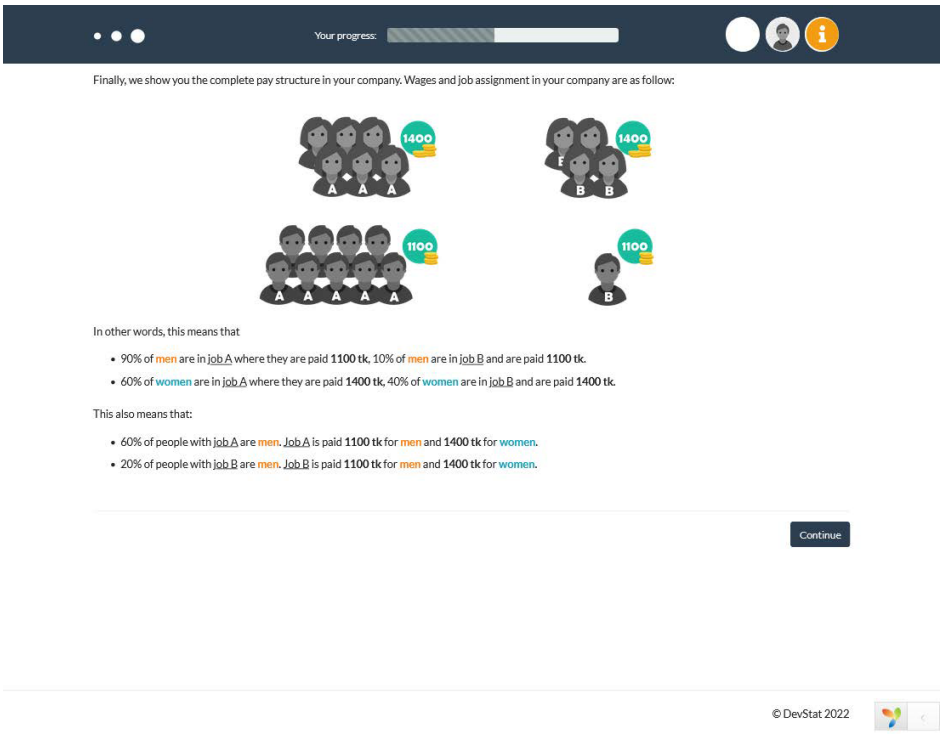


Figure 4: Pay Discrimination Information Presentation. *Source: DevStat*

Employees were then asked to perform the main task, which consisted of the same string task as the preliminary task; however, this time they were asked to correctly classify as many as 240 strings in 10 minutes. It's important to note that employees were made aware that their pay would not depend on their performance on the main task, but only on the wage that was assigned to them by the employer after the preliminary task. Once employees completed their main task, they could choose to complain if they believed the firm had unfairly assigned jobs and wages. Employees were also informed that complaining and thus requesting a change in job allocation could lead to a wage increase for those who rightfully complain and deserve the raise. However, if the complaint is unfounded and unsuccessful, their wage would be reduced.

2.2 Experimental treatments

The experiment employed a 2x2 factorial design, resulting in a total of four distinct treatments to examine the effects of different types of discrimination within a workplace setting (see Table 1). The treatments were organized based on two factors: type of discrimination (job discrimination or pay discrimination) and the group discriminated against (men or women). For clarity, we name the treatments as follows: JW for Job discrimination against Women, JM for Job discrimination against Men, PW for Pay discrimination against Women, and PM for Pay discrimination against Men. Participants were randomly assigned to the four treatments, stratifying by country (Germany or Romania) and self-reported gender, aiming for an approximately 50/50 gender balance within each treatment in both national subsamples.

	Discriminated Group: Women	Discriminated Group: Men
Job discrimination	JW	JM
Pay Discrimination	PW	PM

Table 1: Experimental treatments

In treatments involving pay discrimination (PW and PM), the discriminated group (DG) received lower pay than the privileged group (PG), regardless of their performance (and consequently, independently of the job assigned). The privileged group (PG) received 1400 tokens, while the discriminated group (DG) received 1100 tokens.

In treatments involving job discrimination (JW and JM), the likelihood of participants being assigned to the better job position, Job B (which pays 2000 tokens, as opposed to the 1000 tokens paid by Job A), was influenced by whether they were part of the discriminated group (DG). Specifically, the discriminated group (DG) faced higher requirements to be assigned to Job B. Employees' performance during the preliminary task determined their classification into three performance groups: those in the bottom 60% were categorized as Low Skills, those between 60% and 90% were classified as Medium Skill, and those above 90% were considered High Skills. Within the discriminated group (DG), only High Skill participants (those who performed in the top 10%) were assigned to Job B. Meanwhile, in the privileged group, both Medium Skills and High Skill employees (who represent the top 40% of performing individuals in the group) were given Job B[9].

Figure 5 shows the resulting breakdown of job allocation and wage structure across treatments. We show job allocation and pay for the DG and for the PG under both Job and Pay discrimination.

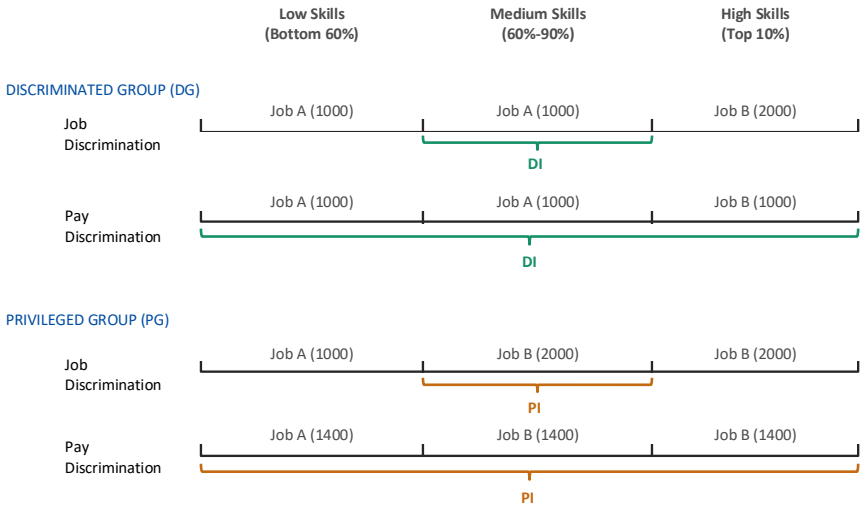


Figure 5: Wage structure and Job allocation by type of discrimination, group (discriminated or privileged) and skills level.

[9]Participants who played the employer role simply applied the deterministic rule shown in Figure 5: for each employee they matched gender, treatment, and skill tier to the prescribed job and token amount.

2.3 Discriminated and Privileged Individuals

In our experimental study, it is important to distinguish between the discriminated group (DG), discriminated individuals (DI), privileged group (PG), and privileged individuals (PI). Participants are assigned to the discriminated group (DG) or the privileged group (PG) according to their gender as specified in the experimental treatments. However, being part of the discriminated group (DG) does not necessarily mean facing discrimination, which involves being treated differently in terms of wage or job allocation compared to participants with similar skills in the privileged group (PG). Nor does belonging to the privileged group (PG) necessarily mean benefiting from the privilege and obtaining a wage or job position different from individuals with similar skills in the discriminated group (DG).

Discriminated individuals (DI) are participants within the DG who, given their skill levels, would have received a different wage or job position if they had belonged to the PG. Symmetrically, privileged individuals (PI) are participants in the PG who, given their skill levels, would have received a different wage or job position if they had belonged to the DG.

In the context of job discrimination, Low Skills (bottom 60%) and High Skills (top 10%) individuals in the DG receive job assignments (Job A and Job B, respectively) that align with their performance levels and are the same as those in the PG. Thus, they are not actually discriminated against. Conversely, Medium Skills (60-90%) individuals within the DG are unfairly assigned to Job A, while their counterparts in the PG are assigned to Job B. This discrepancy identifies the medium skill individuals within the DG as the discriminated individuals (DI) in job discrimination.

In the case of pay discrimination, the DG systematically receives lower pay than the PG, regardless of job assignment or skill level. Therefore, all individuals within the DG are considered discriminated individuals (DI).

Symmetrically, in job discrimination treatments (JW and JM), the Medium Skills individuals in the PG who are assigned to Job B, while their counterparts in the DG are assigned to Job A, are considered privileged individuals (PI). In pay discrimination treatments (PW and PM), all individuals in the PG, regardless of skill level, are privileged individuals (PI) because they receive higher pay than their counterparts in the DG for the same job assignments and skill levels.

Figure 5 outlines DI and PI across groups, skill levels, and types of discrimination.

2.4 Complaints and requests for correction

After completing the main task and being informed about their wage, participants could choose either to not complain and keep their job and wage, or to complain, which meant submitting a request for wage or job allocation correction, if they believed the company had unfairly assigned jobs or wages to some of the participants (not necessarily themselves). See Figure 6 and Figure 7 for the exact wording. If the employee was right to complain, their performance would be reevaluated, and they would be assigned the wage and job they actually deserved, specifically, the wage and job position that participants in the other group (PG or DG) were receiving for the same level of skills. Otherwise, if participants were wrong to complain (meaning that no discrimination was present), they would lose half of their wage. Participants were clearly explained all the potential outcomes arising from complaining[10].

[10]Part of the employers' task was to apply the rule described above: approve a complaint when it signalled discrimination for that employee's gender, treatment, and skill tier; otherwise reject it.

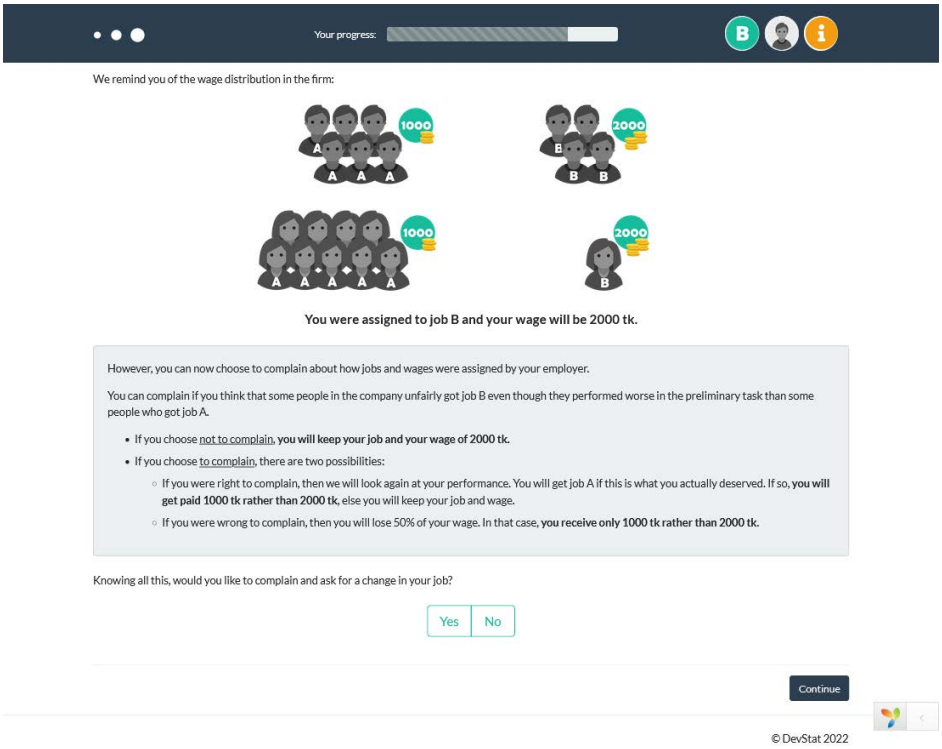


Figure 6: Job Discrimination Complaint Screen. *Source: DevStat*

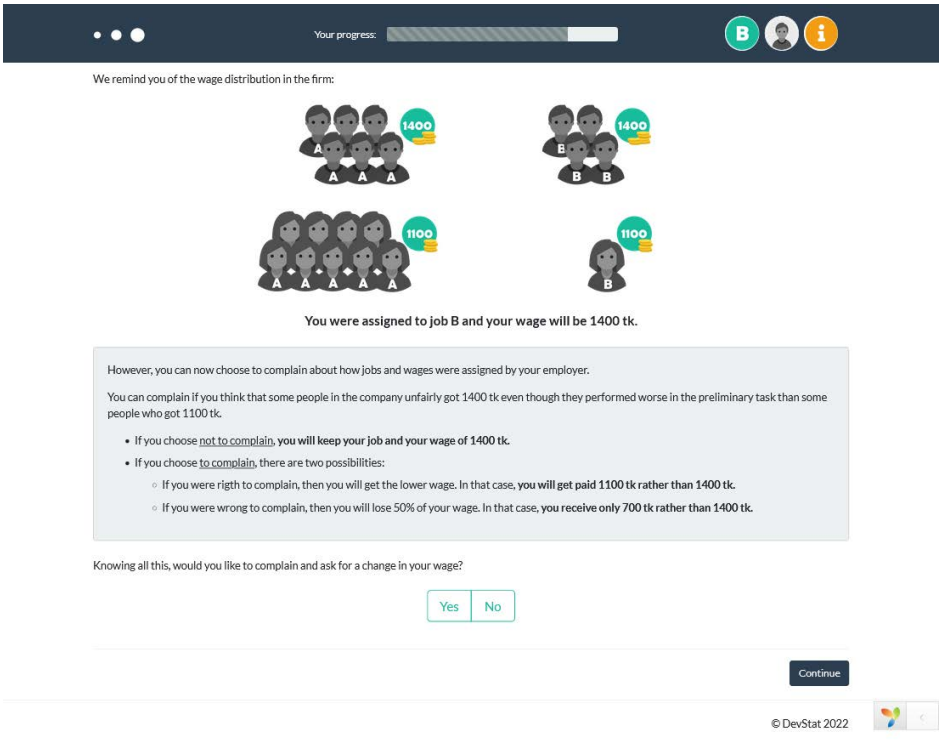


Figure 7: Pay Discrimination Complaint Screen. *Source: DevStat*

Participants were never actually wrong to complain since, by design, there was always a form of discrimination (either job discrimination or pay discrimination). Therefore, none of the participants had their wage cut in half. However, the complaint resulted in a change of job position or wage only in cases where, for the same skill level, participants in the privileged group (PG) and participants in the discriminated group (DG) were paid differently or had different jobs allocated. This means that complaining and denouncing discrimination resulted in changes in job allocation or wage only for those who were either privileged individuals (PI) or discriminated individuals (DI). Specifically, after complaining, DI would receive the better job or higher wage (depending on the type of discrimination they were facing), and PI would receive the worse job or lower wage.

Therefore, denouncing discrimination was generally costly only for PI, as they would lose their privileged position and receive a less favourable job or lower wage. On the other hand, DI benefited from complaining as it led to an improved job position or higher wage[11].

2.5 Post-Experimental Questionnaire

In the post-experimental questionnaire, participants were asked for feedback on the experiment, such as guessing what the experiment was about and whether, in their opinion, the experiment entailed deception. There were also additional questions about the task, including whether they enjoyed it, whether they were bored by it, and whether they got tired of it. All these responses were averaged to provide a summary variable called *Motivation*. Additionally, participants were asked to explain their reasoning behind a complaint or a request to change job allocation (*Reason change*) by selecting from a set of predefined options. They also reported their beliefs about the employer's gender (*Gender Employer*).

They were also asked about sociodemographic characteristics, such as their current work situation, employment status (full-time, part-time, or self-employed), including the sector (private or public), and the size of the company they work for.

Finally, risk attitudes (*Risk taking*), confrontational attitudes (*Less confrontation*), views on competition (*Competition*), and income equality (*Income equality*) were assessed.

As a last question, participants' views on gender roles were elicited using a Likert scale (*Prejudice variable*). See Appendix B for the phrasing used in the post-experimental questionnaire.

2.6 Sample selection and characteristics

The experiment was conducted in two European Union countries: Germany and Romania. The total sample consisted of 3,279 participants, with 1,622 participants from Germany 1657 participants from Romania[12]. These countries were selected because they represent contrasting examples of gender pay disparities within the EU. Germany has one of the highest unadjusted Gender Pay Gaps (20.1%), while Romania has one of

[11]It is important to note that participants, being unaware of their actual skill level relative to others, did not have full knowledge of the outcome of their complaint. From a rational standpoint, a complaint by members of the Privileged Group (PG) would be justified only if they believed that, as a group, they outperformed those in the Discriminated Group (DG), thus validating their higher wages. Alternatively, complaints could also stem from a strong commitment to fairness and equality, where individuals in the PG were willing to accept a lower wage to address and correct discrimination, even at their own expense.

[12]Irrespective of the level of complaints by women, the sample size was designed to have sufficient power to identify a difference of 10% in the rate of complaints between men and women in the same situation. This means that if the real difference is 10% (e.g., women complain at a 30% rate and men at a 40% rate), then there is an 80% probability that we will find a significant (5% level of significance) difference between the rates for men and women with a one-sided test of proportion.

the lowest (2.2%), according to Eurostat data (2018). Additionally, Germany and Romania represent two different EU regions: Central Europe and Eastern Europe, respectively.

To ensure that participants were paying the required attention during their participation, a series of attention and quality checks were implemented throughout the experimental process. These checks were designed to assess comprehension and focus. Data from participants who failed these checks were excluded, as well as observations from individuals who completed the experiment too quickly, i.e., faster than 95% of the participants in the pilot. Also, participants spending more than three times the mean duration of the pilot (126.6 min) were eliminated since they likely faced some kind of interruptions, and their discontinued participation was also considered invalid.

These participants were replaced within the same strata of sex, age, and education to preserve the sample's representativeness. See Table 2 for the descriptive statistics by country[13].

Table 2: Descriptive statistics by country

	Germany (N=1622)	**Romania (N=1657)**	**Overall (N=3279)**
Gender			
Women	806 (49.7%)	829 (50.0%)	1635 (49.9%)
Men	816 (50.3%)	828 (50.0%)	1644 (50.1%)
Age			
Mean (SD)	40.8 (11.0)	38.1 (10.3)	39.4 (10.7)
Median [Min, Max]	38.0 [18.0, 64.0]	37.0 [18.0, 64.0]	37.0 [18.0, 64.0]
Sector			
Public	584 (36.0%)	446 (26.9%)	1030 (31.4%)
Private	817 (50.4%)	1150 (69.4%)	1967 (60.0%)
Self employed	221 (13.6%)	61 (3.7%)	282 (8.6%)
Work Status			
Employed Full-time	1245 (76.8%)	1507 (90.9%)	2752 (83.9%)
Employed Part-time	235 (14.5%)	90 (5.4%)	325 (9.9%)
Self-employed	142 (8.8%)	51 (3.1%)	193 (5.9%)
Other	0 (0%)	9 (0.5%)	9 (0.3%)
Company Size			
1–9 employees	140 (8.6%)	291 (17.6%)	431 (13.1%)
10–49 employees	417 (25.7%)	368 (22.2%)	785 (23.9%)
50–249 employees	556 (34.3%)	390 (23.5%)	946 (28.9%)
250+ employees	509 (31.4%)	608 (36.7%)	1117 (34.1%)
Education			
Primary	94 (5.8%)	93 (5.6%)	187 (5.7%)
Secondary	854 (52.7%)	855 (51.6%)	1709 (52.1%)
Post-secondary	674 (41.6%)	709 (42.8%)	1383 (42.2%)

Within the experimental design, each treatment comprised of at least 800 employees, balanced by gender with at least 400 men and 400 women. See Table 3 for the sample distribution by gender, country and treatment.

[13]See Appendix C for descriptive statistics of psychological variables by country and gender.

Table 3: Sample distribution by gender, treatment and country

Country	Gender	Treatment				
		JW	PW	JM	PM	Total
Germany	Women	200	200	203	203	**806**
	Men	204	206	203	203	**816**
	Total	**404**	**406**	**406**	**406**	**1622**
Romania	Women	217	210	202	200	**829**
	Men	212	200	202	214	**828**
	Total	**429**	**410**	**404**	**414**	**1657**
Total	Women	417	410	405	403	**1635**
	Men	416	406	405	417	**1644**
	Total	**833**	**816**	**810**	**820**	**3279**

Out of the total sample, about half of the participants were assigned to the discriminated group (1634 individuals, comprising 807 men and 827 women). Of the people assigned to the discriminated group, half were in the pay discrimination treatment (827) and the other half in the job discrimination treatment (822).

Participants were paid on average 5€in Germany and 2.1€in Romania. These payments depended on the salary they received, which in turn depended on their performance, the discrimination they faced, and whether or not they chose to file a complaint.

3 Results

This section presents the findings of our study. First, we describe the main outcome variables and their definitions. We then proceed to present the results of the hypothesis tests, discussing each hypothesis in turn. Finally, we summarize the main conclusions derived from our analysis.

3.1 Outcome variables

The primary outcome variables in this analysis are the rate of complaints (*Change*), the effort level in the main task (*Main Correct*), the *Effort Ratio.* and the time spent on the main task (*Main Time*). *Change* measures whether participants complained about job or wage allocation after completing the main task or not. The effort level in the main task is defined as the number of strings that participants correctly classified in the main task. Additionally, the Effort Ratio, defined as (*Main Correct / Preliminary Correct*), represents the number of strings correctly assigned in the main task relative to the number of strings correctly assigned in the preliminary task. The Percentage Correct, defined as (*Main Correct / Main Answered*), indicates the proportion of correctly classified strings out of the total number of strings answered in the main task.

