The London School of Economics and Political Science

Essays in Applied Microeconomics

Martina Zanella

A thesis submitted to the Department of Economics
for the degree of Doctor of Philosophy

4 February 2022

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 38,900 words.

Statement of conjoint work

I confirm that Chapter 2 was jointly co-authored with Jay Euijung Lee, and I contributed 50% of this work.

I confirm that Chapter 3 was jointly co-authored with Jay Euijung Lee, and I contributed 50% of this work.

Acknowledgments

I am extremely grateful to my supervisors, Nava Ashraf, Oriana Bandiera, and Robin Burgess, for offering their time, invaluable guidance, and support throughout this long journey. I learned a lot working with them, and my research benefited immensely from their inputs. I am deeply grateful also to the other members of the LSE faculty who have dedicated their time to help me improve my work during meetings or seminars, including Tim Besley, Gharad Bryan, Jeremiah Dittmar, Maitreesh Ghatak, Camille Landais, Alan Manning, Rachael Meager, Guy Michaels, Steve Pischke, Daniel Reck, Mark Schankerman, and Daniel Sturm. A special acknowledgment goes to Judith Shapiro for priceless guidance and support. Lastly, I want to mention Eliana La Ferrara and Selim Gulesci, who inspired me to start this journey.

I would like to thank Lubala Chibwe, Mike Rose, Nic Warner, Anna Watmuff, Charlotte Rough, Mark Wilbor, and Emma Taverner for excellent administrative and IT help.

I feel very lucky to have crossed paths with a lot of beautiful people along the way, Andrea Alati, Fabio Bertolotti, Adrien Bussy, Svetlana Chekmasova, Amanda Dahlstrand, Alexia Delfino, Gaia Dossi, Sacha Dray, Dita Eckardt, Ruth Fortmann, Friedrich Geiecke, Tillman Hoenig, Takafumi Kawakubo, Krittanai Laohakunakorn, Jay Euijung Lee, Virginia Minni, Niclas Moneke, Marco Pelosi, Wolfgang Ridinger, Kilian Russ, Vincenzo Scrutinio, Xuezhu Shi, Alessandro Sforza, Tommaso Sonno, John Spray, and Celine Zipfel. Thank you for the helpful suggestions, the advice, the emotional support, and the laughs we shared. You made this Ph.D. journey infinitely better.

Words cannot express how grateful I am to Viola, Martina, Luca, Anna, and Nicola. I don't think this would have been possible without their friendship. Special thanks go also to my friends Alessandra, Edoardo, Lorenzo, Francesco, Giacomo, Luca, Andrea, and Francesca for the fun times, the empathy, the words of wisdom, for keeping me grounded and, more importantly, sane!

None of this would have been possible without the support of my family. I am immensely grateful for their love, encouragement, and for being the most reliable and supportive people I could have ever hoped for.

Finally, my deepest gratitude goes to Giovanni for being my rock, tirelessly walking beside me every step of the way, even when it was not easy to do. Thank you for never stopping believing in me.

Abstract

This thesis consists of three chapters in which I study how to design effective policies to address inequality in performance and opportunities in education and the labor market.

In the first chapter, I explore the effect of minority status in explaining disparities in performance. Stereotypes shape the distribution of traits across occupations and majors, influencing payoffs from economic choices. As a consequence, the individuals that we observe in the minority are often individuals who decided to bear the cost of making a choice against stereotypes, a condition that might matter on its own. This margin has remained mostly unexplored due to the difficulty of finding a setting characterized by an independent variation in the two dimensions, which often overlap in real-world environments. In my paper, I disentangle these two effects by combining a choice with well-defined stereotypes (university major) with variation in peer identity across small, exogenously formed classes within the same course. Evidence from the performance of 14,000 students in an elite university indicates that those who go against stereotypes do not suffer from being in the minority, but they impose negative externalities on those who select on stereotypes. This might explain why the majority upholds stereotypes and why targeting minorities to foster inclusion might not be enough and even backfire.

The second and third chapters focus on affirmative action policies, one of the primary policy recommendations to fight the under-representation of women in decision-making bodies. The context is South Korean municipal councils, where gender quotas were introduced shocking a status quo where women were nearly absent.

Chapter two explores the effect of the policy on parties' selection of candidates by exploiting the discontinuity in the intensity of the quota at specific cut-offs of council size. Quotas were implemented in only one of the two independent election arms, leaving space for adjustment in selecting candidates in the arm unaffected by the policy. We find that higher gender quotas in the constrained arm induce municipalities to elect fewer women in the unconstrained arm. However, this pattern gradually reverses over time. The reversal is driven by parties learning about women's competence after having experienced a female councilor. This paper highlights the risk of gender quotas being not effective or even counterproductive if they are introduced before attitudes have changed sufficiently to accommodate them.

The third chapter presents the preliminary findings of a project exploring how group interactions, the decision-making process, and its outcomes change after the introduction of gender quotas. South Korean municipal councils are required to publish transcripts of each meeting, allowing us to speak to the evolution of group dynamics by analyzing rich text data spanning >150,000 meet-

ings. We find that equality in numbers did not immediately translate into equality in voice in decision-making. The women introduced by the quota start their term less vocal than men, even when compared to rookie men equally lacking experience. However, the gap between rookie men and women nearly fully closes by the end of the term, suggesting that differences in talent are not why women are less vocal and that rookie women gradually gain influence as councilors work together during the years.

Contents

List of Figures

List of Tables

Chapter 1

Stereotypical Selection

1.1 Introduction

In recent years, a variety of policies targeting minority groups have been implemented to address the lack of diversity in learning environments and in the workplace.[1] Will these policies be successful in leveling the playing field? Does targeting minorities represent the most effective use of our resources? Minority status is not only related to under-representation, but also to stereotypical expectations. By influencing payoffs from economic choices, stereotypes shape the distribution of groups across fields (Akerlof and Kranton, 2000). As a consequence, the individuals we observe in a numerical minority in a field are often individuals who decided to bear the cost of making a choice against stereotypes ("stereotypical minority"). As such, they might react very differently to the composition of the environment compared to individuals who made choices in line with stereotypes.[2] Assessing whether this is the case is crucial to design effective policies as this makes findings from experiments, where selection is shut down by design, not necessarily generalizable to real-world environments. However, this is challenging, as we would need a setting with some independent variation in the two dimensions, numerical minority and stereotypical minority.

In this paper, I overcome this challenge by studying a real-world setting that combines a choice with well-defined stereotypes with exogenous variation in peer identity. This allows me to observe the stereotypical minority group in the field (for example, women in Mathematics) in the numerical majority and vice-versa. By studying 14,313 students across 10 academic years and 16 departments at an elite UK university, I assess whether selection into fields, and in particular whether individuals made a choice of major in line or against stereotypes regarding gender skills and roles, interacts

[1] Within the Economic discipline alone, examples of such initiatives are the launch of the Committee on the Status of Minority Groups and the Committee on the Status of Women in the Economics Profession by the American Economics Association, or the Minorities in Economics Committee launched by the European Economics Association, or the multitude of programs implemented by universities in support of minority groups, such as experimental study groups promoted by MIT (Russell, 2017), special training programs for women enrolled in STEM at Harvard Business School, the Undergraduate Women in Economics Challenge at Harvard University (Avilova and Goldin, 2018), workshops to mentor young female assistant professors promoted by the American Economic Association (Blau et al., 2010), or provision of role models in mandatory first-year courses (Porter and Serra, 2020).

[2] Recent literature on selection into occupations and majors provides evidence that stereotypes regarding group-specific skills, the composition of the environment, and identity considerations affect beliefs about expected returns, driving individuals out of counter-stereotypical occupations (Akerlof and Kranton, 2000; Oxoby, 2014; Kugler et al., 2021; Del Carpio and Guadalupe, 2018). Furthermore, individuals incorporate their preferences towards the fraction of women and men in the occupation when making choices regarding the type of field to specialize in (Pan, 2015; Card et al., 2008).

with the effect of class gender composition on performance for students once they are enrolled in their first year of bachelor.

I define stereotypical and counter-stereotypical selection leveraging the variety of programs offered by the university. Lab and field experiments document the existence of stereotypically-female and stereotypically-male disciplines in the academic setting: women are believed to be worse than men in Mathematics and Science while being better at reading (Carlana, 2019; Reuben et al., 2014; Ellemers, 2018; Lane, 2012). This translates into men and women making very different educational choices (Bertrand, 2020; Bertrand, 2011), which is what I observe in my setting. While 49% of the population of students enrolled in undergraduate programs at the university are women, they represent the minority of students enrolled in departments such as Mathematics, Economics and Finance, and the majority of students enrolled in programs related to disciplines such as Anthropology, Sociology, International relations.[3] In my analysis I will then indicate as "stereotypical" the choice of students who select in departments characterized by a high share of students of their own gender, while a "counter-stereotypical" choice is that of a student who selects in a department where they are the stereotypical minority, i.e. there is a low share of students of the same gender.[4]

I then assess how the choice of major impacts the effect of the gender composition of the environment on students' performance, by exploiting the exogenous allocation of students to classes conditional on scheduling constraints. Once students enroll in a program of study, they attend on average four courses in their first academic year. For each course, they attend classes (seminars), where they are divided into small groups of maximum 15-20 people solely based on their course selection and scheduling constraints. Thus, I can identify the causal effect of class gender composition by comparing the performance of the *same student* across the courses that they attend during their first year, exploiting the within-student exogenous variation in the share of same-gender classmates that they are exposed to after controlling for course and student fixed effects.[5]

Since students attend several courses outside their department of enrollment and first-year courses are large courses, students who are enrolled in the same program of study are not necessarily allocated to the same classes in all the courses they attend. So, men and women who are enrolled in the same program of study will not be exposed to the same within-student variation in class composition. Thus, I can estimate the effect of class gender composition on performance for students who made stereotypical and counter-stereotypical choices independently. Furthermore, this allows me to observe a variation in class composition for each student that goes beyond the

[3]This is in line with the distribution of men and women in academic fields across countries. While women have on average a higher probability to complete a bachelor's degree or equivalent in Education, Humanities, and Social Sciences (except Economics), they still have a significantly lower probability to enrol and graduate in math-intensive and science fields in most countries (Beede et al., 2011; McNally, 2020; OECD, 2020).

[4]As a proxy of stereotypical selection, I use the average share of men and women enrolled in undergraduate programs in each department over the period 2008-2017. The relationship between stereotypes and gender segregation of the workforce has been documented in psychology (e.g. Garg et al., 2018), and the share of females as an indicator of friendliness/stereotypes of the sector/field has been widely used in the literature (e.g. Bostwick and Weinberg, 2021; Kugler et al., 2021; Hebert, 2020). The results are robust to other proxies of stereotypical selection. See Section 1.5.2

[5]To provide evidence on the validity of the assumptions, I perform three checks. (i) I provide evidence that class allocation does not predict students' individual characteristics. (ii) I provide evidence that the variation in the share of females in the class is not related to the variation in a number of predetermined student characteristics. (iii) I compare the observed allocation to 1000 simulated unconstrained random allocation to provide evidence in support of the assumption that students of different gender do not select systematically different courses among the first-year optional courses. More details can be found in Section 1.4.

variation within the program of enrollment. Hence, I can exploit substantial common support in the variation of the share of same-gender classmates between students who made stereotypical choices and students who made counter-stereotypical choices of major.

I present three sets of results. First, I find that the students whose performance is negatively affected when their gender is outnumbered are those who chose to enroll in programs in line with stereotypes regarding gender skills, such as women in Humanities and men in Math and Science-intensive fields. On the other hand, students who selected against stereotypes do not appear to need to be surrounded by students of their own gender to perform well, but if anything perform better when outnumbered. Interestingly, the performance of students who are enrolled in departments with a balanced gender ratio is not affected by class composition. In terms of the magnitude of the effect, the paper estimates that for students who are enrolled in stereotypical departments, a 10% increase in the share of same-gender classmates *increases* course grades by 0.325 points (2.0% of a standard deviation). On the other hand, for students enrolled in departments stereotypically associated with their gender, a 10% increase in the share of same-gender classmates *decreases* final course grades by 0.288 points (1.76% of a standard deviation).

In line with the strand of literature that theorizes that individuals internalize stereotypes and the composition of the environment when making occupational and educational choices (e.g. Goldin, 2014; Oxoby, 2014; Pan, 2015; Del Carpio and Guadalupe, 2018; Kugler et al., 2021), these results suggest that these factors impact students who made different choices to a different extent once they are in the environment. Consistently with this, I find that the students who are enrolled in stereotypical fields that benefit the most from being surrounded by same-gender classmates are those that come from countries with low Gender Gap Index, i.e. countries with more unequal gender norms. At the same time, students coming from these countries seem to be those that display the strongest negative effect on performance when surrounded by same-gender classmates if they are enrolled in counter-stereotypical majors (in this case the average effects are not significantly different, but they become significant for men).

The estimated patterns do not appear to be gender-specific. The diversity of the environment (47% of students are Asian, while 37% are White) and the variety of programs offered by the institution allow me to replicate the analysis defining class composition and stereotypical selection along ethnic lines. The results confirm the findings for gender, suggesting that the estimated patterns are related to selection based on stereotypes moderating the effect of being under-represented rather than group-specific dynamics.

The second result concerns gender differences in the effect of stereotypical selection. While the effect is present for both men and women, it appears to be stronger for men. Indeed, while men who make stereotypical choices are affected by the composition of the environment regardless of the measure of stereotypical selection used, women display non-linearities in the effect. To capture non-linearities I define stereotypical and counter-stereotypical selection using a definition that becomes gradually more stringent.[6] Women whose performance suffers when in a numerical

[6] I start by considering students who are enrolled in the top five departments with the highest average share of same-gender students as students making stereotypical choices, while students who are enrolled in the five departments with the lowest share of same-gender students are considered to be students who made a choice not in line with stereotypes regarding gender roles and skills. I then replicate the same analysis by defining stereotypical and

minority are those who selected in departments characterized by the highest share of females among undergraduate students, while the effect disappears when less extreme choices are considered.

Lastly, the third set of results speaks to the margins through which selection into fields moderates the effect of being in a numerical minority. I rely on a theoretical framework to rationalize the channels traditionally used to explain how being in a numerical minority affects performance. I micro-found the model by exploiting evidence from a novel survey and rich administrative data. I assume that students invest in effort to maximize their performance in exams and their social image (Bursztyn et al., 2019; Bursztyn and Jensen, 2017; Bursztyn et al., 2017; Austen-Smith and Fryer, 2005; Akerlof and Kranton, 2000), with different weights depending on the extent to which effort is visible to their peers. In this setting, numerical minority affects performance via two channels. First, studies in psychology and sociology show that we prefer to affiliate with those who share our attitudes and beliefs or demographic traits (Inzlicht and Good, 2006). This is rationalized in the theoretical framework assuming that being in a numerical minority increases the marginal cost of effort for students by preventing them from benefiting from academic and emotional support.[7] Second, experiments in psychology and sociology show that being in a minority primes stereotyped aspects of individuals' identity in the mind of the outnumbered (Steele and Aronson, 1995; Spencer et al., 2016). I follow Bordalo et al. (2019) and I assume that students do not have perfect information regarding their ability and the ability of their peers. Their beliefs are affected by stereotypical distortions, which are stronger the more group identity is top of mind. By increasing the salience of group identity, being in a numerical minority increases the strength of stereotypical distortions, boosting (positive stereotypes) or depressing (negative stereotypes) beliefs on ability and, as a consequence, affecting expected returns to effort.

The model predicts that the relevance of these two channels is lower if stereotypical associations are weaker and the cost of interacting with the opposite gender is lower. This might be the case for students who made counter-stereotypical choices, since recent studies find that stereotypes and the composition of the environments are internalized in educational and occupational choices (e.g. Goldin, 2014; Oxoby, 2014; Pan, 2015; Del Carpio and Guadalupe, 2018; Kugler et al., 2021) and that these students self-selected in a field dominated by the opposite gender in spite of the negative stereotypes regarding their group.

I test this by exploiting teaching assistants' evaluations of students' participation in class during the term. In line with the findings of lab and field experiments that relaxing minority status increases participation and willingness to contribute to discussions for individuals in counter-stereotypical domains (Coffman and Shurchkov, 2021; Born et al., 2020; Chen and Houser, 2019; Bordalo et al., 2019; Coffman, 2014), the model predicts that, in absence of stereotypical selection, relaxing minority status induces students enrolled in counter-stereotypical fields to participate more in class.[8] However, consistently with selection playing a role, I find no evidence of an effect of class

counter-stereotypical choices by considering the top and bottom three and two departments in terms of share of same-gender students among undergraduates.

[7] According to the education literature, social interaction and mutual assistance are factors directly related to persistence in college (Tinto, 1975).

[8] According to the model, relaxing minority status reduces the strength of the stereotypical distortions that were inducing students to underestimate their absolute and relative ability. Furthermore, it reduces the marginal cost of effort by increasing the sense of comfort, belonging, and emotional reciprocity.

composition on participation for students who are enrolled in counter-stereotypical departments.

In particular, students who are enrolled in counter-stereotypical departments seem to be less affected by gender stereotypes. I find evidence that the performance of students who are enrolled in counter-stereotypical departments is not affected by the gender of the class teacher, contrary to the performance of students who made a stereotypical choice of major. Since the literature provides evidence that class teachers act as role models, helping breaking stereotypes (e.g. Breda et al., 2021; Porter and Serra, 2020; Olsson and Martiny, 2018), this result supports the theory that students who are part of the stereotypical minority group are indeed less subject to the effect of stereotypes. Furthermore, the results of an implicit association test (Greenwald et al., 1998) to elicit students' associations female-Humanistic and male-Scientific confirm that students who self-selected into a counter-stereotypical field in spite of the negative stereotype regarding their group indeed hold weaker implicit stereotypical associations.

These results have important policy implications. Taken together, they suggest that initiatives targeting minorities to address imbalances in performance are not going to be enough to level the playing field, especially in environments where the presence of individuals from the minority group is the result of a selective process or strategic choices. Targeting minorities in these environments might even backfire. The paper shows that the composition of the environment can explain differences in performance between groups. However, the effect is due to the (stereotypical) majority group benefiting from being over-represented, rather than the (stereotypical) minority group suffering from being under-represented. As a consequence, by enforcing a more balanced gender ratio, we could reduce imbalances in performance, but this would decrease average performance at the same time.[9] Furthermore, as the evidence points toward the stereotypical majority group upholding stronger stereotypical associations, initiatives targeting minorities might risk contributing to perpetuating stereotypes. By making groups' differences salient, they could reinforce stereotypical distortions in the mind of the majority group.

My work contributes to the literature that aims at assessing the effect of stereotypes on performance in stereotypically congruent and in-congruent domains and their interaction with the composition of the group (e.g. Coffman and Shurchkov, 2021; Karpowitz and Stoddard, 2020; Born et al., 2020; Bordalo et al., 2019; Chen and Houser, 2019; Bordalo et al., 2016; Coffman, 2014). This paper is the first to show that results in a natural field setting might differ from experimental evidence due to the fact that individuals who self-select into counter-stereotypical fields are less affected by the composition of the environment and stereotypes. Furthermore, by being able to replicate the analysis along ethnic lines within the same setting, the paper can also speak to the extent to which the results found are due to gender-specific dynamics, rather than under-representation more in general.

My paper relates also to fields experiments and observational studies that estimate the effect of the gender composition of the environment on performance in higher education (Shan, 2020; Zölitz and Feld, 2018; Griffith and Main, 2019; Booth et al., 2018; Huntington-Klein and Rose, 2018; Hill, 2017; Oosterbeek and van Ewijk, 2014; Giorgi et al., 2012). To the best of my knowledge, the evidence to date does not provide a conclusive picture, probably due to the different settings

[9] Evidence provided with a counterfactual scenario keeping students' characteristics constant

and fields studied, and the different sources of variation exploited. Both the methodology and the analysis of my paper build on this work. However, by studying the effect of the composition of the environment on performance across multiple fields within the same setting and exploiting the same source of variation, my paper can shed light on the importance of the field and the selectivity of the environment in explaining differences in results. Moreover, thanks to the rich administrative data and the survey, I can speak to the mechanisms underlying the results in an unusually rich way. The remainder of the chapter is organized as follows. The next section explains the institutional setting. Section 3 describes the data and descriptive statistics. Section 4 presents the empirical strategy and the tests in support of the identifying assumptions. Section 5 illustrates the main results, and Section 6 presents the theoretical framework and the evidence in support of the different mechanisms underlying the results. In Section 7, the analysis is replicated for ethnicity. Finally, Section 8 concludes.

1.2 Institutional Setting

LSE is characterized by a very diverse and multicultural environment: 49% of undergraduate students are women, and roughly 45% are Asian (Chinese, Indian, and other ethnic minorities).

The university offered a very diverse portfolio of courses during the academic years considered: 44 undergraduate degree programs across 16 academic departments. These programs span different disciplines from Mathematics, Economics, Finance, and Statistics, to Anthropology, Sociology, International Relations, and International History. The variation in gender composition across departments is significant. We can observe majority-male departments, where men represent >65% of the students enrolled in undergraduate programs, alongside majority-female departments, characterized by >75% of women (Figure 1.1).

The academic system is homogeneous across programs belonging to different disciplines as students take on average 4 courses during each academic year. Each degree program has its list of core courses – either compulsory or offering a very constrained choice – and a list of elective courses. During the first year, which is the focus of this paper, most of the courses are compulsory, and students have very limited possibilities to choose electives[10]. Students might be required (or allowed) to attend courses outside their department or program of enrolment, particularly so during the first year, when students attend several courses outside their own department[11].

For each course that students take, they attend lectures and *classes*. Lectures are followed by all the students enrolled in the course and are taught by a lecturer. Classes are taught by teaching assistants and students are divided in small groups of maximum 15-20 students. In total, students attend 10 classes each term, one per week (generally one hour long). The empirical strategy of the paper exploits students' assignment to these small groups and their composition.

Each academic year is composed of three terms. Michaelmas term (September- December) and Lent term (January-April) are 10 weeks long and are teaching terms during which students attend lectures. The third term is called Summer term (May - June). During this term students primarily

[10]Students are generally allowed to choose at most one course
[11]See Table 1.A3 in Appendix 2.10.1 for more detailed information

take exams.

1.2.1 Assignment of students to classes

Students are allocated to classes before the beginning of the academic year when each one receives a schedule for lectures and classes.

Class allocation can be considered as good as random conditional on scheduling constraints. According to the university's timetable office, only two criteria are used to allocate students to classes. First, students cannot attend more than 4 consecutive hours of teaching. Second, the timetable office considers potential clashes with other courses students attend during the academic year.

Students officially allocated to the same class represent the peers each student will engage with during classes for the whole year. Once students are assigned to a class, they can only change under exceptional circumstances and via an official request that has to be approved. Furthermore, the official allocation of students into classes reflects very closely actual attendance since classes are compulsory and teaching assistants are required to register students' participation in class.[12]

In the analysis, I am going to exploit this conditionally random class allocation and assess whether students' performance is affected by being in a numerical minority by defining treatment (composition of the environment) based on the characteristics of all the students (undergraduate, general course, intercollegiate, exchange, and postgraduate) officially allocated to the same class at the beginning of the academic year.

1.3 Data and Descriptive Statistics

The paper combines four sources of data: the university's administrative records, the class register, human resources' records, and additional information gathered through an online survey.

Administrative records

The university's administrative records contain information on academic history and performance in each course for the universe of students enrolled in undergraduate courses at LSE during the academic years 2008/09 - 2017/18.

The administrative records also contain some individual background characteristics for students, such as gender, age at enrolment, country of origin, ethnicity, term time accommodation, and the previous schools attended.

Lastly, the university collects and stores students' application information. I have the complete list of all the qualifications that students submit when filing their applications for all the students enrolled in undergraduate programs at the university between 2011 and 2017. Furthermore, for the academic years 2007/08 - 2019/202, the database contains information on the total number of

[12]Whenever students are absent without reason two consecutive times, they receive an "ammunition", and their advisor is contacted. When students fail to attend too many classes, they can be denied access to the final exam. Unofficial changes are possible, but this type of switching is usually limited to one or two sessions. It is difficult to obtain reliable numbers on unofficial switching. From my own experience and consultation with teaching staff, these instances are minimal, given that students have to notify their teaching assistant to be recorded as present when attending a different class

students who applied, got accepted, or were rejected to each program, divided by gender, ethnicity, and country of birth.

Class register

The class register contains information on lectures and classes allocation for each student, professor, and teaching assistant. I use this information to (i) identify all the students that attend classes in the same class group and construct a measure of class composition based on students' demographic characteristics; (ii) match students' class group allocation with teachers' class group allocation and construct a matched teacher-student database.

On top of this information, the class register contains data on attendance and performance in class as the university requires teaching assistants to keep track of students' attendance each week and evaluate their participation and overall class performance (problem set grades and overall class performance assessment). [13]

Human Resources records

The university stores background information (gender and ethnicity) for all the people employed by the institution. I was able to recover background information and match them with the class register for a subsample of teachers who were employed at the university with an official contract over the years considered.

Survey

Information from administrative records is complemented with data gathered from an online survey. The survey was administered to all the students enrolled in undergraduate programs at LSE during the academic year 2020/2021. The questionnaire was administered via email. Participants agreed to take part in the survey and signed an informed consent, in which it was explained that the survey was part of a research project aimed at understanding the determinants of academic and labour market performance for underrepresented and disadvantaged groups, and specifically, the role of minority status on academic performance. The students were informed that if they agreed to participate, they would be asked a few questions regarding demographics, their experience as a student, and their beliefs/attitudes regarding stereotypes related to gender and ethnicity. The time to complete the survey was around 10 minutes and students were entered in a draft for three amazon vouchers. 498 students completed the survey, representing 10% of the overall population. The survey includes questions on demographic characteristics, educational experience, social networks, explicit attitudes towards gender-specific skills, and an Implicit Association Test eliciting the association male-Scientific, female-Humanistic. More information on the survey can be found in Appendix 1.11.6.

[13] These data are stored to prove visa requirements and are used by academic mentors, professors, and teaching assistants to write reference letters for students.

1.3.1 Sample Selection

The full sample of students that are enrolled in undergraduate programs at LSE between 2008 and 2018 consists of 14,389 students. I restrict the sample for the analysis to undergraduate students enrolled in their first year that are attending courses for the first time.

The focus on first year courses is motivated by three factors. First, students' choices are strongly restricted in first year courses, allowing to minimize the problem of selection into courses. Second, during the first year students who attend different programs and departments have several courses in common. This provides a variation in the composition of the courses that each student attends during the first year that goes beyond the variation in the composition of the students enrolled in each program. Furthermore, it allows me to observe significant variation in the composition of classes within each course, given that first-year courses are large and each one is characterized by numerous students and numerous classes. Third, this is the first time that students meet each other. This is the perfect moment to study how behavior and performance are affected by the environment as students have an empty information set regarding each others. Thus, their beliefs on group specific performance will mostly rely on priors rather than the signals that they get from observing their classmates' behavior and performance in class and exams.[14]

A few programs are characterized by half-year courses that are scheduled for the second term. To limit the problem of endogenous choices of courses in the second term (they could be influenced by the class composition experienced in the first term), the sample is restricted to first-year courses that students attend during the first term, half-unit or full unit courses, excluding half-unit courses that students attend only in the second term (9.2% of observations).

The sample is restricted to class groups with a size between 6 and 28 students (excluding smallest and largest 0.5% of observations) to exclude unreasonable and unlikely class sizes.[15] Furthermore, I exclude from the sample all the students who changed class group during the first term, as in this case I cannot identify the initial exogenous class allocation. This happens 5.09% of times.[16] Lastly, I exclude class groups where I observe more than 50% of students changing group. These occurrences correspond to class group "restructuring" - for instance when a class group gets cancelled and students are re-shuffled to other class groups (2.5% of course-year-class group level observations).

Lastly, since the empirical strategy relies on class group fixed effects and student fixed effects, the sample is restricted to courses with at least two class groups and students for which I can observe, after the sample selection explained above, a test score in at least two courses: 54.603 course-year-class group level observations, corresponding to information on 14.313 students (99% of the original sample), who attend on average 3.8 courses in their first year.[17]

As we can see from Table 1.2, the sample is made of 512 courses, with an average course size of 138.44 students, split in 9.36 class groups of 13.43 students each.

[14] As a potential extension, second or third year data could be used to study belief updating after having observed the performance of female or male classmates during the first year, but at the moment this is outside the scope of this paper.

[15] The university regulations fix a cap to 17 students per class group. There are frequent exceptions, but class groups with more than 28 students seems very unlikely, and thus are excluded from the sample.

[16] In Appendix 1.11.1 I present the results of tests that show that the decision to change class group is not driven by the composition of the group.

[17] Excluding an interdisciplinary course that carries no credits and that is excluded from the analysis.

1.3.2 Main outcome variables

Overall course performance

Students are required to get 4 credits during each year of undergraduate programs by attending courses that are worth 0.5 or 1 credit. At the end of each course students get a grade between 0 and 100 that assesses what they learned during the course. This grade can be the result of a final exam, or the combination of final exam and essays or take-home assignments during the year. A student is deemed to have failed a course if his grade is below 40, a third class honor is awarded if the grade is between 40 and 49, a lower second class honor if the grade is between 50 and 59, an upper second class honor if the grade is between 60 and 69 and a first class honor if the grade is 70 or above. Students can also decide to drop-off from the course by not attending the final year exam or not submitting a part of the summative assessments. In this case, a grade of 0 is assigned by the university.

Final grades represent a good measure of performance for two reasons. First, every piece of assignment (exams, essays or take home assignments) is based on absolute grading. Grades are based on the performance of each student in the exam, without any ranking of students or curving of grades.[18]. Second, LSE is characterized by a blind marking system: students do not indicate their names in any type of assessment, but only a candidate number which is secret to examiners and only known to the student.[19] This allows me to disregard potential confounders such as discrimination.

First-year course grades are characterised by substantial variation. This is mainly due to the fact that not all first-year courses count for the final degree classification to the same extent. A student registered on a BA or BSc programme who has completed the first year of the programme and who has passed assessments in courses to the value of at least three credits will be eligible to progress to the second year.[20] The "year-one average" counts for 1/9 of the final classification grade and is calculated by adding together and averaging the best six out of eight grades in first year courses. All first-year one-unit credits will be counted twice, and any half unit credit is counted once to make a total of eight first year grades.

Table 1.1 shows some descriptive statistics regarding first-year grades. The average grade is 60.32, with a standard deviation of 16.35 points. Students do not seem to perform in the same way in all courses, as the within-student standard deviation is 8.61.

Class participation

In order to test the model predictions, I analyse data on teachers' evaluation of students' participation in class. Teaching assistants are required to assess students' participation in class at the end of each term. Participation grades are based on the contributions that each student gives in class during the ten weeks of the term. Each teaching assistant is asked: "please mark the student's overall participation in class during the term" and can give a score from 0 to 3, where 0 stands for

[18]For reference: https://www.lse.ac.uk/social-policy/Current-Students/BScProgrammesMarkingframe.pdf

[19]For reference: https://info.lse.ac.uk/Assessment-Toolkit/Marking-and-moderation, https://info.lse.ac.uk/current-students/challenging-results-and-appeals

[20]With the only exception of the Bachelor of Laws in which students progress to the second year if they have passed all four credits

"No contribution", 1 for *"Occasional contribution"*, 2 for *"Reasonable and alert interest shown"*, and 3 for *"Lively interest and frequent contributions"*.