Another key outcome variable is the reason for change (*Reason Change*), which indicates the main reason participants complained (or did not) . Reasons for requesting a change included believing they deserved a better wage or job given their performance, perceiving discrimination against other participants, being paid less than others of the same gender, being paid less than others of the opposite gender, and wanting to take the chance of increasing their wage. Reasons for not requesting a change were perfectly symmetrical and included believing they did not deserve a better wage or job given their performance, perceiving no

discrimination against other participants, being paid more than others of the same gender, being paid more than others of the opposite gender, and not wanting to risk losing 50% of their wage. This variable captures both egoistic reasons, such as increasing one's own salary or getting a better job, and altruistic reasons, which can be considered forms of allyship within or between groups.

Other important outcome variables include perceptions of fairness regarding their own job allocation (*Own Fairness*) and the perceived fairness of others' job allocations (*Other Fairness*). Additionally, the analysis considers levels of satisfaction with their own wage allocation (*Own Satisfaction*) and participants' satisfaction with others' wage allocations (*Other Satisfaction*). These variables measure participants' satisfaction and perceived fairness regarding job allocation and wage distribution overall.

Participants' levels of motivation during the task (*Motivation*), a composite measure of emotions experienced by participants (*Negative Feelings*), and the perceived closeness participants feel towards other employees of their gender (*Closeness*) are also examined. The *Motivation* variable reflects how motivated participants were during the task, with higher values indicating greater motivation. The *Negative Feelings* variable is an aggregate measure of emotions[14], where positive emotions were reverse-coded to create a consistent index representing negative emotions. This transformation ensured that higher values indicate worse overall feelings towards the wage or job allocation and wage structure. Closeness assesses how close participants feel towards other employees of their gender.

The transformed variables and the phrasing of all the questions are discussed in more detail in Appendix B.

3.2 Illustrative predictions for complaint rates

Figure 8 provides an illustrative overview of how complaint rates might look under a set of strong simplifying assumptions. We label these as "stylized predictions" because participants in our experiment did *not* have full information about their actual skill level and, thus, face some uncertainty about whether a complaint might succeed.

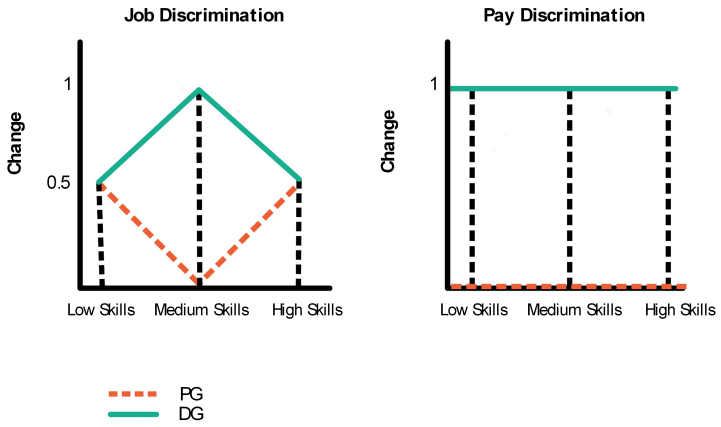


Figure 8: Illustrative predictions of complaint rates under stylized assumptions.

Specifically, the lines in Figure 8 assume:

[14]The *Negative Feelings* variable is a composite measure derived from the following negative emotions: Anger, Sad, Mad, Downhearted, Irritated, Depressed, Frustrated, Overlooked/Passed over/Unrecognized, Anxious, Worried, and Afraid. Positive emotions included in the composite measure, which were reverse-coded, are Happy, Cheerful, and Satisfied.

(i) **Full information:** Participants have perfect knowledge of their own skill level relative to others.

(ii) **Perfect rule knowledge:** Participants know the employer's allocation rule and that any discrimination they encounter is fully determined by this rule.

(iii) **Zero cost of complaining:** Other than the possibility of being reassigned or losing wages if the complaint is unsuccessful, there is no disutility associated with the act of complaining itself (e.g., no other cost).

(iv) **Random tie-breaking:** When participants are indifferent (i.e., if complaining yields the same expected payoff as not complaining), they toss a fair coin to decide whether to complain.

In this stylized setting, under job discrimination, participants with low or high skills should be indifferent between complaining or not as their complaints won't be followed by a change in wage or job allocation. However, they should always complain when they are discriminated individuals (DI), as complaining leads to a higher salary or better job allocation. Figure 8 (orange, dashed line) shows the theoretical rate of complaints for individuals in the Privileged Group (PG), always under selfishness and full information. In the case of job discrimination, participants with low or high skills should be indifferent between complaining or not. Privileged individuals (PI), on the other hand, should never complain because it is costly for them to do so, as they would receive a lower salary or worse job allocation.

In our actual experiment, participants did not have perfect information about their relative performance, and there may be non-monetary or psychological costs to complaining. Therefore, these curves should be read only as providing a *theoretical benchmark* for how self-interested participants *might* behave if they were fully informed.

3.3 Hypotheses Testing

H1: Discrimination Increases Complaints and Reduces Effort

As observed from the tests in Table 4, overall H1 is partially verified. The discriminated group (DG) complains 10 percentage points more than the privileged group (PG) (p-value<0.01). In Table 5, it is evident that under pay discrimination, participants in DG complain more than those in PG (+17 percentage points, p-value<0.01). Under job discrimination, the discriminated group complains more often than the privileged group (+3 percentage points), but this difference is not statistically significant.

Table 4: Results of Tests for H1: Comparison of Complaint Rates and Effort Levels between Discriminated Group (DG) and Privileged Group (PG)

	PG	DG	p-value
Change			
No	1166 (71.53%)	1013 (61.43%)	<0.0001
Yes	464 (28.47%)	636 (38.57%)	
Main Correct	143.75 (46.01)	144.49 (46.34)	0.3243
Main Time	428.69 (218.00)	421.54 (223.07)	0.7127
Effort ratio	2.29 (1.33)	2.40 (3.54)	0.9385
Percentage correct	0.75 (0.21)	0.75 (0.21)	0.1720

Note: Statistical tests were performed as follows: Pearson's Chi-squared test for Change and Wilcoxon rank sum tests for Main Correct, Main Time, Effort Ratio, and Percentage Correct. In the table, the means and standard deviations (SD) are presented for Main Correct, Main Time, Effort Ratio, and Percentage Correct. The number of Yes and No responses along with their percentages are presented for Change.

Concerning the effort level, there was no significant difference between the discriminated group and the privileged group in both pay discrimination and job discrimination treatments.

Notably, we also found that the time spent on the task is the same for both the discriminated group and the privileged group. Additionally, the effort ratio is the same between DG and PG. This indicates that all measures of effort are similar between the two groups.

Table 5: Results of Tests for H1: Comparison of Complaint Rates and Effort Levels between Discriminated Group (DG) and Privileged Group (PG) By type of discrimination

	Job discrimination			**Pay discrimination**		
	PG	**DG**	**p-value**	**PG**	**DG**	**p-value**
Change						
No	520 (63.34%)	496 (60.34%)	0.2112	646 (79.85%)	517 (62.52%)	<0.0001
Yes	301 (36.66%)	326 (39.66%)		163 (20.15%)	310 (37.48%)	
Main Correct	143.03 (43.71)	143.03 (46.65)	0.5440	144.48 (48.24)	145.94 (46.00)	0.4501
Main Time	421.89 (216.02)	414.69 (225.14)	0.7886	435.59 (219.91)	428.35 (220.91)	0.7570
Effort ratio	2.35 (1.69)	2.32 (1.28)	0.4084	2.24 (0.81)	2.49 (4.83)	0.4572
Percentage correct	0.74 (0.21)	0.74 (0.21)	0.4915	0.76 (0.21)	0.76 (0.21)	0.2060

Note: Statistical tests were performed as follows: Pearson's Chi-squared test for Change and Wilcoxon rank sum tests for Main Correct, Main Time, Effort Ratio, and Percentage Correct. In the table, the means and standard deviations (SD) are presented for Main Correct, Main Time, Effort Ratio, and Percentage Correct. The number of Yes and No responses along with their percentages are presented for Change.

In Table 6, we present the reasons that participants provided for complaining or not complaining. The reasons provided are significantly different ($p < 0.001$) between those who complained and those who did not complain. Notably, among the participants who did complain, the majority cited merit—the belief that they deserved more than what they received given their performance—as the reason for complaining and requesting a change in job allocation or wage. Additionally, 16% of those who complained indicated an altruistic reason for their complaint, stating that the employer discriminated against other people.

Table 6: Results of Tests for H1: Reasons Given for Asking (or Not Asking) for a Change/Complaint

Reason Change	**Change**		Total	p-value
	No	Yes		

Merit	559 (25.65%)	681 (61.91%)	1240 (37.82%)	<0.0001
Discrimination others	787 (36.12%)	183 (16.64%)	970 (29.58%)	
Comparison own gender	147 (6.75%)	40 (3.64%)	187 (5.70%)	
Comparison other gender	74 (3.40%)	84 (7.64%)	158 (4.82%)	
Take chance/risk	612 (28.09%)	112 (10.18%)	724 (22.08%)	
Total	2179 (66.45%)	1100 (33.55%)	3279 (100.00%)	

Note: Statistical test was performed using Pearson's Chi-squared test.

H2: Discrimination Lowers Satisfaction, Willingness to Exert Effort, and Enjoyment

As the results from the tests reported in Table 7 suggest, Hypothesis 2 is partially verified. The discriminated group reported lower perceived fairness in their job allocation and lower satisfaction with their wage compared to the privileged group (p-value<0.001). Specifically, when asked about Own fairness and Own satisfaction, the discriminated group consistently provided lower ratings.

Furthermore, the discriminated group also reported lower perceived fairness and satisfaction when assessing other participants' job allocations and wages, although these differences were less significant than their assessments of their own job positions and wages.

Regarding the second part of Hypothesis 2, there was no significant difference in the level of motivation performing the task.

Table 7: Results of Tests for H2: Comparison of fairness, satisfaction, motivation and negative feelings between Discriminated Group (DG) and Privileged Group (PG)

	PG	**DG**	p-value
Own Fairness	3.00 (0.79)	2.87 (0.78)	<0.0001
Own Satisfaction	3.03 (0.77)	2.81 (0.79)	<0.0001
Other Fairness	2.52 (0.83)	2.47 (0.82)	0.0595
Other Satisfaction	2.54 (0.86)	2.46 (0.84)	0.0052
Motivation	3.18 (0.63)	3.14 (0.67)	0.2363

Note: Statistical tests were performed as follows: Wilcoxon rank sum tests for Own Fairness, Own Satisfaction, Other Fairness, Other Satisfaction, and Motivation. In the table, the means and standard deviations (SD) are presented for Own Fairness, Own Satisfaction, Other Fairness, Other Satisfaction, and Motivation.

3.3.1 H3: Pay Discrimination Provokes More Complaints and Lowers Effort More Than Job Discrimination for the DG

Hypothesis 3 is not verified, as the discriminated group (DG) does not complain more under pay discrimination than under job discrimination. Nor does the DG exert more effort under pay discrimination than it does under job discrimination, as presented in Table 8.

Table 8: Results of Tests for H3: Comparison of Complaint Rates and Effort Levels for DG between Pay Discrimination and Job Discrimination

	Job discrimination	**Pay Discrimination**	p-value
Change			
No	496 (60.34%)	517 (62.52%)	0.3644
Yes	326 (39.66%)	310 (37.48%)	
Main Correct	143.03 (46.65)	145.94 (46.00)	0.1987
Main Time	414.69 (225.14)	428.35 (220.91)	0.3212
Effort ratio	2.32 (1.28)	2.49 (4.83)	0.4474
Percentage correct	0.74 (0.21)	0.76 (0.21)	0.1164

Note: Statistical tests were performed as follows: Pearson's Chi-squared test for Change and Wilcoxon rank sum tests for Main Correct, Main Time, Effort Ratio, and Percentage Correct. In the table, the means and standard deviations (SD) are presented for Main Correct, Main Time, Effort Ratio, and Percentage Correct. The number and percentage of responses for Change are also presented.

Table 9: Results of Tests for H3: Comparison of Complaint Rates and Effort Levels for PG between Pay Discrimination and Job Discrimination

	Job discrimination	**Pay Discrimination**	p-value
Change			
No	520 (63.34%)	646 (79.85%)	<0.0001
Yes	301 (36.66%)	163 (20.15%)	
Main Correct	143.03 (43.71)	144.48 (48.24)	0.2930
Main Time	421.89 (216.02)	435.59 (219.91)	0.2356
Effort ratio	2.35 (1.69)	2.24 (0.81)	0.4686
Percentage correct	0.74 (0.21)	0.76 (0.21)	0.0368

Note: Statistical tests were performed as follows: Pearson's Chi-squared test for Change and Wilcoxon rank sum tests for Main Correct, Main Time, Effort Ratio, and Percentage Correct. In the table, the means and standard deviations (SD) are presented for Main Correct, Main Time, Effort Ratio, and Percentage Correct. The number and percentage of responses for Change are also presented.

However, it's interesting to note that the difference in complaints between the discriminated group (DG) and the privileged group (PG) is much larger under pay discrimination (17 pp) than under job discrimination (3pp). Additionally, there are overall more complaints under job discrimination, with 40% of the DG and 37% of the PG complaining, compared to 37% of the DG and 20% of the PG under pay discrimination. This suggests that job discrimination leads to more complaints overall. For more details, see Table 8 and Table 9.

Figure 9 shows the rate of complaints under pay discrimination and job discrimination by performance (skill level). It appears that under job discrimination, what matters is if the participant is assigned to Job A, not if they are part of the Discriminated Group. For the group of participants with low skills, there is no difference in the complaint rate between PG and DG, which seems to suggest that individuals try their chances, given that they have nothing to lose. For the medium skills group, the complaint rate of DI is not higher than the complaint rate of low skill participants who got Job A, which implies that DI are not

particularly aware of being actually discriminated against and that they should get Job B. For the high skill group, the Discriminated Group does not seem aware that complaining is not costly for them (indeed, they complain at the same rate as PG). Under pay discrimination, we see that the DG complains more than the PG, especially when in the high skills group.

The higher rate of complaints in job discrimination thus seems to be determined by low skill individuals in PG who take their chances, even though they should be aware they have no chance of getting their complaint accepted.

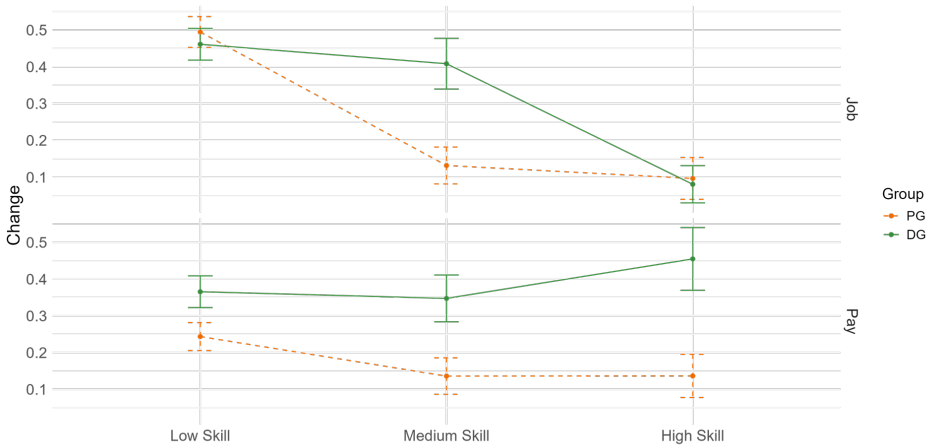


Figure 9: Rate of Complaint by Treatment, Group and Skill Level

The previous results are confirmed by the logistic regression analysis in Table 10. Under job discrimination, being in the Discriminated Group does not affect the likelihood of complaining, while being assigned to Job B (high wage) significantly reduces the chances of filing a complaint. Under pay discrimination, on the other hand, being in the discriminated group and getting paid less matters, especially so when participants are assigned to Job B.

Thus, it seems that under job discrimination, participants may complain because they received the worse job position or low wage, not because they recognize themselves as deserving of a better job position (i.e., they recognize being DI). In contrast, under pay discrimination, people with high skills (who are assigned Job B) seem to recognize that they are deserving of a higher wage or better job position.

However, participants do not provide reasons for their complaints that necessarily align with this observed behavior (see Table 11). Indeed, even though participants in the DG who choose to complain in the pay discrimination treatments indicate merit as the main reason for their complaints, which aligns with our expectations, own merit is emphasized even more in job discrimination, which is not expected given how DI do not seem to complain more than low skills participants who are allocated in Job A. This might be because performance is used to justify discrimination in Job discrimination treatments. Participants may be particularly sensitive to performance outcomes, and they might struggle to accept or acknowledge lower performance, as it could conflict with their self-perception or ego. In addition, those in the DG who complain under pay discrimination often mention being paid less than the other gender. Conversely, being paid less than others of the same gender is very seldom mentioned under job discrimination.

Participants in the DG who do not complain mainly do not see discrimination or do not think they merit better job or pay. The risk factor is, however, also important. This is surprising for pay discrimination because it highlights that participants in the DG did not realize they had nothing to lose from complaining.

Table 10: Logistic Regression Results for Job and Pay Discrimination (Dependent Variable: Change)

Dep. Variable	**Change**											
Subsample	**Job Discrimination**			**Pay Discrimination**			**Job Discrimination**			**Pay Discrimination**		
	Log-Odds	**CI**	**p**	**Log-Odds**	**CI**	**p**	**Log-Odds**	**CI**	**p**	**Log-Odds**	**CI**	**p**
Intercept	-0.02	-0.19 – 0.15	0.797	-1.13	-1.34 – -0.93	<0.001	-2.01	-2.39 – -1.66	<0.001	-1.38	-1.55 – -1.21	<0.001
JobB	-1.99	-2.40 – -1.60	<0.001	-0.71	-1.10 – -0.34	<0.001						
DG	-0.19	-0.42 – 0.03	0.092	0.56	0.30 – 0.82	<0.001	-0.43	-1.26 – 0.30	0.276			
JobB*DG	-0.24	-1.09 – 0.53	0.565	1.11	0.58 – 1.65	<0.001						
Low Pay							1.99	1.60 – 2.40	<0.001	0.87	0.64 – 1.09	<0.001
Low Pay*DG							0.24	-0.53 – 1.09	0.565			
N	1643			1636			1643			1636		
R^2	0.102			0.046			0.102			0.037		

Note: Logistic regression analysis was conducted with "Change" (indicating whether a participant filed a complaint) as the dependent variable. The results are reported as log-odds with 95% confidence intervals.

Table 11: Results of Tests for H3: Reasons Given for Asking (or Not Asking) for a Change/Complaint in DG; Pay Discrimination vs. Job Discrimination

Change	Yes			No		
Discrimination	Job	Pay		Job	Pay	
Reason Change			p-value			p-value
Merit	228 (69.94%)	178 (57.42%)		137 (27.62%)	129 (24.95%)	
Discrimination others	43 (13.19%)	47 (15.16%)		188 (37.90%)	204 (39.46%)	
Comparison own gender	6 (1.84%)	13 (4.19%)	<0.0001	34 (6.85%)	29 (5.61%)	0.2400
Comparison other gender	16 (4.91%)	49 (15.81%)		14 (2.82%)	7 (1.35%)	
Take chance/risk	33 (10.12%)	23 (7.42%)		123 (24.80%)	148 (28.63%)	

Note: Statistical tests were performed using Pearson's Chi-squared test.

Concerning PG, the main reasons provided for complaining are based on their own merit for both pay discrimination and job discrimination. This suggests that many PG participants do not actually seem to realize that they are in the PG group and that they are always allocated to the wage/job that corresponds to their skill level. In addition to this (and to further confirm it), in job discrimination, 15% of participants say they complain to take their chances. Notably, in pay discrimination, more than 27% of participants cite discrimination against others as a reason for complaining (which can be interpreted as a form of allyship), while in job discrimination, this is true for less than 15% of participants in the PG group (see Table 12).

Table 12: Results of Tests for H3: Reasons Given for Asking (or Not Asking) for a Change/Complaint in PG; Pay Discrimination vs. Job Discrimination

Change	Yes			No		
Discrimination	Job	Pay		Job	Pay	
Reason Change			p-value			p-value
Merit	171 (56.81%)	104 (63.80%)		113 (21.73%)	180 (27.86%)	
Discrimination others	48 (15.95%)	45 (27.61%)		183 (35.19%)	212 (32.82%)	
Comparison own gender	19 (6.31%)	2 (1.23%)	<0.0001	33 (6.35%)	51 (7.89%)	0.0096
Comparison other gender	17 (5.65%)	2 (1.23%)		18 (3.46%)	35 (5.42%)	
Take chance/risk	46 (15.28%)	10 (6.13%)		173 (33.27%)	168 (26.01%)	

Note: Statistical tests were performed using Pearson's Chi-squared test.