Teachers' evaluations of students participation in class provide an indication of class dynamics and students interactions. However, since they are grades that teachers give to students, they are not objective measures of students' participation. Thus, results can be considered as indicative of class dynamics conditional on the assumption that teaching assistants do not discriminate or evaluate underrepresented groups differently when they are in a minority.

Even if assessment is in principle compulsory, not all the teaching assistants give a feedback to students. Attrition increases during the academic year: in Michaelmas term, participation is missing for 16.75% of course-year-class group level observations, while 40.68 % are missing in Lent term. For this reason, the analysis with participation grades as outcome variable is restricted to Michaelmas term. In Appendix 1.11.1, I provide evidence that missing participation information is due to teaching assistants providing no information for everybody in the class and do not depend on individual characteristics or the composition of the class group.

Given the sample restriction, participation grades will shed light on the dynamics that characterize classes in the first half of the academic year. While on the one hand this can represent a limitation since exam grades are the result of students' effort during the whole academic year, on the other hand first-term measures are more indicative of how students' behavior is affected by stereotypical distortions and priors, since this is the first time students meet and interact with each other.

The sample for the analysis on class dynamics consists of students for which I can observe a participation grade in at least two class groups: 44.771 course-year-class group level observations, corresponding to information on 13.424 students (93.8% of the sample). Table 1.1 shows that the average participation grade received by students is 2.04, with a between-student standard deviation of 0.85, and a within-student standard deviation of 0.58. Figure 1.3 shows a histogram of the within-student participation gap. I define the participation gap as the difference between student i's highest and lowest grades across all the seminars they attend during Michaelmas term of the first year. The variation is significant since the difference between the highest and the lowest participation grade for the median student is 1, and 30% of students experience a difference of at least 2 between the highest and the lowest participation grade.

1.3.3 Additional key variables

Ex-ante measure of ability

I construct a measure of ability at entry based on the information that students provide during the admission process. The university bases its admission decisions on personal statement, academic achievement, and references. Every program has minimum entry requirements, which are publicly available and clearly stated in the guidelines. They are based on A-level qualifications or equivalents. The A Level is a subject-based qualification conferred as part of the General Certificate of Education, as well as a school leaving qualification offered by the educational bodies in the United Kingdom and the educational authorities of British Crown dependencies to students completing secondary or pre-university education. Students typically study three A levels in different subjects, and the majority of universities set their entry requirements according to this measure. A levels are graded on a scale of A*/A/B/C/D/E. Each A-level grade is worth 30 QCA (Qualifications and Curriculum Authority) points.

Following Campbell et al. (2019), I construct a measure of individual ability at entry based on the best three exam results among the A-level qualification scores students declared when applying to LSE. Some students take courses that are equivalents to A-Levels.[21] In these cases I calculate their A-Level-equivalent scores based on the university conversion tables for foreign students.

Using this criterion, I am able to construct a qualification score for a total of 9.449 students, 90.57% of the students enrolled in undergraduate programs at LSE between academic years 2011/12 and 20117/18.[22] Figure 1.4 displays the resulting qualification score. Due to the high demand for places, the mean qualification score of students enrolled in undergraduate programs at LSE is high: 503.9 QCA points[23] with a standard deviation of 34.26[24].

Information on previously attended schools

The university administrative records contain information on the school students attended before enrolling at LSE. I match these data with the UK Government register of schools and colleges. This register contains information on schools and colleges in England, Wales, Scotland, Northern Ireland, or overseas establishments of UK institutions. I recover information on the characteristics of the previous school of enrolment for 63% of the students in my sample.

I use information such as whether the student attended a single-sex vs mixed-sex school, or an independent vs state school in the mechanisms and robustness sections. Mixed- vs single-sex school information is used as a proxy of the extent to which students are comfortable in engaging with other gender students, while whether students attended independent or state schools is used to characterise students' social background.

Global Gender Gap Index (World Economic Forum)

The university administrative records contain information on students' country of origin. I match this information with data on the World Economic Forum Overall Global Gender Gap Index for the country between 2006 and 2018.[25] The Global Gender Gap Index measures the level of gender equality for 130 countries around the world. To do this, it ranks countries according to calculated gender gaps between women and men in four key areas: health and survival probability, education attainment, economic participation and opportunities, and political empowerment and representation.

I am able to match the student's country of origin with the Global Gender Gap Index for the country for 92.44% of students in the sample. I use this information as a proxy for the strength of

[21] Appendix 1.11.2, displays the correlation of qualification scores constructed using different methodologies. The measures are strongly correlated with each others.

[22] I don't have admission information for students who enrolled in LSE before 2011

[23] This corresponds to a score in between a person that got two A-levels with A* and one A-level with A, and a person that has one A-level with A* and two A-levels with A.

[24] This corresponds to one A-level grade, i.e. 30 QCA points.

[25] Source: The World Bank data

stereotypes regarding gender skills and roles in the student's country of origin.

1.4 Empirical Strategy

The only two criteria that the timetable office uses to allocate students to classes for each course are scheduling constraints and limits on the number of consecutive hours of teaching. Thus, class allocation for each course can be considered as good as random, conditional on the combination of courses that students attend during the year. This exogenous class allocation allows me to identify the effect of being in a minority on students' performance. I leverage this identification strategy to assess how selection into majors interacts with the effect of being in a minority. To do so, I exploit two key features of the environment. First, LSE offers a very diverse portfolio of programs, which allows to define a measure of stereotypical selection. Second, class allocation is orthogonal to the department of enrollment, and students attend several courses outside their department. This ensures that substantial common support in class composition and within-student variation in class composition across programs of enrollment.

1.4.1 Causal effect of being in a numerical minority

I estimate the causal effect of being in a minority by comparing the performance of the same student i across the courses c that they attend in the same academic year a, where they have been assigned to classes g with exogenous peers' characteristics, net of course fixed effects. I estimate the following specification:

$$y_{iacg} = \alpha_{ac} + \alpha_i + \beta \times SLM_{iacg} + \epsilon_{iacg} \tag{1.1}$$

where the independent variable y_{iacg} is students' course grades. SLM_{iacg} is the *share of students like me*, i.e. the share of same gender classmates: the share of females for women and the share of males for men. β captures the effect of an increase in the share of same gender students in the class[26].

Course $\times$ year fixed effects (α_{ac}) are essential for obtaining exogenous variation in peer characteristics. First, several courses are common to different programs of enrolment, thus course fixed effects capture differences in overall course gender composition, which would lead to a systematically different probability of being exposed to same gender peers for the same student across courses. For example, John is a student enrolled in a BSc in Economics. John attends a Mathematics course, MA100, and an Economics course, EC100. Since the Economics course is compulsory also for students enrolled in BSc in Accounting, Government, Economic History, etc., EC100 will be attended by a higher number of women with respect to MA100. Thus, John will be more likely to have more women in class in economics rather than in mathematics. However, once we control for the overall course composition, the share of females in each class within the course is as good as random since class allocation is orthogonal to individual characteristics. Thus, the share of fe-

[26]The approach follows the empirical strategy of Anelli and Peri (2017), Feld and Zölitz (2017), and Brenøe and Zölitz (2020).

males that John is exposed to is exogenous once we control for course fixed effects. Second, course fixed effects control for all the characteristics that make courses different from each others and might affect the performance of students in the course. For instance, Statistics and Mathematics courses might be less conducive to interactions and discussions with respect to courses as History or International relations. Or else, courses have different types of assessment and different levels of difficulty. Third, course fixed effects control for other potential characteristics that might make students attending different courses different and that can be correlated with the overall course gender composition and might affect performance in the course.

Individual fixed effects α_i are also included in the specification, allowing to address multiple concerns. First, they capture students' course selection and scheduling constraints. This accounts for the fact that students who are enrolled in the same program and/or make the same constrained choices are more likely to be assigned to the same class, since they share similar schedules for lectures. Second, individual fixed effects control for the fact that students enrolled in different programs have a different probability of being exposed to same gender classmates due to the different gender composition of programs (and the combination of courses that are part of each program). Third, they control for observable and unobservable individual characteristics that are student-specific, common across courses and determine program or course selection, and as a consequence performance. For example, John is particularly interested in economics, passionate about history and struggles in more mathematical intensive disciplines. Hence, he chooses the BSc in Economics and attends EC100, MA100, ST102, and chooses EH101 as elective. Given his passions, he performs better in EH101 and EC100, where by chance he is allocated to female-majority classes, and he struggles in MA100 and ST102, where the classes he has been allocated to are by chance more male dominated. This example illustrates that without controlling for student fixed effects, we would be mistakenly attributing John's lower performance to class composition rather than person-specific skills and attitudes.

Lastly, by controlling for student fixed effects, I estimate the effect of class composition by exploiting a within-student variation in the share of same gender classmates across courses rather than a within-course variation in class composition. Hence, this estimation strategy does not rely on the assumption that classes are independent from each other, which would be crucial if the effect of class composition was estimated by comparing the performance of students enrolled in the same course who are allocated to classes with a different composition of classmates (within-course variation). Assuming class independence would be problematic in this setting as we observe the same student in different classes and there might be spillovers across courses. Students' performance in a course might not only be affected by their classmates in the course, but also by their experience in the classes of the other courses they attend. I discuss the extent to which spillovers affect the estimates in Section 1.5.2.

Lastly, discrimination, or differential treatment of minority students during the examination process, does not represent a concern in this setting, since grading of exams, essays, group projects, etc. is characterized by blind marking.[27]

[27] Any piece of assessment is marked by a first marker and a second marker, in most of the cases without seeing the first marker's grades/comments. Where there are any differences in mark, the two markers discuss and agree the final mark. The process is completely blind: students do not indicate their names in the exams, but only a

Under the assumption of no zero-sum game, I can identify the effect of being in a minority on the performance of men and women separately. The assumption of non-zero sum game does not seem unreasonable in the LSE setting as marking is based on absolute grading, so grades are based on the performance of each student in the exam, without any ranking of students.[28]

Hence, Specification 1.1 allows to identify the effect of being in a minority on students' performance β, exploiting the variation from the average share of same gender classmates that students experience in the courses they attend, where they have been assigned to classes g with exogenous classmates' characteristics. The underlying assumptions are that (i) conditional on attending a course, students are assigned to classes in a way that is orthogonal to their individual characteristics, conditional on scheduling constraints (student fixed effects). (ii) There are no factors that are correlated with class gender composition that differentially affect men and women. In Sections 1.4.4, I provide evidence towards the validity of these assumptions.

1.4.2 Stereotypical choices

I define *"stereotypical selection"* the choice of a student who is enrolled in a program in line with gender norms and stereotypes. Following the approach widely used in the literature of inferring female friendliness or stereotypes in the sector/field from the share of females (for example Hebert, 2020; Bostwick and Weinberg, 2021; Kugler et al., 2021), I use as a proxy for stereotypical selection the average share of women and men enrolled in undergraduate programs in each department over the period 2008-2017. A man's choice is considered in line with stereotypes and gender norms if he is enrolled in a program that belongs to majority male departments, e.g. Mathematics, Statistics, Economics, Finance. On the other hand, it is considered as against stereotypes and gender norms if he is enrolled in a department such as Anthropology and Sociology, which are majority female departments. In the same way, a woman's choice is considered as more in line with stereotypes the higher is the share of women in her department of enrollment between 2008-2017. A graphical description of the measure of stereotypical selection used can be found in Figure 1.9.

The average share of women enrolled in undergraduate programs in the department over the period 2008-2017 can be considered a good measure of the extent to which students' choices are in line (against) stereotypes and gender norms for four reasons: (i) the distribution of women and men across departments barely changed between 2008 and 2017 (Figures 1.A1); (ii) it closely mirrors the distribution of men and women across subjects in the UK Higher Education system (Figure 1.A2); (iii) it is the result of men and women applying to systematically different programs, rather than the School's selection process (Figure 1.A4) ; (iv) it reflects stereotypes regarding group-specific skills and roles. In line with lab and field experiments in the academic setting documenting a widespread belief that women are worse than men in mathematics and science while being better at reading (Carlana, 2019; Reuben et al., 2014; ellemers2018; Lane, 2012)), majority male departments are primarily characterized by Math-intensive programs, while majority female

candidate number which is secret to examiners and only known to the student.

[28] Regarding participation in class, grades are based on class interactions over a 10 weeks period. A zero-sum game assumption would require that each week 50 minutes class is fully saturated, i.e. students are racing to answer or ask questions, so that each student's intervention will take from other students' participation time. Although this is possible, from personal experience and consultations with teaching staff, this is quite unlikely, especially in first year courses.

departments are characterized by programs related to Humanities.[29] Figure 1.A5 in Appendix 2.10.1 confirms that these associations are indeed in the mind of LSE students. When asked regarding the distribution of men and women and the performance of women relative to men across departments, students of both gender confirmed that they are aware that some disciplines are primarily targeted by men while others by women. Furthermore, their beliefs regarding the relative performance of the two genders is indeed in line with the stereotype.

I exploit two definitions of stereotypical selection based on the average share of men and women enrolled in each department over the period 2008-2017, a continuous measure and a categorical measure. Exploiting the continuous measure, I estimate the following specification:

$$y_{iacg} = \alpha_{ac} + \alpha_i + \beta_1 \times SLM_{iacg} + \beta_2 \times SLM_{iacg} \times STS_i + \epsilon_{iacg} \tag{1.2}$$

where SLM_{iacg} is the *share of students like me*, i.e. the share of same gender classmates that student i experiences in class g, course c and academic year a. STS_i stands for *stereotypical selection*, measured as the average share of same gender students in student i's department of enrolment across academic years 2008-2017. β_2 provides an indication of the extent to which being surrounded by same gender classmates has a different effect on performance for students who made choices that are more in line with gender stereotypes compared to students who made a choice against stereotypes. A positive coefficient implies that the performance of students who are enrolled in departments that are stereotypically-congruent (women in Humanities and men in Mathematics) is more sensitive to the composition of the class compared to the performance of students who are enrolled in departments that are stereotypically associated with the opposite gender. A negative coefficient, on the other hand, implies that the students whose performance is more affected by being in a minority are those who are enrolled in departments stereotypically associated with the opposite gender (women in Mathematics and men in Humanities). The standard errors are clustered at the class level.

The continuous measure of stereotypical selection relies on the assumption that the effect of choices in line or against stereotypes have a linear effect. In order to relax this assumption, I perform a second analysis. I define a categorical measure of stereotypical selection, where I categorize choices using a definition that becomes gradually more stringent. I start by considering students who are enrolled in the top five departments with the highest average share of same gender students as students making stereotypical choices, while students who are enrolled in the five departments with the lowest share of same gender students as students who made a choice not in line with gender stereotypes. I then replicate the same analysis by considering the top and bottom 2 and 3 departments in terms of share of same gender students.[30] I estimate the following specification:

$$y_{iacg} = \alpha_{ac} + \alpha_i + \beta_c \times SLM_{iacg} + \beta_n \times SLM_{iacg} \times NS_i + \beta_s \times SLM_{iacg} \times SS_i + \epsilon_{iacg} \tag{1.3}$$

[29]The relationship between stereotypes and race and gender segregation in the workforce has been documented in psychology(e.g. He et al., 2019; Garg et al., 2018).

[30]When I consider top and bottom 5 departments, I am considering the top and bottom tercile of departments given that the total number of departments in 16.

where the *share of students like me* (SLM_{iacg}) is the share of same gender classmates that student i experiences in class g, course c and academic year a. *stereotypical selection* (SS_i) and *neutral selection* (NS_i) are a dummy equal to one if the student is enrolled in the top 2,3,and 5 departments with the highest share of same gender students among undergraduates, and a dummy equal to one for students who are not enrolled in neither top 2,3, and 5 nor bottom 2,3, and 5 departments respectively. This way, β_c provides us with an estimate of the effect of a change in the share of same gender classmates for students enrolled in counter-stereotypical departments, while β_n and β_s provide us with a measure of the extent to which the effect of a change in gender composition of classmates is different if a student is enrolled in a "neutral" or stereotypical department. By estimating the effect with a progressively more restrictive measure of stereotypical and counter-stereotypical choices, this measure allows me to capture potential non-linearities in the effect. The standard errors are clustered at the class level.

1.4.3 Variation in class composition

Figure 1.5 and Table 1.3 present descriptive evidence on the variation in the share of the same gender classmates. Figure 1.5 shows the variation in the share of same gender classmates and the within-student variation in the share of same gender classmates, which is obtained by taking the difference between the highest and lowest share of same gender classmates for each student.

The average share of same gender students in the classroom is 0.561, with a standard deviation of 0.162. Table 1.3 shows that the within student standard deviation in the share of same gender classmates is 0.112. Hence, controlling for course and students fixed effects, the residual variation represents 68.75% of the initial raw between-student variation.[31]

The impact of selection into majors on the effect of being in a minority on performance can be estimated thanks to the fact that the allocation of students to classes is independent on the program of enrolment. As a matter of fact, once students enroll into a program, they attend the courses that are scheduled for their first year, and in each of them, they are assigned to classes considering only scheduling constraints. Furthermore, during the first year, students attend common courses, often outside their department of enrolment. This implies that they can be assigned to classes with students that do not belong to their program. As it can be seen from Table 1.2, the average share of same program peers is equal to 0.49. Lastly, given that first-year courses are large and the allocation of students to classes is as good as random conditional on scheduling constraints, students who are in the same class in a course are not necessarily allocated to the same class in other courses. Thus, I can observe a variation in class composition for each student that goes beyond the variation within the program of enrollment. This allows me to exploit a substantial common support in class composition and within student variation in class composition across programs of enrollment.

Figure 1.10 and Figure 1.11 display the within-student variation in the share of same gender classmates and the share of same gender classmates for students enrolled in stereotypical and counter-stereotypical departments. The within student variation in the share of same gender

[31] Following Olivetti et al. (2020)

classmates is obtained by taking the difference between student i's highest and lowest share of same gender students across all the classes attended during the first year. We can see that both the within-student variation in the share of same gender classmates and the share of same gender classmates for students enrolled in different types of department share significant common support.

1.4.4 Validity of the identification strategy

Specification 1.1 allows me to identify the effect of being in a numerical minority on students' performance under two assumptions: (i) students are assigned to classes in a way that is orthogonal to their individual characteristics, conditional on scheduling constraints (student fixed effects); (ii) there are no factors that are correlated with class gender composition that differentially affect men and women.

I perform three checks to test that students are not systematically assigned to particular classes. First, I test that class group allocation does not predict students' individual characteristics. Second, I test that being a woman does not predict the share of same gender peers in class. Third, I perform a series of balance checks to provide evidence that the variation in the share of same gender classmates is not related to the variation in predetermined student characteristics.

In addition, I test that selection into courses does not systematically differ for men and women by performing a permutation test where I simulate 1000 unconstrained random allocations and test that the distribution of the variation in the observed share of female students across classes is not statistically different from the distribution obtained with the simulated allocation.

Test 1: Class group allocation does not predict students' characteristics

Following Feld and Zölitz (2017) and Braga et al. (2016), I test that class group allocation does not predict students' characteristics to make sure that students with different characteristics are not systematically assigned to particular classes. I follow the specification below:

$$y_{ig} = \sum_{g=1}^{n_g} \alpha_g \times G_{i,g} + \sum_{p=1}^{n_a} \gamma_p \times PC_{i,p} + \epsilon_{ig},\ \forall a, c \tag{1.4}$$

where the dependent variables are pre-determined individual characteristics of student i enrolled in course c in year t allocated to class g: gender, ethnicity, age at enrollment. The explanatory variables are dummies for each class group g in course c in academic year a. The dummy for class group g is equal to one if student i is assigned to class group g and zero otherwise ($G_{gi} = 1(i\text{'s } group{=}g)$). The allocation is constrained by the fact that students that attend more than one course in the same term cannot attend two classes at the same time, so I control for a dummy for each course that each student takes during the academic year ($\sum_{p=1}^{n_a} PC_{i,p}$). I run one regression for each combination of course $c \times$ academic year a.[32] The sample for each regression consists in all the students enrolled in course c in the same academic year a that didn't change class group during Michaelmas term.[33] Furthermore, the sample is restricted to all the courses that have at

[32] All first year courses plus 15 second year courses, among which 7 are language courses. Students can choose second year courses as elective courses. These correspond to 0.57% of the observations

[33] I am excluding students that changed class group since I can't observe their initial allocation. Tests of randomness of the decision to change class can be found in the appendix.

least 2 class groups.

I test that class dummies are jointly significantly different from zero:

$$H_0 : \alpha_g = 0, \, \forall g = \{1; n_c\}$$

i.e. students who take the same combination of courses are allocated to classes independently from their individual characteristics.

According to Murdoch et al. (2008), the p-values of this test should be uniformly distributed with mean 0.5 if class allocation is random. In this case, courses have different sizes, the number of classes is not fixed, and class belonging to the same course might be of different size. In order to check that also in this case the p-values converge to a uniform distribution with mean 0.5, I Perform a Monte-Carlo simulation where I randomly allocate the students enrolled in first-year courses that didn't change class group during the term to class groups 1000 times, under the assumptions that course size, number of seminars and class size is equal to the observed one. Figure 1.A6 shows the result of the simulation. Indeed p-values converge to a uniform distribution with mean 0.5. This represents the reference point against which I will check under which conditions the observed allocation of students can be considered as good as random.

Figure 1.6 shows the p-values obtained from the tests of joint significance of the class dummies for the observed sample. Results are consistent with conditional random allocation. For the regressions regarding gender and age at entry, less or equal to 5% of tests display a p-value smaller than 0.05, and less than 10% of tests display a p-value smaller than 0.10, in line with what we would obtain if students were randomly allocated to classes. Regarding ethnicity, slightly more than 5% of tests display a p-value smaller than 0.05, but less than 10% of tests display a p-value smaller than 0.10 when we test whether class composition can predict the probability of being white. Regarding being Asian, slightly less than 5% of tests display a p-value smaller than 0.05, and less than 10% of tests display a p-value smaller than 0.10.[34] More details on the randomization tests and the simulation performed can be found in appendix.

Test 2: Being a woman does not predict the share of same gender classmates

Following Feld and Zölitz (2017), Brenøe and Zölitz (2020), and Guryan et al. (2009), I test that female students are not systematically assigned to classes where there are more same gender peers, conditional on having a certain share of same gender peers among the students enrolled in a course. The specification used is the following:

$$SGP_{-i,acg} = \alpha_{ca} + \beta \times F_i + \gamma \times SGP_{-i,ca} + \sum_{a=1}^{n_a} \sum_{p=1}^{n_{p,a}} \delta_{p,a} \times PC_{i,ap} + \epsilon_{iacg} \tag{1.5}$$

[34] I aggregated Chinese, Indian and Other Asians in a unique category called Asian in order to have a big enough sample to be able to perform a meaningful test. We need a big enough sample of students that share characteristic t to be able to test if class allocation is as good as random. As a matter of fact, if the number of students is too small, even if allocated randomly, class allocation might still predict students' characteristics. Let me consider the extreme case in which there is only one Chinese student enrolled in the course, class assignment will predict students' ethnicity even if the Chinese student is allocated randomly to the class.

For each course c in academic year a, I test if being a woman (F_i) predicts the gender composition of peers in the class g (SGP_{iacg}). I control for the share of same gender peers enrolled in the course ($SGP_{-i,ca}$) to clean for the fact that if there are more women than men enrolled in the course, the probability of being in a class group with a student of the same gender will be higher for female students than for male students. Lastly, I control for a series of dummies that are equal to one for all the students that attend the course in the same year, and zero otherwise ($\sum_{a=1}^{n_a} \sum_{p=1}^{n_{p,a}} PC_{i,ap}$). This is because the allocation is constrained by the fact that students attend more than one course in the same term and they cannot attend two classes at the same time. I perform this test for all the courses where there are at least two classes, and standard errors are clustered at the class level.

Table 1.4 reports the result of the above regression. Students are not systematically assigned to class group based on their sex. As a matter of fact, women do not have a higher probability to be in class with same gender peers than men.

Test3 : Balance checks

The tests presented above provide evidence that students are as good as randomly allocated to class groups, conditional on their course choices. In Table 1.5, I produce an array of "balancing tests" to assess whether the variation in the share of same gender students in the class a student is allocated to is related to the variation in a number of predetermined student characteristics. I do so by testing that the share of females in the class is not systematically correlated with ethnicity, previous school characteristics, age at entry, and qualification score at entry. As shown in the table, only one of the estimated correlations appears to be significantly different from zero at 10% level of significance for the sample of analysis. This represents 7% of the tests performed. As expected when running a large number of regressions testing multiple hypotheses, some coefficients are statistically significant. In the absence of a systematic relation between the share of female classmates and other individual characteristics, we would expect 10% of coefficients to be statistically significant at the 10% significance level, 5% at the 5% level, and 1% at the 1% level simply as a result of chance. This is consistent with my results, providing supportive evidence that the estimated effects are due to a change in class group composition along gender lines, rather than other unobserved factors that are correlated with gender and student outcomes and that could confound our estimates.

Test 4: Exogeneity of scheduling constraints

Following Lavy and Schlosser (2011), I perform a permutation test. I Perform a Monte-Carlo simulation where I randomly allocate the students enrolled in first-year courses that didn't change class during the term to classes 1000 times, under the assumptions that course size, number of seminars, and class size is equal to the observed one. This is an unconstrained random allocation, that does not take into account the constraints deriving from the fact that students cannot be allocated to certain classes due to clashes with other courses they attend during the same academic year. As such, this represents a test for the assumption that students of different gender do not select systematically different courses among the first-year optional courses.

For each course c in academic year a, I compute the within-course variation in the share of female students (WCV_{gca}) across classes for both the observed sample allocation and the 1000 simulated random allocations:

$$WCV_{gca} = Class\ share\ of\ female\ students_{gca} - Course\ share\ of\ female\ students_{ca},\ \forall g, c, a, t$$

I then perform a Two-sample Kolmogorov-Smirnov test to test the null hypothesis that the observed distribution of the statistics ($Diff_o$) and the distribution of the statistics for the randomly allocated simulated samples ($Diff_s$) are not significantly different:

$$Ho : Diff_o = Diff_s$$

The results of the test are shown in Figure 1.7. We fail to reject the null hypothesis that the distribution of the observed statistics and the distribution of the simulated statistics are different at every significance level (1, 5, and 10%).

Alternatively, I perform a Two-sample Kolmogorov-Smirnov test to test the null hypothesis that the observed distribution of the within-course standard deviation in the share of females and the randomly simulated distribution are not significantly different. The results can be found in Figure 1.8. They confirm that the observed within-course variation in the proportion of females across classes resembles the variation that would result if the gender composition of each class was randomly generated.

1.5 Results

1.5.1 Effect of stereotypical selection

Table 1.6 displays the results of Specification 1.2 in (Column 1) and Specification 1.3 in Columns (2)-(4). The coefficient of *Share of students like me* is negative and significant, while the coefficient of the interaction between the share of same gender students and the measure of conformism to gender roles and stereotypes is positive and significant in all specifications. This indicates that selection against stereotypes moderates the importance of being in a minority in shaping students' performance. The students whose performance is negatively affected by being in a minority are students who chose to enrol in programs in line with the stereotypes regarding gender skills and roles, i.e. women in Humanities and men in Math and Science-intensive fields. On the other hand, students who selected against stereotypes do not suffer from being in a minority, but if anything perform better when there are few students similar to them in the class (this can be seen from the negative and significant coefficient for *Share of students like me*). Interestingly, the performance of students who are enrolled in departments with a balanced gender ratio is not affected by class composition. This hints towards the fact that social identity considerations and preferences for the composition of the environment affect students way before they sit in the classroom. If these factors are internalized in decisions regarding the field of occupation, they impact students enrolled in different departments to a different extent.

In line with the estimated effect using the continuous definition, the effect of an increase in the share of students like me decreases when we use a less strict definition of stereotypical and counter-stereotypical choices.

In terms of magnitude of the effect, the paper estimates that for students who are enrolled in stereotypical departments, such as a man who is enrolled in Economics, Finance or Mathematics, and a woman enrolled in Anthropology, Sociology or International Relations, a 10% increase in the share of same gender classmates increases course grades by 0.325 points, equivalent to 2.0% of a standard deviation (Column 3). On the other hand, for a man enrolled in Anthropology, Sociology or International Relations and a woman enrolled in Mathematics, Economics and Finance, a 10% increase in the share of same gender classmates decreases final course grades by 0.288 points, i.e. 1.76% of a standard deviation (Column 3).

Table 1.7 provides evidence of the effect at different points of the grade distribution. I define a series of dummies that are equal to one if grades are greater or equal to 40, 50, 60, and 70, respectively. These represent important thresholds for students. A student is deemed to have failed a course if his grade is below 40. The student has achieved a third class honor if the grade is between 40 and 49, a lower second class honor if the grade is between 50 and 59, an upper second class honor if the grade is between 60 and 69 and a first class honor if the grade is 70 or above. Lastly, I define a dummy equal to one if students drop-out from the course and zero otherwise. Being in a minority does not affect the probability of dropping out from the course. The effect on average grades is driven by being in a minority affecting performance in the course, with the effect increasing at the top of the grades distribution.

1.5.2 Is this due to stereotypical selection?

In the previous section, I provided evidence that students who are enrolled in stereotypical fields are negatively affected by being in a minority in the class, while this is not the case for other students. In particular, students who are enrolled in counter-stereotypical departments appear to benefit from being surrounded by students with different characteristics (in some specifications). In this subsection, I provide evidence that these patterns are due to the effect of selection into fields, rather than other potential confounders. In order to do so, I start by providing evidence that the results are not driven by differences in support in the share of same gender classmates and within-student variation in the share of same gender classmates. Second, I show that the results are not driven by differences in teaching assistants or peers' characteristics for students enrolled in different departments. Third, I test whether the results could be explained by spillovers across courses or mechanical effects. Lastly, I provide evidence that the heterogeneity in effect for students enrolled in different departments is related to patterns of selections linked to stereotypes regarding gender skills and norms.

Differences in support of share of same gender classmates

In Section 1.4.2, I provided evidence of significant common support in the share of same gender classmates and within student variation in the share of same gender classmates across students enrolled in different departments. Despite the significant common support, the share of same gender classmates variation is different for students enrolled in different departments, and in particular it is higher for students who are enrolled in stereotypical departments. This could be able to explain the different effects of class composition estimated for students enrolled in different departments if the effect of being in a minority is non linear. Figure 1.A7 shows that there is no evidence of non-monotonicity in the effect of a change in the share of same gender classmates. The estimates of a local polynomial regression are plotted for each value of the share of same gender classmates[35]. This provides evidence in support to the fact that the results cannot be explained by potential differences in the effect of being in a minority across departments due to students being exposed to a different variation in the share of same gender classmates or a different overall share of same gender students.

Difference in teachers and peers' characteristics across departments

Students enrolled in different departments might be exposed to different teachers or peers with different characteristics. In particular, as there are more female students in majority-female departments, there might be more female teachers in these departments. Furthermore, academic regulations are different across programs: the number of compulsory courses are different and the same is true for courses that students can attend outside the program or department of enrollment. Furthermore, math-intensive departments might attract students with different characteristics with respect to disciplines that are related to Humanities. In order to explore if this might be affecting my results, I estimate Specifications 1.2 and 1.3 controlling for teaching assistant fixed effects, share of same ethnicity and same background students (defined as students coming from State funded or Independent schools), share of same program students, and average peers ability, defined based on qualification scores at entry. The results are robust to the introduction of these controls.

Spillovers across courses and mechanical effects

The estimated effects could be overestimated in the presence of spillovers across courses. For instance, being in a minority in a course might induce students to perform worse in the course, but also to perform better in other courses because there are more students of their gender in the class, despite the share being low. This could rationalize the stronger effect for students enrolled in stereotypical departments considering that the within-student variation in the share of same gender classmates for these students is greater than the within-student variation for students enrolled in counter-stereotypical departments.

In order to test if this is the case, I re-estimate the effects by comparing the performance of students who are enrolled in the same course, but have been as good as randomly assigned to classes with a different share of same gender classmates (within-course variation). This identification strategy exploits the exogenous allocation of students across classes within the same course conditional on scheduling constraints. Under the assumption that assignment to classes is independent across courses conditional on scheduling constraints (students who are assigned to a class with a low share of same gender classmates in a course are not systematically assigned to classes

[35]The Figure displays residual course grades and share of same gender classmates obtained by regressing them on course $\times$ year fixed effects and student fixed effects as in Specification 1.1

with a higher share of same gender students in other courses), the effect estimated by exploiting the within-course variation represents a counterfactual to assess the extent to which spillovers across courses induce to overestimate the effect when exploiting the within-student variation.

However, spillovers across courses could also induce us to overestimate the effects when exploiting the within-course variation. For instance, if being in a minority in the Economics course has such a demotivating effect that students perform worse not only in Economics, but also in the other courses they attend, not controlling for student fixed effects would lead to overestimate the effect.

Table 1.A2 displays the results of the two specifications. Column (1) displays the result of Specification 1.2, where I exploit the within-student variation, while Columns (3) and (4) display the results obtained by exploiting the within-course variation. Column (4) exploits one observation per student, which controls for the fact that observations for the same student in different courses are not independent. We can see that the estimates exploiting the within student-variation are smaller in magnitude with respect to the estimated effects obtained by exploiting the within-course variation, confirming that spillover effects are not leading to overestimate the impact of stereotypical selection.

Lastly, Column (2) displays the result of Specification 1.2 without controlling for course $\times$ year fixed effects, confirming that the results are not driven by mechanical effects arising because, controlling for course fixed effects forces the average grades in the course to be equal to 0.

Evidence on stereotypical selection

To support the hypothesis that the results are due to *stereotypical* selection, and in particular whether students made a stereotypical or counter-stereotypical selection when choosing their major, I provide three pieces of evidence.

First, I perform a placebo test by estimating specifications 1.2 and 1.3 controlling for the interaction between the measure of stereotypical selection and the share of same ethnicity classmates, share of same program classmates, share of same background classmates, and peers' average ability at entry. Table 1.A3 shows that the results are robust to the introduction of these controls. Furthermore, none of the coefficients of the interaction between the measure of stereotypical selection and other peers' characteristics are significant, indicating that the coefficient of stereotypical selection is capturing the heterogeneity in the effect of being in a minority caused by students who selected into different fields having different characteristics along gender specific dimensions.

Second, Table 1.A4 shows that the results are robust to other proxies of stereotypical selection related to gender stereotypes: the share of women and men among applicants to undergraduate programs at LSE, the share of men and women enrolled in different subjects in higher education in UK, and the share of men and women among the staff working in each subject in higher education in UK[36].

Lastly, Table 1.A6 provides suggestive evidence that stereotypes play a role. I exploit the Global Gender Gap Index (GGI) of the country of origin of the student as a proxy for the strengths of

[36] details regarding the measures used can be found in Table 1.A5.

gender roles and norms for the student.[37] Table 1.A6 displays the estimates of Specification 1.1 where I interact the share of same gender classmates with the GGI of each student's country of origin. I split students in three categories based on the terciles of the GGI of students' countries of origin. A higher tercile implies a higher GGI, indicating that the student comes from a country with more equal gender norms. I estimate this specification for the sample of students who are enrolled in counter-stereotypical fields (Column 1) and for the sample of students who are enrolled in stereotypical departments (Column 2)[38]. Focusing on students who are enrolled in stereotypical departments, those who benefit the most from being surrounded by same gender classmates are students who come from countries with low GGI, i.e. students who come from countries with more unequal gender roles. However, students coming from countries with low GGI seem to be those who benefit the most from being in a minority if they made a counter-stereotypical selection, even though the effects are not significant. The effect is driven by men, for whom also the effects for students who made counter-stereotypical choices become sizable [39] This provides suggestive evidence that selection based on stereotypes regarding gender roles and norms plays a role in moderating the effect of the gender composition of the classmates on performance.

1.5.3 Gender differences in the effect of stereotypical selection

Is the effect of stereotypical selection different for men and women? In order to give an answer to this question I estimate Specifications 1.2 and 1.3 interacting the share of students like me and the share of students like me $\times$ stereotypical selection with a dummy equal to one if the student is a woman and a dummy equal to one if the student is a man. Table 1.8 displays the results.

The coefficient of the interaction between the share of same gender classmates and stereotypical selection in Column (1) is positive and significant for men, indicating that those men who made a choice of major in line with stereotypes benefit more from having classmates of their own gender compared to men who made choices against stereotypes. The effect for women is smaller in magnitude and not statistically significant. However, Columns (2)-(4) show that when we define the measure of stereotypical selection to include only departments that are very male and female dominated (bottom and top 2 and 3), the coefficients become significant also for women. Furthermore, the coefficient is not statistically significantly different from the effect estimated for men using the same definition of stereotypical selection. On the other hand, the effects estimated for men and women become statistically significantly different when we consider the definition of stereotypical selection that includes bottom and top five departments. This indicates that both men and women are affected by the composition of the environment, but men's performance is affected more with respect to women's performance since the effect is significant and strong also for students who

[37]Guiso et al. (2008) provides evidence that the Country's GGI index is significantly correlated with gender gaps in math and reading. In particular, countries with lower GGI have bigger math gender gaps in favor to men, while countries with high GGI have bigger reading gender gaps in favor to women.

[38]I am using here the top and bottom 5 departments definition of stereotypical selection. This is the most conservative definition of stereotypical and counter-stereotypical, which also allows me to have a big enough sample size to estimate significant effects.

[39]The stronger effect for men is consistent with the evidence in the mechanisms section that the effect for men making stereotypical choices can be reconciled with stereotypes playing a role. On the other hand, through the lenses of the model, the effect for women can be rationalized with relaxing numerical minority decreasing the cost of engaging in out-group interactions. Furthermore, if we look at application patters (Section 6.3), the observed gender segregation across departments seems to be mostly due to men rather than women. More men apply to departments that end up being male dominated, while we don't see such a strong pattern for women.

didn't enroll in extreme departments in terms of gender composition.

Columns (2)-(4) also provide a quantifiable estimate for the effect of being in a minority on performance for men and women who are enrolled in female and male fields. Let us consider Column 3, for example. For men who are enrolled in male fields (stereotypical selection), a 10% increase in the share of men increases course grades by 0.376 points, i.e. 2.3% of a standard deviation in course grades. On the other hand, women who are enrolled in male fields (counter-stereotypical selection), do not suffer from being surrounded by men. As a matter of fact, a 10% increase in the share of women (equivalent to a 10% increase in the share of men) decreases course grades by 0.234 points, i.e. 1.43% of a standard deviation in course grades. In a mirroring way, for women who are enrolled in female fields (stereotypical selection), a 10% increase in the share of women increases course grades by 0.217 points, i.e. 1.33% of a standard deviation in course grades. On the other hand, men who are enrolled in female fields (counter-stereotypical selection) do not suffer from being surrounded by women, as a 10% increase in the share of men (equivalent to a 10% increase in the share of women) decreases course grades by 0.435 points, i.e. 2.67% of a standard deviation in course grades.

This implies that the composition of the environment can explain the performance gaps we observe across fields, but the effect is due to the (stereotypical) majority group benefiting from being over-represented, rather than the (stereotypical) minority group suffering from being under-represented. To illustrate this, we can simulate what would happen to students' performance in the counterfactual scenario of a more balanced gender ratio in male fields. In male fields[40], the average share of women enrolled in undergraduate programs is 35.6% and the gender gap in performance in first-year courses is negative: male students perform better than female students enrolled in the same course by 2.43 points on average. The estimates indicate that if we increased the share of women in the class by 10%, holding the characteristics of students enrolled in these departments constant, the gender gap would decrease by 5.9%.[41] Hence, a more balanced gender ratio would equalize performance across genders. However, this effect would be the combination of two negative effects: a negative effect for men (equivalent to 15.5% of the gap), and a negative effect for women enrolled in these fields (equivalent to 9.6% of the gap). As the only students who benefit from being surrounded by same gender peers are students who made choices of major in line with stereotypes and belong to the majority group (in this case men), a more balanced gender ratio in male-fields would reduce differences in performance across groups. However, at the same time, it would lower average performance, as neither men nor women benefit from a higher share of women.

1.6 Beyond gender-specific dynamics

Are these patterns specific to gender or do they concern being in a minority more in general? In order to provide an answer to this question, I replicate the analysis along ethnic lines. The

[40] I am considering as reference the definition used above that refers to Column 3, the 3 departments with the highest share of men.

[41] This is obtained by summing the effect for men (0.376/2.43=15.5% of the gap) and women (0.234/2.43=9.6% of the gap). If we re-weight the effect by the share of men and women in the field we obtain a reduction of 4.03% of the gap (0.376/2.43*0.544=8.42% of the gap for men, while 0.234/2.43*0.456=4.39% of the gap for women)

institution is characterized by an exceptionally diverse environment: only 37% of students are White, while 47% of the population of undergraduates are are Asian (primarily Chinese and South Asian). Furthermore, also ethnic groups are distributed unequally across departments. The pattern along ethnic lines is exactly the opposite with respect to the distribution of traits that characterizes departments along gender lines. The fields where ethnic minorities (in this case Asian students) represent the majority of enrolled students are Math and Science-intensive fields, which are also the fields where women are under-represented. As for gender, this reflects stereotypes regarding group-specific skills and roles. As a matter of fact, lab and field experiments in the academic setting document a widespread belief that Asians are better at math and science with respect to white students (see for example the review by Spencer et al., 2016).

In this section I am going to study if stereotypical selection along ethnic lines affects the effect of classmates' ethnic composition on students' course performance. I am going to estimate the following specification:

$$y_{iacg} \;=\; \alpha_{ac} + \alpha_i + \sum_{e=1}^{3} E_{i,e} \times [\beta_{1,e} \times SLM_{iacg} + \beta_{2,e} \times SLM_{iacg} \times STS_i] + \epsilon_{iacg} \quad (1.6)$$

where students are divided in three groups $E_{i,e} = 1$(i's ethnic group = e), Asian, White, and Other students, which is a residual category that includes other ethnicity students and students who provided no information on their ethnicity[42]. The *share of students like me* (STS_{iacg}) is the share of same ethnicity classmates that student i experiences in class g, course c and academic year a, and *stereotypical selection* (STS_i) is the average share of same ethnicity students in student i's department of enrolment across academic years 2008-2017. Figure 1.19 provides a graphical description of stereotypical selection along ethnic lines. The average share of same ethnicity students in the department of enrolment for the residual group is calculated as one minus the share of Asian and White students enrolled in each department[43]. $\beta_{2,e}$ provides an indication of the extent to which being surrounded by same ethnicity classmates has a different effect on performance for students who made choices that are more in line with stereotypes regarding ethnic specific skills and norms with respect to students who made a choice against stereotypes, for each ethnic group e. The standard errors are clustered at the class level.

This exercise is based on the assumption that students are not systematically allocated to different classes because they belong to particular ethnic groups. I provide evidence for the validity of this assumption by replicating for ethnicity the same tests performed for gender. Evidence and a detailed explanation on the results can be found in Appendix 1.11.5.

Table 1.11 displays the results of this exercise. Stereotypical selection moderates the effect of being in a minority also along ethnic lines. Both White and Asian students benefit from being surrounded by more classmates of their ethnicity when they made a stereotypical selection of

[42]This categorization is motivated by the necessity of having large enough groups to perform the analysis. Table 1.A19 in Appendix 1.11.5 provides evidence that pulling students coming from different Asian countries together is not such a strong assumption given that the performance of Asian students coming from different countries is affected by an increase in the share of classmates coming from other Asian countries in the same qualitative way as an increase in the share of Asian classmates who were born in the same subset of Asian countries (dividing students in Chinese, India, Other Asian).

[43]The results are robust to restricting the sample to only Asian and White students (Table 1.A7 for evidence)

major. This is robust to controlling for the share of same gender classmates (Column 2). These results confirm that the estimated patterns concern the interaction between being in a minority and selection into fields, and are not only related to gender-specific dynamics.

1.7 Mechanisms

How does stereotypical selection intervene to moderate the effect of being in a minority on performance?

According to the evidence of lab and field experiments in psychology, economics and sociology, we can expect being in a minority to affect students' performance via two main channels. First, experiments in psychology show that people prefer to affiliate with those who share their attitudes and beliefs or demographic traits, since we have positive affective responses for those who are similar to us, and we also expect increased comfort and trust when interacting with them (Inzlicht and Good, 2006). So, minority status could affect performance by reducing opportunities for social interactions and mutual academic assistance, which are factors directly related to persistence in college (Tinto, 1975).

The second channel through which we can expect minority status to impact students' performance is by affecting expectations regarding returns and confidence about personal achievement through its effect on beliefs about self and others' ability. Previous work in psychology, sociology and economics suggests that differences in performance can be partially explained by biased beliefs regarding own-self and others' ability caused by stereotypes (e.g. Bordalo et al., 2019; Spencer et al., 2016; Coffman, 2014). For instance, Bordalo et al. (2019) shows that people tend to overestimate their own ability and the ability of other people of their gender in categories that are judged to be stereotypically congruent with their group (stereotypically male domains for men or stereotypically female domains for women), and underestimate their own ability and the ability of other people of their gender in categories that are judged to be stereotypically in-congruent with their group. The effect is stronger when individuals are paired with people of the other gender (Bordalo et al., 2019) or when individuals are in groups where their gender is over-represented (Karpowitz and Stoddard, 2020; Born et al., 2020; Chen and Houser, 2019). As a matter of fact, stereotyping is exacerbated by minority status, which reminds people about their social identity and their belonging to the group that makes them distinct (Hoff and Pandey, 2006; Inzlicht et al., 2006). In the education setting lab and field experiments document a widespread belief that women are worse than men in math and science, while are better at literature (Carlana, 2019; Reuben et al., 2014; Ellemers, 2018). Thus, according to this channel, being in a minority is expected to depress beliefs for female students in math-intensive fields, where on the contrary it should inflate believed relative ability for men. On the other hand, being in a minority is expected to depress male students beliefs regarding relative ability in fields related to humanities, where on the contrary it should inflate women's believed relative ability.

However, these predictions do not take selection into account, as they are based on experiments that abstract from considering selection. Recent literature on selection into occupations and majors provide evidence that negative stereotypes regarding group-specific skills affect beliefs about expected returns, driving individuals out of counter-stereotypical occupations (Kugler et al., 2021; Del Carpio and Guadalupe, 2018; Oxoby, 2014). Furthermore, individuals incorporate their preferences toward the fraction of women and men in the occupation when making choices regarding the type of field to specialize in (Pan, 2015). As a consequence, students who selected in a field dominated by the opposite gender in spite of the negative stereotypes regarding their group (women in math-intensive fields and men in humanities) might have weaker stereotypical associations and might face lower costs in interacting with students of the opposite gender, being, as a consequence, less affected by minority status.

I test the extent to which this is the case by building a theoretical framework. I embed the key channels indicated by the experimental literature as the main reasons why minority status affects performance in a theoretical framework and I derive predictions regarding the effect of relaxing minority status for LSE students in absence of selection. I then enrich the framework to show how traits related with stereotypical selection could lead to explain what we observe in the data.

1.7.1 Theoretical Framework

Students are risk neutral and choose educational investments (e_i) to maximize $U(e_i, a_i^b, a_{-i}^b, g, s_g, \beta_i^T, \beta_i^S)$

$$\max_{e_i} \quad \overbrace{\beta_i^P[f(a_i^b, e_i)]}^{\textbf{Course Performance}} \quad + \overbrace{\beta_i^S g(e_i, a_i^b - a_{-i}^b, g)}^{\textbf{Image concerns}} - c(e_i, \delta_i s_g) \tag{1.7}$$

Students' choice of effort depends on two factors (following Bursztyn et al., 2019 and Ashraf et al., 2014: (i) learning motives -$f(.)$, students benefit from investing more effort since they want to maximize their course performance, and (ii) image concerns -$g(.)$, since students care about appearing smart in front of their peers. The latter will be more important (higher β_i^S) when effort is visible to the peers. In the case of undergraduate students at LSE, an example of public investments in education is participation in class. For instance, students might refrain from participating to avoid appearing not smart enough in front of their peers, or they might participate only to appear as smart[44]. The classical assumptions of returns to schooling are assumed: $f_a > 0$, $f_{aa} < 0$, $f_e > 0$,$f_{ee} < 0$, thus course performance is increasing and concave in ability and effort. In the same way, I assume that the utility that students derive from social image is increasing and concave in effort: $g_e > 0, g_{ee} < 0$. Peers' ability enters students' utility function because students care about being of higher ability than their peers rather than having higher grades per se (Ashraf et al., 2014): $g(a_i^b - a_{-i}^b > 0) > 0$, $g(a_i^b - a_{-i}^b < 0) < 0$, $g_{a_i^b - a_{-i}^b} \geq 0$[45]. The benefit of effort in terms of image is higher the higher is students' relative ability: $g_{e, a_i^b - a_{-i}^b} \geq 0$. This implies that students of higher perceived relative ability get higher utility from image if they make higher investments in effort, and that conditional on investing in effort students with higher perceived relative ability get higher utility. This is a key assumption of the model and evidence for its validity can be found in Section 1.7.2. Lastly, following Spence (1974), the cost function is convex and increasing in

[44] Evidence regarding the importance of image concerns for participation is discussed in the section 1.7.2

[45] This is a framework where social comparison enters additively in the utility function (Ashraf et al., 2014; Kandel and Lazear, 1992)

effort: $c_e > 0$ and $c_{ee} > 0$. I model beliefs on ability building on the framework of Bordalo et al. (2019). Students hold imperfect information regarding their ability and the ability of their peers. Beliefs about their and their peers' ability in the subject rely on stereotypes and are affected by stereotypical distortions.

$$a_i^b = A_g + \mu_i + \theta_i \sigma(s_g)(A_g - A_{-g}) \tag{1.8}$$

$$a_i^b - a_{-i}^b = (A_g - A_{-g})(1 - s_g)(1 + 2\theta_i \sigma(s_g)) \tag{1.9}$$

A_g is the average ability of group g, and μ_i is individual specific ability, such that $E_i(\mu_i) = 0$ and $E_i(A_g + \mu_i) = A_g$. Stereotypes driven distortions are defined as $\theta_i \sigma(s_g)(A_g - A_{-g})$. The distortion in beliefs that derive from stereotypes contains a kernel of truth since stereotypes exaggerate true differences in ability between groups. If a student belongs to the "better" performing group, stereotypical distortions will lead him/her to overestimate his/her own absolute ability. On the contrary, if a student belongs to the "worse" performing group, stereotypical distortions will lead him/her to underestimate his/her own absolute ability. The size of the bias depends on how much stereotypes are top of mind ($\sigma(s_g)$) for the student. If stereotypes are top of mind, differences in ability between the two groups are exaggerated. Changes in class composition change the extent to which stereotypes are salient in the students' minds. In particular, stereotyping is exacerbated by minority status, $\sigma_{s_g} \leq 0$, i.e. the extent to which stereotypes are top of mind for a member of group g is decreasing in the size of his/her group s_g.[46].

The second channel through which class composition affects students' choices of effort is through their costs. This is in line with minority status limiting opportunities for social interactions. In particular, I assume that the cost of effort decreases in the share of own group students, $C_{s_g} \leq 0$ and $C_{es_g} \leq 0$. This assumption implies that students feel more comfortable participating in class when they are surrounded by students that are "similar" to them. This can be due to the fact that we expect increased comfort, trust and positive affective responses from those who are similar to us. But this can also have to do with potential language barriers. Regarding private forms of effort, such as hours spent studying, this might happen because class composition influences students networks formation, and as a consequence their academic support.

I assume that individuals are heterogeneous along two key dimensions: the strength of stereotypical associations θ_i and their cost of interacting with students of the opposite gender δ_i. The former determines the strength of stereotypical distortions, while the latter determines the extent to which students benefit from being surrounded by students similar to them in terms of marginal cost of effort. These are the dimensions through which stereotypical selection might intervene to moderate the effect of being in a minority on performance.

[46]More details on the definition and derivations of beliefs on ability can be found in Appendix 1.11.4

Effect of changes in class composition on effort

Given the assumptions on concavity of g and f and convexity of the cost functions in effort, there exist at least an interior solution such that

$$e_i^* : \beta_i^P f_e + \beta_i^S g_e - C_e = 0 \tag{1.10}$$

Given the equilibrium first order conditions described above, the effect of a change in class composition is described by the following expression:

$$\frac{\partial e_i^*}{\partial s_g} = -\frac{\beta_i^P \overbrace{f_{ea}}^{?} \frac{\partial a_i^b}{\partial s_g} + \beta_i^S \overbrace{g_{e,a-a_{-i}}}^{\geq 0} \frac{\partial a_i^b - a_{-i}^b}{\partial s_g} - \delta_i \overbrace{c_{es_g}}^{\leq 0}}{\underbrace{\beta_i^P f_{ee} + \beta_i^S g_{ee} - c_{ee}}_{\leq 0}} \tag{1.11}$$

The effect of a change in class composition on effort depends on three components, a learning component, an image component and a cost of effort component. In order to illustrate how these channels interact in shaping the overall effect, I am going to explain what happens when relaxing numerical minority generates a positive shock on beliefs regarding absolute and relative ability.

(i) Learning component - the effect depends on the sign of f_{ae}, i.e. whether ability and effort are complement or substitute in learning. If ability and effort are complement, students will be induced to invest more in effort, as the positive shock on absolute ability increases the marginal returns of an additional hour of effort on performance in the course. On the other hand, if ability and effort are substitute, students are induced to reduce the amount of effort they invest in the course as they can obtain the same grade with a lower level of effort.

(ii) Image component - the prediction is straightforward: students invest more in effort since they receive a confidence boost. When students believe they are relatively better than their peers, the benefits of investing in effort deriving from image are higher, as they get higher utility for each level of effort.

(iii) Cost of effort component - relaxing numerical minority will have an effect on the cost of effort, and in particular, it will induce student to invest more when the share of same gender people increases. This is due to the fact that the model assumes that the marginal cost of effort is lower when students expect and/or receive more emotional and academic support since they are surrounded, and can engage, with a bigger pool of students at a lower cost.

Thus, the three channels combined imply that, if ability and effort are complement, a positive shock on beliefs regarding relative and absolute ability induces students to invest more, since image and learning components reinforce each others. On the contrary, if effort and ability are substitute in learning, the effect depends on the weight that learning and image concerns have in the utility function.

Whether students receive a positive or negative shock on beliefs regarding relative ability when relaxing minority status depends on the nature of stereotypes in the field. I assume that men perform better on average ($A_M - A_W > 0$) in stereotypically male departments, while women perform better on average ($A_W - A_M > 0$) in stereotypically female departments, so that stereotypical distortions carry indeed a kernel of truth. This is consistent with what we observe at the LSE and with the definition of stereotypical choices used in the paper.[47]

$$\frac{\partial a_i^b}{\partial s_g} = \overbrace{\theta_i \sigma_{s_g}}^{\leq 0}(A_g - A_{-g}) \tag{1.12}$$

$$\frac{\partial E(a_i^b - a_{-i}^b)}{\partial s_g} = (A_g - A_{-g})\overbrace{[-1 - 2\theta_i\sigma + 2\theta_i(1 - s_g)\sigma_{s_g}]}^{\leq 0} \tag{1.13}$$

If students are enrolled in departments in line with stereotypes ($A_g - A_{-g} > 0$), an increase in the share of students of their gender depresses their beliefs regarding absolute and relative ability. An increase in the share of same gender classmates reduces the extent to which stereotypes are top of mind for the student, thus weakening the stereotypical distortions that were leading him/her to overestimate his/her own absolute ability. At the same time, an increase in the share of same gender classmates increases the average perceived ability of the peers, depressing the student's perceived relative ability. This for instance would be the case for men (women) in math-intensive (humanities) departments when the share male (female) students in class increases.

On the other hand, if students are enrolled in counter-stereotypical departments ($A_g - A_{-g} < 0$), an increase in the share of same gender classmates improves their beliefs regarding absolute and relative ability. As a matter of fact, an increase in the share of same gender classmates, by reducing the extent to which stereotypes are top of mind,reduces stereotype threat, which was leading them to underestimate their own absolute ability. In the same way, an increase in the share of same gender classmates decreases the average perceived ability of the peers, improving the student's perceived relative ability. This is the case for women (men) in math and science-intensive (humanities) fields when the share of women (men) in class increases.

$$\begin{aligned}
&\text{if } (A_g - A_{-g}) > 0 \rightarrow \frac{\partial e_i^*}{\partial s_g} = -\frac{\beta_i^P \overbrace{f_{ea}}^{?} \overbrace{\frac{\partial a_i^b}{\partial s_g}}^{\leq 0} + \beta_i^S \overbrace{g_{e,a-a_{-i}} \frac{\partial E(a_i^b - a_{-i}^b)}{\partial s_g}}^{\leq 0} - \delta_i \overbrace{c_{es_g}}^{\leq 0}}{\underbrace{\beta_i^P f_{ee}^e + \beta_i^S g_{ee} - c_{ee}^e}_{\leq 0}} \\
&\text{if } (A_g - A_{-g}) < 0 \rightarrow \frac{\partial e_i^*}{\partial s_g} = -\frac{\beta_i^P \overbrace{f_{ea}}^{?} \overbrace{\frac{\partial a_i^b}{\partial s_g}}^{\geq 0} + \beta_i^S \overbrace{g_{e,a-a_{-i}} \frac{\partial E(a_i^b - a_{-i}^b)}{\partial s_g}}^{\geq 0} - \delta_i \overbrace{c_{es_g}^e}^{\leq 0}}{\underbrace{\beta_i^P f_{ee}^e + \beta_i^S g_{ee} - c_{ee}^e}_{\leq 0}}
\end{aligned} \tag{1.14}$$

Prediction 1 (Counter-stereotypical departments). *Relaxing minority status increases effort for students that are are enrolled in counter-stereotypical departments ($A_g - A_{-g} < 0$) whenever effort and ability are complement, or effort and ability are substitutes but the effect on image and cost of effort prevails on the effect on learning.*

[47] Figure 1.A8 displays the raw gender gap in course performance for students enrolled in different departments calculated as the difference in course performance between men and women enrolled in the same course in the same year. Women enrolled in female-majority departments perform better than men, while the gender gap is negative for students enrolled in male-majority departments. The sample contains all the courses attended by students enrolled in undergraduate programs in the department during the bachelor.

Prediction 2 (**Stereotypical departments**). *For students enrolled in fields that are stereotypically congruent with the group identity ($A_G - A_{-G} > 0$), relaxing minority status generates an increase in effort for students if the negative effect on image is small and ability and effort are substitute. On the other hand, relaxing minority status generates a reduction in effort if the positive effect on the cost of effort is small and ability and effort are complement.*

It can be seen from equations (14) that the effect of relaxing minority status on effort choices is stronger the stronger is the effect that relaxing minority status has on beliefs regarding relative and absolute ability and the stronger is the effect on the marginal cost of effort. Equations (12) and (13) show that the effect of relaxing numerical minority on beliefs on ability is stronger the stronger are stereotypical associations for the students (θ_i). On the other hand, the strength of the effect of relaxing minority status on the marginal cost of effort depends on δ_i, the individual cost of interacting with students of the opposite gender. Thus, the model delivers a third prediction:

Prediction 3 (**Stereotypical selection**). *The effect of changes in class composition on effort choices are stronger the stronger are stereotypical associations for students (θ_i) and the higher is the individual cost of interacting with students of the opposite gender (δ_i).*

1.7.2 From Theory to Empirics

In order to test the model predictions, I exploit rich administrative data and a novel survey. LSE requires teaching assistants to evaluate students' participation in class during the term. Lab and field experiments provide clear evidence that, by reducing the daunting effect of negative stereotypes, relaxing numerical minority increases participation and willingness to contribute to discussions for individuals in counter-stereotypical domains (Coffman and Shurchkov, 2021; Karpowitz and Stoddard, 2020; Chen and Houser, 2019; Bordalo et al., 2019; Coffman, 2014). I exploit teachers' evaluations of students' participation in class as a counterfactual to estimate the effect of a change in group composition on participation and willingness to contribute to discussions in a real-life setting where selection against stereotypes plays a role[48].

Furthermore, participation in class can be considered as a form of public effort, since it is observable to peers. Thus, we can expect choices of participation to be motivated primarily by image concerns rather than learning motives. Figure 1.12 displays the relationship between exam grades and participation in class (Panel A), and participation and individual and peers' ex-ante ability (Panels B and C), controlling for student and class group fixed effects. Participation is beneficial for students' learning, since we can see that it is positively correlated with exam grades: students who participate more in the first term obtain higher exam grades at the end of the year. However, the students who participate the most are high ability students, and participation appears

[48] These data have two caveats. First of all, they are not objective measures of students' participation and performance, but they are grades that teachers give to students. Thus, results can be considered as indicative of class dynamics conditional on the assumption that teaching assistants do not discriminate or evaluate underrepresented groups differently when they are in a minority. Secondly, they only concern Michaelmas term (due to attrition problems, I consider only Michaelmas term grades, not Lent term evaluations). As a consequence, they are able to shed light on the dynamics that characterize classes in the first half of the academic year. This on the one hand represents a limitation, since exam grades are the result of students' effort during the whole academic year. On the other hand, we can argue that first term measures are more indicative of how students' behavior is affected by stereotypical distortions, since this is the first time students meet and interact with each other.

to be negatively correlated with the ability of the peers in the class group. This is in line with participation in class being primarily motivated by image concerns and in particular with the idea that students care about being better than their peers. This is confirmed also by the evidence gathered through the survey. Figure 1.13 displays that students are significantly more comfortable answering questions in class rather than asking questions. Furthermore, they admitted that the worry of not looking “smart enough” has refrained them from participating at least once. On the contrary they disagree with the statement that they refrained from participating at least once for the fear of appearing ”too know-it-all”.