H4: Satisfaction with Wage and Job Varies by Type of Discrimination

Hypothesis 4 is partially verified as we found that there is indeed lower satisfaction with wage among the discriminated under pay discrimination compared to job discrimination. However, for participants in the DG,

there is no difference in fairness rating of job allocation between pay discrimination and job discrimination. Additionally, there are also no differences for DG in satisfaction with the wages of other individuals and the fairness of other individuals' job allocations. See Figure 10.

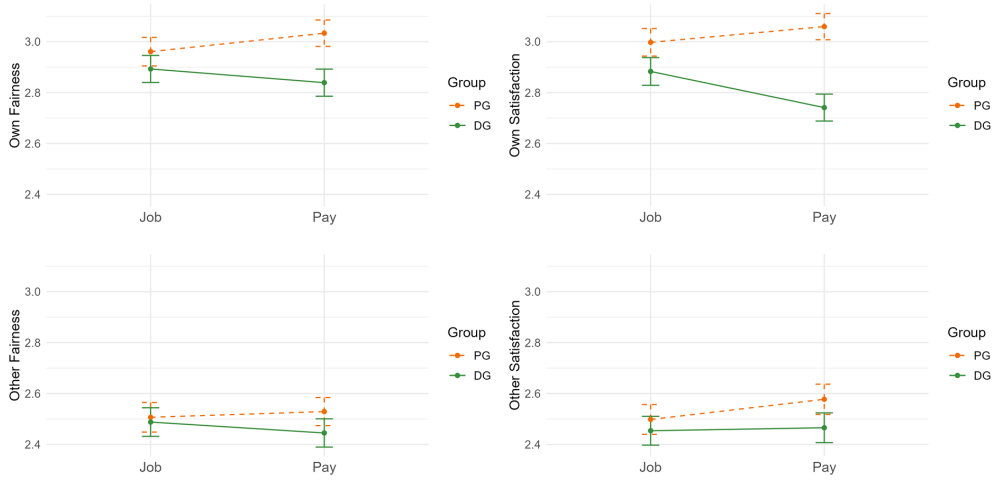


Figure 10: Comparison of fairness, satisfactions between Pay discrimination and Job discrimination by group

H5: Women Are Less Likely to Seek Pay Raises or Job Changes, More Likely to Reduce Effort

Hypothesis 5 is not verified, as there are no gender differences in complaints (neither for PG nor DG). Women in DG, contrary to our hypothesis, put in higher effort, as measured by Main Correct (but this is also true for PG), Main Time, and Percentage Correct. Interestingly, against H5, men seem to react to discrimination by reducing the time spent on tasks, while women actually spend more time. Testing is reported in Table 13.

Table 13: Results of Tests for H5: Comparison of Complaint Rates and Effort Levels for PG and DG between Genders

	PG			**DG**		
	Women	**Men**	p-value	**Women**	**Men**	p-value
Change						
No	585 (72.40%)	581 (70.68%)	0.4417	516 (62.39%)	497 (60.46%)	0.4203
Yes	223 (27.60%)	241 (29.32%)		311 (37.61%)	325 (39.54%)	
Main Correct	147.62 (47.49)	139.95 (44.20)	0.0018	146.31 (49.53)	142.66 (42.83)	0.0078
Main Time	425.80 (220.67)	431.54 (215.44)	0.4796	449.09 (210.87)	393.83 (231.54)	<0.0001
Effort ratio	2.26 (1.04)	2.33 (1.55)	0.5055	2.34 (1.28)	2.47 (4.85)	0.1265
Percentage correct	0.76 (0.20)	0.75 (0.21)	0.3713	0.77 (0.20)	0.73 (0.21)	0.0012

Note: Statistical tests were performed as follows: Pearson's Chi-squared test for Change and Wilcoxon rank sum tests for Main Correct, Main Time, Effort Ratio, and Percentage Correct. In the table, the means and standard deviations (SD) are presented for Main Correct, Main Time, Effort Ratio, and Percentage Correct. The number and percentage of responses for Change are also presented.

H6: Recognition of Discrimination Depends on Its Type

H6 coincides with H4, and hence we find it partially verified. Specifically, we verify that lower pay is more likely to be ascribed to discrimination by the discriminated group under pay discrimination rather than job discrimination. For DG, lower satisfaction with wage under pay discrimination than under job discrimination can be seen as ascribing the lower wage to discrimination. However, the first part of the hypothesis, that a worse job assignment is more likely to be ascribed to discrimination under job discrimination than pay discrimination, is not verified. See Figure 10.

H7: Women Are Less Likely Than Men to Attribute Job or Wage Disparities to Discrimination

H7 posits that women are less likely than men to attribute their assignment to Job A (particularly under job discrimination) or their lower pay (especially under pay discrimination) to discrimination. However, this hypothesis is not supported by the data. Specifically, there is no observed gender difference in perceived fairness of own job allocation in the Discriminated Group (DG), and this remains true even when examining job discrimination and pay discrimination separately[15]. Contrary to expectations, women in the DG report lower satisfaction with their wages compared to men. Additionally, there is no significant gender difference in the perceived fairness of other job allocations or in reported satisfaction with other participants' wages within the DG. These results are illustrated in Figure 11.

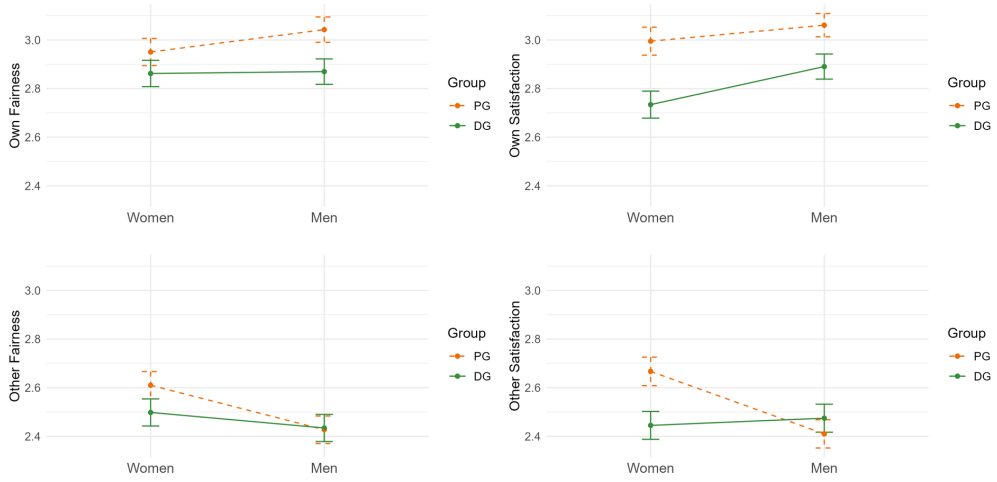


Figure 11: Comparison of fairness, satisfactions between Genders by group

H8: Discrimination Increases Negative Emotions, Especially for Women

H8 is verified in the sense that the discriminated group experienced more negative emotions concerning wage/job allocation, as indicated by higher "Negative Feeling" scores. Additionally, there was no significant difference in perceived closeness to own gender. Participants in the discriminated group are also less likely to perceive their employer as being of the same gender as themselves. See Table 14.

[15]Robustness checks by types of discrimination can be provided upon request.

Table 14: Results of Tests for H8: Comparison of negative feelings, Closeness and Gender Employer between Discriminated Group (DG) and Privileged Group (PG)

	PG	**DG**	p-value
Negative Feelings	3.10 (1.15)	3.55 (1.18)	<0.0001
Closeness	3.12 (0.71)	3.12 (0.71)	0.9368
Gender Employer			
Opposite	571 (35.03%)	803 (48.70%)	<0.0001
Same	1059 (64.97%)	846 (51.30%)	

Note: Statistical tests were performed as follows: Pearson's Chi-squared test for Gender Employer and Wilcoxon rank sum tests for Negative Feeling, and Closeness. In the table, the means and standard deviations (SD) are presented for Negative Feeling, and Closeness. The number and percentage of responses for Opposite Gender Employer are also presented.

As seen in Figure 12, the second part of H8 is verified as well. Women report higher levels of negative feelings when discriminated against than men do.

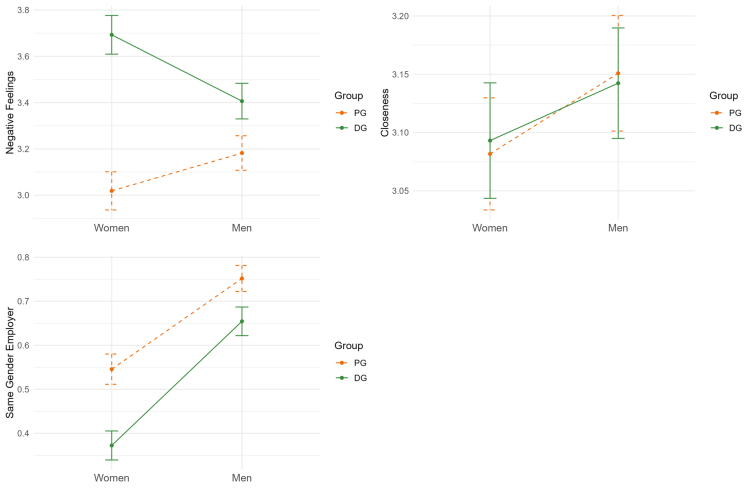


Figure 12: Comparison of negative feelings, Closeness and Gender Employer between Genders

4 Conclusions

Summary of results

Our research findings offer critical insights into the impact of discriminations (in a context of pay transparency) on individual behaviour on the workplace, intertwining three dimensions: differences between those affected by discrimination and those who are not, different types of discrimination and gender differences in reaction to discrimination. We found that the likelihood to complain is higher under job discrimination than pay discrimination. The Discriminated Group (DG) is more likely to complain about discrimination than the Privileged Group (PG) when there is pay discrimination. However, both DG and PG are equally likely to complain about their job assignment under job discrimination if they did not get the higher-paid job. This suggests that people may complain for the wrong reasons, such as receiving a lower salary or worse job and taking a chance to improve it, rather than recognizing unfair pay or job allocation based on performance.

Neither type of discrimination resulted in lower effort by the DG. The DG reported lower perceived fairness of job allocation and lower satisfaction with wages, but this did not affect their overall motivation to perform the task. Satisfaction with own wage allocation was lower under pay discrimination, but there was no difference in perceived fairness of job allocation between pay and job discrimination.

We did not find gender differences in complaint rates among the DG. However, we observed that women exerted more effort under discrimination, spending more time on tasks and completing more strings than men. Women in the DG were less satisfied with their wage allocations compared to men and reported more negative feelings than their male counterparts. In general, participants in the DG reported more negative feelings than those in the PG.

Discussion

In terms of policy relevance, we find that pay transparency works well for both pay discrimination and job discrimination, in the sense that participants who are discriminated against complain more than those who are not, and there are no negative effects in terms of effort, performance, or time spent on the task. This aligns with the aims of the Pay Transparency Directive[16], which requires the provision of information on average pay levels, broken down by sex, for categories of employees doing the same work or work of equal value, as well as objective and gender-neutral criteria for determining pay and pay progression. Expanding this transparency to include details about job allocations, such as the percentage of men and women at different levels of the hierarchy, could help workers recognise not only pay discrimination but also job discrimination. This fits especially well with the directive, as pay discrimination can be viewed as unequal pay for the same job, while job discrimination represents unequal job assignments for work of equal value. This is particularly relevant in a labour market that still presents glass ceilings and gender gaps in promotions. To some extent, this is a direct step further from making promotion rules transparent to ensure they are gender neutral. Job discrimination might result from non-gender-neutral promotion rules or the failure to apply gender-neutral promotion rules properly. Hence, providing figures on job allocations might be useful to see the results and effectiveness of those rules. It is important that the information provided is comprehensible and that workers are informed about what they are looking at and potentially guided through the information provided. More detailed data can be, however, more costly for firms, equality bodies, organizations, etc.

[16](Directive (EU) 2023/970 of the European Parliament and of the Council of 10 May 2023 to Strengthen the Application of the Principle of Equal Pay for Equal Work or Work of Equal Value between Men and Women through Pay Transparency and Enforcement Mechanisms (Text with EEA Relevance), 2023).

In real life, we might expect the role of gender and the effect of gender on complaint rates to be different and more relevant than those observed here in this experiment. Complaints are "easier" to make in an experimental setup like ours, where the influence of gender roles and the observation of participants' behaviour are minimized. Gender roles are inherently social roles, so they are more likely to emerge in a social setting where mutual observation of behaviour occurs. Public observability, in fact, has been shown to affect behaviour due to concerns about reputation, social image, and fear of backlash [Andreoni and Bernheim, 2009, Eagly, 1987, Rege and Telle, 2004]. For instance, single women are less likely to signal ambition or leadership when they believe their responses will be observable by peers [Bursztyn et al., 2017]. Additionally, women tend to report lower self-assessments (ranking their performance lower) when both their performance and self-assessment are observable, a pattern not seen in men [Ludwig et al., 2017]. Public observability also reduces women's likelihood of initiating negotiations, but not men's, though the effect is small and the statistical significance is weak [Ren et al., 2022][17]. It follows that in an alternative experiment featuring a more public and threatening environment for lodging complaints, the results might differ and align more closely with our hypotheses, with women being less likely than men to complain, women exerting more effort, and so on.

Another issue in terms of experimental validity relates to observability. Although our laboratory design lets us observe individual performance (in a controlled environment where a form of discrimination is deliberately included in every treatment), real-world output is a mix of ability, effort and luck and any discrimination that does occur is usually subtler and much harder to document. Consequently, both the act of complaining and the proof required to judge a complaint's legitimacy are far less transparent in practice, posing thorny feasibility and legality questions for any enforcement mechanism. This is particularly relevant as we found that non-discriminated participants (those in the privileged group) complain a lot, especially in cases of job discrimination. On one hand, this could be an expression of "allyship," intended as actions taken in support of the discriminated group by people in a privileged position. However, this does not seem to be the case, as the reasons indicated for complaining by the privileged group in cases of job discrimination are mainly merit and taking a chance, rather than concern for the discrimination of others (and thus, allyship)[18]. This suggests that participants do not realize they are privileged, possibly due to wrong inference from the data presented on the wage/job structure in the firm. The frequent complaints from non-discriminated individuals present a significant challenge, as they can burden both the administrative processes and practical resources. The Pay Transparency Directive[16] encourages the involvement of equality bodies, organizations, associations, and workers' representatives to help workers who wish, as stated in the Directive, "to seek to exercise their right to equal pay." However, evaluating whether complaints are grounded is not necessarily easy for workers' representatives or equality bodies. It follows that one of the main concerns should be how to deal with workers potentially misusing the complaint channel, with clear policies that make it possible to reject complaints from non-discriminated individuals. This involves careful documentation and application of criteria for promotion or assignment to better paid positions in firms, and clear communication of pay and job information to help workers interpret them.

[17] It is worth noting, however, that Buser et al. [2021] found no evidence that public observability affects gender differences in the willingness to compete.

[18] Note also that, by design, a successful complaint adjusts only the complainer's own wage or job; the broader allocation in the firm remains unchanged, a feature that may further encourage self-interested rather than solidaristic complaints.

Conclusions

The main takeaways from our study are that, under pay transparency, discriminated employees are more likely to complain compared to non-discriminated ones, without this negatively impacting their effort, performance, or time committed to tasks. This suggests that by providing information on average pay and job allocations across genders, employees are better positioned to identify and challenge discrimination. However, it's crucial to carefully manage how this information is presented to prevent baseless complaints from non-discriminated workers. Our findings indicate that these individuals often fail to recognize their privileged position and mistakenly perceive themselves as discriminated against. Such undue complaints pose a significant bureaucratic challenge as they require resources to verify their validity.

Interestingly, we did not observe gender differences in the rates of complaints. Yet, women under discrimination exerted more effort and reported greater dissatisfaction with their wages and more negative emotions compared to their male counterparts. Since gender roles typically become more pronounced in settings where behavior is publicly observable, further research might explore if gender differences in complaint behavior emerge under pay transparency when complaints must be lodged in a more public and potentially punitive setting.

In essence, pay transparency can serve as a powerful tool against both pay and job discrimination, reinforcing the principle that everyone should receive equal pay for equal work or work of equal value.

Nonetheless, it also underscores the importance of designing and implementing transparency measures thoughtfully, especially in terms of information provision, to ensure they enhance workplace fairness effectively without leading to unintended consequences.

Acknowledgments

We would like to thank Ginevra Marandola and Hannah Nohlen for their early involvement in the project. We also extend our gratitude to Yolanda Gómez and the DevStat team for their contributions.

References

George A. Akerlof and Janet L. Yellen. The Fair Wage-Effort Hypothesis and Unemployment*. *The Quarterly Journal of Economics*, 105(2):255–283, May 1990. ISSN 0033-5533. doi: 10.2307/2937787. URL https://doi.org/10.2307/2937787.

Emily T. Amanatullah and Michael W. Morris. Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of Personality and Social Psychology*, 98(2):256–267, 2010. ISSN 1939-1315. doi: 10.1037/a0017094.

Emily T. Amanatullah and Catherine H. Tinsley. Punishing female negotiators for asserting too much... or not enough: Exploring why advocacy moderates backlash against assertive female negotiators. *Organizational Behavior and Human Decision Processes*, 120(1):110–122, January 2013. ISSN 0749-5978. doi: 10.1016/j.obhdp.2012.03.006. URL https://www.sciencedirect.com/science/article/pii/S0749597812000416.

James Andreoni and B. Douglas Bernheim. Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, 77(5):1607–1636, 2009. ISSN 1468-0262. doi: 10.3982/ECTA7384. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA7384.

LINDA Babcock and SARA Leschevar. *Women Don't Ask*. Princeton University Press, stu - student edition edition, 2003. ISBN 978-0-691-08940-9. URL http://www.jstor.org/stable/j.ctt7rh37.

Marianna Baggio and Ginevra Marandola. Employees' reaction to gender pay transparency: an online experiment. *Economic Policy*, 38(113):161–188, January 2023. ISSN 0266-4658. doi: 10.1093/epolic/eiac066. URL https://doi.org/10.1093/epolic/eiac066.

Michael Baker, Yosh Halberstam, Kory Kroft, Alexandre Mas, and Derek Messacar. Pay Transparency and the Gender Gap. *American Economic Journal: Applied Economics*, 15(2):157–183, April 2023. ISSN 1945-7782, 1945-7790. doi: 10.1257/app.20210141. URL https://pubs.aeaweb.org/doi/10.1257/app.20210141.

Morten Bennedsen, Elena Simintzi, Margarita Tsoutsoura, and Daniel Wolfenzon. Do Firms Respond to Gender Pay Gap Transparency? *The Journal of Finance*, 77(4):2051–2091, 2022. ISSN 1540-6261. doi: 10.1111/jofi.13136. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.13136.

Morten Bennedsen, Birthe Larsen, and Jiayi Wei. Gender wage transparency and the gender pay gap: A survey. *Journal of Economic Surveys*, 37(5):1743–1777, 2023. ISSN 1467-6419. doi: 10.1111/joes.12545. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/joes.12545.

Marianne Bertrand. Coase Lecture – The Glass Ceiling. *Economica*, 85(338):205–231, 2018. ISSN 0013-0427. URL https://www.jstor.org/stable/26746388.

Barbara Biasi and Heather Sarsons. Flexible Wages, Bargaining, and the Gender Gap*. *The Quarterly Journal of Economics*, 137(1):215–266, February 2022. ISSN 0033-5533. doi: 10.1093/qje/qjab026. URL https://doi.org/10.1093/qje/qjab026.

Francine D. Blau and Jed Devaro. New Evidence on Gender Differences in Promotion Rates: An Empirical Analysis of a Sample of New Hires. *Industrial Relations: A Journal of Economy and Society*, 46(3): 511–550, 2007. ISSN 1468-232X. doi: 10.1111/j.1468-232X.2007.00479.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-232X.2007.00479.x.

Francine D. Blau and Lawrence M. Kahn. The U.S. Gender Pay Gap in the 1990S: Slowing Convergence. *ILR Review*, 60(1):45–66, October 2006. ISSN 0019-7939. doi: 10.1177/001979390606000103. URL https://doi.org/10.1177/001979390606000103.

Francine D. Blau and Lawrence M. Kahn. The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3):789–865, September 2017. ISSN 0022-0515. doi: 10.1257/jel.20160995. URL https://www.aeaweb.org/articles?id=10.1257%2Fjel.20160995&trk=article-ssr-frontend-pulse_x-social-details_comments-action_comment-text.

Clément Bosquet, Pierre-Philippe Combes, and Cecilia García-Peñalosa. Gender and Promotions: Evidence from Academic Economists in France. *The Scandinavian Journal of Economics*, 121(3):1020–1053, 2019. ISSN 1467-9442. doi: 10.1111/sjoe.12300. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/sjoe.12300.

Hannah Riley Bowles, Linda Babcock, and Lei Lai. Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, 103(1):84–103, May 2007. ISSN 0749-5978. doi: 10.1016/j.obhdp.2006.09.001. URL https://www.sciencedirect.com/science/article/pii/S0749597806000884.

Victoria L. Brescoll. Who Takes the Floor and Why: Gender, Power, and Volubility in Organizations. *Administrative Science Quarterly*, 56(4):622–641, December 2011. ISSN 0001-8392. doi: 10.1177/0001839212439994. URL https://doi.org/10.1177/0001839212439994.