Evidence for students who made a counter-stereotypical choice of major

When social image represents the primary concern and effort is public, the model predicts that relaxing minority status induces students who are enrolled in counter-stereotypical departments to participate more. This occurs as relaxing minority status reduces stereotypical distortions, which were inducing students to underestimate their absolute and relative ability, and allows them to benefit from a reduction in the marginal cost of effort.

Figure 1.14 shows the results of Specification 1.3 on participation in class defining stereotypical selection based on the bottom and top 5 departments categorization.[49] There is no evidence of an effect of changes in class composition on participation for students who are enrolled in counter-stereotypical fields.

This result indicates that, contrary to experimental evidence that relaxing numerical minority increases participation and willingness to contribute to discussions for individuals in counter-stereotypical domains (Coffman and Shurchkov, 2021; Karpowitz and Stoddard, 2020; Chen and Houser, 2019; Bordalo et al., 2019; Coffman, 2014), these channels do not play a role in real world environments where selection is important.

First, we can exclude that the absence of effect is due to participation being unaffected by class composition, as students who made choices of major in line with stereotypes display significant effects. Second, the evidence cannot be reconciled with stereotypes playing a role. This zero effect is consistent with experimental findings that minority status impacts performance of minorities in counter-stereotypical domains through stereotype threat (stereotypes inducing the “worse” performing group to participate less by depressing their believes regarding their relative ability) if we assume that the daunting effect of stereotypes is so strong that not even a change in class group composition can improve students’ beliefs regarding their ability and induce them to participate in class. However, this does not seem to be the case. Table 1.A8 displays the gender gap in class participation across departments. We observe a significantly smaller gender gap in participation grades in female fields with respect to male fields, indicating that students who chose a counter-stereotypical major are more vocal on average with respect to same gender students who chose a major in line with stereotypes.

[49]This is the definition that I use when discussing the mechanisms as it is the least restrictive and most conservative definition of stereotypical and counter-stereotypical choices, and the one that allows me to have a big enough sample size to deter significant effects.

Evidence for students who made a stereotypical choice of major

Figure 1.14 shows that, contrary to the evidence for students who made counter-stereotypical choices, men and women who are enrolled in stereotypical majors modify their participation with changes in class composition. However, while women participate more in classes where there are more women, men participate less when surrounded by more men. This indicates that the effects for men and women who are enrolled in stereotypical departments are driven by different channels.

Interpreted through the lenses of the model, these results imply that the effect for women is driven by them benefiting from greater comfort and support when surrounded by more women. On the other hand, the effect for men can be reconciled with positive stereotypes regarding men's performance in stereotypically male departments inducing them to be less confident when surrounded by men, participating less to the discussion.

This is consistent with suggestive evidence indicating that the cost of participating for women is higher. The left panel of Figure 1.17 shows that women have lower participation grades with respect to men in the same class for every level of ability at entry. Furthermore, evidence from the auxiliary survey in the right panel of Figure 1.17 displays that women declare that they are significantly less comfortable asking and answering questions in class with respect to men.

An additional piece of suggestive evidence in line with these findings is presented Table 1.9, where I estimate the effect of an increase in the share of same gender classmates on participation for students whose previous school was a single-sex school and for students whose previous school was a mixed sex school. I estimate Specification 1.3 interacting the share of students like me with a dummy equal to one if the student's previous school was a single sex school. I restrict the analysis to the subsample of students for which I have information regarding their previous school[50]. Whether students studied in a single vs mixed- sex school is used as a proxy for the strength of the cost of interacting with opposite gender peers. The underlying assumption is that students who studied in schools with only students of their own gender would be less used and less comfortable interacting with opposite gender students.

Table 1.9 presents results for the overall sample of students in Column (1), students who made counter-stereotypical choices of major in Column (2), and students who made stereotypical choices of major in Column (3).[51] In line with the effect for women being driven primarily by the cost of interacting with opposite gender students, increasing the share of female classmates always increases participation for women, but the effect is stronger for women who studied in single sex schools. On the contrary, for men, increasing the share of male classmates reduces participation, consistently with a reduction in perceived relative ability when men are surrounded by other men in stereotypically male departments. However, this effect disappears for men who studied in single sex schools, who, according to our assumption, are the men for which the cost of interacting with the opposite gender is stronger. For these men the negative effect of stereotypes is counteracted by them benefiting from a lower share of women in class reducing the cost of participating.

[50]Students whose previous school was in UK (or oversees but related to a UK institution) and for whom I managed to match the information with the UK government register of school and colleges

[51]The definition of stereotypical selection based on the bottom and top 5 departments categorization. This is the definition that I use when discussing the mechanisms as it is the least restrictive and most conservative definition of stereotypical and counter-stereotypical choices, and the one that allows me to have a big enough sample size to deter significant effects.

In light of this finding, the results on course performance could be reconciled through the lenses of the model with men in stereotypically male departments underestimating their relative ability due to an increase in the share of men, as a consequence investing less in participation (image channel), but investing more in learning to improve their course performance. This would imply that effort and relative ability are substitute in the learning function. This is plausible if we consider that students have a fixed amount of time to allocate to effort across different courses, and as a consequence they might want to allocate less time to study to courses where they believe their subject specific ability is higher or their peers are on average less competitive. Another potential explanation is that students' effort decisions are also driven by a competitive component as in Azmat et al. (2019), and, as a consequence, their optimal choice of effort is lower when they are in a class where they believe they are relatively better.

1.7.3 Do students internalize stereotypes and environment composition in their choices?

In the previous sections, I provided evidence that the channels that have been indicated as the main drivers of the effect of being in a minority on performance in experimental settings do not apply to students who made choices against stereotypes. In this last section, I shed light on whether students who made different choices of major are heterogeneous along traits consistent with them internalizing environment composition and stereotypes when making their choice.

First, the unbalanced gender composition of departments seems to reflect preferences or strategic behavior of students at the application stage. As a matter of fact, the patterns in applications across departments exactly mirror the distribution of men and women enrolled in different programs (Figure 1.A4). This gender segregation in applications is mostly driven by men. Figure 1.A9 shows the number of applicants for each offer made in the same year by each department by gender. For the same number of available places a disproportionately higher number of men apply to departments that end up being male-dominated. This pattern does not appear to be as pronounced for women, which could reconcile why I find a weaker effect for women.

Second, I provide evidence that students who made different choices of major differ in the strength of stereotypical associations and the cost of interacting with opposite gender students. Recent literature on selection into occupations and majors provide evidence that individuals incorporate stereotypes (Kugler et al., 2021; Del Carpio and Guadalupe, 2018; Oxoby, 2014) and the fraction of women and men in the occupation (Pan, 2015) when making choices regarding the type of field to specialize in. As a consequence, students who select in a field dominated by the opposite gender in spite of the negative stereotypes regarding their group (women in math-intensive fields and men in humanities) might have weaker stereotypical associations and might face lower costs in interacting with students of the opposite gender. This could rationalize why students who made choices against stereotypes would be less affected by minority status, as the model predicts that the effect of changes in class composition on effort choices are stronger the stronger are stereotypical associations for students (θ_i) and the higher is the individual cost of interacting with students of the opposite gender (δ_i).

Evidence on the strength of stereotypical associations is provided in Figure 1.15, which displays the results of an implicit association test (Greenwald et al. 1998) to elicit students' associations between female-humanistic and male-scientific[52]. A score of 0 indicates no association between male-Scientific and female-Humanistic. A positive score indicates that the student unconsciously associates women with humanities and men with science and math. Lastly, a negative score indicates that the student unconsciously associates men with humanities and women with science and math. Students on average associate men with math and science and women with humanities, but this implicit association is significantly stronger for students who enrolled in departments in line with stereotypes[53].

In line with this finding, Table 1.10 provides suggestive evidence that the performance of students who are enrolled in counter-stereotypical departments is not affected by the gender of the class teacher, contrary to the performance of students who made a stereotypical choice of major. Since class teachers act as role models, breaking stereotypes regarding gender roles and skills (see for example Breda et al., 2021; Porter and Serra, 2020; Olsson and Martiny, 2018; Carrell et al., 2010), this is an additional piece of evidence that points in the direction of students who are enrolled in counter-stereotypical departments having lower stereotypical associations, and as a consequence being less affected by gender stereotypes.

Regarding the cost of interacting with opposite gender students, I rely on the evidence from the auxiliary survey provided in Figure 1.16. Students who are enrolled in counter-stereotypical fields nominate significantly more people of the opposite gender when asked about their friends, people they study with, and people they ask questions on the material to. While this might be partially explained by a different availability of students of the opposite gender to interact with, the difference between students enrolled in stereotypical and counter-stereotypical departments is significantly bigger when it concerns students to ask questions on the material to. This seems to indicate that students who are enrolled in counter-stereotypical departments have a lower cost of engaging with opposite gender peers, allowing them to benefit from academic support from their peers independently on their gender. However, additional research is needed to be able to conclusively say that this is due to differences in the cost of engaging with students of the opposite gender rather than, for example, the availability of peers that are perceived as high ability and that students believe can help them in understanding the course material.

Lastly, differences in ability, and in particular stereotypical minorities being positively selected, could be an alternative potential explanation for why their performance is less affected by the composition of the environment. However, differences in ability do not seem to be able to explain the results. First, the results on performance show that students who made choices against-stereotypes are indeed affected by the composition of the environment. They do not suffer from being in a numerical minority, but they benefit from it. Second, students who are enrolled in stereotypical and counter-stereotypical departments have equal potential and ability at entrance. Even though Math-intensive departments are the most competitive and selective, with the higher

[52] Details on the survey can be found in Appendix 1.11.6

[53] The result is confirmed also when I restrict the sample to first year students. Furthermore, the very same pattern can be found when analysing the data from a gender-scientific implicit association test carried out by Project Implicit. The evidence can be found in appendix 1.11.6.

overall rejection rates, there is no evidence of a gender gap in qualifications at entry for students who are enrolled in different departments. Figure 1.18 shows the distribution of qualification scores for men and women across departments. There is no significant difference in average qualification scores at entrance between men and women, nor in the distributions of qualification scores at entry for men and women enrolled in the same type of departments.

1.8 Discussion

This paper aims at empirically testing if selection into fields plays a role in moderating the effect of being in a numerical minority on performance. I consider a particular type of selection: whether individuals make choices in line or against stereotypes regarding their group. This is relevant as minority status and selection against stereotypes often overlap, as stereotypes shape the distribution of groups across fields by influencing payoffs from economic choices (Akerlof and Kranton, 2000). As a consequence, those individuals that we observe in a numerical minority are often individuals who decided to bear the cost of making a choice against stereotypes. In contrast, individuals who belong to the majority group are often individuals who made a choice in line with stereotypes.

This paper provides evidence that, as individuals internalize social identity considerations and the composition of the environment when making their choices, these factors affect people belonging to the same group (for example, gender or ethnicity) in a different way depending on the environment where we observe them. In particular, selection against stereotypes significantly moderates the detrimental effect associated with being in a minority in counter-stereotypical domains that is documented by the experimental literature.[54] As a matter of fact, the only students whose performance is negatively affected by being under-represented are students who made stereotypical choices, such as men in Mathematics or women in Sociology, and, as such, are part of the majority group in the field. On the contrary, students who made choices against stereotypes, such as women in Mathematics or men in Sociology, do not perform worse when in classes surrounded by different gender peers. Students who make choices in line with stereotypes also hold stronger implicit stereotypical associations and seem to be more affected by stereotypes, hinting towards the majority group being the perpetrator of stereotypes.

These results have important policy implications. The counterfactual simulation presented in Section 1.5.3 shows that as students internalize stereotypes and the composition of the environment when making their choices of major, the only students who benefit from being surrounded by same gender peers are students who made choices of major in line with stereotypes and belong to the majority group. As a consequence, enforcing a more balanced gender ratio (with policies such as quotas) would reduce differences in performance across groups, decreasing average performance at the same time. Moreover, the evidence indicates that students who belong to the majority group are the ones that uphold the strongest implicit stereotypical associations. Hence, by making differences between groups salient and reducing the possibility for the majority group to update their beliefs through experiential learning, initiatives targeting minorities, such as networking events, mentoring

[54] For example Coffman and Shurchkov (2021); Born et al. (2020); Chen and Houser (2019); Bordalo et al. (2019); Coffman (2014); Spencer et al. (2016); Spencer et al. (1999); Steele and Aronson (1995)

programs, or quotas, might reinforce stereotypical distortions in the mind of the majority group, contributing to perpetrate stereotypes.

This is particularly important when the environment is very selective, and the presence of a person from an under-represented group is the result of a series of strategic choices. This is very likely to be the case at the LSE. Application to undergraduate programs is centralized in the UK, and students can only apply to 5 programs (across all universities in the UK) in the same year. Furthermore, applications at LSE are very competitive: 75% of students who apply are rejected every year. Thus students who apply at LSE are very motivated and very confident.

The LSE setting is well-suited to answer the paper's research question as it provides an ideal environment to shed light on the effect of selection based on stereotypes on minority status. Furthermore, it allows to disregard discrimination, thanks to the blind grading system. However, the paper's findings cannot be easily generalized to other educational environments since LSE and its students are not comparable to the average university. Given the competitive nature of the environment and the strategic decision-making process that applying to undergraduate programs in the UK entails, these findings might be informative for selective and competitive working environments. For instance, the paper's findings might inform policies addressing the under-representation of women in decision-making bodies or leading positions. In this regard, the results on participation in class become particularly informative, as this outcome can be indicative regarding what happens in work environments where social interactions between colleagues occur.

The discussion so far abstracted from considering the effect that a more balanced gender ratio could have on encouraging participation of minorities into these fields. This represents another important avenue for research. Since this paper provides evidence in support of students internalizing the composition of the environment and stereotypes when making their choice of major, measures such as affirmative action policies could also impact major choices of future cohorts, through their effect on the composition of the environment and the pool of minorities students in counter-stereotypical fields.

References

Akerlof, G. A. and R. E. Kranton (2000, 08). Economics and Identity*. *The Quarterly Journal of Economics 115*(3), 715–753.

Anelli, M. and G. Peri (2017). The effects of high school peers' gender on college major, college performance and income. *The Economic Journal*.

Ashraf, N., O. Bandiera, and S. S. Lee (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior Organization 100*(C), 44–63.

Austen-Smith, D. and J. Fryer, Roland G. (2005, 05). An Economic Analysis of "Acting White"*. *The Quarterly Journal of Economics 120*(2), 551–583.

Avilova, T. and C. Goldin (2018, May). What can uwe do for economics? *AEA Papers and Proceedings 108*, 186–90.

Azmat, G., M. Bagues, A. Cabrales, and N. Iriberri (2019). What you don't know... can't hurt you? a natural field experiment on relative performance feedback in higher education. *Management Science 65*(8), 3714–3736.

Bandiera, O., V. Larcinese, and I. Rasul (2010). Heterogeneous class size effects: New evidence from a panel of university students*. *The Economic Journal 120*(549), 1365–1398.

Beede, D. N., T. Julian, D. Langdon, G. McKittrick, B. Khan, and M. Doms (2011). Women in stem: A gender gap to innovation. *Sexuality & Gender Studies eJournal*.

Bertrand, M. (2011). Chapter 17 - new perspectives on gender. Volume 4 of *Handbook of Labor Economics*, pp. 1543–1590. Elsevier.

Bertrand, M. (2020, May). Gender in the twenty-first century. *AEA Papers and Proceedings 110*, 1–24.

Blau, F. D., J. M. Currie, R. T. A. Croson, and D. K. Ginther (2010, May). Can mentoring help female assistant professors? interim results from a randomized trial. *American Economic Review 100*(2), 348–52.

Booth, A. L., L. Cardona-Sosa, and P. Nolen (2018). Do single-sex classes affect academic achievement? an experiment in a coeducational university. *Journal of Public Economics 168*, 109–126.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016). Stereotypes. *Quarterly Journal of Economics 131*(4), 1753–1794.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019). Beliefs about gender. *American Economic Review 109*(3), 739–73.

Born, A., E. Ranehill, and A. Sandberg (2020). Gender and willingness to lead: Does the gender composition of teams matter? *The Review of Economics and Statistics 0*(ja), 1–46.

Bostwick, V. and B. Weinberg (2021). Nevertheless she persisted? gender peer effects in doctoral stem programs. *Journal of Labor Economics 0*(ja), null.

Braga, M., M. Paccagnella, and M. Pellizzari (2016). The Impact of College Teaching on Students' Academic and Labor Market Outcomes. *Journal of Labor Economics 34*(3), 781–822.

Breda, T., J. Grenet, M. Monnet, and C. van Effenterre (2021, March). Do Female Role Models Reduce the Gender Gap in Science? Evidence from French High Schools. PSE Working Papers halshs-01713068, HAL.

Brenøe, A. A. and U. Zölitz (2020). Exposure to more female peers widens the gender gap in stem participation. *Journal of Labor Economics 38*(4), 1009–1054.

Bursztyn, L., G. Egorov, and R. Jensen (2019, 05). Cool to be Smart or Smart to be Cool? Understanding Peer Pressure in Education. *The Review of Economic Studies 86*(4), 1487–1526.

Bursztyn, L., T. Fujiwara, and A. Pallais (2017, November). 'acting wife': Marriage market incentives and labor market investments. *American Economic Review 107*(11), 3288–3319.

Bursztyn, L. and R. Jensen (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics 9*(1), 131–153.

Campbell, S., L. Macmillan, R. Murphy, and G. Wyness (2019, August). Inequalities in student to course match: evidence from linked administrative data. CEP Discussion Papers dp1647, Centre for Economic Performance, LSE.

Card, D., A. Mas, and J. Rothstein (2008, 02). Tipping and the Dynamics of Segregation*. *The Quarterly Journal of Economics 123*(1), 177–218.

Carlana, M. (2019, 03). Implicit Stereotypes: Evidence from Teachers' Gender Bias*. *The Quarterly Journal of Economics 134*(3), 1163–1224.

Carrell, S. E., M. E. Page, and J. E. West (2010, 08). Sex and Science: How Professor Gender Perpetuates the Gender Gap*. *The Quarterly Journal of Economics 125*(3), 1101–1144.

Chen, J. and D. Houser (2019). When are women willing to lead? the effect of team gender composition and gendered tasks. *The Leadership Quarterly 30*(6), 101340.

Coffman, Katherine B., C. B. F. and O. Shurchkov (2021). Gender stereotypes in deliberation and team decisions. Harvard Business School Working Paper No. 19-069.

Coffman, K. B. (2014). Evidence on Self-Stereotyping and the Contribution of Ideas. *The Quarterly Journal of Economics 129*(4), 1625–1660.

Del Carpio, L. and M. Guadalupe (2018, October). More Women in Tech? Evidence from a field experiment addressing social identity. CEPR Discussion Papers 13234, C.E.P.R. Discussion Papers.

Delfino, A. (2021). Breaking Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs. IZA Discussion Paper 14083.

Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology 69*(1), 275–298. PMID: 28961059.

Feld, J. and U. Zölitz (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics 35*(2), 387–428.

Garg, N., L. Schiebinger, D. Jurafsky, and J. Zou (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences 115*(16), E3635–E3644.

Giorgi, G. D., W. G. Woolston, and M. Pellizzari (2012). Class size and class heterogeneity. *Journal of the European Economic Association 10*(4), 795–830.

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

Greenwald, A. G., D. E. McGhee, and J. L. K. Schwartz (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology 74 6*, 1464–80.

Griffith, A. L. and J. B. Main (2019). First impressions in the classroom: How do class characteristics affect student grades and majors? *Economics of Education Review 69*, 125–137.

Guiso, L., F. Monte, P. Sapienza, and L. Zingales (2008). Culture, gender, and math. *Science 320*(5880), 1164–1165.

Guryan, J., K. Kroft, and M. J. Notowidigdo (2009, October). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics 1*(4), 34–68.

He, J. C., S. K. Kang, K. Tse, and S. M. Toh (2019). Stereotypes at work: Occupational stereotypes predict race and gender segregation in the workforce. *Journal of Vocational Behavior 115*, 103318.

Hebert, C. (2020). Gender stereotypes and entrepreneur financing. 10th miami behavioral finance conference, available at ssrn: https://ssrn.com/abstract=3318245 or http://dx.doi.org/10.2139/ssrn.3318245.

Hill, A. J. (2017). The positive influence of female college students on their male peers. *Labour Economics 44*, 151–160.

Hoff, K. and P. Pandey (2006, May). Discrimination, social identity, and durable inequalities. *American Economic Review 96*(2), 206–211.

Huntington-Klein, N. and E. Rose (2018). Gender peer effects in a predominantly male environment: Evidence from west point. *AEA Papers and Proceedings 108*, 392–395.

Inzlicht, M., J. Aronson, C. Good, and L. McKay (2006, May). A particular resiliency to threatening environments. *Journal of Experimental Social Psychology 42*(3), 323–336.

Inzlicht, M. and C. Good (2006). How environments can threaten academic performance, self-knowledge, and sense of belonging. *In S. Levin and C. van Laar (Eds.), The Claremont symposium on Applied Social Psychology. Stigma and group inequality: Social psychological perspectives*, 129–150.

Kandel, E. and E. P. Lazear (1992). Peer pressure and partnerships. *Journal of Political Economy 100*(4), 801–817.

Karpowitz, Chris, J. P. and O. Stoddard (2020). Strength in numbers: A field experiment on gender, influence and group dynamics. IZA Discussion Paper No. 13741.

Kugler, A. D., C. H. Tinsley, and O. Ukhaneva (2021). Choice of majors: are women really different from men? *Economics of Education Review 81*, 102079.

Lane, K. A., G. J. X. . D.-L. E. (2012). Implicit science stereotypes mediate the relationship between gender and academic participation. *Sex Roles: A Journal of Research 66(3-4)*, 220–234.

Lavy, V. and A. Schlosser (2011, April). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics 3*(2), 1–33.

McNally, S. (2020). Gender differences in tertiary education: What explains stem participation. CEP Discussion Paper No. 1721.

Murdoch, D. J., Y.-L. Tsai, and J. Adcock (2008). P-values are random variables. *The American Statistician 62*(3), 242–245.

OECD (2020). *Education at a Glance 2020.*

Olivetti, C., E. Patacchini, and Y. Zenou (2020). Mothers, Peers, and Gender-Role Identity. *Journal of the European Economic Association 18*(1), 266–301.

Olsson, M. and S. E. Martiny (2018). Does exposure to counterstereotypical role models influence girls' and women's gender stereotypes and career choices? a review of social psychological research. *Frontiers in Psychology 9*, 2264.

Oosterbeek, H. and R. van Ewijk (2014). Gender peer effects in university: Evidence from a randomized experiment. *Economics of Education Review 38*, 51–63.

Oxoby, R. J. (2014). Social inference and occupational choice: Type-based beliefs in a bayesian model of class formation. *Journal of Behavioral and Experimental Economics 51*, 30–37.

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

Porter, C. and D. Serra (2020, July). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics 12*(3), 226–54.

Reuben, E., P. Sapienza, and L. Zingales (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences 111*(12), 4403–4408.

Russell, L. (2017). Can learning communities boost success of women and minorities in stem? evidence from the massachusetts institute of technology. *Economics of Education Review 61*, 98–111.

Shan, X. (2020). Does minority status drive women out of male-dominated fields? Working paper.

Spencer, S. J., C. Logel, and P. G. Davies (2016). Stereotype threat. *Annual Review of Psychology 67*(1), 415–437. PMID: 26361054.

Spencer, S. J., C. M. Steele, and D. M. Quinn (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology 35*(1), 4–28.

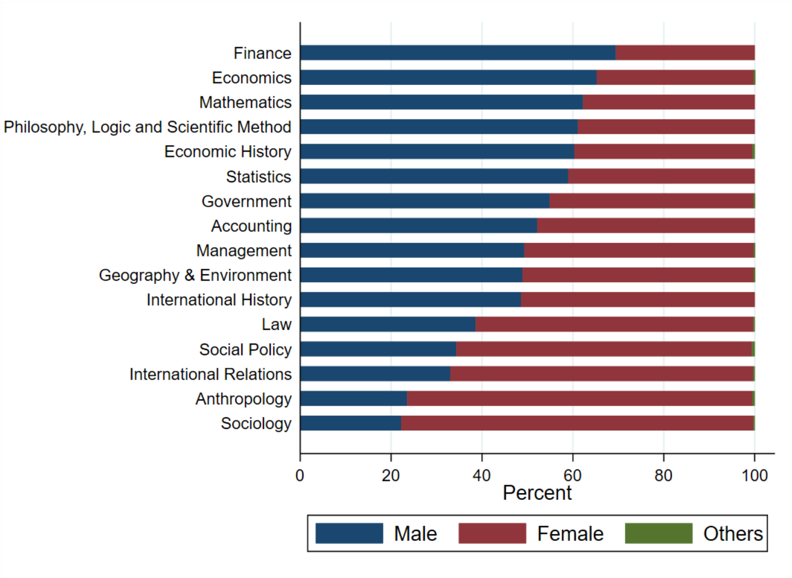
Steele, C. and J. Aronson (1995, November). Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology 69*(5), 797–811. Copyright: Copyright 2018 Elsevier B.V., All rights reserved.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research 45*(1), 89–125.

Zölitz, U. and J. Feld (2018). The effect of peer gender on major choice. ECON - Working Papers No. 270, Department of Economics - University of Zurich.

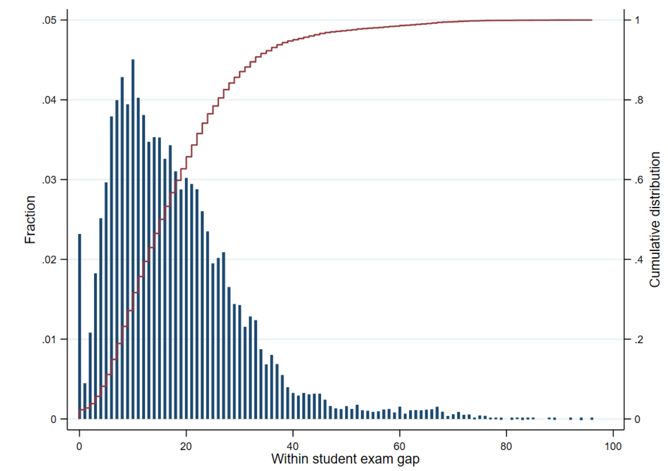
1.9 Figures

Figure 1.1: Departments gender composition



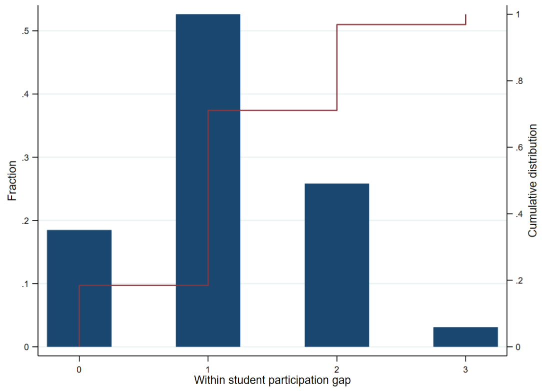
Notes: This figure illustrates the gender composition of each department. It is constructed based on the number of first year undergraduate students that are enrolled in bachelor programs offered by the department between 2008/09 and 2017/18. The figure shows the average across academic years 2008/09 to 2017/18.

Figure 1.2: Within-student course performance gap



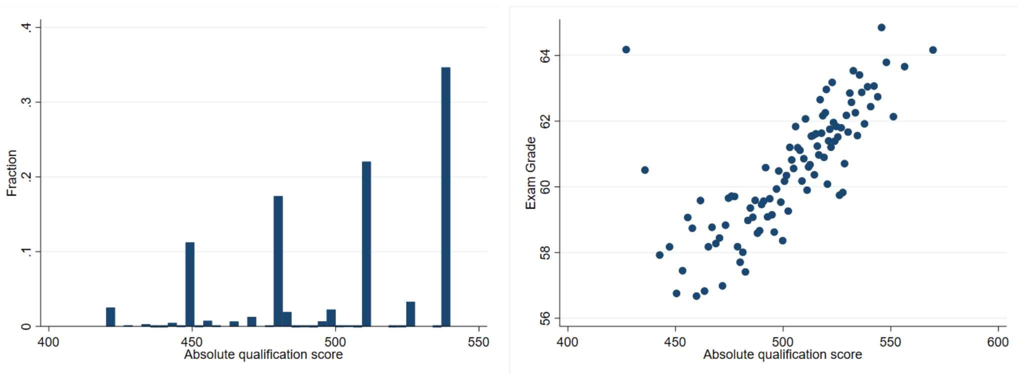
Notes: Following Bandiera et al., 2010. The figure displays the within student exam gap. This is defined as the difference between student i's highest and lowest exam mark across all the courses attended during Michaelmas term of the first year. The sample considered in the graph is the sample in analysis: 54.603 course-year-class group level observations, corresponding to information on 14.313 students.

Figure 1.3: Within-student class participation gap



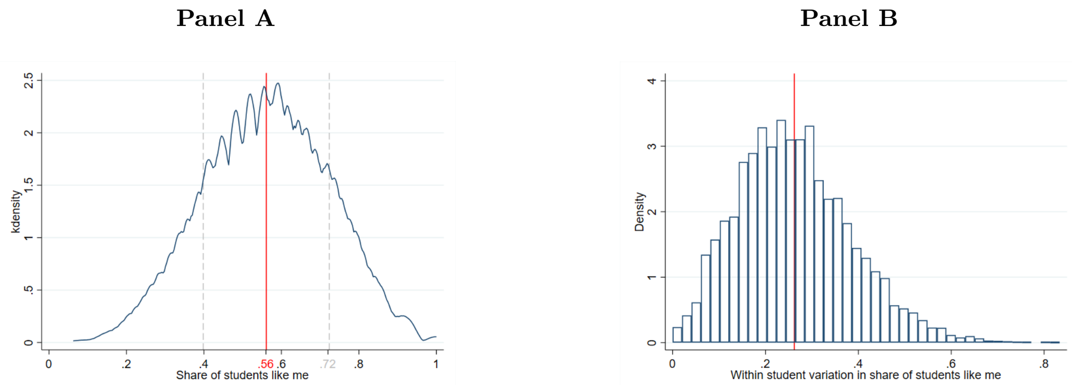
Notes: Following Bandiera et al., 2010. The figure displays the within student participation gap. This is defined as the difference between student i's highest and lowest participation grade across all the classes attended during Michaelmas term of the first year. The sample consist of the sample of analysis restricted to all the students for which I can observe a participation grade in at least two class groups: 44.771 course-year-class group level observations, corresponding to information on 13.485 students.