Emily Breza, Supreet Kaur, and Yogita Shamdasani. The Morale Effects of Pay Inequality*. *The Quarterly Journal of Economics*, 133(2):611–663, May 2018. ISSN 0033-5533. doi: 10.1093/qje/qjx041. URL https://doi.org/10.1093/qje/qjx041.

Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais. 'Acting Wife': Marriage Market Incentives and Labor Market Investments. *American Economic Review*, 107(11):3288–3319, November 2017. ISSN 0002-8282. doi: 10.1257/aer.20170029. URL https://www.aeaweb.org/articles?id=10.1257/aer.20170029.

Thomas Buser, Eva Ranehill, and Roel van Veldhuizen. Gender differences in willingness to compete: The role of public observability. *Journal of Economic Psychology*, 83:102366, March 2021. ISSN 0167-4870. doi: 10.1016/j.joep.2021.102366. URL https://www.sciencedirect.com/science/article/pii/S0167487021000064.

René Böheim and Sarah Gust. The Austrian Pay Transparency Law and the Gender Wage Gap, 2021. URL https://papers.ssrn.com/abstract=3812077.

Gary Charness, Ramón Cobo-Reyes, Juan A. Lacomba, Francisco Lagos, and Jose Maria Perez. Social comparisons in wage delegation: experimental evidence. *Experimental Economics*, 19(2):433–459, June 2016. ISSN 1573-6938. doi: 10.1007/s10683-015-9448-x. URL https://doi.org/10.1007/s10683-015-9448-x.

Andrew E. Clark, David Masclet, and Marie Claire Villeval. Effort and Comparison Income: Experimental and Survey Evidence. *ILR Review*, 63(3):407–426, April 2010. ISSN 0019-7939. doi: 10.1177/001979391006300303. URL https://doi.org/10.1177/001979391006300303.

Katherine Baldiga Coffman. Evidence on Self-Stereotyping and the Contribution of Ideas*. *The Quarterly Journal of Economics*, 129(4):1625–1660, November 2014. ISSN 0033-5533, 1531-4650. doi: 10.1093/qje/qju023. URL https://academic.oup.com/qje/article/129/4/1625/1853974.

Rachel Croson and Uri Gneezy. Gender Differences in Preferences. *Journal of Economic Literature*, 47(2): 1–27, 2009. doi: 10.1257/jel.47.2.1.

Zoë Cullen. Is Pay Transparency Good? *Journal of Economic Perspectives*, 38(1):153–180, February 2024. ISSN 0895-3309. doi: 10.1257/jep.38.1.153. URL https://www.aeaweb.org/articles?id=10.1257%2Fjep.38.1.153.

Zoë Cullen and Ricardo Perez-Truglia. How Much Does Your Boss Make? The Effects of Salary Comparisons. *Journal of Political Economy*, 130(3):766–822, March 2022. ISSN 0022-3808. doi: 10.1086/717891. URL https://www.journals.uchicago.edu/doi/full/10.1086/717891.

Sirus H. Dehdari, Emma Heikensten, and Siri Isaksson. What Goes Around (Sometimes) Comes Around: Gender Differences in Retaliation, April 2019. URL https://papers.ssrn.com/abstract=3378279.

Arindrajit Dube, Laura Giuliano, and Jonathan Leonard. Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior. *American Economic Review*, 109(2):620–663, February 2019. ISSN 0002-8282. doi: 10.1257/aer.20160232. URL https://www.aeaweb.org/articles?id=10.1257/aer.20160232.

Alice H. Eagly. *Sex Differences in Social Behavior: A Social-role interpretation*. New York, April 1987. ISBN 978-0-203-78190-6. doi: 10.4324/9780203781906.

EIGE. Gender Composition of CEOs, Executives, and Non-Executives in the Largest Listed Companies, 2024. URL https://dgs-p.eige.europa.eu/data/view?code=wmidm_bus_bus__wmid_comp_compex&orderby=2013-B1&orderdir=dsc.

Deborah Thompson Eisenberg. Money, Sex, and Sunshine: A Market-Based Approach to Pay Discrimination. *Arizona State Law Journal*, 43(3):951–1020, 2011. URL https://heinonline.org/HOL/P?h=hein.journals/arzjl43&i=959.

Paula England, Andrew Levine, and Emma Mishel. Progress toward gender equality in the United States has slowed or stalled. *Proceedings of the National Academy of Sciences*, 117(13):6990–6997, March 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1918891117. URL https://pnas.org/doi/full/10.1073/pnas.1918891117.

Eurofound. *Understanding the gender pay gap – What role do sector and occupation play? – Employment and labour markets*. Publications Office of the European Union, 2021. doi: 10.2806/939946.

Eurostat. Gender pay gap in unadjusted form. Dataset, 2024. URL https://ec.europa.eu/eurostat/databrowser/view/sdg_05_20/default/table?lang=en&category=t_labour.t_earn.

Christine L Exley and Judd B Kessler. The Gender Gap in Self-Promotion. *The Quarterly Journal of Economics*, 137(3):1345–1381, July 2022. ISSN 0033-5533, 1531-4650. doi: 10.1093/qje/qjac003. URL https://academic.oup.com/qje/article/137/3/1345/6513425.

Christine L. Exley and Kirby Nielsen. The Gender Gap in Confidence: Expected but Not Accounted For. *American Economic Review*, 114(3):851–885, March 2024. ISSN 0002-8282. doi: 10.1257/aer.20221413. URL https://www.aeaweb.org/articles?id=10.1257/aer.20221413.

Antonio Filippin and Paolo Crosetto. A Reconsideration of Gender Differences in Risk Attitudes. *Management Science*, 62(11):3138–3160, 2016. ISSN 0025-1909. URL https://www.jstor.org/stable/44012237.

Nickolas Gagnon, Kristof Bosmans, and Arno Riedl. The Effect of Unfair Chances and Gender Discrimination on Labor Supply, 2020. URL https://papers.ssrn.com/abstract=3529014.

Claudia Goldin. A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4): 1091–1119, 2014. doi: 10.1257/aer.104.4.1091.

Robert Grasser, Andrew H. Newman, and Grazia Xiong. The Effect of Horizontal Pay Transparency on Employee Motivation When Pay Dispersion Is Performance Based and Non-Performance Based, November 2023. URL https://papers.ssrn.com/abstract=3802785.

Stefan Grimm. Effects of choice observability on risk taking: The role of norms. *Journal of Behavioral and Experimental Economics*, 80:34–46, June 2019. ISSN 2214-8043. doi: 10.1016/j.socec.2019.03.003. URL https://www.sciencedirect.com/science/article/pii/S2214804318304749.

Andreas Gulyas, Sebastian Seitz, and Sourav Sinha. Does Pay Transparency Affect the Gender Wage Gap? Evidence from Austria. *American Economic Journal: Economic Policy*, 15(2):236–255, May 2023. ISSN 1945-7731, 1945-774X. doi: 10.1257/pol.20210128. URL https://pubs.aeaweb.org/doi/10.1257/pol.20210128.

Simon Gächter and Christian Thöni. Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis. *Journal of Economic Behavior & Organization*, 76(3):531–543, December 2010. ISSN 0167-2681. doi: 10.1016/j.jebo.2010.08.008. URL https://www.sciencedirect.com/science/article/pii/S0167268110001538.

Ingrid Haegele. The Broken Rung: Gender and the Leadership Gap, April 2024. URL http://arxiv.org/abs/2404.07750.

Madeline E. Heilman. Gender stereotypes and workplace bias. *Research in Organizational Behavior*, 32:113–135, January 2012. ISSN 0191-3085. doi: 10.1016/j.riob.2012.11.003. URL https://www.sciencedirect.com/science/article/pii/S0191308512000093.

Madeline E. Heilman, Suzette Caleo, and Francesca Manzi. Women at Work: Pathways from Gender Stereotypes to Gender Bias and Discrimination. *Annual Review of Organizational Psychology and Organizational Behavior*, 11(Volume 11, 2024):165–192, January 2024. ISSN 2327-0608, 2327-0616. doi:

10.1146/annurev-orgpsych-110721-034105. URL https://www.annualreviews.org/content/journals/10.1146/annurev-orgpsych-110721-034105.

Iñigo Hernandez-Arenaz and Nagore Iriberri. A Review of Gender Differences in Negotiation. In *Oxford Research Encyclopedia of Economics and Finance*. December 2019. ISBN 978-0-19-062597-9. doi: 10.1093/acrefore/9780190625979.013.464. URL https://oxfordre.com/economics/display/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-464.

John Hurley, Sara Grubanov-Boskovic, Martina Bisello, Carlos Vacas-Soriano, Marta Fana, and Enrique Fernández-Macías. European Jobs Monitor 2021: Gender gaps and the employment structure. 2021.

Alice M. Isen and Johnmarshall Reeve. The Influence of Positive Affect on Intrinsic and Extrinsic Motivation: Facilitating Enjoyment of Play, Responsible Work Behavior, and Self-Control. *Motivation and Emotion*, 29(4):295–323, December 2005. ISSN 0146-7239, 1573-6644. doi: 10.1007/s11031-006-9019-8. URL http://link.springer.com/10.1007/s11031-006-9019-8.

Linda K. Kaye and Charlotte R. Pennington. "Girls can't play": The effects of stereotype threat on females' gaming performance. *Computers in Human Behavior*, 59:202–209, June 2016. ISSN 0747-5632. doi: 10.1016/j.chb.2016.02.020. URL https://www.sciencedirect.com/science/article/pii/S0747563216300723.

Andreas Leibbrandt and John A. List. Do Women Avoid Salary Negotiations? Evidence from a Large-Scale Natural Field Experiment. *Management Science*, 61(9):2016–2024, September 2015. ISSN 0025-1909. doi: 10.1287/mnsc.2014.1994. URL https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2014.1994.

Sandra Ludwig, Gerlinde Fellner-Röhling, and Carmen Thoma. Do women have more shame than men? An experiment on self-assessment and the shame of overestimating oneself. *European Economic Review*, 92:31–46, February 2017. ISSN 0014-2921. doi: 10.1016/j.euroecorev.2016.11.007. URL https://www.sciencedirect.com/science/article/pii/S0014292116302197.

Corinne A. Moss-Racusin and Laurie A. Rudman. Disruptions in Women's Self-Promotion: The Backlash Avoidance Model. *Psychology of Women Quarterly*, 34(2):186–202, June 2010. ISSN 0361-6843. doi: 10.1111/j.1471-6402.2010.01561.x. URL https://doi.org/10.1111/j.1471-6402.2010.01561.x.

Raviv Murciano-Goroff. Missing Women in Tech: The Labor Market for Highly Skilled Software Engineers. *Management Science*, 68(5):3262–3281, May 2022. ISSN 0025-1909. doi: 10.1287/mnsc.2021.4077. URL https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2021.4077.

Markus M. Möbius, Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. Managing Self-Confidence: Theory and Experimental Evidence. *Management Science*, 68(11):7793–7817, November 2022. ISSN 0025-1909. doi: 10.1287/mnsc.2021.4294. URL https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2021.4294.

Hannah-Hanh D. Nguyen and Ann Marie Ryan. Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6):1314–1334, 2008. ISSN 1939-1854. doi: 10.1037/a0012702.

Muriel Niederle and Lise Vesterlund. Do Women Shy Away From Competition? Do Men Compete Too Much?*. *The Quarterly Journal of Economics*, 122(3):1067–1101, August 2007. ISSN 0033-5533. doi: 10.1162/qjec.122.3.1067. URL https://doi.org/10.1162/qjec.122.3.1067.

Muriel Niederle and Lise Vesterlund. Gender and Competition. *Annual Review of Economics*, 3:601–630, 2011. URL http://hdl.handle.net/10986/9120.

Tomasz Obloj and Todd Zenger. The influence of pay transparency on (gender) inequity, inequality and the performance basis of pay. *Nature Human Behaviour*, 6(5):646–655, May 2022. ISSN 2397-3374. doi: 10.1038/s41562-022-01288-9. URL https://www.nature.com/articles/s41562-022-01288-9.

Tyler G. Okimoto and Victoria L. Brescoll. The Price of Power: Power Seeking and Backlash Against Female Politicians. *Personality and Social Psychology Bulletin*, June 2010. doi: 10.1177/0146167210371949. URL https://journals.sagepub.com/doi/abs/10.1177/0146167210371949.

Claudia Olivetti and Barbara Petrongolo. The evolution of gender gaps in industrialized countries. *Annual review of Economics*, 8(1):405–434, 2016.

Michael Ransom and Ronald L. Oaxaca. Intrafirm Mobility and Sex Differences in Pay. *ILR Review*, 58(2): 219–237, January 2005. ISSN 0019-7939. doi: 10.1177/001979390505800203. URL https://doi.org/10.1177/001979390505800203.

Maria Recalde and Lise Vesterlund. Gender Differences in Negotiation and Policy for Improvement. Technical Report w28183, National Bureau of Economic Research, Cambridge, MA, December 2020. URL http://www.nber.org/papers/w28183.pdf.

María P. Recalde and Lise Vesterlund. Gender Differences in Negotiation: Can Interventions Reduce the Gap? *Annual Review of Economics*, 15(1):633–657, September 2023. ISSN 1941-1383, 1941-1391. doi: 10.1146/annurev-economics-092022-115353. URL https://www.annualreviews.org/doi/10.1146/annurev-economics-092022-115353.

Paul Redmond and Seamus Mcguinness. The Gender Wage Gap in Europe: Job Preferences, Gender Convergence and Distributional Effects. *Oxford Bulletin of Economics and Statistics*, 81(3):564–587, 2019. ISSN 1468-0084. doi: 10.1111/obes.12282. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/obes.12282.

Mari Rege and Kjetil Telle. The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics*, 88(7):1625–1644, July 2004. ISSN 0047-2727. doi: 10.1016/S0047-2727(03)00021-5. URL https://www.sciencedirect.com/science/article/pii/S0047272703000215.

Yufei Ren, Lin Xiu, and Amy B. Hietapelto. Dare to ask in front of others? Women initiating salary negotiations. *Journal of Economic Psychology*, 92:102550, October 2022. ISSN 0167-4870. doi: 10.1016/j.joep.2022.102550. URL https://www.sciencedirect.com/science/article/pii/S0167487022000629.

Nina Roussille. The Role of the Ask Gap in Gender Pay Inequality. *The Quarterly Journal of Economics*, 139(3):1557–1610, August 2024. ISSN 0033-5533. doi: 10.1093/qje/qjae004. URL https://doi.org/10.1093/qje/qjae004.

Laurie A Rudman. Self-promotion as a risk factor for women: the costs and benefits of counterstereotypical impression management. *Journal of personality and social psychology*, 74(3):629, 1998. ISSN 1939-1315.

Steven J. Spencer, Christine Logel, and Paul G. Davies. Stereotype Threat. *Annual Review of Psychology*, 67(1):415–437, January 2016. ISSN 0066-4308, 1545-2085. doi: 10.1146/annurev-psych-073115-103235. URL https://www.annualreviews.org/doi/10.1146/annurev-psych-073115-103235.

Claude M. Steele and Joshua Aronson. Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5):797–811, 1995. ISSN 1939-1315. doi: 10.1037/0022-3514.69.5.797.

Jenny Säve-Söderbergh. Gender gaps in salary negotiations: Salary requests and starting salaries in the field. *Journal of Economic Behavior & Organization*, 161:35–51, May 2019. ISSN 0167-2681. doi: 10.1016/j.jebo.2019.01.019. URL https://www.sciencedirect.com/science/article/pii/S0167268119300290.

Richard G. Trotter, Susan Rawson Zacur, and Lisa T. Stickney. The new age of pay transparency. *Business Horizons*, 60(4):529–539, July 2017. ISSN 0007-6813. doi: 10.1016/j.bushor.2017.03.011. URL https://www.sciencedirect.com/science/article/pii/S0007681317300356.

Lotte Vermeulen, Elena Núñez Castellar, Dirk Janssen, Licia Calvi, and Jan Van Looy. Playing under threat. Examining stereotype threat in female game players. *Computers in Human Behavior*, 57:377–387, April 2016. ISSN 0747-5632. doi: 10.1016/j.chb.2015.12.042. URL https://www.sciencedirect.com/science/article/pii/S0747563215303137.

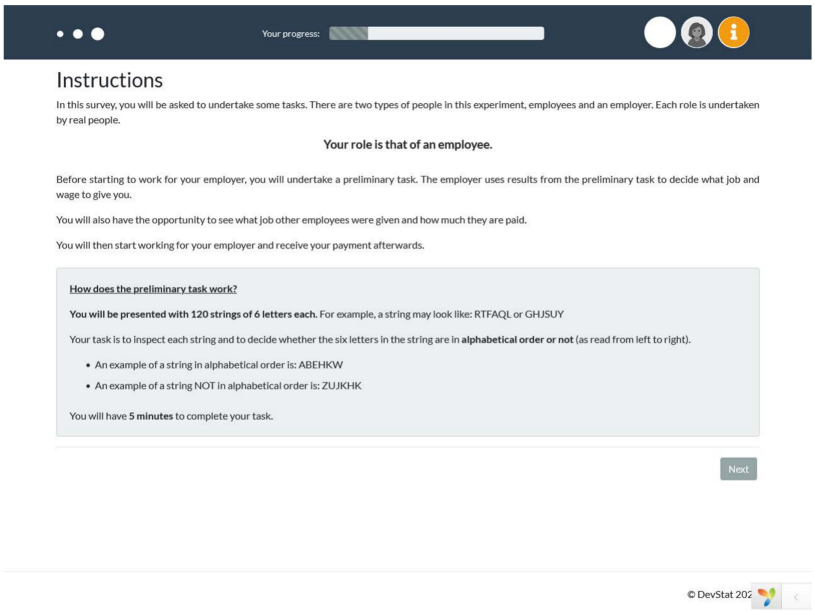
List of abbreviations and definitions

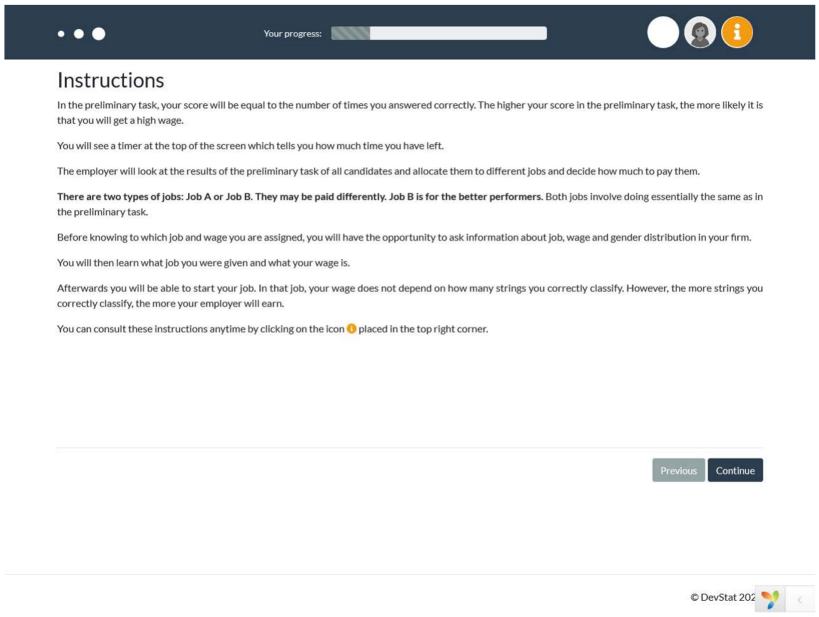
Abbreviations	Definitions
JOB A	Basic job, normally entails lower pay.
JOB B	Job for those who perform better, normally entails higher pay.
PRIVILEGED GROUP (PG)	A group that has higher chances of getting a better paid job or higher salary.
DISCRIMINATED GROUP (DG)	A group that has lower chances of getting a better paid job or higher salary.
SKILLS LEVEL	LOW SKILL (LS)– Bottom 60% MEDIUM SKILL (MS) – Between 60 and 90% HIGH SKILL (HS) – Above 90%
DISCRIMINATED INDIVIDUAL (DI)	Participants within the DG who, given their skill levels, would have received a different wage or job position if they had belonged to the PG.
PRIVILEGED INDIVIDUAL (PI)	Participants in the PG who, given their skill levels, would have received a different wage or job position if they had belonged to the DG.
CHANGE	Rate of complaints.
MAIN CORRECT	Measure of effort, number of strings that participants correctly classified in the main task.
EFFORT RATIO	Number of strings correctly assigned in the main task relative to the number of strings correctly assigned in the preliminary task.
PERCENTAGE CORRECT	Proportion of correctly classified strings out of the total number of strings answered in the main task.
REASON CHANGE	Reasons why participants complained or requested (or did not request) a change.
OWN FAIRNESS	Perceptions of fairness regarding own job allocation
OWNS SATISFACTION	Satisfaction with own wage allocation
OTHER FAIRNESS	Perceived fairness of others' job allocations
OTHER SATISFACTION	Satisfaction with others' wage allocations
MOTIVATION	Participants' levels of motivation during the task
NEGATIVE FEELINGS	Composite measure of negative emotions experienced by participants
CLOSENESS	How close participants feel towards other employees of their gender.