Figure 1.4: Qualification scores at entry



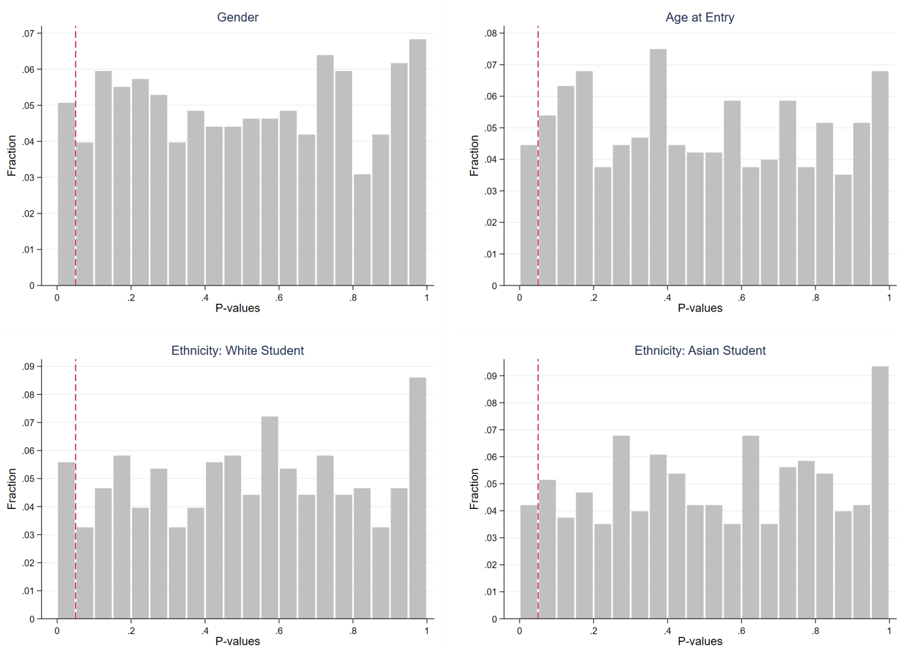
Notes: Panel A displays qualification scores at entry for 9.449 students, 90% of the students enrolled in undergraduate programs in LSE between academic years 2011/12 and 2017/18 (students for which I have information regarding qualifications at entry). Following Campbell et al. (2019), I construct a qualification score based on the best three exam results among the A-level qualification scores students declared when applying to LSE. Some students take courses/programs that are equivalents to A-Levels. In these cases I calculate their A-Level equivalence scores based on the university conversion tables for foreign students. A levels are graded on a scale of A*/A/B/C/D/E. Each A-level grade is worth 30 QCA (Qualifications and Curriculum Authority) points. Panel B displays the correlation between qualification scores at entry and course grades controlling for qualification type, individual characteristics (program of enrolment, gender, ethnicity, social background), and course fixed effects

Figure 1.5: Within-student variation in share of same gender classmates



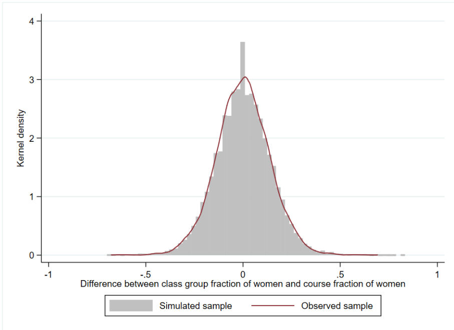
Notes: Panel A displays the distribution of the share of students like me, defined as the share of same gender classmates. The red vertical line indicates the average, while the two grey dashed lines indicate the average plus and minus one standard deviation. Panel B displays the distribution of the within student variation in the share of students like me. This is defined as the difference between the maximum and minimum share of same gender classmates experienced by each student during the first academic year. The red line indicates the average variation.

Figure 1.6: Identification: Observed sample P-values



Notes: This graph show the p-values of a test of joint significance of class dummies from a regression of gender (age at entry, a dummy for being White, and a dummy for being Asian, one at the time) on class dummies and dummies for each one of the courses taken by students in the academic years in analysis. A total of 454 regressions were performed. The sample is restricted to the courses that have at least 2 classes. The test is performed on all the students that attend courses attended by the students in the sample of analysis and that did not change classes (which are the students for whom I can recover the initial class allocation). More information on the text can be found in Appendix 1.11.3.

Figure 1.7: Identification: Allocation simulation - Difference in share of female in class and course

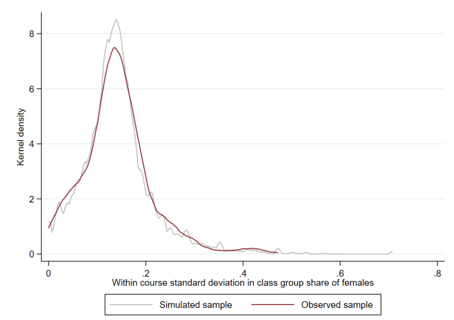


Two-sample Kolmogorov-Smirnov test

Smaller group	D	P-value
0:	0.0130	0.193
1:	-0.0116	0.271
Combined K-S:	0.0130	0.384

Notes: The Figure displays in dark grey the distribution of the difference between the share of women in each class and the share of women in the course for 1000 simulations of an unconstrained (not considering scheduling constraints) random allocation of students to classes. This is compared with the same statistics for the observed allocation (in red). The result of a two-sample Kolmogorov-Smirnov test shows that the distribution of the observed statistics is not significantly different from the simulated statistics.

Figure 1.8: Identification: Allocation simulation - Within-course std. dev. in class groups share of females

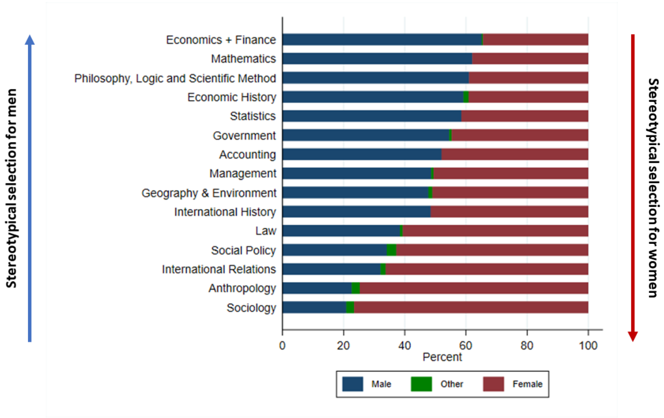


Two-sample Kolmogorov-Smirnov test

Smaller group	D	P-value
0:	0.0286	0.410
1:	-0.0272	0.446
Combined K-S:	0.0286	0.764

Notes: The Figure displays in dark grey the distribution of the within course standard deviation in the class share of females for 1000 simulations of an unconstrained (not considering scheduling constraints) random allocation of students to classes. This is compared with the same statistics for the observed allocation (in red). The result of a two-sample Kolmogorov-Smirnov test shows that the distribution of the observed statistics is not significantly different from the simulated statistics.

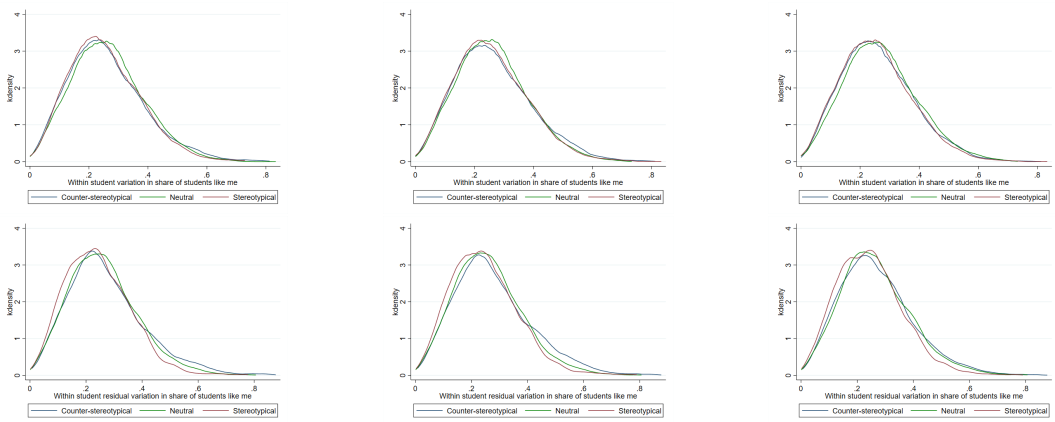
Figure 1.9: Stereotypical selection - Definition



Notes: The Figure displays the distribution of men and women enrolled in the first year of undergraduate programs across departments. The statistics is the average of academic years 2008-2017. The blue and red arrows at the sides display the direction of choices of majors that would be in line with stereotypes and gender norms for men and women respectively.

Figure 1.10: Within-student variation in share of same gender classmates across departments

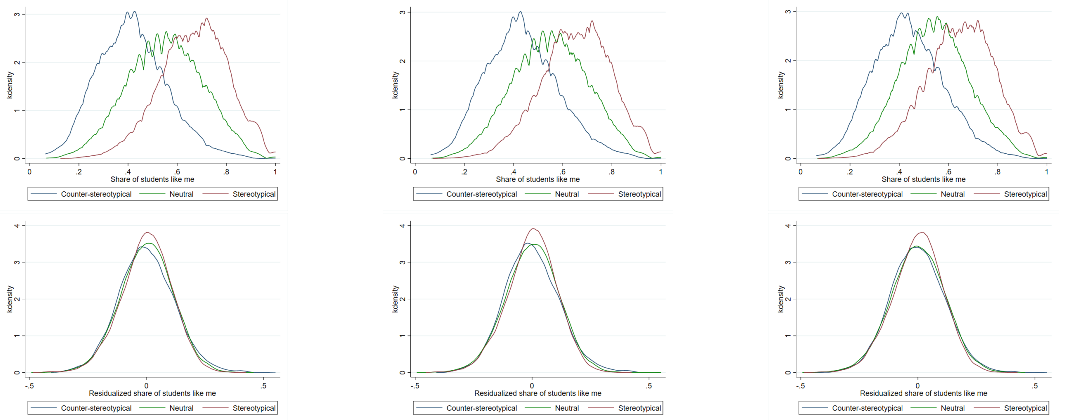
Panel A: Top & Bottom 2 **Panel B: Top & Bottom 3** **Panel C: Top & Bottom 5**



Notes: The figure displays the within student variation in the share of same gender classmates (top) and the within student variation in the residual share of same gender classmates (bottom) for students enrolled in stereotypical and counter-stereotypical departments, according to the three definitions used. The within student variation in the share of same gender classmates is obtained by taking the difference between student i's highest and lowest share of same gender classmates across all the classes attended during the first year. The within student variation in the residual share of same gender classmates is obtained by taking the difference between student i's highest and lowest residual share of same gender classmates across all the classes attended during the first year, obtained after regressing the share of same gender classmates on course $\times$ year fixed effects. Kolmogorov-Smirnov Test of equality in distribution for Stereotypical vs Against-stereotypes P-value: Variation - 0.526, 0.208, 0.139; Residual variation - 0.000, 0.000, 0.000.

Figure 1.11: Share of same gender classmates across departments

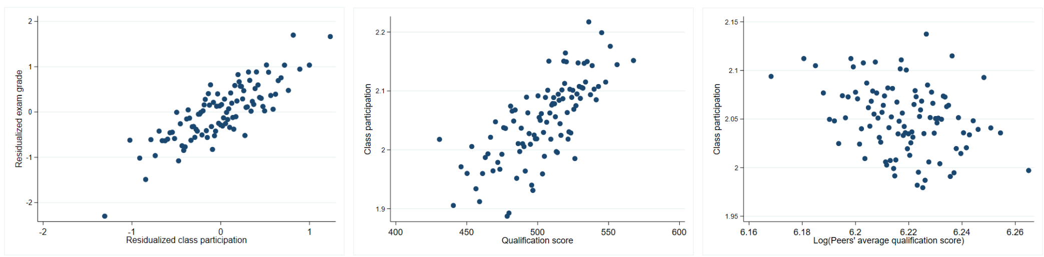
Panel A: Top & Bottom 2 **Panel B: Top & Bottom 3** **Panel C: Top & Bottom 5**



Notes: The figure displays the share of same gender classmates and the residual share of same gender classmates for students enrolled in stereotypical and counter-stereotypical departments, according to the three definitions used. The residual share of same gender classmates is obtained by taking the residuals of the regression of same gender classmates on course × year fixed effects and student fixed effects. Kolmogorov-Smirnov Test of equality in distribution for Stereotypical vs Against-stereotypes P-value: Share - 0.000, 0.000, 0.000; Residual share - 0.000, 0.000, 0.000.

Figure 1.12: Mechanisms: Class participation

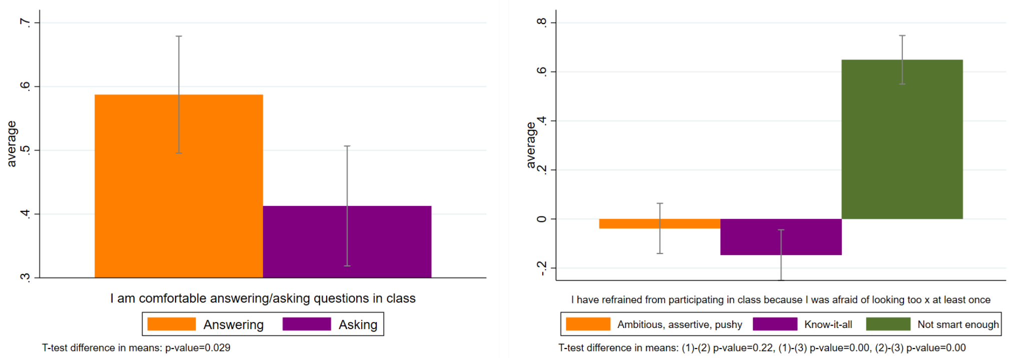
Panel A Panel B Panel C



Notes: Panel A displays a binned scatterplot of residualized exam grades on residualized participation in class. Each variable is the residual of a regression on class group and students fixed effects. Panel B displays a binned scatterplot of participation grades on students' ability controlling for class group and individual characteristics (gender, program of enrollment, school of origin, ethnicity). Panel C displays a binned scatterplot of participation grades on class group peers' ability controlling for class size, course and students fixed effects.

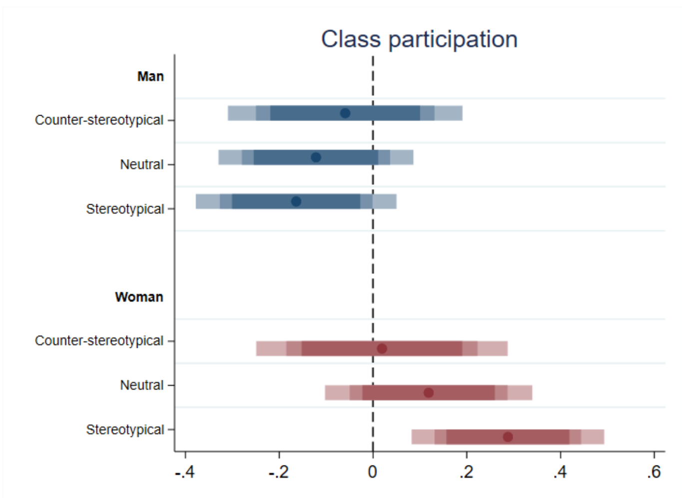
Figure 1.13: Mechanisms: Image concerns and class participation

Panel A Panel B



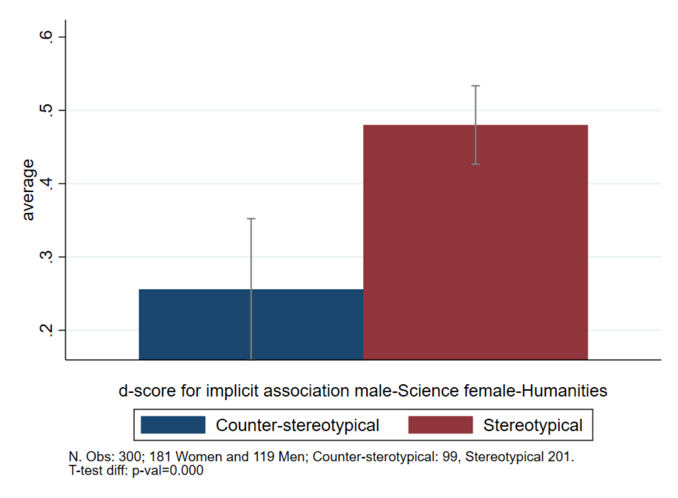
Notes: Following Bursztyn et al. (2017). Panel A displays the answers to the following questions: "Thinking about the courses you have taken in your career at LSE, how much do you agree or disagree with the following statements? (-2: strongly disagree, +2: strongly agree) (i) I am comfortable answering questions or contributing to the discussion in class; (ii) I am comfortable answering questions or contributing to the discussion in class". Panel B displays the answers to the following questions: "Thinking about the courses you have taken in your career at LSE, how much do you agree or disagree with the following statements? (-2: strongly disagree, +2: strongly agree) (i) I have refrained from participating in class because I was afraid of looking too ambitious, assertive, or pushy at least once (ii) I have refrained from participating in class because I was afraid of looking too "know-it-all" at least once (iii) I have refrained from participating in class because I was afraid of looking "not smart enough" at least once".

Figure 1.14: Mechanisms: Stereotypical selection - Class participation



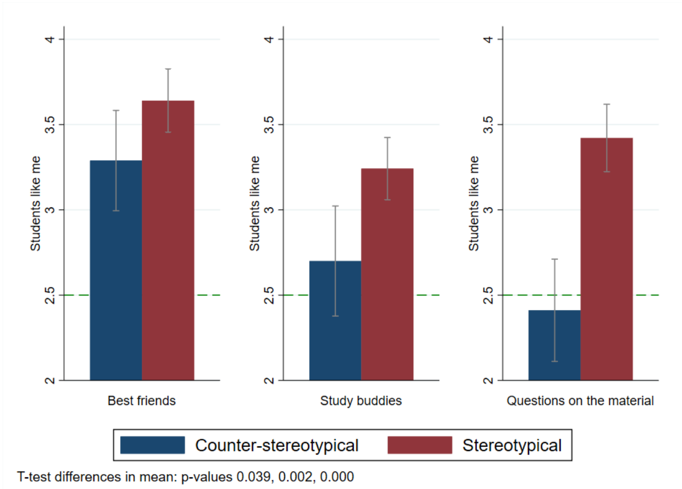
Notes: The Figure displays the results of specification 1.3. The outcomes variable are participation grades. Stereotypical and counter-stereotypical departments are defined based on the top and bottom 5 departments in terms of share of students like me enrolled in the department. Standard errors are clustered at the class group level.

Figure 1.15: Mechanisms: Scientific-Male, Humanistic-Female IAT



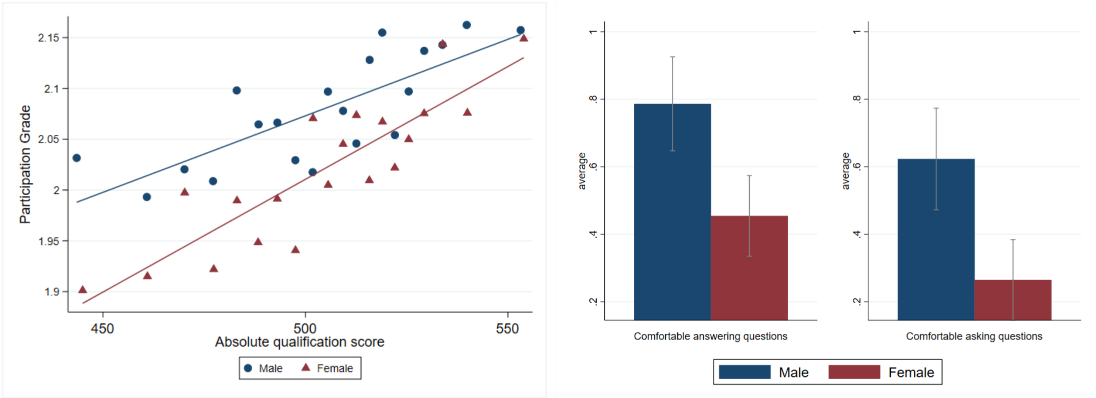
Notes: Results of an implicit association test (Greenwald et al. 1998) to elicit students' associations between female-humanistic and male-scientific. A score of 0 indicates no association between male -scientific and female-humanistic; a positive score indicates that the student associates women with humanities and men with science and math; lastly a negative score indicates that the student associates men with humanities and women with science and math.

Figure 1.16: Mechanisms: Social network



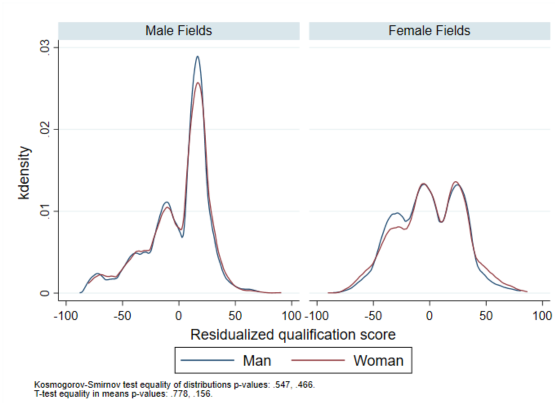
Notes: The figure displays students' replies to the following questions: "Thinking about 5 of your best friends/people you study with/people you ask questions on the material to, how many of them are women?. The answers were standardized in order to display on the y-axis the number of same gender people they nominated. An answer equal to 2.5 indicates that they interact with peers independently on their gender. The P-values for the T-test for the difference in means between Stereotypical and Counter-stereotypical are: 0.039, 0.002, 0.000. The p-value for the difference in mean between "best friends vs questions on the material" and "study buddies and questions on the material" for students enrolled in counter-stereotypical departments are 0.000 and 0.156. The p-values for the difference between the gaps for "best friends vs questions on the material" and "study buddies and questions on the material" are 0.061 and 0.008.

Figure 1.17: Mechanisms: Cost of participating



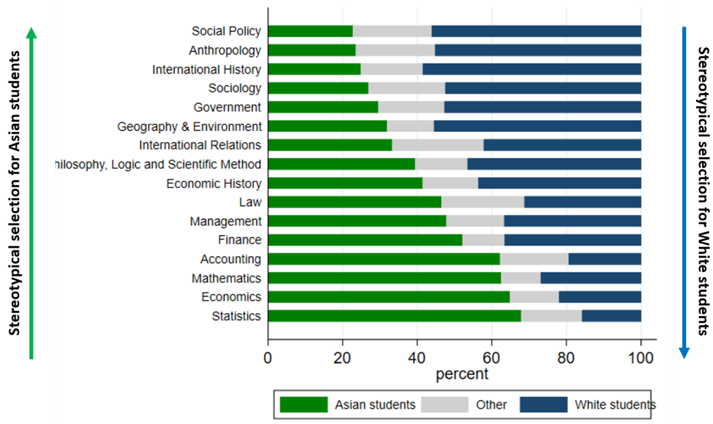
Notes: Panel A displays a binned scatterplots of class participation controlling for class group fixed effects and individual characteristics (program of enrollment, school of origin, ethnicity). Panel B displays the answers to the survey questions: "Thinking about the courses you have taken in your career at LSE, how much do you agree or disagree with the following statements? (-2: strongly disagree, +2: strongly agree) (i) I am comfortable answering questions or contributing to the discussion in class; (ii) I am comfortable answering questions or contributing to the discussion in class".

Figure 1.18: Qualification scores at entry by department and gender



Notes: The graph displays the residualized qualification score, residuals of regression of qualification score on qualification type and program x year fixed effects, for men and women enrolled in Male and Female fields. Male and Female fields are the 5 departments with the highest share of men and women among undergraduates students, respectively. The statistics include the undergraduate students enrolled in LSE between 2011 and 2017 for whom I was able to reconcile information at entry with qualifications requirements at LSE: 9.449 students, 90% of the students enrolled in undergraduate programs in LSE between academic years 2011/12 and 20117/18. Following Campbell et al. (2019), I construct a measure of individual quality based on the best three exam results among the A-level qualification scores students declared when applying to LSE. Some students take courses that are equivalents to A-Levels. In these cases I calculate their A-Level equivalence scores based on the university conversion tables for foreign students. A levels are graded on a scale of A*/A/B/C/D/E. Each A-level grade is worth 30 QCA (Qualifications and Curriculum Authority) points. P-values of two-sided Kosmogorov-Smirnov tests of equality of distributions: Male Fields - 0.547, Female Fields - 0.466. P-values of ttes of equality in means: Male Fields - 0.778, Female Fields - 0.156.

Figure 1.19: Stereotypical selection - Definition ethnicity



Note: The figure displays the distribution of Asian, White, and other ethnicity students enrolled in undergraduate programs across departments. The statistics is the average of academic years 2008-2017. The green and blue arrows at the sides display the direction of choices of majors that would be in line with stereotypes and norms for Asian students and White students respectively. Other students include other ethnicity students and students who don't disclose their ethnic group. Other ethnicity students is a residual category that includes students of other ethnic groups, but also students who provided no information regarding their ethnicity.

1.10 Tables

Table 1.1: Descriptive Statistics: Outcome measures

	N.	Mean	SD
Panel A: Course grades			
Raw	54603	60.32	16.35
Residual grades after controlling for Course FEs	54603	0.00	15.84
Residual grades after controlling for Course and Student FEs	54603	0.00	7.94
Panel B: Participation grades			
Raw	44771	2.04	.85
Residual grades after controlling for Course FEs	44771	0.00	.82
Residual grades after controlling for Course and Student FEs	44771	0.00	.56

Notes: Panel A - The sample of the analysis on performance: students for which I can observe an exam grade in at least two courses: 54.603 course-year-class group level observations, corresponding to information on 14.313 students. The residual variation is obtained by taking the residuals of a regression of exam grades on course fixed effects, and course and student fixed effects respectively. Panel B - The sample is the sample of analysis restricted to all the students for which I can observe a participation grade in at least two class groups: 44.771 course-year-class group level observations, corresponding to information on 13.485 students. The residual variation is obtained by taking the residuals of a regression of participation grades on course fixed effects, and course and student fixed effects respectively.

Table 1.2: Descriptive Statistics: Students' characteristics

	N.	Mean	SD	Min	Max
Students' Characteristics:					
Females	14313	.49	.50	0	1
Age at entry	14313	18.55	1.224551	16	56
White	14313	.36	.48	0	1
Asian	14313	.47	.50	0	1
Black	14313	.04	.20	0	1
Other	14313	.04	.19	0	1
Missing	14313	.09	.28	0	1
Single sex school	9154	.33	.47	0	1
Qualification score at entry	9449	503.90	34.27	420	540
Classes and courses:					
Course size	512	138.44	163.91	12	1011
N. classes per course	512	9.36	10.77	1	69
Class size	4886	13.43	2.45	4	23
N. classes per student	14313	3.81	0.60	2	7
Class composition:					
Share of same gender class mates	54603	.56	.16	.06	1
Share of co-ethnic class mates	54603	.35	.18	.04	1
Share of same program class mates	54447	.49	.32	.04	1
Average peers qualification score at entry	35935	501.89	15.57	420	540
Outcomes:					
Course Grade	54603	60.32	16.35	0	100
Pr(complete course)	54603	.97	.17	0	1
Participation grade	44771	2.04	.85	0	3

Notes: The descriptive statistics concern the sample of the analysis on performance: students for which I can observe an exam grade in at least two courses: 54.603 course-year-class group level observations, corresponding to information on 14.313 students. The maximum number of classes for each students are 7. Students for which we observe more than 4 classes are 5%, 85% of which attend 5 courses. These are students who attend language courses or half-unit courses during Michaelmas term.

Table 1.3: Descriptive Statistics: Variation in share of same gender classmates

	N.	Mean	SD
Raw	54603	0.56	0.162
Residual share after controlling for Course FEs	54603	0.00	0.158
Residual share after controlling for Student and Course FEs	54603	0.00	0.112

Notes: Following Olivetti et al., (2020). The table displays the share of same gender students in class and the residual share of same gender students in class obtained by taking the residuals of a regression of the class share of same gender classmates on course fixed effects, and course and student fixed effects respectively.

Table 1.4: Identification: Effect of gender on the share of same gender peers

	Share of same gender peers	
	(1)	(2)
Panel A: Gender		
Female	0.003	0.001
	(0.004)	(0.004)
Course leave-out-mean	0.993***	0.994***
	(0.016)	(0.018)
Observations	54603	44771
Course fixed effects	Y	Y
Parallel course dummies	Y	Y

Notes: Standard errors are clustered at the class group level. For each course c in academic year t I test that individual characteristics cannot predict the share of same gender peers in the class g the student was assigned to. The dependent variable is the share of same gender peers. $leave\text{-}out\ mean_{ict}$ is the share of same gender peers (dependent variable) in the course for each student. Parallel course dummies are a series of dummies that are equal to one for all the students that attend the course in the same year, and zero otherwise. The omitted category are men. Column (1) sample is the sample of the course performance analysis, while column (2) restrict the sample to the analysis on participation.

Table 1.5: Identification: Effect of other characteristics on the share of same gender peers

	Independent Variable: Share of females	
Dependent Variables	(1)	(2)
White	-0.005	-0.004
	(0.015)	(0.017)
Asian	0.014	0.005
	(0.016)	(0.017)
Other Ethnicity	-0.003	-0.002
	(0.009)	(0.010)
Unknown Ethnicity	-0.005	0.002
	(0.010)	(0.011)
Independent School	-0.029*	-0.036**
	(0.015)	(0.017)
State School	0.014	0.011
	(0.014)	(0.016)
Other School	0.015	0.025
	(0.017)	(0.019)
Mixed School	0.005	0.000
	(0.017)	(0.019)
Single Sex	-0.006	-0.015
	(0.014)	(0.016)
Not applicable	0.001	0.015
	(0.017)	(0.018)
Age at entry	-0.027	-0.051
	(0.040)	(0.044)
Qualification Score at entry	-0.374	-0.604
	(0.873)	(0.956)
Course Fixed Effects	Y	Y
Parallel Course Dummies	Y	Y
N	54603	44771
N. Tests performed	12	12
N. Tests significant at 1%	0	0
N. Tests significant at 5%	0	1
N. Tests significant at 10%	1	1
Share Tests significant at 1%	0	0
Share Tests significant at 5%	0	0.08
Share Tests significant at 10%	0.08	0.08
Total N. Tests performed	24	
Total Share Tests significant at 1%	0	
Total Share Tests significant at 5%	0.04	
Total Share Tests significant at 10%	0.08	

Notes: Standard errors are clustered at the class group level. For each course c in academic year t I test that individual characteristics cannot predict the share of female in the class g the student was assigned to. Column (2) restricts the sample to the sample for the analysis on participation. Each row corresponds to a separate regression where the independent variable is the share of females in the class and the dependent variable is the individual characteristic, controlling for course fixed effects, a dummy for each course the student attends during the academic year, and a female dummy. When the dependent variables are age at entry and qualification score at entry, a dummy for missing values is included in the regression.

Table 1.6: Main Results: Stereotypical selection

	Course grade			
	Continuous	Top and Bottom 2	Top and Bottom 3	Top and Bottom 5
	(1)	(2)	(3)	(4)
Share of students like me	-5.937***	-3.555***	-2.884***	-1.167
	(1.701)	(1.232)	(1.042)	(0.725)
Share of students like me × Neutral		3.757***	3.154***	1.028
		(1.300)	(1.132)	(0.954)
Share of students like me × Stereotypical selection	12.390***	7.759***	6.138***	3.447***
	(3.166)	(1.555)	(1.313)	(0.944)
Course fixed effects	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y
Observations	54603	54603	54603	54603
Mean Dependent Variable	60.320	60.320	60.320	60.320
	(16.345)	(16.345)	(16.345)	(16.345)

Notes: This table provides evidence of the results of Specification 1.2 in Column (1) and Specification 1.3 in Columns (2)-(4). In Column (1) stereotypical selection is defined as the average share of same gender students in the student i department of enrolment across academic years 2008-2017. In Columns (2) I consider as students making stereotypical choices, students who are enrolled in the top two departments with the highest average share of same gender students, while students who enrolled in the two departments with the lowest share of same gender students, as students who made a choice not in line with stereotypes regarding gender roles and skills. In Columns (3) and (4), I replicate the same analysis by defining stereotypical and counter stereotypical choices by considering the top and bottom 3 and 5 departments in terms of share of same gender students among undergraduates. The outcome variable is course grades, with incomplete courses coded as 0. Standard errors are clustered at the class group level.

Table 1.7: Main Results: Stereotypical selection - Grades distribution

	Course grade	Pr(Drop-out)	Grade $\geq$ 40	Grade $\geq$ 50	Grade $\geq$ 60	Grade $\geq$ 70
	(1)	(2)	(3)	(4)	(5)	(6)
Share of students like me	-5.937***	0.015	-0.056*	-0.131***	-0.198***	-0.089
	(1.701)	(0.016)	(0.031)	(0.047)	(0.066)	(0.056)
Share of students like me × Stereotypical Selection	12.390***	-0.023	0.117**	0.276***	0.386***	0.207**
	(3.166)	(0.030)	(0.058)	(0.086)	(0.124)	(0.103)
Course fixed effects	Y	Y	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y	Y	Y
Observations	54603	54603	54603	54603	54603	54603
Mean Dependent Variable	60.320	0.029	0.931	0.843	0.628	0.232
	(16.345)	(0.167)	(0.254)	(0.364)	(0.483)	(0.422)

Notes: This table provides evidence of the results of Specification 1.2. The outcome variable is course grades, with incomplete courses coded as 0 in Column (1). In Column (2) the dependent variable is a dummy equal to one if students drop-out from the course and zero otherwise. In Columns (3) to (6) the outcome variables are a series of dummies that are equal to one if grades are greater or equal to 40, 50, 60, and 70, respectively. Standard errors are clustered at the class group level.

Table 1.8: Main Results: Stereotypical selection - Gender differences

	Course grade			
	Continuous	Top and Bottom 2	Top and Bottom 3	Top and Bottom 5
	(1)	(2)	(3)	(4)
Female x Share of students like me	-2.432	-2.982**	-2.344*	-0.449
	(2.169)	(1.332)	(1.215)	(1.022)
Female x Share of students like me × Neutral		3.722**	2.985**	0.811
		(1.462)	(1.380)	(1.368)
Female x Share of students like me × Stereotypical Selection	5.252	5.536***	4.509***	1.186
	(3.866)	(2.120)	(1.635)	(1.196)
Male x Share of students like me	-10.508***	-5.910*	-4.348**	-2.238**
	(2.535)	(3.135)	(2.006)	(0.957)
Male x Share of students like me × Neutral		5.555*	4.233**	1.573
		(3.195)	(2.107)	(1.298)
Male x Share of students like me × Stereotypical Selection	21.279***	10.655***	8.112***	5.664***
	(4.838)	(3.311)	(2.256)	(1.319)
Course fixed effects	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y
Observations	54603	54603	54603	54603
Mean Dependent Variable	60.320	60.320	60.320	60.320
	(16.345)	(16.345)	(16.345)	(16.345)
Gender differences in effect:				
Share of students like me	-8.008**	-2.876	-1.974	-1.790
	(3.254)	(3.346)	(2.314)	(1.390)
Share of students like me × Stereotypical Selection	15.906***	5.071	3.576	4.463**
	(6.041)	(3.710)	(2.703)	(1.760)
Effect for students who made a stereotypical selection:				
Female		2.554	2.165*	.737
		(1.633)	(1.106)	(0.639)
Male		4.745***	3.765***	3.426***
		(1.069)	(1.008)	(0.895)

Notes: This table provides evidence of the results of Specification 1.2 (Column 1) and Specification 1.3 (Column 2-4) interacting Share of students like me and Share of students like me × Stereotypical Selection by a dummy equal to one if the student is female and a dummy equal to one if the student is male. In column (1) stereotypical selection is defined as the average share of same gender students in the student's department of enrolment across academic years 2008-2017. In column (2) I consider as students making stereotypical choices, students who are enrolled in the top two departments with the highest average share of same gender students, while students who enrolled in the two departments with the lowest share of same gender students as students who made a choice not in line with stereotypes regarding gender roles and skills. In Columns (3) and (4), I replicate the same analysis by defining stereotypical and counter stereotypical choices by considering the top and bottom 3 and 5 departments in terms of share of same gender students among undergraduates. The outcome variable is course grades, with incomplete courses coded as 0. Standard errors are clustered at the class group level. The additional tests at the bottom of the table display a test of the differences between the estimated coefficient for men and women for each specification (Gender differences in effect), and the estimated effect of an increase in the share of same gender classmates for students who made a stereotypical selection: sum of the coefficient of the Share of students like me and Share of students like me interacted with Stereotypical Selection for men and women (Effect for students who made a stereotypical selection).

Table 1.9: Mechanisms: Participation grade and cost of same-group interactions

	Participation Grade		
	All	Counter-Stereotypical	Stereotypical
	(1)	(2)	(3)
Panel A: Woman			
Share of students like me	0.185**	0.074	0.155
	(0.074)	(0.147)	(0.116)
Single Sex × Share of students like me	-0.094	-0.242	0.234
	(0.104)	(0.206)	(0.175)
Panel B: Man			
Share of students like me	-0.159**	-0.088	-0.190*
	(0.069)	(0.144)	(0.108)
Single Sex × Male x Share of students like me	0.217**	0.390	0.273
	(0.110)	(0.252)	(0.172)
Observations	29062	6596	12153
Course Fixed Effects	Y	Y	Y
Student Fixed Effects	Y	Y	Y
Effect if Single Sex = 1			
Woman	0.091	-0.168	0.389**
	(0.093)	(0.178)	(0.162)
Man	0.058	0.302	0.083
	(0.100)	(0.225)	(0.153)

Notes: This table provides evidence of the results of Specification 1.2, interacting Share of students like me with a dummy equal to 1 if the student attended a single-sex school. The sample is restricted to students for which I have information on previous schools attended. In Column (1) I am considering all students, in Column (2) students who are enrolled in counter-stereotypical departments, in Column (3) students who are enrolled in stereotypical departments. The definition of Stereotypical selection is based on top and bottom 5 departments in terms of share of same gender students enrolled.

Table 1.10: Mechanisms: Gender of Teaching assistant

	Course Grade			
	Continuous	Top and Bottom 2	Top and Bottom 3	Top and Bottom 5
	(1)	(2)	(3)	(4)
Same gender TA	-1.309*	-0.461	-0.417	-0.002
	(0.775)	(0.558)	(0.462)	(0.376)
Same gender TA × Neutral		0.192	0.105	-0.421
		(0.594)	(0.512)	(0.457)
Same gender TA × Stereotypical selection	2.635*	2.105***	1.669***	0.664
	(1.461)	(0.764)	(0.629)	(0.538)
N	26669	26669	26669	26669
Course Fixed Effects	Y	Y	Y	Y
Student Fixed Effects	Y	Y	Y	Y

Notes: This table provides evidence of the results of Specification 1.2, interacting Share of students like me with a dummy equal to 1 if the teacher's gender is the same as the student's gender. The sample is restricted to classes where I have information on the gender of the teaching assistant. In Column (1) I am considering all students, in Column (2) students who are enrolled in counter-stereotypical departments, in Column (3) students who are enrolled in stereotypical departments. The definition of Stereotypical selection is based on top and bottom 5 departments in terms of share of same gender students enrolled.

Table 1.11: Ethnicity Results: Stereotypical selection

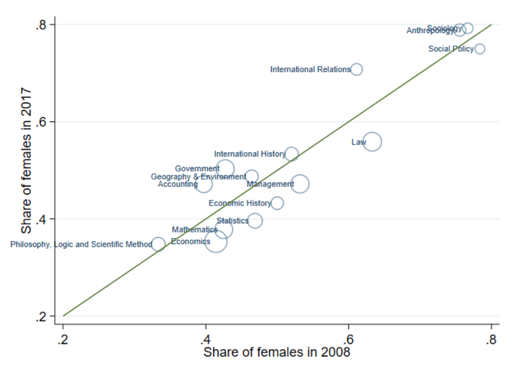
	Course grade	
	(1)	(2)
Panel A: White students:		
Share of students like me	-2.129	-2.122
	(1.828)	(1.827)
Share of students like me × Stereotypical Selection	9.144**	9.143**
	(4.065)	(4.062)
Panel B: Asian students:		
Share of students like me	-5.203***	-5.205***
	(1.723)	(1.723)
Share of students like me × Stereotypical Selection	12.899***	12.911***
	(3.375)	(3.375)
Panel C: Other students:		
Share of students like me	-11.288**	-11.242**
	(4.507)	(4.504)
Share of students like me × Stereotypical Selection	59.670***	59.438***
	(22.131)	(22.117)
Share of same gender classmates		0.551
		(0.368)
Course fixed effects	Y	Y
Student fixed effects	Y	Y
Observations	54603	54603

Notes: This table provides evidence of the results of Specification 1.6. The outcome variable is course grades, with incomplete courses coded as 0. Stereotypical selection is defined as the average share of same ethnicity students in the student i department of enrolment across academic years 2008-2017. In Column (2) I additionally control for the share of same gender classmates. Standard errors are clustered at the class group level.

1.11 Appendix

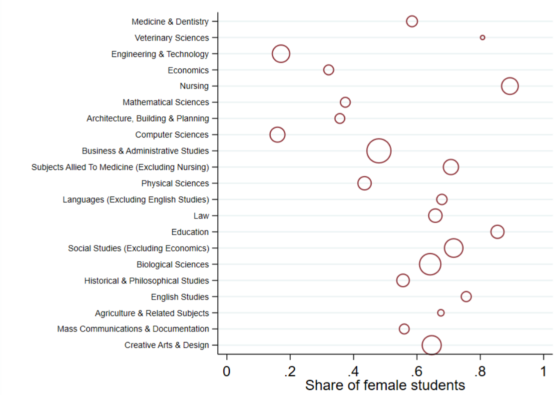
1.11.1 Additional Tables and Figures

Figure 1.A1: Departments gender composition over the years



Notes: This figure illustrates the change in gender composition of LSE departments between 2008/09 and 2017/18. It is constructed based on the number of first year undergraduate students that are enrolled in bachelor programmes offered by each department in the two academic years.

Figure 1.A2: Enrollment UK Higher Education - Gender and subject



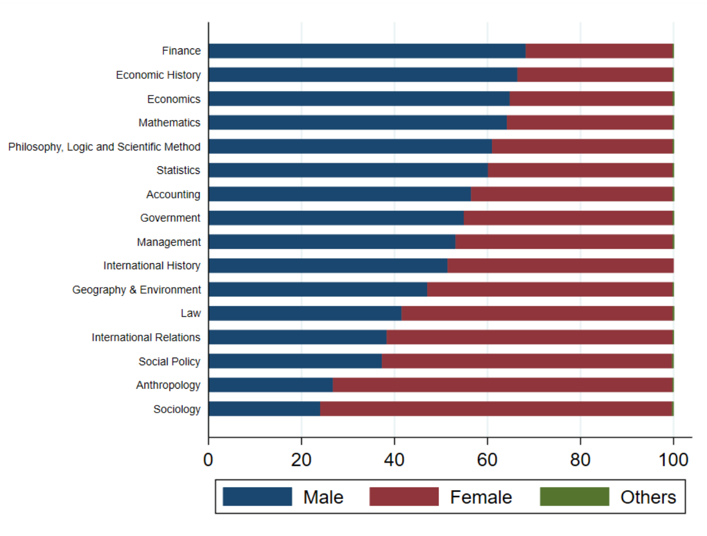
Notes: Information on the share of women among the overall population of undergraduate students enrolled in UK Universities. Source: HESA Higher Education Student Data, academic year 2018-2019.

Figure 1.A3: Share of courses outside the department of enrolment

	Academic year										
Department of Enrolment	**2008/09**	**2009/10**	**2010/11**	**2011/12**	**2012/13**	**2013/14**	**2014/15**	**2015/16**	**2016/17**	**2017/18**	**2018/19**
Accounting	0,75	0,75	0,75	0,75	0,75	0,80	0,80	0,80	0,80	0,80	0,75
Anthropology	0,44	0,44	0,44	0,44	0,44	0,38	0,38	0,38	0,38	0,38	0,50
Economic History	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,67	0,67	0,57	0,61
Economics	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75
Finance									0,80	0,80	0,80
Geography & Environment	0,50	0,50	0,50	0,44	0,38	0,38	0,41	0,44	0,44	0,44	0,44
Government	0,67	0,67	0,67	0,67	0,67	0,67	0,67	0,55	0,65	0,55	0,58
International History	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,88	0,67	0,78	0,56
International Relations	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,50
Law	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Management	0,79	0,79	0,79	0,79	0,67	0,67	0,67	0,67	0,67	0,67	0,71
Mathematics	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
Mathematics, Statistics										0,25	0,25
Philosophy, Logic and Scientific Method	0,58	0,58	0,64	0,58	0,58	0,58	0,58	0,67	0,67	0,67	0,67
Social Policy	0,52	0,52	0,52	0,52	0,52	0,52	0,52	0,55	0,43	0,38	0,38
Sociology	0,50	0,50	0,50	0,50	0,50	0,25	0,25	0,25	0,25	0,40	0,25
Statistics	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75

Notes: The table displays the share of courses outside the department of enrollment among courses in program for first year undergraduate students. Information are gathered from LSE Course Guides and Program Regulations for each academic year. When students can choose more than one course, the course is considered as outside the department if one of the choices is a course outside the department of enrolment. E.g. EC100 is considered outside the department for all the students who are enrolled in programs of study that do not fall under the Economics department.

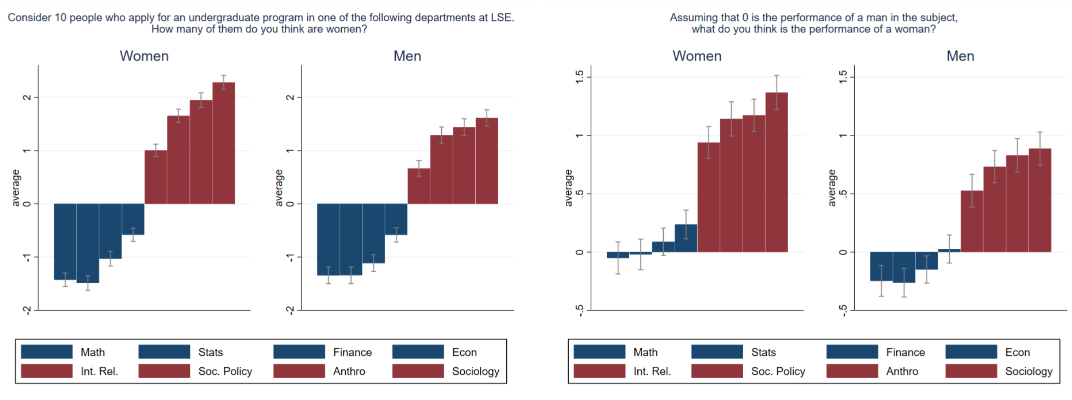
Figure 1.A4: Gender composition of applicants



Notes: This figure illustrates the gender composition of each department. It is constructed based on the number of men and women that apply for bachelor programs at the university every year. The figure shows the average across academic years 2007/08 to 2019/20.

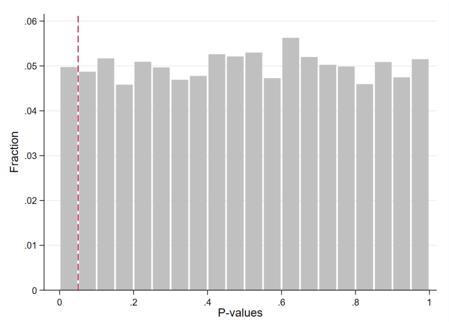
Figure 1.A5: Explicit beliefs

Panel A: Number of applicants   Panel B: Performance



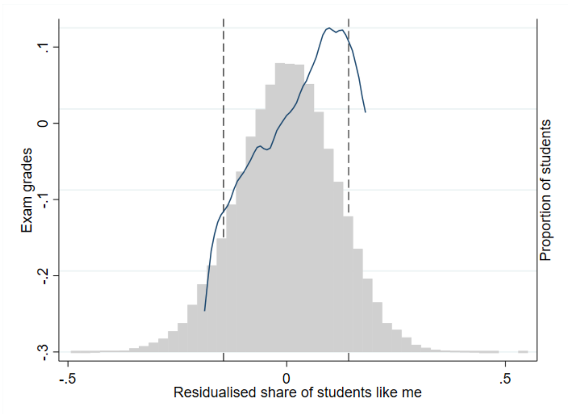
Notes: Panel A displays the answers to a question regarding the application of men and women across departments at the LSE. Panel B displays answers to a question regarding the performance of women relative to men across departments. Departments appeared in a random order to students in the survey. Questions built on Carlana, 2019 and Delfino, 2021.

Figure 1.A6: Identification: Simulated samples P-values



Notes: This graph show the p-values of a test of joint significance of the class dummies from a regression of gender on class dummies for each first year course in each academic year for each of the 1000 randomly simulated allocations. The sample is restricted to the courses that have at least 2 classes. The test is performed on all the students that attend first year courses and that did not change classes (which are the students for whom I can recover the initial class allocation).

Figure 1.A7: Robustness: Minority effect linearity



Notes: Local polynomial plot of the relationship between residualised course grades on the residualised share of same gender classmates. These are obtained by regressing course grades and the share of samge gender classmates on course and student fixed effects, respectively. The vertical lines display the 5st and 95th percentile of the share of same gender classmates. The grey histogram displays the support of the residualised share of same gender classmates.

Table 1.A1: Robustness: Peer characteristics and TA fixed effects

	Course grade				
	(1)	(2)	(3)	(4)	(5)
Panel A: Continuous definition					
Share of students like me	-5.937***	-4.033**	-5.999***	-3.175*	-6.549***
	(1.701)	(1.933)	(1.701)	(1.703)	(2.094)
Share of students like me × Stereotypical Selection	12.390***	8.757**	12.515***	6.953**	13.191***
	(3.166)	(3.619)	(3.164)	(3.171)	(3.905)
Panel B: Top & Bottom 2:					
Share of students like me	-3.555***	-2.173	-3.610***	-2.420**	-4.524***
	(1.232)	(1.353)	(1.232)	(1.230)	(1.606)
Share of students like me × Stereotypical Selection	7.759***	5.570***	7.803***	5.186***	8.848***
	(1.555)	(1.765)	(1.554)	(1.552)	(2.023)
Panel B: Top & Bottom 3:					
Share of students like me	-2.884***	-2.135*	-2.926***	-1.914*	-3.626***
	(1.042)	(1.153)	(1.042)	(1.045)	(1.303)
Share of students like me × Stereotypical Selection	6.138***	4.911***	6.165***	3.962***	7.081***
	(1.313)	(1.485)	(1.312)	(1.321)	(1.663)
Panel B: Top & Bottom 5:					
Share of students like me	-1.167	-0.649	-1.204*	-0.596	-1.182
	(0.725)	(0.813)	(0.724)	(0.722)	(0.888)
Share of students like me × Stereotypical Selection	3.447***	2.447**	3.489***	2.208**	3.187***
	(0.944)	(1.077)	(0.942)	(0.940)	(1.176)
Course fixed effects	Y	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y	Y
TA fixed effects	X	Y	X	X	X
Share of same ethnicity students	X	X	Y	Y	X
Share of same school students	X	X	Y	Y	X
Share of same program students	X	X	X	Y	X
Classmates' average qualification score at entry	X	X	X	X	Y
Observations	54603	51622	54603	54447	35883

Notes: This table provides evidence of the results of Specification 1.2 (Panel A) and Specification 1.3 (Panel B). In Panel A stereotypical selection is defined as the average share of same gender students in the student's department of enrolment across academic years 2008-2017. In Panel B I consider as students making stereotypical choices, students who are enrolled in the top two departments with the highest average share of same gender students,while students who enrolled in the two departments with the lowest share of same gender students, as students who made a choice not in line with stereotypes regarding gender roles and skills. I replicate the same analysis by defining stereotypical and counter stereotypical choices by considering the top and bottom 3 and 5 departments in terms of share of same gender students among undergraduates (Top Bottom 3, Top Bottom 5). The outcome variable is course grades, with incomplete courses coded as 0. Standard errors are clustered at the class group level. Column (1) contains the basic specification, Column (2) includes teaching assistants fixed effects, Column (3) includes controls for the share of same ethnicity and same previous school classmates, Column (4) adds the share of same program students to Column (3) controls, Column (5) includes a control for the classmates' average qualification score at entry.

Table 1.A2: Robustness: Spillovers and mechanical effects

	Course Grade			
	(1)	(2)	(3)	(4)
Share of students like me	-5.321***	-4.598**	-9.757***	-9.308**
	(1.566)	(1.818)	(2.067)	(4.511)
Stereotypical Selection			-6.365***	-6.294
			(2.093)	(4.538)
Share of students like me × Stereotypical Selection	11.116***	9.886***	18.920***	18.321**
	(2.892)	(3.399)	(3.802)	(8.211)
Observations	54603	54603	54603	14313
Course × year FEs	Y	X	Y	Y
Student FEs	Y	Y	X	X
Parallel Courses	X	X	Y	Y

Notes: This table provides evidence of the results of Specification 1.2 in Column (1). Column (2) displays the results of Specification 1.2 with no course fixed effects. Columns (3) and (4) display the results obtained by exploiting the within-course variation: Specification 1.2 with no student fixed effects, but with controls for the other courses attended by each student during the academic year. Column (3) includes all the observations, while (4) exploits one observation per student. Standard errors are clustered at the class level.

Table 1.A3: Robustness: Placebo tests

	Course grade		
	(1)	(2)	(3)
Share of students like me	-5.937***	-6.006***	-3.097*
	(1.701)	(1.701)	(1.701)
Share of students like me × Stereotypical Selection	12.390***	12.534***	6.919**
	(3.166)	(3.165)	(3.170)
Share of co-ethnic classmates		1.987	
		(1.860)	
Share of co-ethnic classmates × Stereotypical Selection		-0.537	
		(3.422)	
Share of same program classmates			4.707***
			(0.967)
Share of same program classmates × Stereotypical Selection			-2.291
			(1.770)
Course fixed effects	Y	Y	Y
Student fixed effects	Y	Y	Y
Observations	54603	54603	54447

Notes: This table provides evidence of the results of Specification 1.2 in Column (1). Column (2) displays the results of Specification 1.2 with additional controls for Stereotypical selection interacted with the share of same ethnicity classmates (Column 2), adn the share of same program classmates (Column 3). Standard errors are clustered at the class level.

Table 1.A4: Robustness: Alternative measures of stereotypical Selection

	Course grade			
	LSE Undergraduates	LSE Applications	UK HESA Undergraduates	UK HESA Staff
	(1)	(2)	(3)	(4)
Share of students like me	-5.937***	-4.890***	-7.409***	-4.125***
	(1.701)	(1.812)	(1.739)	(1.401)
Share of students like me × Stereotypical Selection	12.390***	10.410***	15.313***	8.934***
	(3.166)	(3.411)	(3.265)	(2.645)
Course fixed effects	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y
Observations	54603	54603	54603	52623

Notes: This table provides evidence of the results of Specification 1.2 using the share of same gender students among students enrolled in undergraduate programmes at LSE as proxy for stereotypical selection in Column (1). Columns (2)-(4) display the results of Specification 1.2, using different proxies for stereotypical selection: the share of same gender students among students who applied to undergraduate programmes at LSE (Column 2), the share of same gender students among students enrolled in undergraduate programmes in UK universities (Column 3), the share of same gender staff working in UK universities in each field (Column 4). Information on the data used to create the alternative proxies of stereotypical selection can be found in Table 1.A5. Standard errors are clustered at the class level.

Table 1.A5: Departments categorization

	LSE Undergraduate Enrolment	LSE Undergraduate Applications	UK HESA Undergraduate Enrolment	UK HESA Staff
	(1)	(2)	(3)	(4)
Finance	0.31	0.32	0.42	0.43*
Economics	0.35	0.35	0.33	0.30
Mathematics	0.38	0.36	0.39	0.23^{+}
Philosophy, Logic and Scientific Method	0.39	0.39	0.47	0.29
Economic History	0.39	0.33	0.53^{+}	
Statistics	0.41	0.40	0.43	0.23^{+}
Government	0.45	0.45	0.47*	0.37^{x}
Accounting	0.48	0.43	0.46	0.43*
Management	0.51	0.48	0.47	0.43*
Geography & Environment	0.51	0.53	0.56	0.40
International History	0.51	0.49	0.53^{+}	0.42
Law	0.61	0.58	0.63	0.51
Social Policy	0.66	0.62	0.69	0.65
International Relations	0.67	0.62	0.47*	0.37^{x}
Anthropology	0.76	0.73	0.74	0.51
Sociology	0.78	0.76	0.75	0.55

Notes: Columns (1)-(4) display the share of females among students enrolled in undergraduate programmes at LSE, students who applied to undergraduate programmes at LSE, students enrolled in undergraduate programmes in UK universities, staff working in UK universities respectively. In columns (1)-(3) the share is calculated as the average of the share of females in each subject across academic years 2008-2017. In column (4) the reported share is the share across academic years 2014-2018. *,+,x are symbols used to indicate that data come from aggregate statistics since the department isn't present as a separate voice.

Table 1.A6: Robustness: Global Gender Gap Index

	Course Result	
	Counter-Stereotypical	Stereotypical
	(1)	(2)
Panel A: All		
Share of students like me	-0.758	3.463***
	(1.376)	(1.101)
GGI Tercile=2 × Share of students like me	0.633	-0.761
	(1.863)	(1.474)
GGI Tercile=3 × Share of students like me	0.918	-3.275**
	(1.984)	(1.526)
Panel B: Men		
Share of students like me	-4.746*	5.260***
	(2.443)	(1.560)
GGI Tercile=2 × Share of students like me	4.965*	-0.588
	(2.985)	(2.215)
GGI Tercile=3 × Share of students like me	3.392	-5.234**
	(2.988)	(2.216)
Panel C: Women		
Share of students like me	0.306	0.112
	(1.617)	(1.307)
GGI Tercile=2 × Share of students like me	-0.671	-0.007
	(2.465)	(1.657)
GGI Tercile=3 × Share of students like me	1.099	0.315
	(2.891)	(1.785)
N	11290	20255
Course Fixed Effects	Y	Y
Student Fixed Effects	Y	Y

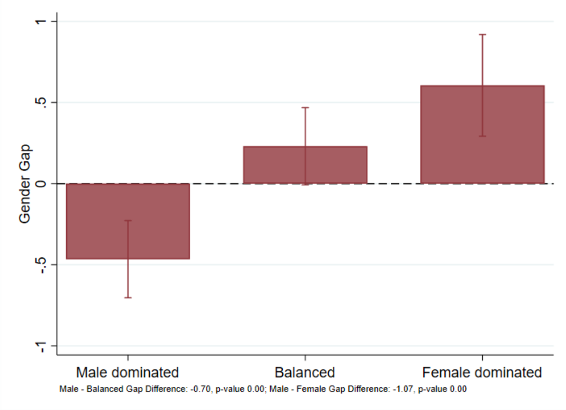
Notes: This table provides evidence of the results of Specification 1.1, interacting Share of students like me with a dummy for female and a dummy for male, and with a dummy equal to 1 if the student's country of origin belongs to the second tercile of the GGI distribution or the third tercile. The reference category are students whose country of origin belongs to the first tercile of the Gender Gap Index distribution (more unequal countries). The sample is restricted to students for which I have information on the Gender Gap Index of the country of origin. In Column (1) I consider students who are enrolled in counter-stereotypical departments, in Column (2) students who are enrolled in stereotypical departments. The definition of Stereotypical selection is based on top and bottom 5 departments in terms of share of same gender students enrolled. Standard errors are clustered at the class level.

Table 1.A7: Ethnicity Robustness: Stereotypical selection - No Other

	Course grade	
	(1)	(2)
Panel A: White students:		
Share of students like me	-1.856	-1.844
	(1.839)	(1.838)
Share of students like me × Stereotypical Selection	9.138**	9.123**
	(4.078)	(4.074)
Panel B: Asian students:		
Share of students like me	-5.739***	-5.736***
	(1.728)	(1.728)
Share of students like me × Stereotypical Selection	13.266***	13.272***
	(3.376)	(3.375)
Share of same gender classmates		0.771*
		(0.400)
Course fixed effects	Y	Y
Student fixed effects	Y	Y
Observations	45471	45471

Notes: This table provides evidence of the results of Specification 1.6. The outcome variable is course grades, with incomplete courses coded as 0. Stereotypical selection is defined as the average share of same ethnicity students in the student i department of enrolment across academic years 2008-2017. In Column (2) I additionally control for the share of same gender classmates. The sample is restricted to students who declared Asian or White as their ethnic group. Standard errors are clustered at the class group level.

Figure 1.A8: Gender gap: Course performance



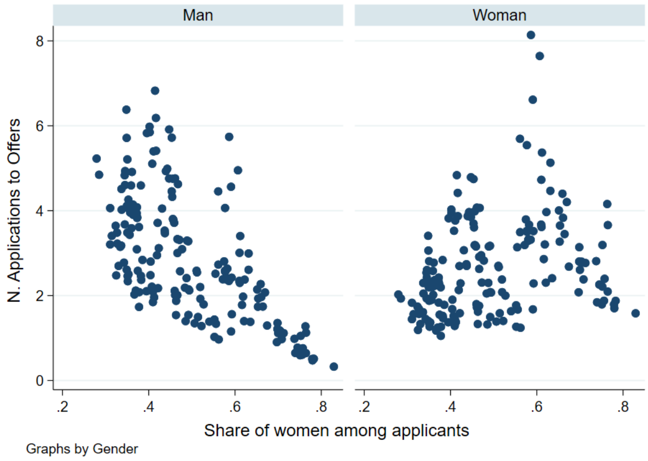
Notes: The gender gap is calculated as the difference in course performance between women and men enrolled in the same course in the same year. The sample contains all the courses attended by students enrolled in undergraduate programs in the department during the years of bachelor. Departments are categorized based on the top and bottom 5 definition of stereotypical selection.