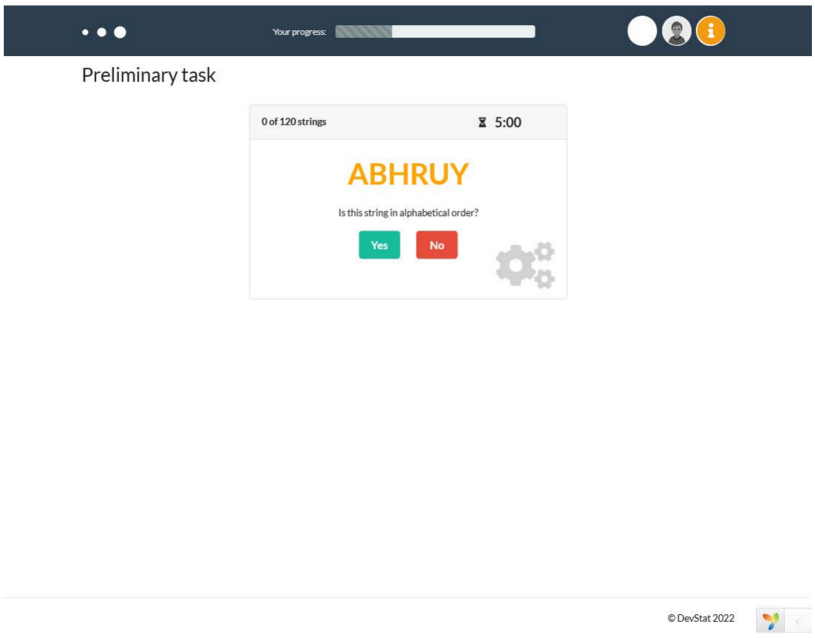
Appendix A

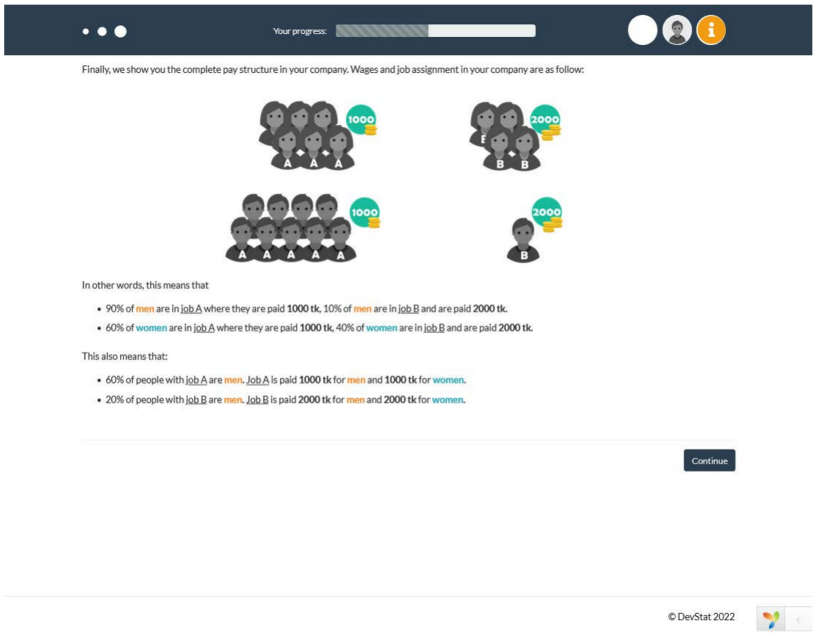
Experiment Process Flow with Screenshots

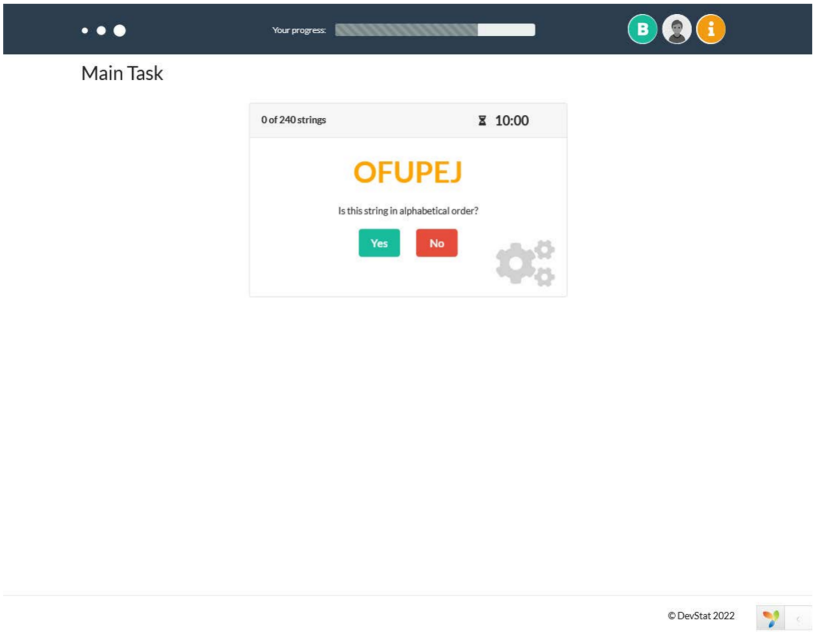
In Appendix A, we reproduce a representative subset of screenshots that trace the chronological progression of a participant's experience in our experiment. These screenshots specifically depict the scenario of a male participant encountering job discrimination against men.

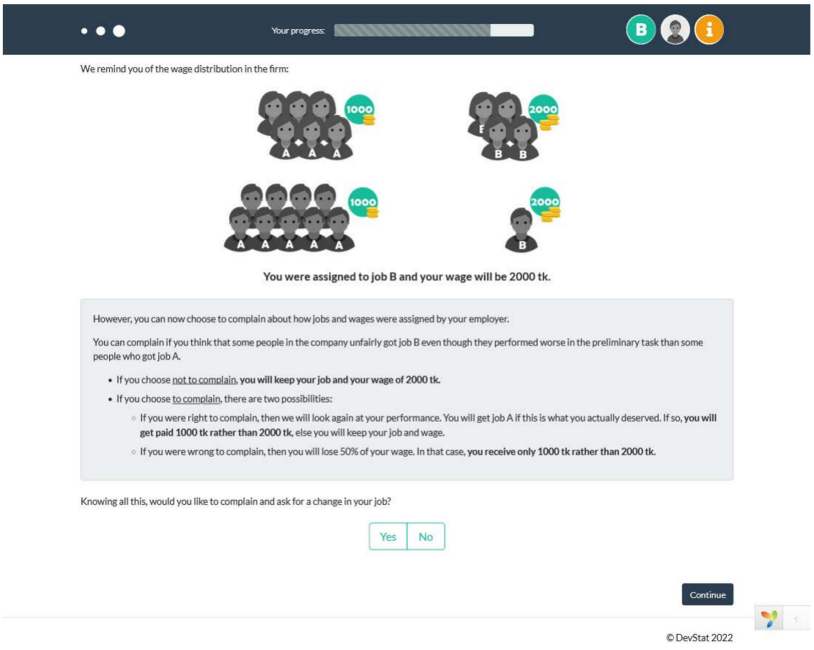


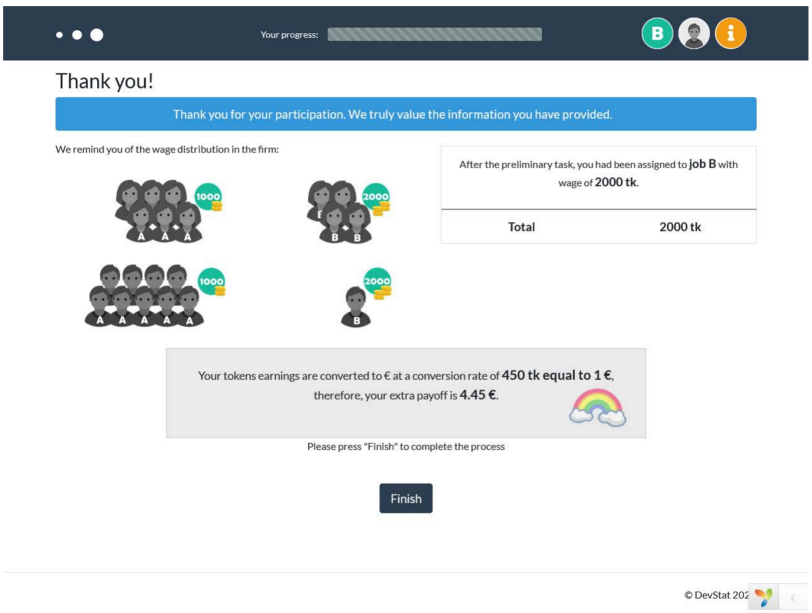



Your progress:
Preliminary task
0 of 120 strings
5:00
ABHRUY
Is this string in alphabetical order?
Yes
No
© DevStat 2022


Your progress:
Finally, we show you the complete pay structure in your company. Wages and job assignment in your company are as follow:
1000
2000
1000
2000
In other words, this means that
90% of men are in job A where they are paid 1000 tk. 10% of men are in job B and are paid 2000 tk.
60% of women are in job A where they are paid 1000 tk. 40% of women are in job B and are paid 2000 tk.
This also means that:
60% of people with job A are men. Job A is paid 1000 tk for men and 1000 tk for women.
20% of people with job B are men. Job B is paid 2000 tk for men and 2000 tk for women.
Continue
© DevStat 2022


Your progress:
Main Task
0 of 240 strings
10:00
OFUPEJ
Is this string in alphabetical order?
Yes
No
© DevStat 2022


Your progress:
We remind you of the wage distribution in the firm:
1000
2000
1000
2000
You were assigned to job B and your wage will be 2000 tk.
However, you can now choose to complain about how jobs and wages were assigned by your employer.
You can complain if you think that some people in the company unfairly got job B even though they performed worse in the preliminary task than some people who got job A.
If you choose not to complain, you will keep your job and your wage of 2000 tk.
If you choose to complain, there are two possibilities:
If you were right to complain, then we will look again at your performance. You will get job A if this is what you actually deserved. If so, you will get paid 1000 tk rather than 2000 tk, else you will keep your job and wage.
If you were wrong to complain, then you will lose 50% of your wage. In that case, you receive only 1000 tk rather than 2000 tk.
Knowing all this, would you like to complain and ask for a change in your job?
Yes
No
Continue
© DevStat 2022



Appendix B

Codebook and transformed variables

Variable	Range of values	Description
Pre.Answered	Integer [0,120]	Number of strings answered in the preliminary task
Pre.Correct	Integer [0,120]	Number of correct strings in the preliminary task
Pre.Time	Integer	Time spends in the preliminary task
Pre.Last	Integer	Time of last answer in the preliminary task
Job.Assigned	Integer [1,2]	Job assigned to subject: 1: Job A - 2: Job B
Confidence1	Integer [0,120]	Out of the Pre.Answered strings you did, how many strings do you think you classified correctly?
Confidence2	Integer [1,4]	On average, how well do you think other candidates performed? 1: Better than me 2: Slightly better than me 3: Slightly worse than me 4: Worse than me
Confidence3	Integer [1,4]	On average, which gender do you think performed best in this task? 1: Your gender better than Opposite gender 2: Your gender slightly better than Opposite gender 3: Your gender slightly worse than Opposite gender 4: Your gender worse than Opposite gender
Choice	Integer [1,2]	Please select which information you wish to receive: 1: I want to see the average pay for discriminated gender and for not discriminated gender and how many discriminated gender and not discriminated gender there are in the firm. 2: I want to see the average pay for job A and job B and how many people got job A and job B in the firm.
Belief	Integer [1.2]	Please now tell us what job you think you have been assigned to: 1: Job A - 2: Job B

Own fairness	Integer [1,4]	Do you think you were fairly allocated to Job A/Job B? 1: Certainly no - 4: Certainly yes
Own Satisfaction	Integer [1,4]	Please rate your level of satisfaction with your wage 1: Very unsatisfied - 4: Very satisfied
Other fairness	Integer [1,4]	Do you think some people got job B even though they performed worse than you in the preliminary task? / Do you think some people got job A even though they performed better than you in the preliminary task? 1: Certainly no - 4: Certainly yes
Other satisfaction	Integer [1,4]	Do you think some people are paid more than you even though they performed worse than you in the preliminary task? / Do you think some people are paid less than you even though they performed better than you in the preliminary task? 1: Certainly no - 4: Certainly yes
emotion.1 to emotion.14	Integer [1,7]	When seeing the job you were assigned to as well as the wage, job, and gender distribution in your firm, to what extent did you experience these emotions? Anger, Sad, Mad, Downhearted, Irritated, Depressed, Happy, Frustrated, Cheerful, Overlooked/Passed over/Unrecognized, Satisfied, Anxious, Worried, Afraid 1: Not at all 2: Slightly 3: Somewhat 4: Moderately 5: Quite a bit 6: Very much 7: An extreme amount
Closeness	Integer [1,4]	How close do you feel towards other your gender employees 1: Very close 2: Somewhat close 3: Not very close 4: Not at all close
Main.Answered	Integer [0,240]	Number of strings answered for the main task
Main.Correct	Integer [0,240]	Number of correct strings for the main task
Main.Time	Integer	Time spends in the main task
Main.Last	Integer	Time of last answer in the main task
Change	Integer [1,2]	Knowing all this, would you like to complain? 1: Yes - 2: No
Successful	Integer [-1,0,1]	If the change was successful -1: Not applicable (no change decision) 0: No 1: Yes – and change of job/wage 2: Yes – but no change of job/wage
Fb.1	Varchar	What do you think this experiment was about?
Fb.2	Integer [0,1]	Was it hard for you to understand what you had to do in this experiment? 1: Yes - 2: No
Fb.3	Integer [1,2]	Do you think there was any part in this experiment where we lied to you? 1: Yes - 2: No
Motivation.4.1	Integer [1-4]	I wanted to challenge myself by doing my best 1: Disagree - 4: Agree
Motivation.4.2	Integer [1-4]	I wanted to make my employer happy by working hard 1: Disagree - 4: Agree
Motivation.4.3	Integer [1-4]	I enjoyed working on the task 1: Disagree - 4: Agree
Motivation.4.4	Integer [1-4]	I was tired of doing the task 1: Disagree - 4: Agree

Motivation.4.5	Integer [1-4]	I was bored of doing the task 1: Disagree - 4: Agree
Reason change	Integer [1-6]	Change = Yes What was the main reason for requesting a change of wage/job at the end of the experiment? Because I think I deserved a better wage/job given my performance Because I think my employer discriminated against other participants Because I was paid less than other your gender Because I was paid less than opposite gender Because I wanted to take the chance of increasing my wage Change = No What was the main reason for not requesting a change of wage/job at the end of the experiment? Because I think I did not deserve a better wage/job given my performance Because I think my employer did not discriminate against other participants Because I was paid more than other your gender Because I was paid more than opposite gender Because I did not want to take the risk of losing 50% of my wage
Gender Employer	Integer [1,2]	What gender do you think your employer is? 1: Women - 2: Men
Risk Taking	Integer [1-4]	On a scale from 1 to 4, do you generally want to take risks or do you try to avoid taking risks? 1: I try to avoid taking risks - 4: I want to take risks
Less confrontation	Integer [1-4]	On a scale from 1 to 4, how likely are you to confront people who behave badly towards you? 1: Very likely - 4: Very unlikely
Income Equality	Integer [1-4]	How would you place your views on this scale? 1: Incomes should be made more equal 4: There should be greater incentives for individual effort
Competition	Integer [1-4]	How would you place your views on this scale? 1: Competition is good - 4: Competition is harmful
Fq.4.1	Integer [1-4]	When a mother works for pay, the children suffer 1: Disagree - 4: Agree
prejudice.4.2	Integer [1-4]	On the whole, men make better political leaders than women do 1: Disagree - 4: Agree
prejudice.4.3	Integer [1-4]	A university education is more important for a boy than for a girl 1: Disagree - 4: Agree
prejudice.4.4	Integer [1-4]	On the whole, men make better business executives than women do 1: Disagree - 4: Agree
prejudice.4.5	Integer [1-4]	Being a housewife is just as fulfilling as working for pay 1: Disagree - 4: Agree
prejudice.4.6	Integer [1-4]	When jobs are scarce, men should have more right to a job than women. 1: Disagree - 4: Agree
prejudice.4.7	Integer [1-4]	If a woman earns more money than her husband, it's almost certain to cause problems. 1: Disagree - 4: Agree
Total.Profit	Integer	Subjects Total Profit in tokens
Time.Spent	Integer	Total time spent

Several variables in the dataset were transformed to derive new measures.

The Effort Ratio (Effort_ratio) was calculated by dividing the number of correct answers in the main task (Main.Correct) by the number of correct answers in the preliminary task (Pre.Correct), provided Pre.Correct is greater than zero. The Percentage Correct (Perc_correct) was derived by dividing Main.Correct by

Main.Answered.

For psychological measures, the Prejudice variable was calculated as the average of responses to several items (Prejudice.4.1 to Prejudice.4.7) measuring gender prejudice, where higher values indicate greater prejudice.

To measure feelings and motivation, composite variables were created. The Feeling variable is an aggregate measure of various emotions experienced by participants. It was computed as the average of several negative emotions (Emotion.1, Emotion.2, Emotion.3, Emotion.4, Emotion.5, Emotion.6, Emotion.8, Emotion.10, Emotion.12, Emotion.13, Emotion.14), and the reparameterized values of positive emotions (7 minus the values of Emotion.7, Emotion.9, Emotion.11). This transformation ensured that higher values consistently represented worse feelings. The Motivation variable captures participants' motivation levels and was calculated by averaging responses to three positive motivation items (Motivation.4.1, Motivation.4.2, Motivation.4.3) and two reparameterized negative motivation items (5 minus the values of Motivation.4.4, Motivation.4.5). As Motivation increases, it indicates that individuals are more motivated.

Appendix C

Psychological variables

Table C1: Descriptive statistics by country and gender: Psychological variables

	Germany		**Romania**		**Overall**	
	Female (N=806)	**Male (N=816)**	**Female (N=829)**	**Male (N=828)**	**Female (N=1635)**	**Male (N=1644)**
Risk taking						
Mean (SD)	2.57 (1.09)	2.75 (1.02)	2.79 (0.962)	2.88 (0.895)	2.68 (1.03)	2.81 (0.960)
Median [Min, Max]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]
Less Confrontation						
Mean (SD)	2.69 (1.01)	2.69 (0.949)	2.31 (1.11)	2.25 (1.08)	2.50 (1.08)	2.46 (1.04)
Median [Min, Max]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]	2.00 [1.00, 4.00]	2.00 [1.00, 4.00]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]
Income equality						
Mean (SD)	2.81 (1.01)	3.01 (0.916)	3.21 (0.983)	3.26 (0.890)	3.01 (1.02)	3.14 (0.912)
Median [Min, Max]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]	4.00 [1.00, 4.00]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]
Competition						
Mean (SD)	2.47 (0.961)	2.36 (0.994)	2.00 (1.06)	1.87 (1.02)	2.23 (1.04)	2.11 (1.03)
Median [Min, Max]	2.00 [1.00, 4.00]	2.00 [1.00, 4.00]	2.00 [1.00, 4.00]	2.00 [1.00, 4.00]	2.00 [1.00, 4.00]	2.00 [1.00, 4.00]
Prejudice						
Mean (SD)	2.25 (0.827)	2.45 (0.751)	1.87 (0.616)	2.07 (0.626)	2.06 (0.752)	2.25 (0.716)
Median [Min, Max]	2.14 [1.00, 4.00]	2.57 [1.00, 4.00]	1.86 [1.00, 4.00]	2.00 [1.00, 4.00]	1.86 [1.00, 4.00]	2.29 [1.00, 4.00]

Note: The variables presented in the table are measured on a scale from 1 to 4. For "Risk taking," a score of 1 indicates "I try to avoid taking risks" and a score of 4 indicates "I want to take risks." For "Less confron-tation," a score of 1 indicates "Very likely" to confront people who behave badly towards you, while a score of 4 indicates "Very unlikely." For "Income equality," a score of 1 signifies "Incomes should be made more equal," whereas a score of 4 signifies "There should be greater incentives for individual effort." For "Competition," a score of 1 indicates "Competition is good" and a score of 4 indicates "Competition is harmful." The "Prejudice" values are averages calculated based on a series of ques-tions measuring gender prejudice, with higher values indicating higher levels of prejudice towards women

Gender in Teaching: Insights from Five Million Syllabi on Collaboration, Interdisciplinarity, and Reading Selections

Andrea Blasco* Elisa Orlandi†

Abstract

Using the Open Syllabus corpus of roughly five million university syllabi (2000–2019) from more than 3800 institutions, we study how the gender composition of teaching teams is distributed and how it co-varies with course content. Mixed-gender co-teaching is rare (about 3 percent of courses, compared with 10 percent for same-gender teams) and this gap has not narrowed over time. A Monte-Carlo benchmark that reshuffles instructors within each institution–field–year, while preserving the team size of every course, shows that mixed-gender teams form only about half as often as expected under gender-neutral assignment. Team gender composition is also correlated with what appears on the syllabus. Courses taught by mixed-gender teams are, on average, more interdisciplinary than courses taught by solo male instructors, whereas differences among single-gender configurations of teaching teams are modest. All-female teaching teams are associated with a higher share of female-authored assigned readings in the course material than all-male teams, and female-led courses assign slightly newer literature than male-led courses, but this difference decreases over time. These associations remain after controlling for field, country and year.