Table 1.A8: Gender gap: Class participation

	Class participation	
	(1)	(2)
Female	-0.089***	-0.091***
	(0.014)	(0.013)
Female × Balanced	-0.046**	-0.052***
	(0.020)	(0.019)
Female × Female Fields	-0.060***	-0.068***
	(0.021)	(0.021)
Balanced	0.023	0.006
	(0.017)	(0.017)
Female Fields	0.016	-0.009
	(0.024)	(0.025)
Observations	44771	44661
Course fixed effects	Y	X
Class group fixed effects	X	Y

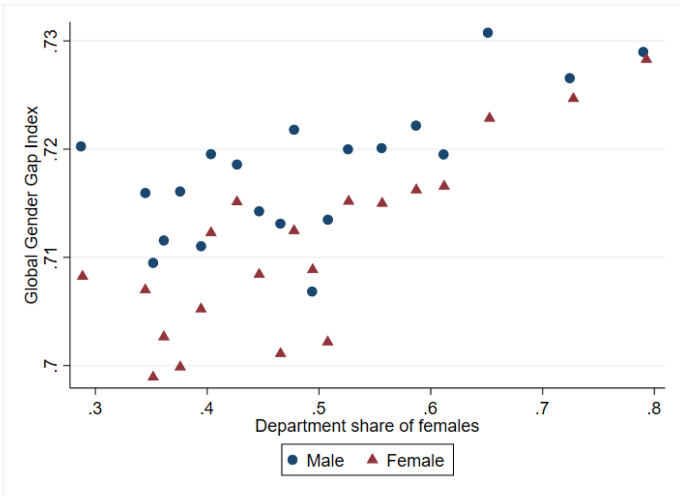
Notes: This table provides evidence of the results of a specification where class participation is regressed on course × year fixed effects (Column 1) the female dummy interacted with a dummy equal to one if the department the student is enrolled in is a balanced department and a dummy equal to one if the department the student is enrolled in is a female dominated department. The reference category are male dominated departments. The definition of female, balanced and male dominated departments is based on top and bottom 5 departments in terms of share of female students among undergraduate students. Standard errors are clustered at the class level. The outcome variable is participation grades. In Column (2) I control for class fixed effects instead of course × year fixed effects.

Figure 1.A9: Number of applications by number of offers - Gender



Notes: Information on undergraduate students applying to LSE in academic years 2007-2018. The y-axis displays the ratio of number of applications to number of offers in each academic year for each department. On the x-axis, departments are ordered based on the share of women among applicants, from lowest to highest.

Figure 1.A10: Selection into fields: Gender Gap Index



Notes: This is a binned scatter-plot that displays the relationship between the gender gap index in student's country of origin (y-axis) and the share of females in the departments the student is enrolled into (x-axis), controlling for year fixed effects. Sample: students enrolled in the first year of undergraduate programs for which I have information on the GGI of the country of origin.

1.11.2 Appendix to data

Students changing class groups

Students have the possibility to change class group during the term. Students are not allowed to change class group whenever they fancy, but changes are allowed only under particular circumstances, i.e. if the student is not able to follow the allocated seminar due to clashes with other courses that arose during the term, or external circumstances. In order to be able to change class group, they have to submit an official request. Considering the sample of analysis, this happens 5.9% of times. If students changed class group in a systematic way, class group allocation would not be exogenous anymore. In order to test that the decision to change class group does not depend on the gender composition of the class group, and that omitting students who changed class group does not generate bias, I perform two tests.

Table 1.A9 and 1.A10 show the results of regressions where a dummy equal to one when a student changed class group is regressed on gender. Columns (2) to (5) include also class group fixed effects, while Columns (3) and (5) include a control for parallel courses, a dummy for each course the student attend in Michaelmas term. In Table 1.A9 the sample is restricted to all the students in analysis (first year undergraduate students who attended the courses for the first time), while Table 1.A10 includes all the students that are allocated to a class group where a student in analysis is allocated, i.e. all the students that will contribute in defining class group gender composition. Both tables are divided into two panels. Panel A considers all the class groups the student has been allocated to during the term. If a student changed class group once during Michaelmas term, the student will be present in the sample twice for the same course. Panel B consider only one class group per student. If a student changed class group, I randomly select one of the groups to which he has been allocated.

In order to test whether the decision to change class group is independent on the group composition, I reconstruct the initial and final class group allocation for the students who changed class group. Since students normally keep the class group to which they have been allocated for all the three terms in the academic year, for 87.68% of students I am able to identify final and initial allocation by assuming that, among the two class groups observed, the final allocation is the class group the student is allocated to in Lent term. Table 1.A11 shows the results of a regression on the sample of all students who change class group in Michaelmas term for which I am able to reconstruct initial and final allocation, of the share of females in the class group on a dummy equal to one when the class group corresponds to the reconstructed initial allocation and zero if it corresponds to the final allocation (Initial allocation), a dummy for whether the student is a female, and the interaction of the two.

We can see from Panel B of Table 1.A10 that women have a slightly higher probability of changing class group. However, this probability is very small, between 0.04 and 0.07 percentage points higher with respect to a man in the same class group[55]. Furthermore, we can see from Table that the decision to change class group is independent to the gender composition of the class group. Since I am excluding all the students who changed class group when I construct the measure of class group composition, the share of female peers will be underestimated, but in an homogeneous way and independently to class group composition.

Lastly, Panel B of Table shows that once I control for parallel courses, the female dummy coefficient becomes insignificant, and Table shows that for the student in analysis, class group changes are independent on class group composition. This means that when I exclude the observation for

[55]To get a correct estimate of this probability we have to look at Panel B, since in Panel A people who changed class group are oversampled

students who change class group, I am creating an unbalanced sample, but this shouldn't generate a bias since, once I control for parallel courses (or equivalently student fixed effects) this does not vary with the gender of the student, and this is orthogonal to class group composition.

Table 1.A9: Decision to change class group - Analysis sample

	Dummy=1 if student changed class group				
	(1)	(2)	(3)	(4)	(5)
Panel A: All class groups					
Female	0.008**	0.008***	0.006**	0.013***	0.010**
	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)
Observations	61696	61696	61696	37584	37584
Panel B: Random class groups					
Female	0.004**	0.003*	0.003	0.006*	0.005
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	58593	58593	58593	34481	34481
class group Fixed Effects		x	x	x	x
Parallel Courses			x		x

Notes: The Table shows the results of a regression where the independent variable is a dummy equal to one if the student changed class group. The sample is restricted to Michaelmas term, first year undergraduate students who attended the courses for the first time. Panel A considers all the class groups the student has been allocated to during the term. If a student changed class group, the student will be present in the sample twice for the same course. Panel B consider only one class group per student. If a student changed class group, I randomly select one of the groups to which he has been allocated. Columns (1)-(3) include all class groups, columns (4) and (5) only class groups where at least one student changed class group. t statistics from standard errors clustered at class group level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A10: Decision to change class group - All students

	Dummy=1 if student changed class group				
	(1)	(2)	(3)	(4)	(5)
Panel A: All class groups					
Female	0.009***	0.011***	0.009***	0.018***	0.015***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)
Observations	76280	76280	76280	47266	47266
Panel B: Random class groups					
Female	0.005**	0.005**	0.004**	0.008**	0.007**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	72031	72031	72031	43017	43017
class group Fixed Effects		x	x	x	x
Parallel Courses			x		x

Notes: The Table shows the results of a regression where the independent variable is a dummy equal to one if the student changed class group. The sample consists of all the students that have been allocated to a class group where a student in analysis has been allocated. Panel A considers all the class groups the student has been allocated to during the term. If a student changed class group, the student will be present in the sample twice for the same course. Panel B consider only one class group per student. If a student changed class group, I randomly select one of the groups to which he has been allocated. Columns (1)-(3) include all class groups, columns (4) and (5) only class groups where at least one student changed class group. t statistics from standard errors clustered at class group level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A11: Decision to change class group - Initial allocation

	class group share of females					
	(1)	(2)	(3)	(4)	(5)	(6)
Initial allocation	0.007	0.007	0.007	0.010	0.010	0.010
	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
Initial allocation × Female	0.005	0.005	0.005	0.003	0.003	0.003
	(0.011)	(0.008)	(0.008)	(0.010)	(0.007)	(0.007)
Female	0.127***	0.072***	0.069***	0.125***	0.067***	0.064***
	(0.008)	(0.006)	(0.006)	(0.007)	(0.005)	(0.005)
Course Fixed Effects		x	x		x	x
Parallel courses			x			x
Included Students	Analysis	Analysis	Analysis	All	All	All
Observations	5222	5222	5222	7040	7040	7040

Notes: The Table shows the results of a regression where the share of females in the class group is regressed on "Initial allocation", a dummy equal to one when the class group corresponds to the reconstructed initial allocation and zero if it corresponds to the final allocation, a dummy for whether the student is a female, and the interaction of the two. The sample consists of all the students who changed class group once in Michaelmas term, for whom I was able to reconstruct the initial and final class group allocation (87.68% of cases). Since students normally keep the class group to which they have been allocated for all the three terms in the academic year, the final allocation is assumed to be the class group the student is allocated to in Lent term. t statistics from standard errors clustered at class group level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Participation grade attrition checks

Teaching assistants are required to give a performance and participation grade to all the students in the seminar. However, even if assessment is in principle compulsory, not all the teaching assistants give a feedback to the students. In Michaelmas term, participation is Missing for 16.75% of course-year-class group level observations, thus, the dataset is an unbalanced panel. However, the reason why class group participation is Missing is because the teaching assistant didn't give a grade to anybody in the class group.

As it can be seen in 1.A12 in 75% of cases, participation is Missing for all the students in the class group, while in the remaining of the cases, it is Missing because the student dropped the course or changed class group before the term ended, or the student never attended the seminar. In Table 1.A13 we can see that women have a slightly lower probability (1 percentage point) of having a participation grade, but this effect disappears when we control for course fixed effects and class group fixed effects. This indicates that women don't have a lower probability of having a participation grade, but the probability that teaching assistants give a participation grade to students is lower in courses where there are more women. The same is true for ethnicity. Regarding age of entry, even controlling for class group fixed effects, an additional year of age at entry decreases the probability of having a participation grade by 0.1 percentage points. Table 1.A14 confirms the fact that the probability of having a participation grade does not depend on the share of females or the share of students from particular ethnic groups in the class group.

Table 1.A12: Attrition checks: Participation grade - Reasons

Fraction of class groups based on presence of participation information

	MT	LT	ST
None or only general course students (compulsory)	0.34	0.19	0.95
Whole class	0.40	0.28	0.01
Whole class, except students who changed class group	0.05	0.13	0.00
Whole class, except students who were never present	0.04	0.06	0.02
Whole class, except students who dropped the course	0.01	0.06	0.00
Whole class, except combination of the above	0.16	0.28	0.02

Table 1.A13: Attrition checks: Participation grade - Individual characteristics

	Pr(Participation grade)		
	(1)	(2)	(3)
Female	-0.011**	-0.003	-0.001
	(0.004)	(0.003)	(0.001)
Age on entry	-0.001	-0.001	-0.002*
	(0.001)	(0.001)	(0.001)
Arab	-0.010	0.016	0.012
	(0.022)	(0.018)	(0.008)
Asian	-0.023*	-0.010	-0.003
	(0.014)	(0.011)	(0.005)
Black	-0.020	-0.005	0.005
	(0.016)	(0.013)	(0.006)
Chinese	-0.028**	-0.002	0.002
	(0.014)	(0.011)	(0.005)
Missing	0.017	-0.008	-0.002
	(0.015)	(0.011)	(0.005)
Mixed	-0.035**	-0.017	-0.000
	(0.015)	(0.012)	(0.005)
White	-0.032**	-0.011	-0.005
	(0.014)	(0.011)	(0.005)
Observations	54603	54603	54603
Course Fixed Effects	X	Y	X
Class group Fixed Effects	X	X	Y

Notes: The Table shows the results of a regression where the independent variable is a dummy equal to one when the student has a participation grade and zero otherwise. The independent variables are background students' individual characteristics. The omitted category for ethnicity is Other. t statistics from standard errors clustered at class group level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robustness: Qualification score at entry

	(1)	(2)	(3)	(4)
Best 3	1			
Best 3 Not-excluded	0.995***	1		
Best 3 Preferred	0.983***	0.987***	1	
Best 3 Most common	0.951***	0.954***	0.954***	1
N.	9449			

Table 1.A14: Attrition checks: Participation grade - Class group composition

	Pr(Participation grade)		
	(1)	(2)	(3)
Panel A: Gender			
Share of females	-0.097***	-0.025	-0.003
	(0.031)	(0.031)	(0.027)
Panel B: Ethnicity			
Share of Indian	0.095*	-0.017	-0.020
	(0.055)	(0.053)	(0.046)
Share of Chinese	-0.027	0.016	0.007
	(0.037)	(0.042)	(0.038)
Share of Black	0.024	-0.093	-0.115
	(0.104)	(0.091)	(0.079)
Share of Other Asian	0.024	-0.008	-0.004
	(0.052)	(0.046)	(0.042)
Share of Other	0.031	-0.023	-0.059
	(0.096)	(0.078)	(0.069)
Share of Missing	0.218***	0.010	-0.013
	(0.041)	(0.052)	(0.046)
Observations	54603	54603	54603
Course Fixed Effects	X	Y	Y
Student Fixed Effects	X	X	Y

Notes: The Table shows the results of a regression where the independent variable is a dummy equal to one when the student has a participation grade and zero otherwise. In Panel A the independent variable is the share of females in the class group. In Panel B the independent variables are the share of students from different ethnic groups in the class group, and the omitted category is the share of white students. t statistics from standard errors clustered at class group level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.11.3 Appendix to identification strategy

class group allocation does not predict students' characteristics

Following Zoelitz and Feld (2017) and Braga et al. (2016), I perform the following regression:

$$y_{itcg} = \sum_{g=1}^{n_c} \alpha_g * 1(i\text{'s } group\ = \ g) + \epsilon_{itcg},\ \forall t, c \tag{1.15}$$

where the dependent variable is a dummy equal to one if student i enrolled in course c in year t allocated to class group g has characteristic t (female, asian, white, age at enrolment), and the independent variables are dummies for each class group g in course c in year t. The dummy for class group g is equal to one if student i is assigned to class group g and zero otherwise. I run one regression for each combination of course c x academic year t to cover all the courses that first year undergraduate students attend in the academic years in sample [56]. The sample for each regression consists in all the students enrolled in course c in the same academic year t that didn't change class group during Michaelmas term [57]. Furthermore, the sample is restricted to all the courses that have at least 2 class groups.

I test that class group dummies are jointly significantly different from zero:

$$H_0 : \alpha_g = 0,\ \forall g = \{1; n_c\}$$

Table 1.A15 shows the results of the tests performed on the observed sample. The total number

[56] All first year courses plus 15 second year courses, among which 7 are language courses. Students can choose second year courses as elective courses. These correspond to 0.57% of the observations

[57] I am excluding students that changed class group since I can't observe their initial allocation. Tests of exogeneity of the decision to change class group can be found in the appendix.

of combinations of courses x academic year are 576, 523 observations belong to first year courses, while 53 to second year courses. Column 'N.' displays the number of performed regressions, column 'P¡0.05' displays the proportion of tests with a p-value smaller than 0.05, column 'P¡0.10' displays the proportion of tests with a p-value smaller than 0.10. Columns (2) - (4) show the results of regression 1.15; columns (5) - (7) show the results of regression 1.15 in which I include a dummy for all the mandatory courses that students attend in the same academic year; columns (8) - (10) show the results of regression 1.15 with dummies for all the elective courses that students attend in the same academic year; and columns (11) - (13) show the results of regression 1.15 controlling for dummies for all the mandatory and elective courses that students attend in the same academic year. We can see that when I add the parallel course dummies to control for clashes the number of performed regressions decrease. This is due to the fact that these specifications are identified out of students that are allocated to the same class and attend the same combination of courses during the year. The more courses I include in the parallel course dummies, the higher becomes the number of courses and class groups in which I don't have enough students that attend the same combination of courses. In particular, when I control for clashes with mandatory and elective courses, I lose 99 observations (combination of courses $\times$ academic years). 85 of these courses correspond to language courses in different academic years. This makes sense if we think that language courses are not part of a bachelor program, but can be attended by students that come from different programs. In the last 3 columns (14-16), I report the results of the tests obtained from the regressions in which I included dummies for mandatory and elective courses (columns 11-13), to which I add the results of the tests for the unconstrained regressions for all the Missing 109 combinations of courses $\times$ academic years.

We can see from the Table that the proportion of tests with a p-value smaller than 0.05 and 0.10 are respectively 5.8% and 12.1% in the unconstrained regressions, and remain slightly above the threshold (5% and 10%) also when I control for mandatory or elective courses separately. When I control for all the courses that the students contemporaneously attend during the year (mandatory and electives), the results of the tests seem to become consistent with random allocation of students into class groups, indicating that the exogenous allocation is conditional on course clashes. As a matter of fact, the fraction of p-values smaller than 5% is exactly 5% and the fraction of p-values smaller than 10% is 9%. In other words, students who take the same combination of courses are allocated to class groups independently from their gender. Lastly, the proportion of p-values below the thresholds remain below 5% and 0% also in columns (13) - (15).

Table 1.A15: Identification: Observed sample P-values - Gender

	Unconstrained			Mandatory			Elective			Mandatory + Elective			Combination		
	N.	P¡0.05	P¡0.10	N.	P¡0.05	P¡0.10	N.	P¡0.05	P¡0.10	N.	P¡0.05	P¡0.10	N.	P¡0.05	P¡0.10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
First year	505	0.061	0.125	505	0.071	0.119	501	0.064	0.113	441	0.050	0.091	505	0.048	0.085
Second year	48	0.02	0.083	48	0.042	0.104	48	0.042	0.063	13	0.077	0.077	48	0.021	0.0625
Total	553	0.0579	0.1211	553	0.0687	0.1175	549	0.0619	0.1093	454	0.0507	0.0903	553	0.0452	0.0832

Notes: The table shows the results of the tests of joint significance of the class group dummies in specification 1.15. The sample consists of all the students that are enrolled in courses attended by first year students in the academic years in analysis. The total number of combinations of courses x academic year are 553, 505 observations belong to first year courses, while 48 to second year courses. Column 'N.' displays the number of performed regressions, column 'P¡0.05' displays the proportion of tests with a p-value smaller than 0.05, column 'P¡0.10' displays the proportion of tests with a p-value smaller than 0.10. Columns (2) - (4) show the results of specification 1.15; columns (5) - (7) show the results of specification 1.15 in which I include a dummy for all the mandatory courses that students attend in the same academic year; columns (8) - (10) show the results of specification 1.15 with dummies for all the elective courses that students attend in the same academic year; and columns (11) - (13) show the results of specification 1.15 controlling for dummies for all the mandatory and elective courses that students attend in the same academic year. In the last 3 columns (14-16), I report the results of the tests obtained from the regressions in which I included dummies for mandatory and elective courses (columns 11-13), to which I add the results of the tests for the unconstrained regressions for all the missing 99 combinations of courses $\times$ academic years.

Table 1.A16 shows the result of specification 1.15 controlling for clashes (mandatory and electives) when the dependent variable is age on entry, a dummy equal to 1 if the student is Asian, and a dummy equal to one if the student is white respectively. The results of the tests are consistent with random allocation of students into class groups, indicating that the exogenous allocation is conditional on course clashes. In other words, students who take the same combination of courses are allocated to class groups independently from their age on entry or ethnicity.

Table 1.A16: Identification: Observed sample P-values - Other charactersitics

	Mandatory + Elective								
	Age on Entry			Asian			White		
	N.	P¡0.05	P¡0.10	N.	P¡0.05	P¡0.10	N.	P¡0.05	P¡0.10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First year	405	0.040	0.094	407	0.044	0.096	409	0.056	0.090
Second year	22	0.136	0.182	21	0.00	0.048	21	0.048	0.048
Total	427	0.045	0.098	428	0.042	0.093	430	0.056	0.088

Notes: The table shows the results of the tests of joint significance of the class group dummies in specification 1.15. The sample consists of all the students that are enrolled in courses attended by first year students in the academic years in analysis. Column 'N.' displays the number of performed regressions, column 'P¡0.05' displays the proportion of tests with a p-value smaller than 0.05, column 'P¡0.10' displays the proportion of tests with a p-value smaller than 0.10. Columns (1) - (9) show the results of specification 1.15 controlling for dummies for all the mandatory and elective courses that students attend in the same academic year. The number of performed regressions are 427, 428 and 430 for Age on entry, Asian dummy and White dummy respectively.

1.11.4 Appendix to theoretical framework

Definition and derivation of beliefs on ability

Following the framework proposed by Bordalo et al., 2019, individual i's belief on individual j and k's abilities are defined by the following expressions, assuming that individual j belongs to group G, while individual k belongs to group $-G$

$$a_j^b = a_{i,j}^b = a_{i \to j}^b = A_g + \mu_j + \theta_i \sigma(s_g)(A_g - A_{-g})$$
$$a_k^b = a_{i,k}^b = a_{i \to k}^b = A_{-g} + \mu_k + \theta_i \sigma_g(A_{-g} - A_g)$$

The strength of stereotypical distortions depends on how much stereotypes are salient for individual i, and on the strength of stereotypical associations for individual i ($\theta_i \sigma_g$).

Let us assume that there is a share s_g of type G students and a share $s_{-g} = (1 - s_g)$ of type $-G$ students in class. Student i's belief on the ability of the peers in the class ($-i$) is defined by the following expression

$$a_{-i}^b = a_{i,-i}^b = E(a_{i \to -i}^b) = A_g s_g + A_{-g}(1 - s_g) + \theta_i \sigma(s_g)(2s_g - 1)(A_g - A_{-g})$$

I assume that $E_{j \in -i}(\mu_j) = 0$ and $E_{k \in -i}(\mu_k) = 0$. This implies that in absence of stereotypical distortions, i's beliefs regarding the average ability of type G students in class is equal to A_G, the average ability of individuals belonging to group G, and i's beliefs regarding the average ability of type $-G$ students in class is equal to A_{-G}, the average ability of individuals belonging to group $-G$.

Student i's belief on their relative ability will be equal to the difference between their belief regarding their own ability and the ability of the peers in the class

$$a_i^b - a_{-i}^b = (A_g - A_{-g})(1 - s_g)(1 + 2\theta_i \sigma(s_g))$$

1.11.5 Appendix to Analysis along ethnic lines

Validity of the identification strategy

In order provide evidence that students belonging to different ethnic groups are not systematically assigned to particular classes, I replicate the tests performed for gender. The first test I perform is discussed in Section . Following Feld and Zölitz (2017, 2018) and Braga et al. (2016), I test that class group allocation does not predict students' individual characteristics by regressing individual characteristics on class dummies for each course in each year. I then test that class group dummies are jointly significantly different from zero. Figure 1.6 shows the p-values obtained from the tests of joint significance of the class group dummies for the observed sample. Regarding the White dummy, slightly more than 5% of tests display a p-value smaller than 0.05, but less than 10% of tests display a p-value smaller than 0.10. Regarding the Asian dummy, slightly less than 5% of tests display a p-value smaller than 0.05, and less than 10% of tests display a p-value smaller than 0.10. More details on the exogeneity tests and the simulation performed can be found in appendix.

In the second test, I test that students are not systematically assigned to class groups where there are more (less) students of their own ethnic group, conditional on having a certain share of same ethnicity peers among the students enrolled in a course. The specification used is the following:

$$\begin{aligned} \textit{Share of same ethnicity peers}_{iacg} = \alpha_{ca} + \beta \times 1(\text{Ethnic group}_i = e) + \gamma \times \textit{leave-out mean}_{ica} + \\ \sum_{a=1}^{n_a} \sum_{p=1}^{n_{p,a}} \delta_{p,a} \times 1(\textit{i took course p}) + \epsilon_{iacg} \end{aligned} \tag{1.16}$$

For each course c in academic year a I test that individual characteristics (Ethnic group) do not predict the composition of peers in the class group g they are assigned to. I control for the share of same ethnicity peers enrolled in the course ($\textit{leave-out mean}_{ica}$). Lastly, I control for a series of dummies that are equal to one for all the students that attend the course in the same year, and zero otherwise ($\sum_{a=1}^{n_a} \sum_{p=1}^{n_{p,a}} \delta_{p,a} \times 1(\textit{i took course p})$). This is because the allocation is constrained by the fact that students attend more than one course in the same term and they cannot attend two classes at the same time. I perform this test for all the courses where there are at least two class groups and standard errors are clustered at the class group level.

Table 1.A17 reports the result of the above regression. Students are not assigned to class group systematically based on their ethnic group. As a matter of fact, Chinese, Asian, White, Black and other minority students do not have a higher probability of being assigned to a class group with more students of their ethnic group.

Table 1.A17: Identification: Effect of ethnicity on share of same ethnicity peers

Share of same ethnicity peers		
	(1)	(2)
Panel A: Ethnicity		
Chinese	0.006	0.007
	(0.004)	(0.004)
Other	-0.004	-0.006
	(0.004)	(0.004)
White	0.003	0.005
	(0.004)	(0.004)
Missing	-0.003	-0.002
	(0.005)	(0.005)
Course level leave-out-mean	0.992***	0.991***
	(0.012)	(0.013)
Observations	54603	44771
Course fixed effects	Y	Y
Parallel course dummies	Y	Y

Notes: Standard errors are clustered at the class group level. For each course c in academic year t I test that individual characteristics cannot predict the share of same ethnicity peers in the class group g (y_{itcg}) the student was assigned to. The dependent variable is the share of same ethnicity peers. Parallel course dummies are a series of dummies that are equal to one for all the students that attend the course in the same year, and zero otherwise. The check consists in testing that the students enrolled in these courses are allocated in class groups independently on their individual characteristics, after considering the other courses they selected. The omitted category is Other Asian. This sample includes also students that are not in their first year of a bachelor program at the university. I am excluding students that changed class group since I can't observe their initial allocation. Tests of exogeneity of the decision to change class group can be found in the appendix. Column (3) to the sample for the analysis on participation.

Lastly, I produce an array of "balancing tests" to study whether the variation in the share of ethnic classmates a student is allocated to is related to the variation in a number of predetermined student characteristics: gender, previous school characteristics, age at entry, and qualification score at entry. As shown in Table 1.A18, only two of the estimated correlations appear to be significantly different from zero for the sample of analysis. They concern the characteristics of the previously attended school.

Effect of changes in ethnic composition of the class of students' performance

I estimate the effect of a change in class composition along ethnic lines on students' performance estimating the following specification:

$$y_{iacg} = \alpha_{ac} + \alpha_i + \sum_{e=1}^{n_{e-1}} [\beta_{1,e} \textit{Share of ethnic group e students}_{iacg} + \\ + \beta_{2,e} \times \textit{Share of ethnic group e students}_{iacg} \times 1(\text{Ethnic group}_i = e)] + \epsilon_{iacg} \quad (1.17)$$

Table 1.A18: Identification: Effect of other characteristics on the share of same ethnicity peers

Dependent Variables	Independent Variable: Share of Asian (1)	(2)	Independent Variable: Share of White (3)	(4)
Female	0.001	-0.004	0.016	0.013
	(0.017)	(0.019)	(0.017)	(0.019)
Independent School	0.032**	0.029*	-0.039**	-0.027
	(0.015)	(0.017)	(0.016)	(0.017)
State School	-0.026*	-0.029*	0.036**	0.035**
	(0.014)	(0.016)	(0.014)	(0.016)
Other School	0.003	0.001	-0.000	-0.002
	(0.010)	(0.011)	(0.010)	(0.012)
Mixed School	-0.008	-0.001	-0.016	-0.007
	(0.016)	(0.017)	(0.017)	(0.019)
Single Sex	0.021	0.015	0.002	0.002
	(0.014)	(0.016)	(0.014)	(0.016)
Not applicable	-0.015	-0.007	0.014	0.005
	(0.016)	(0.017)	(0.016)	(0.017)
Age at entry	0.006	0.002	-0.059	-0.047
	(0.042)	(0.047)	(0.043)	(0.049)
Qualification Score at entry	0.438	0.024	-1.197	-1.361
	(0.871)	(0.938)	(0.916)	(0.989)
Ethnicity dummies	Y	Y	Y	Y
Course Fixed Effects	Y	Y	Y	Y
Parallel Course Dummies	Y	Y	Y	Y
Program of enrollment Fixed Effects	Y	Y	Y	Y
N	54603	44771	54603	44771

Notes: Standard errors are clustered at the class group level. For each course c in academic year t I test that individual characteristics cannot predict the share of Asian or White students in the class g (y_{itcg}) the student was assigned to. Column (2) and (4) restrict the sample to the sample for the analysis on participation. Each row corresponds to a separate regression where the independent variable is the share of Asian or White classmates and the dependent variable is the individual characteristic, controlling for course fixed effects, a dummy for each course the student attends during the academic year, ethnic group dummies, and a dummy for the program of enrollment (not included in the balance checks for gender). When the dependent variables are age at entry and qualification score at entry, a dummy for missing values is included in the regression.

where the unit of analysis is students (i) belonging to ethnic group e, who attends course c, in the academic year a, and who is assigned to class g, and the independent variables y_{iacg} are course grades. $1(\text{Ethnic group}_i = e)$ is a dummy equal to one if the student belongs to ethnic group e. Students are divided in seven groups: White, Chinese, Indian, Other Asian, Black, other ethnic minorities, and students with missing ethnic group. This is the finest possible split that could be performed. Students with missing ethnic information are included in the regression so that the omitted category are White students and the coefficients estimate the effect of an increase in the share of Chinese, Indian, Black, Other Asian classmates with respect to having additional White classmates. *Share of ethnic group e students*$_{iacg}$ is the share of classmates belonging to ethnic group e, thus $\beta_{1,e}$ will capture the effect of an increase in the share of ethnic group e classmates for students in the omitted (reference) group, White students, while $\beta_{2,e}$ will capture the effect of an increase in the share of type e classmates on the gap in course grades between students belonging to ethnic group e and the reference group (White students). The specification follows the main empirical strategy of the paper, thus I include course fixed effects, α_{acg}, and student fixed effects, α_i. The assumptions underlying the identification strategy are the same as for the main empirical strategy. The validity is discussed in the previous section.

Table 1.A19 display the results of Specification 1.17 with and without individual fixed effects (Column 2 and 3 respectively. Minority status affecting performance is not a phenomenon specific to gender, since it is able to explain part of the performance gap also for ethnic minorities. Chinese students' are the group for which performance appears to be more affected by class group composition, and the effect is robust to all the specifications. In particular, the gap in performance between Chinese and White students increases by 0.999 points when the share of Chinese students in the class group increases by 10%. This represents 22% of the raw gap between Chinese and White students in first year exams. Interestingly, Chinese do not benefit only from having more Chinese in the class, but also from having more Indian and Other Asian students with respect to White students in the class group. The effect is qualitatively the same for Indian and Other Asian students: their average performance gap with respect to white students decreases if the proportion of Asian students in the class group is higher. However, the effect is not significant. This is not very surprising for Other Asian students given that the category is a spurious category, including all Asian students who are not Chinese or Indian. Interestingly, the effect for white students mirrors exactly the effect for Asian students: white students' performance is significantly lower in class groups where there is a higher proportion of Asian students. In particular, a 10% increase in the share of Indian students decreases White students' performance by 0.368 points, significant at 1%, a 10% increase in the share of Chinese students in class decreases White students' performance by 0.240 points, significant at 1%, and a 10% increase in the share of Other Asian students decreases White students' performance by 0.192 points, significant at 5%. The proportion of Black students is very low, so it is not possible to estimate precisely an effect for this group.

Given that the effects of increasing the share of Chinese, Indian and Other Asian students are very homogeneous both for Asian students and White students, in the main analyses I aggregate students of Asian background in a composite category to increase the power of the estimates.