Keywords: Education, Gender bias, Teams, interdisciplinarity, novelty, syllabi
JEL Codes: I21, I23, J16

*European Commission, Joint Research Centre, Brussels, Belgium; (andrea.blasco@ec.europa.eu)
†University of Bologna, Bologna, Italy, and European Commission, Joint Research Centre, Brussels, Belgium.

1 Introduction

Universities, vocational schools, and other higher-education institutions are crucial in disseminating knowledge and developing human capital. Instructors shape this role through their choices in teaching subjects, assigned readings, and co-teaching arrangements. These decisions have a significant impact on student learning outcomes and students' career trajectories (1–3). However, they are not always optimal. Instructors can prefer working alone or with colleagues from similar backgrounds, limiting course diversity and interdisciplinarity, reinforcing gender stereotypes on certain subjects (4–6), or even limiting instructors' exposure to different teaching methodologies, hindering their professional growth (7). Despite its significance, research on the underlying factors driving instructors' choices in teaching remains limited primarily due to a lack of data.

While prior research has examined academics' decision-making in research collaborations, less is known about collaboration in teaching practices and, in turn, student outcomes. This study contributes to this literature by analysing a massive dataset of English course syllabi taught between 1990 and 2019 across several countries. This dataset allows for a systematic, quantative analysis of trends over multiple years (8). Each syllabus provides detailed information about the instructors, including their gender, the institution, and the course's content, including its field, assigned readings, and an overall description of its content. This data allows us to systematically examine how instructors of different genders form teams, highlighting patterns consistent with gender homophily and institutional constraints in teaching collaboration. Additionally, by computing various metrics related to teaching content, this data lets us assess how different configurations of the teaching teams — gender composition and size — associate with teaching practices, particularly interdisciplinarity and the assigned readings, controlling for institution, field, academic year, and other potential drivers of content.

Specifically, we aim to address two key questions. First, how often do academics choose to teach alone versus co-teach with a colleague, and to what extent are co-teaching teams composed of mixed-gender pairs? Second, how do the size and gender composition of teaching teams relates to key aspects of teaching: (1) the course's interdisciplinarity, (2) the novelty of assigned readings, and (3) the gender representation of the cited authors. Previous research has shown that these aspects shape students' learning outcomes and career trajectories (1–3), while also influencing instructors' professional development and promotion within their institution (7).

Our results reveal a relation between gender and co-teaching practices, with mixed-gender teams occurring consistently less frequently than same-gender teams across institutions and fields. Furthermore, mixed-gender teams occur much less frequently than expected under a null model that forms teams in a gender-neutral manner while keeping fixed field-related and institutional constraints. These findings underscore a consistent and widespread underrepresentation of mixed-gender collaborations in teaching, that is not fully explained by

unobserved differences across fields or institutions.

Our analysis further reveals a significant and strong association between the gender composition of a teaching team and the course's interdisciplinarity, with mixed-gender teams being more interdisciplinary than all-male teams or courses taught by individual instructors. We also find a significant association of team configurations with the fraction of cited female authors, with courses taught by a female instructor alone citing a higher fraction of female authors compared to courses taught by male instructors, with mixed-gender teams falling in between. Finally, we also find a trend in novelty, with female instructors assigning more recent readings than men, regardless of team size.

Previous research has shown that interdisciplinarity research is less likely to be funded (9), tends to attract fewer citations when it is highly interdisciplinary (10), and is correlated with the probability of publication in academic journals (11). Furthermore, students attending colleges with more interdisciplinary courses tend to earn higher earnings after graduation (3, 12). We extend this work by looking at the association between gender composition and interdiscipilnarity in teaching, showing how mixed-gender teams occur less frequently but tend to be more interdisciplinary.

Our findings also contribute to the literature on gender dynamics in academia, particularly research team formation (13–15), which has shown significant gender homophily – a tendency to collaborate with colleagues of the same gender (16). It also examines gender differences in citation patterns, a driving factor of the persistent gender bias in academia (17), including tenure promotion (18), grant success (19), co-authorships (14), and peer recognition (20, 21). While prior research has largely focused on these areas, we shift attention to the citation gap in teaching. This issue may not only reinforce existing gender bias in academia but also shape students' learning outcomes and future career choices (22) – effects that are less well understood.

We also build on prior research on the underrepresentation of female-authored works in university curricula, which has been shown within specific fields (psychology and international relations) and at a small scale (23–25). Our results reveal a consistent and significant gap in cited works between female and male instructors, only partially addressed within mixed-gender co-teaching.

Finally, studies about team formation are especially relevant to our work. These studies have found a tendency of mixed-gender teams to perform better in various settings (26, 27). In research collaboration, for example, mixed-gender teams often obtain more citations, produce more novel research, publish in more prestigious journals and are more interdisciplinary (26, 28). Although we do not measure team "performance" as we lack data on students outcomes, our results show that mixed-gender co-teaching teams are consistently less likely to form, but, once created, they tend to deliver different outcomes, especially a higher interdisciplinarity, that

previous literature has suggested having an impact on students outcomes as discussed above.

2 Materials and Methods

2.1 Data

We examined a corpus of over six million documents compiled by Open Syllabus (New York, US). This dataset was created through web extractions that identified syllabi from university websites, with a median confidence level of 99.8%. A tagging algorithm extracted key course details, such as the title, field, description, academic year, duration, and language, a list of anonymised instructors, and the assigned readings.[1] While the original dataset included syllabi in 49 languages, most documents (96%) were in English. For simplicity, we focused exclusively on these documents.[2]

The resulting dataset comprised 5.4 million syllabi from approximately 4,000 higher education institutions across fifteen countries from 1990 to 2019. OpenSyllabus classified these syllabi into 69 top-level fields derived from the U.S. Department of Education's CIP code classification.[3] About 2.9 million syllabi (53% of the total) listed readings matched with bibliographic sources, providing additional metadata about authorship information, journal, and publication year. The institution was matched to a list of more than 22,000 entities from the Research Organisation Registry, providing further metadata including the instutition's country and enrollment figures — the institutions in our sample account for over 35 million enrolled students today.[4]

Each syllabus lists one or more instructors, with 76% of the syllabi listing a single instructor, 16% listing two, 4% listing three, and another 4% listing more than three instructors. Instructor gender was determined automatically by OpenSyllabus based on names, resulting in 52% male, 37% female, and 11% unknown categories. After excluding syllabi with unknown gender, the distribution was 58% male and 42% female instructors, which aligns closely with the 45% of female academic staff reported in OECD countries (29). The same inference method was used to determine the gender of the authors listed in the readings, resulting in 32% female and 56% male authors, with only 12% of unknown gender.[5]

[1] The documentation available at: https://docs.opensyllabus.org

[2] Our focus on English-language courses means that, while we have comprehensive data for English-speaking countries such as Canada, Ireland, the United States, and Great Britain, the sample in non-English-speaking countries tends to be more representative of internationally oriented universities. These are typically institutions offering programs in English or advanced-level courses, such as postgraduate programs or disciplines where English is the primary medium of instruction within traditional universities.

[3] The original CIP classification is available online: https://nces.ed.gov/ipeds/cipcode/browse.aspx?y=55

[4] For further information on how OpenSyllabus classified and matched the data, the related documentation is available online at https://docs.opensyllabus.org.

[5] Additional summary tables and figures that detail the structure of the dataset, including the number of syllabi by country and field, the distribution of teaching-team configurations appear in Appendix SI-1.

2.2 Outcome variables

We defined the following key outcome variables to analyse how different team configurations, such as gender and team size, relate to interdisciplinarity and characteristics of assigned readings, such as their publication age and the inclusion of female-authored works.

Table 1: Outcome variables

Name	Definition
Interdisciplinarity	Percentile rank of the course's interdisciplinarity score for the year.
Age of References	Percentile rank of the average publication age of assigned readings.
Ratio of Female Authors	Proportion of women authors in the assigned readings.

Interdisciplinarity

To estimate interdisciplinarity, we measured field overlap using course descriptions written in the syllabi. Following Evans et al. (3, 12), we assigned each course an interdisciplinarity score based on its description. This approach converts descriptions into "bags of words", where word frequencies are normalised with the inverse ratio of the term frequency to document frequency metric. A correlation matrix is then generated across different academic fields to measure the distance between fields. The interdisciplinarity score for each syllabus is computed by taking one minus the weighted average of the pairwise correlations with other syllabi, with weights equal to the conceptual proximity of different fields. This method ensures that syllabi associated with distant fields –either academically or conceptually– are considered more interdisciplinary. To scale this approach for millions of documents, we optimized for efficiency by using random subsamples for academic fields across academic years. The final interdisciplinarity score was averaged across multiple subsamples for robustness. See Supplementary Information (Section SI-3) for details. To ensure robust comparisons in our analysis, we computed the percentile rank of the interdisciplinarity score for each syllabus i:

$$\text{Interdisciplinarity}_i = \text{PR}_{yr}(\text{Interdisciplinarity Score}_i),$$

where PR_{yr} represents the percentile rank function applied to all syllabi within a given year yr.

Readings Selection

To investigate characteristics of the assigned readings, we calculated the following two dimensions: the age of readings (a measure of "novelty") and the proportion of female authors in the assigned readings. First, we

define the *Age of References* variable as the difference between the syllabus year (Year_i) and the average publication year of each assigned reading k:

$$\text{Age of References}_i = \text{PR}_{yr}\left(\text{Year}_i - \sum_{k=1}^{N_i} \text{Publication Year}_k / N_i\right),$$

where PR_{yr} represents the percentile rank function applied to all syllabi within a given year yr. This variable gives a proxy of how recent, or "novel," the readings are.[6] Then, we define the *Ratio of Female Authors* as the proportion of female authors among all authors in the assigned readings:

$$\text{Ratio of Female Authors}_i = \frac{\text{Female Authors}_i + 1}{\text{Female Authors}_i + \text{Male Authors}_i + 2}.$$

Here, we add two pseudo-observations (one for each gender) to stabilize the ratio, preventing extreme values in cases with very few authors. This metric allows us to investigate whether gender and collaboration relate to the representation of female-authored work in teaching. See Table 1 for a recap of the relevant outcome variables.

2.3 Simulated Teaching Collaborations

To examine whether the gender composition of co-teaching teams deviates from what one would expect under gender-neutral matching, we employ the following Monte Carlo approach. Drawing from a methodology developed elsewhere (30), we counted the frequency of courses taught individually and the frequency of gender combinations (male-male, female-male, etc.) of those taught by two instructors, disaggregating these data per field, institution, and academic year. Then, we compared these combinations against those expected by chance in a gender-neutral composition network, where all instructors are switched randomly within a given institution, field, and academic year. This approach matches our assumption that forming teams between institutions and fields is limited, at least within a single academic year, while forming teams within the same field and institution is attainable.

The switching algorithm preserves the total gender counts and the distribution of teams. This ensures that a course with a given number of instructors in the original data will have the same number of instructors in the randomised network. Similarly, an institution with a given number of male and female instructors teaching in each field will have the same number of male and female instructors. The only difference between the randomised and the original data will be the gender composition of the teams. Therefore, in the randomized network, instructors form teams as if they were unaware of the gender. See the details in the Supplementary

[6] While more sophisticated methods to measure novelty are available (30, 31), we opt for a simpler metric. Teaching innovation tends to be more incremental, and computationally intensive novelty indicators are impractical for large-scale datasets like ours.

Information (Section SI-2).

2.4 Regression analysis

We analyse how different teaching team configurations, $j \in \{F, M, MM, MF/FM, FF\}$, where F = female alone, M = male alone, MM = two males, FM/MF = mixed gender, FF = two female instructors, associate with course outcomes across academic years, $t = 1999, \cdots, 2020$. We run the specification separately for each academic year t. The outcomes of interest include: (1) interdisciplinarity, (2) the average age of references, and (3) the proportion of cited female authors. We employ the following linear mixed-effects model:

$$Y_{j,t} = \alpha_t + \text{Team}_{j,t} + \eta_t + \text{Country}_t + \text{Enroll}_t + \delta_t + \epsilon_{j,t}.$$

Where:

- $Y_{j,t}$ is the outcome variable for team configuration j in year t,
- α_t is the intercept estimated within year t,
- $\text{Team}_{j,t}$ is a fixed effect for team configuration,
- η_t is a random intercept for each of the 69 academic fields,
- Country_t and Enroll_t are categorical fixed effects for the institution's country and enrollment size, respectively,
- δ_t is a random intercept for each of approximately 4000 unique institutions,
- $\epsilon_{j,t}$ is the residual error term.

3 Results

3.1 Monte Carlo Benchmark: Observed vs. Expected Team Composition

Figure 1 presents trends in the instructors' gender composition and team size in university courses from 2000 to 2019 for courses with one or two instructors. Panel A shows courses with one instructor, where male-only courses (orange triangles) consistently dominated but declined from around 60% in 2000 to 46% in 2019. Over the same period, female-only courses (green circles) increased from roughly 25% to 36%. Panel B shows courses with two instructors, where male-only courses also decreased—from about 8% in the early 2000s to 6% in 2020, while female-only (green cricles) and mixed-gender (purple squares) instructor

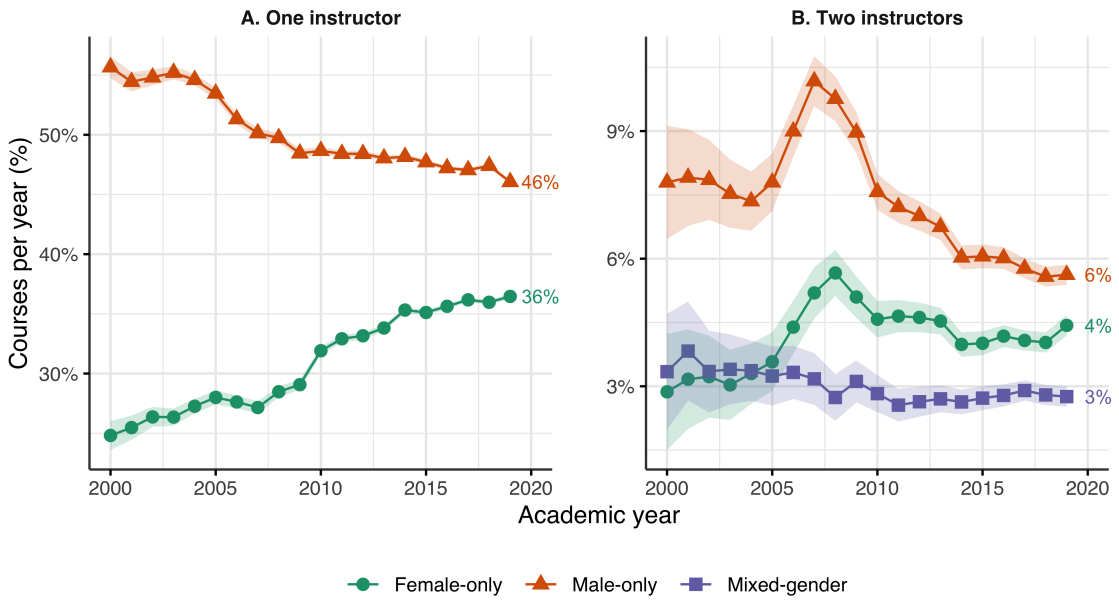


Figure 1: Evolution of Teaching Teams by Gender Composition and Size. Trends in course gender composition from 2000 to 2019 reveal a steady increase in women's participation (both in solo-taught and two-instructor courses), while the proportion of mixed-gender teams has remained relatively stable over time. (A) Percentage of solo-taught courses by instructor gender over time. (B) Percentage of two-instructor courses by gender composition over time. $N = 5.1$ million courses; shaded bands show 95 % CIs.

teams remained relatively stable, each accounting for approximately 4% and 3%, respectively, in recent years. Together, these results reveal a gradual but consistent shift toward greater gender diversity in course instruction over time, particularly a decline in male-only instruction and a rise in female-only instruction, especially in single-instructor courses.[7] Notably, mixed-gender teaching teams (3%) are consistently less common than same-gender teams (10%).

To assess whether gender imbalances in co-teaching teams persist after accounting for differences across fields and institutions, Figure 2 compares the actual size and gender compositions of teaching teams in university courses from 2000 to 2019 (solid green lines) with simulated compositions (dashed orange lines) generated from a gender-neutral Monte Carlo model. This model randomly shuffles instructors while preserving the overall distribution of course loads per year by institution and academic field but ignoring gender-based preferences or constraints, as discussed before (see Section 2).

The results reveal that actual data systematically deviate from simulated expectations: female-female (FF) and male-male (MM) courses occur more frequently in actual data than in simulations by 66% and 20%, respectively, in 2019. By contrast, mixed-gender collaborations (MF/FM)[8] are consistently underrepresented

[7]This overall trend aligns with the increase in women holding academic positions worldwide. In 2000, women constituted approximately 35% of academic staff worldwide. By 2022, this figure had risen to about 44%, according to data from the World Bank; The data are available at: https://data.worldbank.org/indicator/SE.TER.TCHR.FE.ZS

[8]Note that, in the main text, we do not take into account the order in which instructors are listed on the syllabus. Appenidx

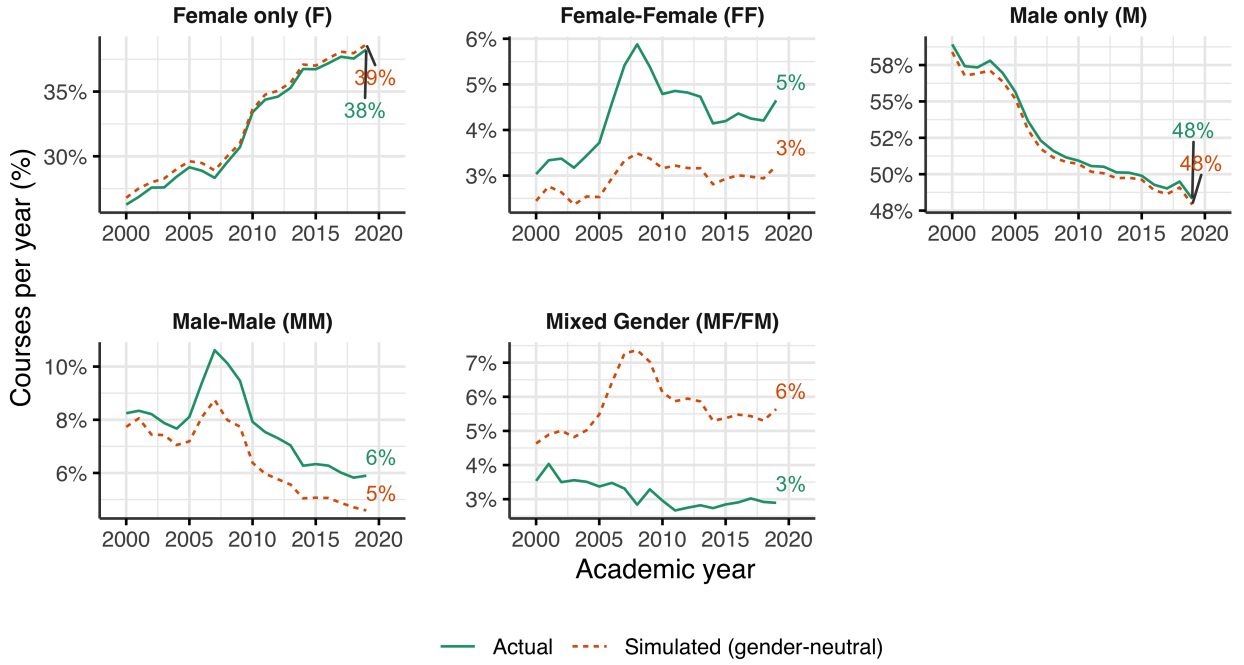


Figure 2: Comparison of gender composition between actual courses and courses simulated with Monte Carlo. Simulations ensure gender neutrality while keeping constant the institutional constraints (i.e., number of teams by size, institution, academic year, and 69 academic fields). Simulated mixed-gender teams consistently exceed observed proportions, highlighting persistent barriers to forming gender-diverse teaching teams.

in the actual data (3%) compared to simulations (6%). These findings suggest that significant social or institutional dynamics—beyond chance—reinforce same-gender pairings, especially among women. Such dynamics may include gender-based homophily, mentoring networks, or departmental assignment practices.

Figure 3 disaggregates the comparison between actual and simulated mixed-gender (MF/FM) courses by academic field. Nearly all fields show clear underrepresentation of mixed-gender teams (relative to expectations) by two to four times. For instance, in Medicine (Health and Welfare), the actual share of mixed-gender collaborations is 4.8% versus an expected 9.1%, indicating a large imbalance. Similar gaps are evident in Law (3.8% actual vs 6.4% simulated) and Linguistic (4.3% vs 6.7%), underscoring the pervasiveness of barriers to mixed-gender teams across different academic fields. Remarkably, we see no differences in these patterns between male- vs female-dominated fields (e.g., Engineering or Accounting vs Nursing or Chemistry), suggesting that the underrepresentation of mixed-gender teams may be a structural feature rather than one driven by field-specific gender imbalances[9].

SI-4 therefore repeats the Monte-Carlo comparison separately for MF and FM teams. The shortfall of mixed pairs is virtually identical in the two directions, indicating that our results are not driven by the instructor who happens to appear first (a position that may reflect alphabetical order, coordination role, or seniority).

[9]As a robustness check, Appendix SI-5 (Figure SI-5) plots each academic field's gap between observed and simulated mixed-gender teams against that field's share of female instructors. The scatter shows no discernible upward or downward pattern, indicating little connection between a field's gender balance and the gap.

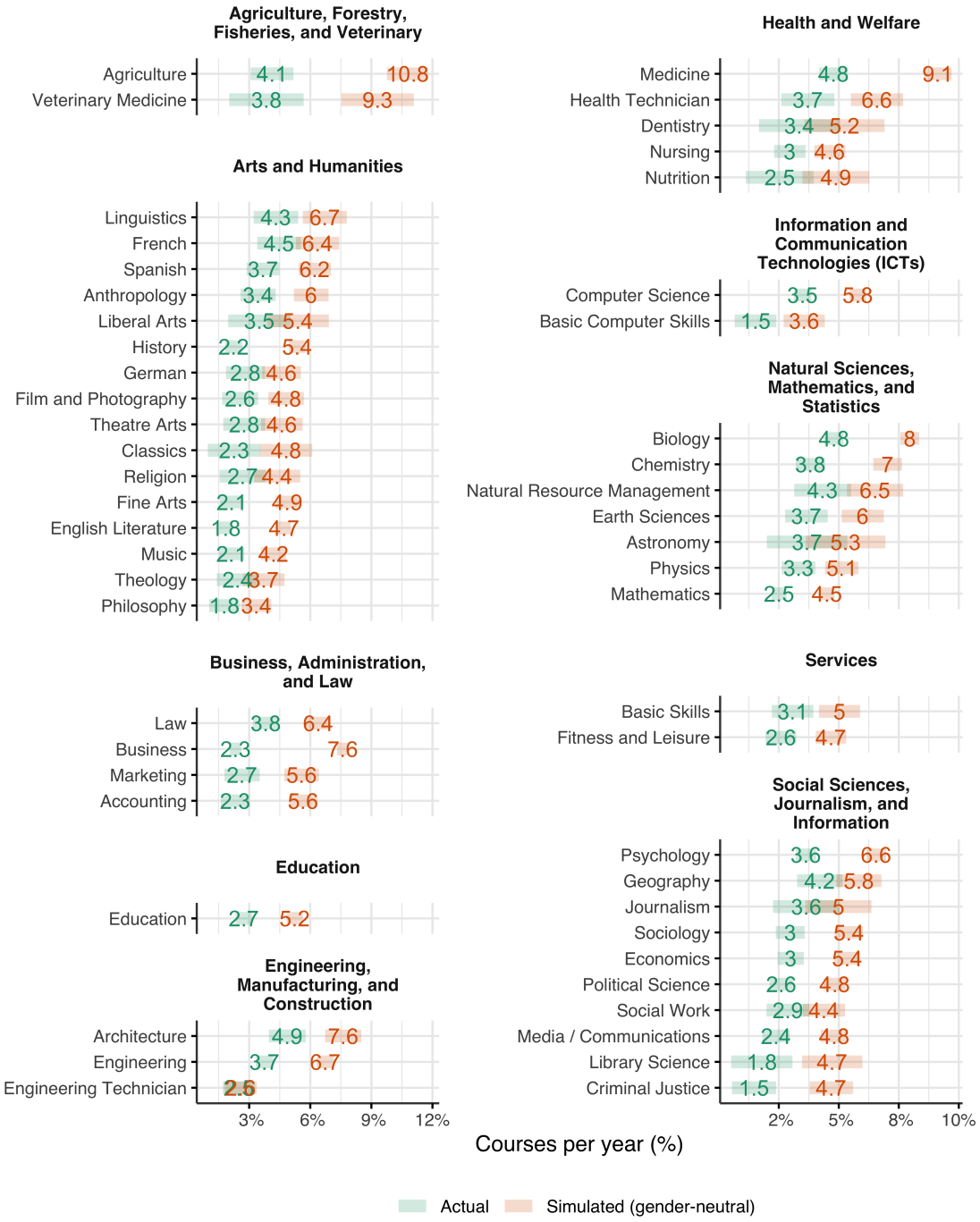


Figure 3: This figure disaggregates the comparison of proportions between actual (green) and simulated (orange) mixed-gender (MF/FM) courses by academic field. The shaded bar indicates 95% confidence level of the proportion.

Figure 4 illustrates that the proportion of mixed-gender teams in the simulations consistently exceeds the observed proportions across various geographic regions, despite notable variation in the magnitude of these differences. In 2019, for example, mixed-gender teams are relatively rare in Great Britain (1%) compared to Canada (4%). However, in the simulations, the proportion of mixed-gender teams is substantially higher in both countries—1% vs 3% and 4% vs 5%, respectively. In 2019, the largest discrepancy between simulated and observed proportions is found in EU countries from 5% to 12%. Overall, these patterns underscore the robustness of the findings across diverse geographic contexts.

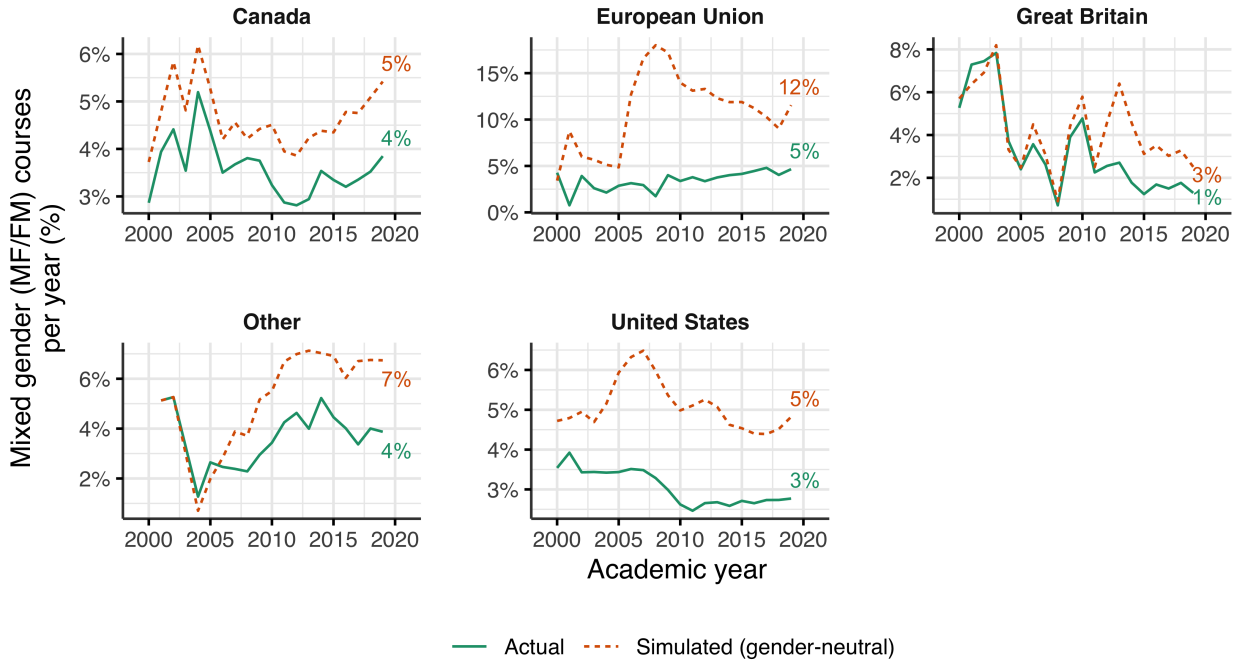


Figure 4: This figure disaggregates the comparison of proportions between actual (green) and simulated (orange) mixed-gender courses (MF/FM) by country. Simulated mixed-gender teams consistently exceed observed proportions in all countries. EU countries are: Austria, Denmark, France, Germany, Ireland, Italy, Netherlands, Poland, Portugal, Spain and Sweden.

3.2 The association with Course Materials

To examine how teaching team size and gender composition are associated with course materials, we analyse three key metrics: the percentile rank of the interdisciplinarity score per year ("interdisciplinarity"), the percentile rank of the average publication year of the readings ("age of readings"), and the share of female authors cited in the assigned readings ("share of female authors"). As discussed above, we estimate linear mixed-effects regressions separately for each academic year, controlling for unobserved differences across academic fields, institutions, and other course-level characteristics.

3.2.1 Interdisciplinarity

Figure 5 illustrates the differences in interdisciplinarity across team configurations. Our results indicate that there are no systematical differences in interdisciplinarity between female and male instructors when they are teaching alone: in recent years women tend to be more interdisciplinary, but this effect is not consistent over time. Conversely, mixed-gender teams tend to exceed in interdisciplinary all-male teams both individual and with two male instructors. This finding is consistent over time. At the same time, we find no consistent evidence of a difference between mixed gender and teams with two female instructors. These results point to a consistent association between gender diversity and interdisciplinarity.

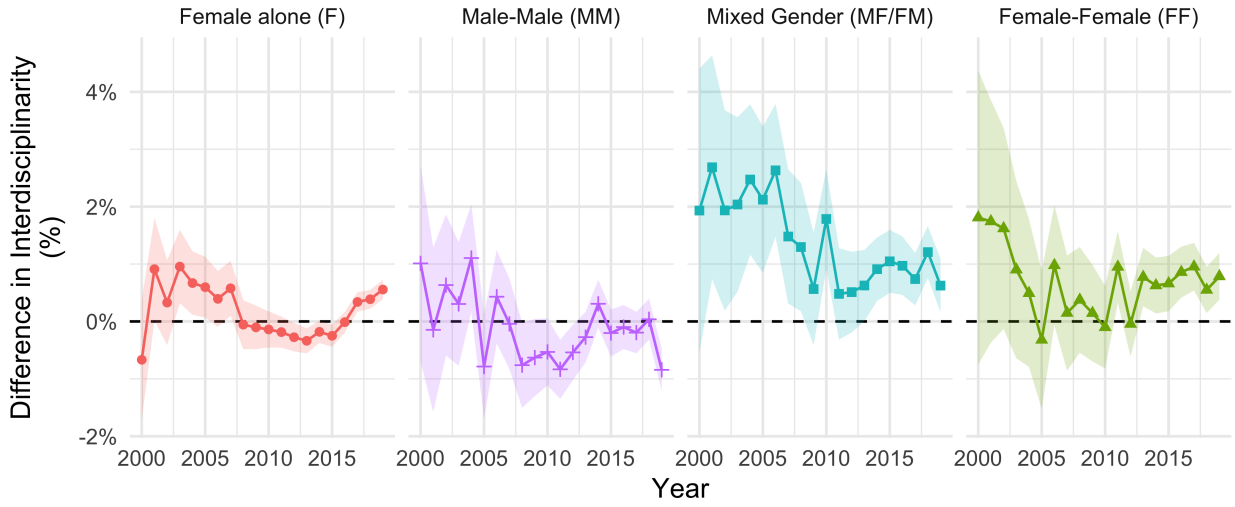


Figure 5: This figure illustrates the yearly difference in interdisciplinarity relative to courses taught by a single male instructor (baseline = 0) for each team configuration. The values are regression coefficients obtained separately for each academic year, with controls for country and field. Interdisciplinarity is expressed as the percentile rank of each course's interdisciplinarity score within its cohort; positive values therefore indicate a greater interdisciplinarity. The shaded area represents 95% confidence intervals.

3.2.2 Citing Women Authors

Figure 6 highlights significant and consistent gender association with the percentage of female authors over total authors cited in course readings per year, controlling for academic field and country.[10] Our results show that all-female courses consistently cite a higher fraction of female authors than all-male ones in the same year, with an effect that decreases over time going from 6% in 2000 to 3% in 2019 (with minor or insignificant differences between courses taught by one vs. two instructors). Mixed-gender courses are somewhat in between, as they cite a larger share of women than all-male courses, but less than all-female courses. Overall, these results indicate a consistent association between gender and cited patterns, suggesting that readings

[10]This analysis excludes syllabi where no readings were matched with the available bibliographic sources and, within the matched readings, excluding the references where the authors' gender remained unidentified.

tend to cluster based on the gender of the instructors and mixed-gender teams could promote greater diversity in assigned readings.

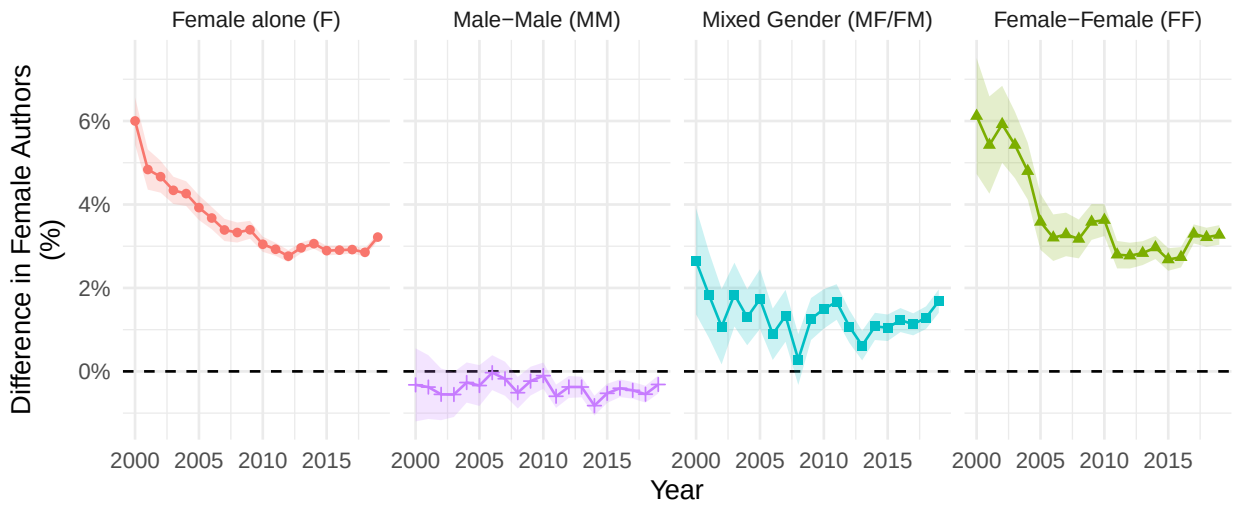


Figure 6: For each team configuration, dots plot the coefficient from a linear regression run separately for each year that compares its share of female-authored readings with the baseline of a solo male-taught course (set to 0). The regressions control for country, field and other covariates. A positive coefficient means the configuration cites a higher percentage of female authors than the male-alone baseline in that year; a negative coefficient means the opposite. The shaded area represents 95% confidence intervals.

3.2.3 Age of Readings ("novelty")

Figure 7 illustrates significant gender differences in the publication age of the selected readings.[11] Between 2000 and 2019, courses taught by a single female instructor tend to assign newer readings than otherwise similar courses taught by a single man in that year, with a difference that goes from -3% in 2000 to -1% in 2019. Similarly, courses led by two women also assign newer material than those taught by two men, with a decreasing difference. Mixed-gender teams tend to assign newer readings than the baseline in few years, but not consistently in all the years. There are no significant differences between courses taught by two male instructors and the single-male baseline. Overall, these findings suggest that women tend to assign more novel readings, although this gender difference tend to decrease over time.

[11] This analysis includes only courses with matched bibliographic data.

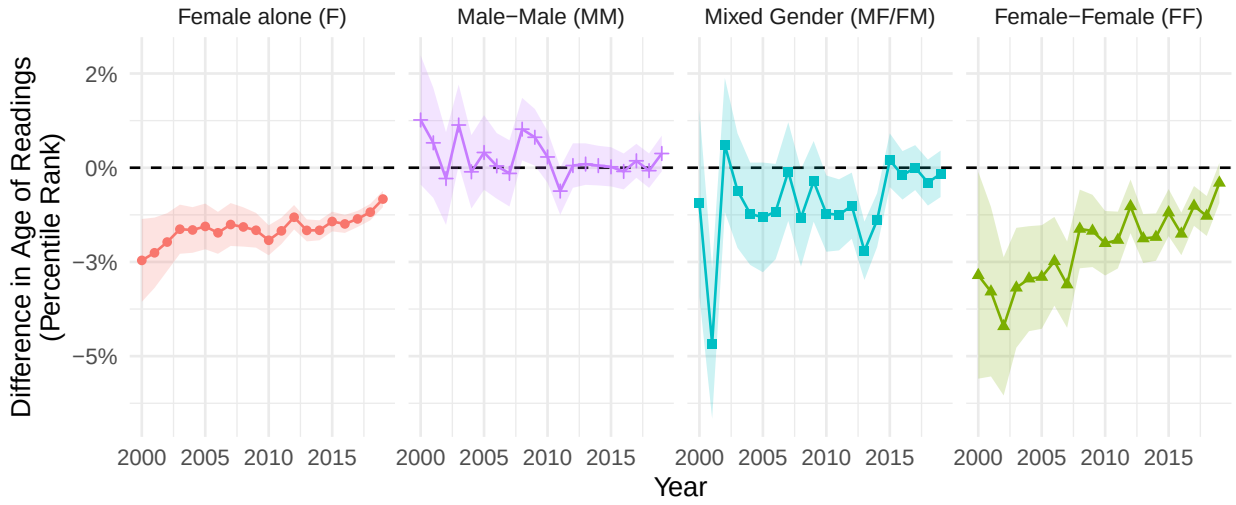


Figure 7: This figure illustrates the yearly difference in the percentile rank of the Age of References relative to courses taught by a single male instructor (baseline = 0), for each team configuration. The Age of Reference variable is the syllabus' year minus the average publication year of the assigned readings. The coefficients are estimated separately for each academic year with controls for country, field and other covariates; negative values indicate that the team configuration assigns newer readings compared with the male-alone baseline in the same year. The shaded area represents 95% confidence intervals.

3.3 Robustness

To ensure the reliability and robustness of our results, we conducted several additional analyses. In addition to reporting results from separate regressions on subsets of data by year, we also estimated mixed-effects regressions using the full dataset. These results are broadly consistent with those presented in the paper.

Furthermore, while the reported model specification is our preferred one, we explored several alternatives. First, we log-transformed our dependent variables – the age of readings, interdisciplinarity score, and count of female authors — to explore potential nonlinear relationships, particularly relevant for disciplines with very old readings (e.g. history).

Second, we employed a quasi-poisson mixed-effect model instead of the reported linear mixed-effect regression to better account for non-linearities in the count of female authors. We also trimmed the data to exclude outliers, such as courses with an unusually high number of readings. These changes yield similar results.

We further tested alternative model specifications such as considering the number of instructors as a numeric variable interacted with the proportion of female instructors to explore a linear association with the dependent variables. While we observe similar results, the linear association could be misleading when extrapolated beyond two instructors. So, we preferred to keep team configurations as fixed effects, as reported here.

Finally, we further explored specifications distinguishing between male-led and female-led courses, using the gender of the first listed instructor as an (imperfect) proxy for seniority. Overall, results appear robust to

these additional analyses.

4 Conclusion

Our analysis of approximately five million syllabi from over 4,000 universities revealed several key findings. First, we show that, while the share of classes taught by female instructors has been increasing over the last twenty years, the fraction of mixed-gender co-teaching classes has remained consistently low (3%) and significantly below that of same-gender teams (10%). This trend might be explained by a limited gender diversity in certain fields or administrative practices that restrict opportunities to form mixed-gender collaborations. But it can also arise from gender preferences in team formation, whereby instructors tend to form teams with the same gender partners ("homophily"), a trend already observed in research collaborations and other academic domains as discussed in the Introduction.

We further show that the underrepresentation of mixed gender teams persists even after accounting for unobserved differences across fields, institutions, and years. Specifically, we employ Monte Carlo simulations that randomly form teams in a gender-neutral manner, allowing us to control for some institutional constraints, such as maintaining constant the course loads across institutions and across 69 academic fields within a given year. By comparing simulated with actual teams, we find that instructors tend to partner with same-gender colleagues more than twice as expected between 2005 and 2019. This result is robust across academic fields, although it is more sizeable in certain fields (Chemistry, Business, Medicine) than others. The underrepresented share of mixed-gender teams is consistent across institutions in different countries, with the strongest effect in the selected EU countries, where only 5% of courses are mixed-gender, compared to an expected 12% under the gender-neutral benchmark.

Our findings further highlight how different team configurations are associated with key aspects of teaching: (1) mixed-gender teams tend to be more interdisciplinary than all-male teams, (2) all-female teams tend to assign more novel readings, and (3) mixed-gender teams tend to cite a higher share of female authors than all-male teams, but less than all-female teams. These associations are not driven by unobserved differences across fields, institutions, and years that we account for in the regressions.

5 Discussion

Our findings underscore a significant and consistent underrepresentation of gender-mixed teaching teams, with potential implications for both instructors and students. Limited mixed-gender collaborations in teaching may limit the exchange of information among instructors of different gender and restrict students' access to

diverse courses. These dynamics, in turn, could hinder interdisciplinary engagement or reinforce structural disadvantages for women within academic networks. Our work thus identifies a potential additional driver of gender bias, complementing existing work on gender disparities in various settings within academia, such as research collaboration, access to funding, and student gender stereotypes.

Our analysis bears several limitations. First, the simulations could be improved by more granular information on the course's subfields that go beyond the 69 top-level fields identified in the dataset. If male and female instructors tend to concentrate in different subfields (e.g., Econometrics and Macroeconomics, within Economics), the observed underrepresentation of mixed-gender teams may reflect a limited availability of instructors of the other gender, especially for advanced courses where teaching teams are more likely to form around narrower areas of expertise. From a policy perspective, this limitation is important as we currently cannot disentangle whether the underrepresentation derives from gender-based preferences towards subfields, limiting availability, or a tendency to seek same-gender partners, homophily. Further research is needed to fully disentangle these underlying mechanisms.

Similarly, another limitation is the lack of data on potential determinants of team formation that operate within institutions, such as instructor experience or academic rank. Hierarchical differences or administrative structures may contribute to the observed gap in cross-gender collaborations. However, our results illustrate that the under-representation of mixed-gender teams persists even when co-teaching teams are disaggregated by male-led and female-led courses (as measured by the order of appearance in the syllabus), suggesting that ordering and by extension simple rank effects cannot wholly account for the gap.

An additional limitation is that syllabi may sometimes be inherited from previous instructors or reflect departmental templates. In these cases, the course content may not be directly chosen by the current teaching team. However, since our analysis focuses on associations rather than causal effects, this does not invalidate our results. Moreover, even inherited syllabi are institutionally endorsed and assigned by the listed instructors, making them a meaningful part of the observed teaching patterns. To the extent that inheritance occurs, it likely introduces random noise rather than systematic bias, making the associations we report more, not less, robust.

Our outcome variables also deserve a bit more discussion. Interdisciplinarity, novelty, and the share of female authors capture interesting dimensions of teaching content, yet they represent only a slice of the complexities inherent in course design and student learning. The three measures are descriptive tools, useful for comparing millions of courses, but they should not be read as universal benchmarks that every syllabus ought to maximise. In certain fields, or in some highly specialised sub-field, recent literature or interdisciplinarity or

gender diversity in materials are neither feasible nor desirable. A seminar on ancient Greek lyric poetry or on biblical Hebrew, for example, rightly focuses on canonical texts; likewise, a graduate course in advanced algebraic topology need not venture far outside mathematics. Certain courses (e.g a studies of classical antiquity) offer few female-authored primary sources, so increasing gender diversity might not be a realistic goal. The variables therefore indicate a course's relative position within its yearly cohort, not a normative target for curriculum design.

This study opens several avenues for future inquiry. A first priority is to observe directly how co-teaching teams are formed, so that voluntary partnering can be distinguished from administrative assignment. Access to departmental workload files or other institutional records would clarify the extent to which the shortfall of mixed-gender teams reflects scheduling rules rather than individual choice. A second line of work would assemble finer-grained data on instructors and courses (covering specific sub-fields and career stages) to test whether the under-representation of mixed teams persists once availability constraints in strongly gender-segregated specialties are fully accounted for. Finally, research that explores instructors' teaming attitudes and constraint through surveys, interviews, or experiments, could shed light on whether gender-related preferences play an independent role. Addressing these questions would move the evidence base beyond the descriptive patterns documented here toward a fuller understanding of who teaches with whom and why.

References

1. I. Bakkenes, J. D. Vermunt, T. Wubbels, Teacher learning in the context of educational innovation: Learning activities and learning outcomes of experienced teachers. *Learning and instruction* **20**, 533–548 (2010).

2. M. H. Vembye, F. Weiss, B. Hamilton Bhat, The effects of co-teaching and related collaborative models of instruction on student achievement: A systematic review and meta-analysis. *Review of Educational Research* **94**, 376–422 (2024).

3. S. Han, J. LaViolette, C. Borkenhagen, W. McAllister, P. S. Bearman, Interdisciplinary college curriculum and its labor market implications. *Proceedings of the National Academy of Sciences* **120**, e2221915120 (2023).

4. N. Ellemers, Gender stereotypes. *Annual review of psychology* **69**, 275–298 (2018).

5. P. Lockwood, "Someone like me can be successful": Do college students need same-gender role models? *Psychology of women quarterly* **30**, 36–46 (2006).

6. S. Canaan, P. Mouganie, The impact of advisor gender on female students' STEM enrollment and persistence. *Journal of Human Resources* **58**, 593–632 (2023).

7. L. R. Buckingham, A. López-Hernández, B. Strotmann, Learning by comparison: The benefits of co-teaching for university professors' professional development in *Frontiers in Education*, (Frontiers Media SA, 2021), p. 776991.

8. A. Nowogrodzki, Mining the secrets of college syllabuses. *Nature* **539**, 125–126 (2016).

9. L. Bromham, R. Dinnage, X. Hua, Interdisciplinary research has consistently lower funding success. *Nature* **534**, 684–687 (2016).

10. V. Larivière, Y. Gingras, On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology* **61**, 126–131 (2010).

11. S. Xiang, D. M. Romero, M. Teplitskiy, Evaluating interdisciplinary research: Disparate outcomes for topic and knowledge base. *Proceedings of the National Academy of Sciences* **122**, e2409752122 (2025).

12. E. D. Evans, Measuring interdisciplinarity using text. *Socius* **2**, 2378023116654147 (2016).

13. S. Milojević, Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences* **111**, 3984–3989 (2014).

14. A. Boschini, A. Sjögren, Is team formation gender neutral? Evidence from coauthorship patterns. *Journal of Labor Economics* **25**, 325–365 (2007).

15. H. B. Love, A. Stephens, B. K. Fosdick, E. Tofany, E. R. Fisher, The impact of gender diversity on scientific research teams: A need to broaden and accelerate future research. *Humanities and Social Sciences Communications* **9**, 1–12 (2022).

16. M. Kwiek, W. Roszka, Gender-based homophily in research: A large-scale study of man-woman collaboration. *Journal of Informetrics* **15**, 101171 (2021).

17. N. Boivin, S. Täuber, U. Beisiegel, U. Keller, J. G. Hering, Sexism in academia is bad for science and a waste of public funding. *Nature Reviews Materials* **9**, 1–3 (2024).

18. K. Weisshaar, Publish and perish? An assessment of gender gaps in promotion to tenure in academia. *Social Forces* **96**, 529–560 (2017).

19. G. Bedi, N. T. Van Dam, M. Munafo, Gender inequality in awarded research grants. *The Lancet* **380**, 474 (2012).

20. H. Sarsons, Recognition for group work: Gender differences in academia. *American Economic Review* **107**, 141–145 (2017).

21. D. Card, S. DellaVigna, P. Funk, N. Iriberri, Gender differences in peer recognition by economists. *Econometrica* **90**, 1937–1971 (2022).

22. D. Card, C. Domnisoru, S. G. Sanders, L. Taylor, V. Udalova, "The impact of female teachers on female students' lifetime well-being" (National Bureau of Economic Research, 2022).

23. L. J. Skitka, Z. J. Melton, A. B. Mueller, K. Y. Wei, The gender gap: Who is (and is not) included on graduate-level syllabi in social/personality psychology. *Personality and Social Psychology Bulletin* **47**, 863–872 (2021).

24. J. Colgan, Gender bias in international relations graduate education? New evidence from syllabi. *PS: Political Science & Politics* **50**, 456–460 (2017).

25. J. K. Harris, M. A. Croston, E. T. Hutti, A. A. Eyler, Diversify the syllabi: Underrepresentation of female authors in college course readings. *PLoS One* **15**, e0239012 (2020).

26. Y. Yang, T. Y. Tian, T. K. Woodruff, B. F. Jones, B. Uzzi, Gender-diverse teams produce more novel and higher-impact scientific ideas. *Proceedings of the National Academy of Sciences* **119**, e2200841119 (2022).

27. J. B. Bear, A. W. Woolley, The role of gender in team collaboration and performance. *Interdisciplinary science reviews* **36**, 146–153 (2011).

28. A. Maddi, Y. Gingras, Gender diversity in research teams and citation impact in economics and management. *Journal of Economic Surveys* **35**, 1381–1404 (2021).

29. OECD, "What is the profile of academic staff and what is the student-academic staff ratio?" in *Education at a Glance 2022: OECD Indicators*, (OECD Publishing, 2022).

30. B. Uzzi, S. Mukherjee, M. Stringer, B. Jones, Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).

31. J. Wang, R. Veugelers, P. Stephan, Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy* **46**, 1416–1436 (2017).

SI Supporting Information

SI-1 Additional Descriptive Statistics

Table SI-1: Top Rows of The Dataset

formation	composition	field	year	country	n	team_gender	team_size
team	f	Accounting	1994	CA	2	Female-only	1
team	f	Accounting	1994	US	1	Female-only	1
team	f	Accounting	1995	US	4	Female-only	1
team	f	Accounting	1997	CA	2	Female-only	1
team	f	Accounting	1997	US	2	Female-only	1
team	f	Accounting	1998	CA	7	Female-only	1

Table SI-2: Descriptives

Variable	**Value**
Syllabi	5,381,496
Syllabi with matched refs	2,927,728
Syllabi with 2+ Instructors	965,291
Institutions	4,061
Institutions (Grid ID)	3,865
Student enrollment (total)	35,327,015
Fields	69
Countries	15
Year (min)	1990
Year (median)	2015
Year (max)	2019

Table SI-3: Syllabi per Country

Country	Syllabi	Syllabi (%)
USA	3 554 098	66.1
Great Britain	583 612	10.9
Canada	389 818	7.3
Italy	299 851	5.6
Poland	107 214	2.0
Netherlands	86 462	1.6
Germany	62 374	1.2
Other	55 285	1.0
Ireland	50 864	0.9
Portugal	42 143	0.8
Sweden	42 097	0.8
Spain	37 777	0.7
Austria	33 278	0.6
Denmark	24 262	0.5
France	5 891	0.1

Table SI-4: Number of Syllabi per Field

Field	N (thousands)	%	Field	N (thousands)	%
Accounting	76.9	1.4	Agriculture	36.2	0.7
Anthropology	51.9	1	Architecture	48.5	0.9
Astronomy	13.9	0.3	Atmospheric Sciences	6.6	0.1
Basic Computer Skills	51.1	0.9	Basic Skills	50.7	0.9
Biology	250.3	4.7	Business	450.4	8.4
Career Skills	13.9	0.3	Chemistry	106.4	2
Chinese	10.6	0.2	Classics	23.3	0.4
Computer Science	310.5	5.8	Construction	10.9	0.2
Cosmetology	6.9	0.1	Criminal Justice	45.5	0.8
Criminology	10.5	0.2	Culinary Arts	9.7	0.2
Dance	14.2	0.3	Dentistry	13.9	0.3
Earth Sciences	49.5	0.9	Economics	130.9	2.4
Education	240.3	4.5	Engineering	215.1	4
Engineering Technician	61.2	1.1	English Literature	357.3	6.6
Film and Photography	49.8	0.9	Fine Arts	112.4	2.1
Fitness and Leisure	93.0	1.7	French	33.5	0.6
Geography	43.3	0.8	German	42.2	0.8
Health Technician	31.9	0.6	Hebrew	3.1	0.1
History	227.4	4.2	Japanese	10.7	0.2
Journalism	19.9	0.4	Law	99.5	1.9
Liberal Arts	16.7	0.3	Library Science	25.0	0.5
Linguistics	32.3	0.6	Marketing	51.6	1
Mathematics	406.2	7.6	Mechanic / Repair Tech	23.4	0.4
Media / Communications	118.8	2.2	Medicine	110.0	2
Military Science	2.9	0.1	Music	101.2	1.9
Natural Resource Management	28.0	0.5	Nursing	98.5	1.8
Nutrition	19.6	0.4	Philosophy	70.8	1.3
Physics	80.4	1.5	Political Science	167.3	3.1
Psychology	222.7	4.1	Public Administration	7.8	0.1
Public Safety	7.4	0.1	Religion	29.9	0.6
Sign Language	6.9	0.1	Social Work	48.8	0.9
Sociology	107.5	2	Spanish	60.5	1.1
Theatre Arts	36.3	0.7	Theology	36.9	0.7
Transportation	4.3	0.1	Veterinary Medicine	11.9	0.2
Women's Studies	7.4	0.1			

Table SI-5: Teaching Team Configurations

Team composition	N (thousands)	%
Male only (M)	2587.6	48.1
Female only (F)	1828.6	34.0
Male-Male (MM)	349.6	6.5
Female-Female (FF)	228.6	4.3
Mixed Gender (MF)	83.4	1.6
Mixed Gender (FM)	67.9	1.3

SI-2 Simulating Gender-Neutral Courses

We adapt the methodology of (30) to see whether same-gender teams form more frequently than expected by chance. We exclude teams with more than ten members to minimize the influence of outliers, and we stabilize team proportions by adding two pseudo-observations to avoid extreme (0 or 1) rates in small samples.

Observed Data For each institution i, field ℓ, and year t, let $M_{i,\ell,t}$ and $F_{i,\ell,t}$ be the numbers of male and female instructors. Each syllabus s has a team of size n_s and an observed gender composition (e.g., all-male, all-female, or mixed).

Randomization We construct a Monte Carlo "gender team composition" network that preserves:

1. The total $M_{i,\ell,t}$ and $F_{i,\ell,t}$ in each institution-field-year.
2. Each syllabus s retains its team size n_s, but we randomly swap instructor genders across syllabi while preserving total male and female counts.

SI-3 Interdiscipline Similarity

We measured interdisciplinarity using text similarity between syllabi as in (12) and (3). We transformed text from course descriptions into "bags of words," with term frequences (TF) normalised using the inverse document frequency (IDF). For each year, we calculated the TF-IDF scores for all syllabi and for 69 academic fields, using concatenated descriptions for the fields. We then computed the weighted average of the cosine similarity, $\cos(i, f)$, between each course i and field f, where the weight is based on the similarity between field f and the course's closest field $f_{\max}$. Specifically, $\cos(f_{\max}, f)$, where $f_{\max}$ is the field that has the highest cosine similarity with course i (i.e., $f_{\max} = \arg\max_f \cos(i, f)$) in that academic year.

Thus, the interdisciplinarity score for course i is:

$$\text{interdisciplinarity score}_i = 1 - \frac{\sum_{f \in (1,69)} \cos(f_{\max}, f) \cdot \cos(i, f)}{\sum_{f \in (1,69)} \cos(f_{\max}, f)}.$$

To reduce computational costs, we "bootstrap" the field-by-field cosine similarity matrix by using a 10% random sample of syllabi for each academic year. We repeated the sub-sampling procedure ten times and averaged the results.

SI-4 Simulating gender-neutral courses with order of appearance

Instructors may not always choose their co-teachers freely; senior faculty might have greater say in team assignments, while junior colleagues might follow departmental decisions. For that reason, the first position on a syllabus can act as a rough signal of hierarchy (even though it may also reflect alphabetical order or coordination duties or even work-load unevenness). And while the main text pools MF and FM courses, to confirm that ordering does not drive our results we rerun the Monte-Carlo benchmark while treating MF (man first, woman second) and FM (woman first, man second) as distinct team types.

Figure SI-4 shows that, for every year from 2000 to 2019, the observed shares of MF and FM teams are each about one percentage point lower than their respective Monte-Carlo expectations, and the two gaps are essentially identical year by year.

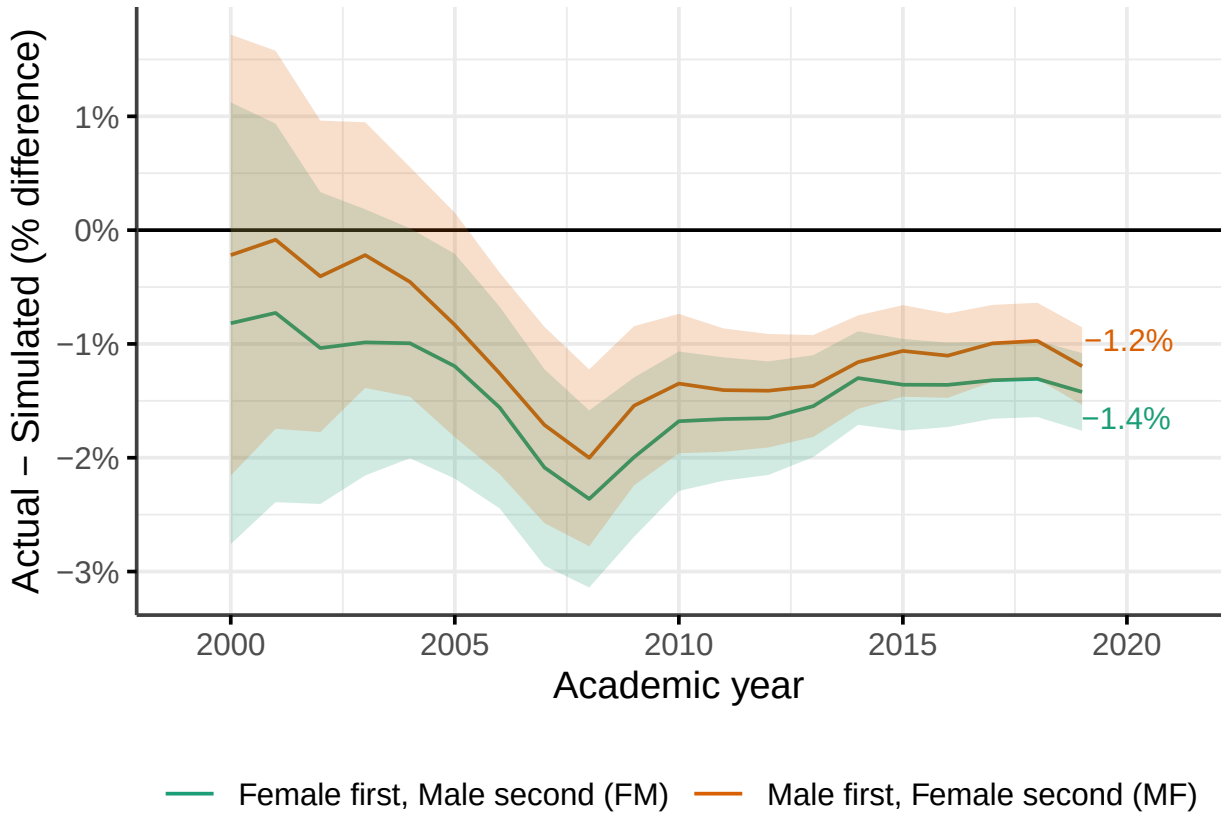


Figure SI-4: Difference between observed and simulated shares of mixed-gender teams, by instructor ordering. "MF" denotes courses in which a man is listed first and a woman second; "FM" the reverse. Dashed lines are the 95% confidence bands of the simulated benchmark. Year-by-year gaps are statistically indistinguishable.

SI-5 Field-level gender balance and the mixed-team gap

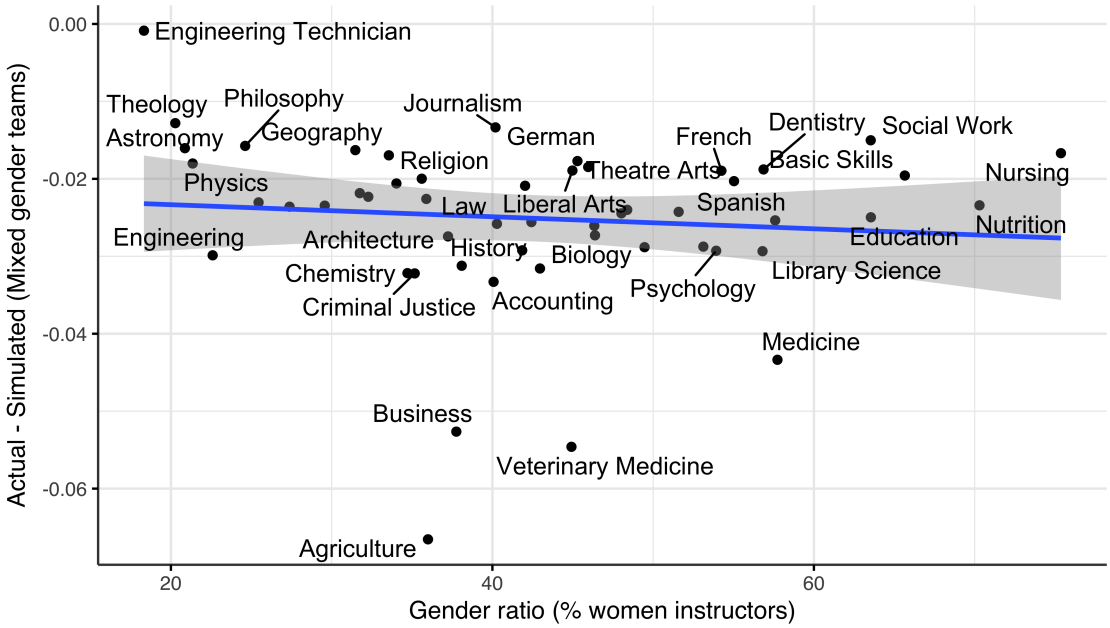


Figure SI-5: Gap between observed and simulated mixed-gender teams by academic field. Each point is a field; the horizontal axis is the percentage of female instructors in that field, and the vertical axis is the difference (actual – simulated) in the share of mixed-gender courses. There is no visible relation between a field's gender balance and the mixed-team gap.