Table 1.A19: Ethnicity Results: Minority effect

	Course grades	
	(1)	(2)
Share of Indian students	-1.121	-3.680***
	(1.451)	(1.141)
Share of Chinese students	-1.711	-2.403***
	(1.119)	(0.876)
Share of Other Asian students	-0.331	-1.912**
	(1.280)	(0.923)
Share of Black students	1.606	1.262
	(2.105)	(1.493)
Indian	0.235	
	(0.933)	
Indian × Share of Indian students	-4.166*	3.145
	(2.437)	(2.091)
Indian × Share of Chinese students	2.810	0.621
	(1.758)	(1.556)
Indian × Share of Other Asian students	-2.210	0.564
	(2.501)	(1.979)
Indian × Share of Black students	0.705	-4.848
	(4.556)	(3.531)
Chinese	-1.447*	
	(0.749)	
Chinese × Share of Indian students	7.815***	9.166***
	(1.991)	(1.830)
Chinese × Share of Chinese students	10.146***	7.526***
	(1.531)	(1.345)
Chinese × Share of Other Asian students	6.707***	6.595***
	(1.977)	(1.593)
Chinese × Share of Black students	3.897	3.545
	(3.415)	(2.812)
Other Asian	-1.349*	
	(0.818)	
Other Asian × Share of Indian students	-0.029	4.629**
	(2.362)	(1.983)
Other Asian × Share of Chinese students	0.316	1.342
	(1.652)	(1.507)
Other Asian × Share of Other Asian students	1.100	1.310
	(2.199)	(1.708)
Other Asian × Share of Black students	-2.772	-2.885
	(3.809)	(2.779)
Black	-1.280	
	(1.344)	
Black × Share of Indian students	-0.631	1.090
	(3.749)	(3.295)
Black × Share of Chinese students	-4.292*	-3.000
	(2.585)	(2.572)
Black × Share of Other Asian students	-2.981	-0.393
	(3.699)	(3.003)
Black × Share of Black students	-0.940	-2.983
	(4.995)	(4.233)
N.	54603	54603
Class size	Y	Y
Course fixed effects	Y	Y
Student fixed effects	X	Y

Notes: The table displays the results of Specification 1.17 in Column (2). In Column (1), the results of the same specification without student fixed effects are presented. The outcome variable is course grades, with non-takers coded as 0. The regressions also contain the share of students with missing information on ethnicity and the share of other ethnic minorities, a dummy for missing information and other residual ethnic minorities, and their interactions with all the other ethnic dummies and share. This is done so that the omitted category are white students, and the increase in the share of each ethnic minority can be read as an increase in the share of the ethnic minorities with respect to white students. Standard errors are clustered at the class group level.

1.11.6 Survey

During the academic year 2020/2021 I administered a complementary online survey to students (June-July 2021). The survey was designed using Qualtrics and was sent to students to their institutional email address through the university system. The survey included 4 sections. In the first section, I asked students some questions regarding their experience at the university, disregarding as much as they could the last year of remote teaching. Following Burztyn and Jensen (2019), I asked students questions regarding the extent to which they think image is important, and what elements contribute to being popular. Moreover, following Bursztyn et al. (2017), I asked students questions regarding how much they felt comfortable participating in class. This section was meant to be used to micro-found the theoretical model.

In the second section, I asked students to perform a double-target Implicit Association Test (Greenwald et al., 1998). I followed Carlana (2019) to design a Gender-Scientific implicit association test to elicit the extent to which students automatically associate Scientific disciplines with men and Humanistic disciplines with women. Subjects are presented with two sets of stimuli. The first set of stimuli are female and male names. Given the multicultural environment, in order to be neutral with respect to language differences and use names that could be identified by every student as clearly referring to male or female, I followed the approached used in the Gender-Science IAT designed by Project Implicit [58]. I used as female and male names words such as Man, Son, Father, Boy, Uncle, Grandpa, Husband, Male, Mother, Wife, Aunt, Woman, Girl, Female, Grandma, Daughter. The second set are words related to Scientific disciplines (e.g., Math, Physics, Engineering, etc.) and Humanistic disciplines (e.g. Literature, History, Humanities, etc.). The IAT is composed of seven rounds. Round 1 and 2 are practice rounds of only female and male names and only Humanistic and Scientific disciplines respectively. In the following five rounds, one word at a time (either a female or male name, or a word associated with Humanistic or Scientific in a random fashion) appears on the screen and individuals are instructed to categorize it to the left or the right according to different labels displayed on the top of the screen. In "hypothesis-inconsistent" ("hypothesis-consistent") rounds individuals categorize to one side of the screen - Humanistic Male (Humanistic Female) and to the opposite side of the screen - Scientific Female (Scientific Male). The order of the two types of rounds was randomized at the individual level. The blocks used to calculate the IAT score (d-score) are rounds 3, 4, 6, and 7. The number of words that need to be categorized is 20 in blocks 3 and 6, and 40 in blocks 4 and 7, as in the standard IAT 7-blocks (Greenwald et al. 2003). The measure of implicit association between gender and Scientific is given by the standardized mean difference score in four types of rounds. The intuition is that people with a greater implicit association of Scientific with men and Humanistic with women take longer to correctly categorize names in the "hypothesis-inconsistent pairings". Thus, the higher and more positive the d-score the stronger is the association between the two concepts. The order of the four types of blocks was randomized at the individual level. The IAT was incorporated in Qualtrics using the ad-hoc approach designed by Carpenter et al. (2021)[59].

[58] Organization of researchers founded by Dr. Tony Greenwald (University of Washington), Dr. Mahzarin Banaji (Harvard University), and Dr. Brian Nosek (University of Virginia). implicit.harvard.edu

[59] An explanation of the approach can be found at this link: https://iatgen.wordpress.com/

The third section of the survey consisted in questions regarding explicit associations. Following Delfino (2020) and Carlana (2019), I asked questions concerning their beliefs regarding the distribution of men and women and the performance of women compared to men in different fields.

Finally, the last section of the survey contained questions regarding demographic information and the students' social network.

Chapter 2

How do political parties respond to gender quotas? Evidence from South Korea

2.1 Introduction

Gender quotas in politics are currently used by 130 countries in the world (International IDEA, 2021). Although countries that have adopted quotas boast a higher share of females in national parliaments on average, quotas have not necessarily been accompanied by less conservative attitudes towards female politicians.[1] Where discrimination and hidden barriers against women persist, gender quotas might prove ineffective or even counterproductive, particularly when a majority of the incumbents are male.

In this paper, we study the reaction of highly male-dominated political parties to the introduction of gender quotas in South Korean municipal council elections over four election cycles. The South Korean setting provides a rare opportunity to study the reactions of affected politicians since the gender quota regulates only one of the two separate arms through which councilors get elected. In South Korea's mixed electoral system, the first group of councilors is elected through a plurality vote in the municipality's constituent wards. The second group is elected by party-list proportional representation. As the gender quota affects only the proportional representation arm ("PR arm"), we can study how parties strategically respond in the unconstrained ward-level plurality vote arm ("ward arm"). This enables us to detect what is typically unobservable: the preference of political parties for female candidates or lack thereof.

For our identification strategy, we use the cross-sectional variation in the number of seats reserved for proportional representation as a measure of the intensity of exposure to the gender quota. The quota imposes that all odd-number candidates in the party list for the PR arm need

[1] Across the world, the countries with gender quotas in national parliaments have an average female share of 26%, relative to 20% for the countries without (Inter-Parliamentary Union, 2021). Yet, the fraction of respondents agreeing to the statement "Men make better political leaders than women do", asked in the World Values Survey, is 47.5% in the countries with quotas - higher than the 47.1% in those without quotas. See Figure 2.A1 in Appendix Section 2.10.1 for the statistics by country.

to be female. Since at least 10% of the councilors are required to be elected through party-list proportional representation, this creates discontinuities in the intensity of the quota's bite at specific cut-offs of council size. We can thus study the effect of the quota on how political parties select candidates in the unconstrained ward arm by comparing parties in the municipalities lying just above and below the cut-offs, i.e. the treated and control municipalities, respectively.

We find that parties initially counteracted the introduction of the quota. In the treated municipalities where quotas caused the election of a higher number of women in the constrained arm (PR arm), fewer women were elected in the unregulated arm (ward arm) in the first cycle after the implementation of the reform. This can be traced back to parties putting forward fewer female candidates in the unconstrained arm, especially when the probability of the winning is higher. A significant reduction in the number of female candidates is observed when the sample is restricted to i) the two biggest parties - parties with the highest probability of getting a seat, ii) favorable ballot positions - candidates positioned high up on the ballot paper for a party tend to garner votes from voters who have a preference for the party but not for any single candidate, and iii) "safe" wards - candidates in wards where the party had a stronghold in the previous election cycle.

However, the pattern gradually reverses over time. We track municipalities over the following three election cycles and find that the treated municipalities gradually increased the number of female ward candidates. By the last election cycle, these municipalities had entirely reversed their initial reaction and, in fact, had a *greater* number of female ward candidates than control municipalities.

What is driving the gradual change in the response to quotas? Two mechanisms are consistent with this finding. The first has to do with dynamically changing attitudes towards women. The political parties might have gradually changed their attitudes towards female candidates from a position of initial aversion to greater favorability, as party members and voters learn about the competency of females through exposure to female councilors. The second mechanism concerns mere mechanical changes in the pool of female candidates available, with no change in parties' perception of women. Parties in treated municipalities might have simply experienced an initial shortage of qualified women, consequently being forced to pull women out of the ward arm to fill the quota in the PR arm. Moreover, the gradual increase in female candidates in subsequent election cycles can be explained by faster growth in the female share of the pool of candidates with political experience, as a direct consequence of the treatment.

Although it is difficult to disentangle the two mechanisms fully, our results suggest the former is more important. First of all, even if the latter mechanism was at play, the number of female councilors in the initially treated municipalities was still very small. Thus, they could not have accounted for the entire rise in female candidates in the later election cycles. Secondly, we see the same reverse in pattern over time for female candidates with zero councilor experience. Lastly, the shift in the selection towards female candidates occurs faster and stronger when the first elected female PR councilors are highly educated. The municipalities where candidate lists become more favorable towards women are those where more able women were elected in the first election cycle, providing suggestive evidence of learning about female councilors' competency.

To clarify how the treatment led to lasting effects on the strategies of individual parties, we conduct a supplementary analysis at the *party* level. We exploit a second source of discontinuity represented by the different probability of winning a PR seat at specific cut-offs of vote share, to compare the strategies of parties that marginally won a councilor in the previous election against those of parties that marginally lost. In doing so, we distinguish the effects by the position of the marginal candidate in the list, which matters for the number and gender of PR councilors elected from within the party. The focus of the exercise is to elucidate the dynamic link of how the experience of an additional female councilor from within a party translates to future selection of candidates by the party.

The learning effect occurs more strongly when the female councilor is "one of your own" rather than a "generic" woman elected in the council. When the marginal candidate was in the first position in the party list, parties that marginally won put forth a higher female share among ward candidates in the following election cycle. Recalling that a female candidate must fill the first position in the party list, close elections around the number-1 PR candidate compare parties that marginally won a woman to parties competing in the same council that did not win any PR candidate. Therefore, a party's future selection of candidates is more favorable towards women when the party gets a woman *from within the party* elected. This suggests that the learning effect is stronger when closer interactions with the woman take place.

Furthermore, we find that parties gradually change their attitudes towards women as they learn about their competence through experiencing them. First, the effect arises from the parties that listed a man rather than a woman in the second position of their list. Thus, the parties that had a prior *preference towards men* are those that update their beliefs on the competency of women upon the experience of female councilors. In addition, we find no difference in the female share among ward candidates in the following election cycle between marginally winning and losing parties, when the marginal candidate was in the second position in the list. Close elections around the number-2 PR candidate compare parties that marginally won two PR councilors with parties who won only one female PR councilor. Thus, what matters for a party's future selection of candidates is not the *share* of women among the party's councilors but whether the party experienced *at least one* female councilor.

This paper shows that quotas might deliver unexpected and undesirable effects when introduced before attitudes have changed sufficiently to accommodate them. However, they can still be effective over time despite initial resistance by the incumbents, even in settings where the target group consists of a tiny minority. With females holding only 3% of seats before the quota, South Korean politics is one such context, but it is not alone. Many countries currently have a very small female representation in politics, including Haiti, Iran, Kuwait, Lebanon, Mali, Nigeria, Qatar, and Yemen.[2] Moreover, gender quotas are increasingly being introduced in other settings such as company boards, where the incumbents are similarly, if not more, male-dominated (Ahern and Dittmar, 2012).

A large literature exists on the consequences of political gender quotas, studied in numerous contexts. This paper's main contribution lies in the fact that the South Korean setting enables

[2]See Figure 2.A2 in the Appendix for a visual demonstration of where in the world there is low female representation.

us to study the strategic responses of parties in unusually rich ways. The parallel voting system with the quota only applying to the proportional representation arm implies a whole other arm is unconstrained. Furthermore, the unconstrained arm of the election system is the way through which around 85% of councilors are elected, and therefore constitutes the more consequential arm. Such a structure of gender quota greatly expands the degree of freedom in which parties can respond, relative to quotas that reserve seats for women (Chattopadhyay and Duflo, 2004; Clayton, 2015), alternate between male and female candidate lists (Besley et al., 2017a), or mandate a minimum share of women in candidate lists (De Paola et al., 2010; Esteve-Volart and Bagues, 2012; Baltrunaite et al., 2014).[3]

Our second contribution is that we can study learning. Before the quotas, women held close to no seats in South Korean municipal councils. The introduction of quotas represented a radical shock to the status quo and, as a consequence, an opportunity for parties to learn about female councilors. Other types of quotas, such as those that mandate a minimum share of women in candidate lists, do not ensure that women end up elected. Therefore, these quotas are likely to be limited in increasing female representation to appreciable levels. For example, Bagues and Campa (2021) find that Spain's adoption of such a quota led to only a small increase in female councilors immediately afterwards and that no further gains were achieved in the medium run. Moreover, Dahlerup and Freidenvall (2013) review that among eight European countries with legislated quotas, France, Greece, Ireland, and Slovenia have no rules on the placement of females in winnable seats.

The remainder of the chapter is organized as follows. Section 3.2 provides a background on the institutional setting of South Korea's municipal council elections. We then describe the data in Section 2.3. Section 2.4 lays out our empirical strategy, and Section 2.5 discusses the results. In Section 2.6, we discuss the pieces of evidence that point towards learning as an explanation for the results in Section 2.5. Finally, Section 2.7 concludes.

2.2 Institutional setting

2.2.1 The role of municipal councils

There are 226 municipal councils in South Korea. Municipal councils represent the legislative branch that works with municipal governments, the executive branch, to oversee local matters. Councils have several legally defined responsibilities, which include reviewing and approving the spending of municipal governments, adopting and revising local bills, monitoring the municipal governments' administrative functions, and examining petitions submitted by residents. Municipal governments administer around a third of South Korea's total public expenditure (Ministry of the Interior and Safety, 2018).

[3]Before describing the richness of the ways in which parties can respond to quotas, it must first be made clear that parties determine the set of candidates running for election. The case is obvious for the PR arm, because one cannot be elected without being a member of a party in that arm. However, also in the ward arm, it is the parties that nominate the candidates to run for election. It is possible that a candidate runs as an independent, but very rarely he or she will win.

2.2.2 Electoral rules and gender quotas

Municipal councils were established during the mid-1990s, and from then, elections have taken place every four years. Seven elections were held so far, with 2018 being the latest election year. Up to the third election in 2002, all councilors were directly elected through plurality vote in single-member constituent wards. It was extremely rare to find candidates affiliated with a political party.

However, major reforms were made to the electoral rules from the fourth election in 2006. They are summarized in Table 2.1. First, the parallel voting system was introduced, where at least 10% of the councilors needed to be elected through party-list proportional representation. Among a total of 7 to 35 seats in a council, the number of proportional seats increased as a step function of the total council size: 1 for councils with up to 10 seats, 2 for those with 11 to 20 seats, 3 for those with 21 to 30 seats, and so on.

Second, the remaining seats were reserved for plurality voting in multi-member constituent wards. Each constituency elected between 2 and 4 councilors, and therefore multiple candidates from the same party could run in the same constituency.[4] Figure 2.3 illustrates what the ballot papers look like for the two arms of the municipal council elections.

Third, a gender quota was put in place: all odd-number candidates in the proportional party list needed to be female. As the numbers of seats reserved for proportional representation are small, most elected councilors turned out to be the first candidates in the lists, and therefore female. As a consequence, the introduction of quotas sharply increased the proportion of female councilors. Municipal councils were severely male-dominated prior to the reform, with only 3% of councilors being female. Due to the introduction of quotas in 2005, female representation in municipal councils reached more than 30% in the last election in 2018. Figure 3.1 illustrates how the female ratio developed in municipal councils over time. The most striking feature is the sharp rise in the female ratio immediately after the reform.

Last, subsidies were offered to parties based on the female ratio among the parties' candidates nationwide. However, it is unlikely that the subsidies affected much of the political parties strategies, particularly at the municipality level. The scale of the subsidies have been criticized for being too low to effectively expand female nomination (Jin, 2018; Kim et al., 2003; Lee, 2003). Indeed, they account for only around 5 to 6% of the total value of election subsidies (National Election Commission, 2018). Therefore, the presence of the subsidies are unlikely to have impacted political parties' selection of candidates.

Amendments to electoral rules continued between the 2006 and 2010 elections. It was stipulated that in either the municipal council elections or the higher-up provincial council elections, there must be at least one female candidate in each general election district. As there are around 250 general election districts, compared to 226 municipalities, a general election district approximately compares to a municipality.[5] Legislative Impact Analysis Reports indicate that most parties chose to satisfy this rule in the municipal council elections, due to the larger number of candidates (Lee,

[4]The maximum number of candidates a party could nominate for a ward equalled the preset number of seats for that ward.

[5]General election districts are divided depending on population size and local representativeness. A large municipality may contain five general election districts, and up to five small municipalities may comprise a general election district.

2019). Selecting which ward to place the female candidate would have been a strategic concern for the political parties.

2.2.3 Background behind the adoption of gender quotas

If some parties had led the move for the reform against opposition from other parties, then we should recognize that parties' strategic responses to the quota might be very heterogeneous in nature. Thus, here we discuss the background behind the adoption of the quota.

Before gender quotas were adopted in the municipal council elections, they were adopted first in the general election for the National Assembly in 2004. The adoption was influenced by increasing demands by women's organizations to raise female representation in politics, which at the time was dramatically behind the international average.[6] As females constitute half the voters, it was in the interest of political parties to put gender quotas forward amongst their election pledges. Moreover, there are views that the adoption of the quota was also a political tactic (Jeon, 2013). Political parties wanted to increase the size of the National Assembly back to what it was before the size cut during the Asian Financial Crisis, and the fact that the majority of the added seats will go to females, with the quota, made for a good excuse to expand the Assembly.

Once the quota was adopted in the general election, it became the natural next step to introduce it in the regional elections. The gender quota in the municipal council election was passed in the National Assembly, led by both major parties. Some argue that there was political motivation behind it, too (Kim, 2005). One new element in the reform was the party nomination system – a ward candidate must be nominated by their party in order to run with the party affiliation – but it was disputed as a ploy to deepen party influence. Political parties used the quota to justify the party nomination system, since the gender quota was embedded in the proportional representation arm where party nomination was essential.

To sum, it is difficult to say that there was a major division among political parties in their support of the gender quota when it was passed.

2.3 Data

Two sources of data are used. First, data related to the execution of the elections are collected by web scraping the website of the National Election Commission. The website posts detailed data on all past elections, including population, candidate information, and vote outcomes. Second, to examine the consequence of the municipal councils' legislative activities, we use the data on municipal governments' expenditures from the Local Finance Disclosure System of the Ministry of the Interior and Safety.

2.3.1 Population

Because ward divisions are centrally determined based on population size, population data is published. The number of residents is available by ward, voting eligibility, gender, and citizenship

[6]See Cho and Kim (2010) for a summary of the major activities of women's organizations.

status. Moreover, the data includes the number of households by ward. This data is used to perform balancing checks in order to validate the identification strategy, which relies on the assumption that municipalities locally around the PR seat thresholds are similar.

2.3.2 Candidates

Various background characteristics of all candidates are also made publicly available by the National Election Commission. These are election arm (ward or PR) classification, election district name, candidate number, party affiliation, name, gender, date of birth, age, occupation, education, and pertinent work experience. Whether a candidate is favored by his or her party is revealed by the election arm and candidate number. Typically, candidates that are deemed less competitive are placed on the PR election arm, and the candidate numbers directly translate to the position on the ballot, in which higher positions attract more votes.

Figure 2.2 illustrates how the female share among candidates have been increasing continually, even when not stipulated by the quota. In particular, plot [b] shows that more females are running in wards as the sole candidates of their parties, and plot [c] shows that more females are taking the highest ballot positions even when multiple same-party candidates are running. Plot [d], on the other hand, shows that more females are taking the even-number party list slots, which would not happen with a strong preference for men.[7]

2.3.3 Votes

The website of the National Election Commission also includes vote counts by ward. These vote counts enable us to see in which wards parties have their strongholds. Therefore, we can categorize wards into safe and contestable ones in the perspective of the political parties. Parties would then allocate their favored and less favored candidates to different wards accordingly.

Moreover, we can learn by which margin the winners won. In the regression discontinuity identification strategy, we rely on the assumption that close victories result in sharp changes in the composition of councilors by party, in an environment where parties enjoy similar degrees of popularity from the voters.

Electoral outcomes determine the gender ratio of the elected councilors. Table 3.1 provides descriptive statistics on the gender composition of councils by election cycle. The table also depicts how the reform in 2005 introduced the PR arm as well as the gender quota in that arm.

2.3.4 Municipal budget

Municipal budget data is used to perform balancing checks, to show that municipalities locally around the PR seat thresholds are similar in terms of economic scale and council performance. The budget of a municipal government reflects the economic prosperity of the municipality, as around a half is sourced from local tax and non-tax revenue. In addition, data is available on the share of the municipality's expenditures spent on running the municipal council (2002-2020).

[7]In the latest general election of 2020, where the same gender quota on the PR arm applies, almost all PR candidates in positions 2, 4, and 6 are male.

There have been numerous accusations in the past of councilors appropriating large sums of the local budget for their private use (Local Decentralization Bureau – Election and Local Council Division, 2019). For instance, they would go on international policy-research trips where the itinerary largely consists of sightseeing. Another example is of councilors ordering member pins made of pure gold. As such, a measure of the performance of a council is the frugality of its operation costs. Newspapers have traditionally included it in their assessments of councils (Jang, 2008).

2.4 Empirical Strategy

2.4.1 Regression discontinuity design around the number of PR seats

To get at the causal effect of the gender quota, we make use of the fact that the gender quota affects municipalities at different intensities depending on the proportion of the PR seats in the council. The number of PR seats increases as a step function of municipal council size, which is pre-determined centrally by the National Election Commission based on population size and regional representativeness. The step function is depicted by the navy dots in Figure 2.4.

The regression discontinuity design compares the characteristics of ward and PR candidates in municipalities on each side of the step function's thresholds, while controlling for council size. In order to account for the fact that there is not just one but many thresholds, we categorize councils into bins based on the proximity to thresholds, as illustrated in Figure 2.4.

Therefore, this strategy estimates the effect of an additional PR councilor *in general*, rather than an additional *female* PR councilor. Nonetheless, while the gender quota does not necessitate that the second PR councilor be male, in practice almost all PR councilors end up being female.[8] This fact is due to PR candidates even in even-number positions frequently being female, and also due to PR councilors frequently being the number-1 candidates of multiple parties.[9] We also check in Section 2.4.3 that an additional PR councilor strongly implies an increase in the number of female PR councilors.

The regression discontinuity specification is given by:

$$Y_{cbt} = \sum_{s=4}^{7} \beta_s \times Treat_{cbt} + f(x_{cbt}) + \delta_b + \gamma_t + \epsilon_{cbt} \tag{2.1}$$

where Y_{cbt} denotes the outcome variable for municipal council c belonging to bin b in election cycle t. The running variable is $x_{cbt} \equiv (\text{council size})_{cbt} - \text{threshold}_b$, with $\text{threshold}_b \in \{10, 20, 30\}$. In addition, $Treat_{cbt} \equiv 1(x_{cbt} \geq 0)$, signifying an additional PR councilor. Therefore, β_s estimates the effect of having an additional PR councilor, pooling all the bins together, in election cycle s. Moreover, the baseline function form of f is linear, and we do not allow for the effect of x_{cbt} to differ to the left and right of the threshhold. The reason for this choice is that making f quadratic or allowing for differential trends on either side of the threshold barely makes a difference.

[8] Table 3.1 shows that among PR councilors, 87% to 98% are female each election.

[9] Appendix Table 2.A1 shows that it is relatively rare to find multiple PR seats getting allocated to the same party.

Another specification, based on treatment status at election cycle 4, is:

$$Y_{cbt} = \sum_{s=4}^{7} \beta_s \times (\mathit{Treat\ at\ cycle}\ 4)_{cb} + f(x_{cb4}) + X_{cbt} + \delta_b + \gamma_t + \epsilon_{cbt} \tag{2.2}$$

where $(\mathit{Treat\ at\ cycle}\ 4)_{cb} \equiv Treat_{cb4}$, and X_{cbt} denote control variables such as council size or the number of ward seats.

The outcomes we consider are the number of ward and PR candidates by gender. A factor to note is that when the outcome variable relates to the ward elections, we change the running variable to $\tilde{x}_{cbt} \equiv$ (number of ward councilors)$_{cbt}$ − (number of ward councilors at the threshold)$_b$, for ease of interpretation.[10]

2.4.2 Contemporaneous treatment vs. treatment at cycle 4

Equations (2.1) and (2.2) estimate the effects of contemporaneous treatment and initial treatment, respectively. In practice, there is barely any difference which specification we use, because the treatment status changes after election cycle 4 for only 3.7% of the councils. We settle on equation (2.2) as our main specification, though. The first reason is that the initial treatment assignment is more exogenous. Upon the first treatment, the treated and control municipalities may evolve on different paths, which would make them no longer balanced at the subsequent election cycles. Secondly, measuring the effect of the initial treatment maintains the same composition of treated municipalities. If the effect of contemporaneous treatment, specified by equation (2.1), varies over time, then it is unclear whether it is due to the small number of councils that are changing their treatment status, or due to the same councils reacting differently to the treatment over time. With equation (2.2), we can safely conclude that it is the latter.

2.4.3 Did the quota bite?

Because we are interested in the consequence of the change in the gender composition of councilors brought about by the quota, it is important to verify that there is a change in the number of *female* PR councilors at the discontinuity thresholds.

Table 2.3 reports the results of regressing (2.2) with the number of female PR councilors as the outcome variable, separately for each bin. While having an additional PR councilor at cycle 4 significantly increases the number of female PR councilors over all the cycles at bins 1 and 2, there is no such effect at bin 3. Moreover, there are very few observations at bin 3. The regression results of Table 2.3 are echoed by Figure 2.5, which shows that the average number of female PR councilors sharply increase at the thresholds of bins 1 and 2, but not at bin 3. Therefore, in the reduced-form results that follow, we restrict the sample to bins 1 and 2.

[10] If we keep the running variable based on council size, then the regression estimates the effect of $Treat = 1$, i.e. having one more PR councilor, while controlling for council size. Then in the regression, the councils with $Treat = 1$ effectively have one fewer ward councilor than those with $Treat = 0$. Therefore, it becomes more difficult to interpret the sign of the coefficient on $Treat$ when the outcome variable relates to ward elections, e.g. the number of female ward councilors or candidates. When the running variable is based on the number of ward councilors, however, we are free from this problem. Changing the running variable this way does not change much else. In fact, the coefficients $\hat{\psi}_0$ and $\hat{\psi}_1$ stay the same, as well as the R-squared value.

We next focus on the treatment effect over time. In both columns (1) and (2) of Table 2.3, the effect of the treatment at cycle 4 remains similar over the election cycles. Because the vast majority (96.3%) of the initially treated municipalities continue to get treated each cycle, the constancy in the coefficients implies that first-stage effect of the treatment - increasing the number of female PR councilors - is constant, too. This constancy implies that the effects on other outcome variables, i.e. the reduced-form treatment effects, should also be constant over time unless the initial treatment leads treatment and control groups on different paths.

The standard errors are clustered by municipality for two reasons. First, the variation of the initial treatment variable is at the level of the municipality. Second, parties formulate strategies chiefly within a municipality, rather than moving around candidates across municipalities. In fact, there are many factors that tie down a candidate to a certain municipality to be nominated in. A candidate is legally required to have been a resident of the municipality they are running in for at least 60 days prior to the election. In addition, as municipal councilors deal with local grass-roots matters, a candidate familiar with the municipality will win more votes *ceteris paribus*. Hence, a candidate usually runs in the municipality they have a connection with, such as their birthplace, long-term residence, or place of education. Moreover, the final say of a party's nomination lies on the head of the municipal branch of the party, so a candidate typically serves the local activities of the party in the municipality they desire to run in for a long time before getting nominated. Finally, once a candidate is nominated in a municipality, they put on a campaign and become known to the residents. So if they were to run again, they would not start over at a new location. For all these factors, rarely do parties move around candidates across municipalities for strategic reasons.

As a way to buttress the validity of the regression discontinuity design, Appendix Section 2.10.2 formally tests and confirms that as council size increases, there is a change in the number of female PR councilors only at the thresholds and at no other point.

2.4.4 Validity of the regression discontinuity design

Balance tests A critical part of the identification strategy is that there are no confounders associated with the treatment status at election cycle 4. We regress equation (2.2) for various pre-determined characteristics, to check that they are balanced to the left and right of the threshold. The sample consists of councils at election cycle 4, and the regression results are presented in Table 2.4.

Panel (A) confirms that the population characteristics are balanced. In particular, the voting age population by gender is no different, alleviating the concern that the preference for female councilors among voters may be different between the treated and control municipalities. In Panel B, columns (8) and (9) refer to the vote share received by each main party in the previous election's PR arm. Columns (10) and (11) show that the initial treatment group is balanced in terms of economic prosperity and council performance. Columns (12) and (13) demonstrate that the structure of the ward election arm is balanced, as there is no difference in the number or size of wards between the treatment and control municipalities.

Bunching Is there a possibility that there is gerrymandering? For example, a council may manipulate its constituent areas to manipulate the council's size and therefore, treatment status. If there is manipulation, one evidence of it would be bunching at the threshold. Figure 2.6 displays the histogram of the frequency of municipalities by council size. Visually, it is hard to say there is bunching around the thresholds of 11 and 12. In addition, it is difficult to formally test for bunching around the threshold, e.g. the McCray (2008) density test, due to the coarseness in the council size variable.[11] However, there are specific electoral rules against gerrymandering.

The division of election constituencies is determined by the Municipal Council Election Committee. The committee is set up in each district, and it consists of up to 11 members appointed by the district mayor among the individuals nominated by the media, legal community, academic community, civic groups, the district council, and District Election Committee. Municipal councilor or party member cannot be in the committee. The committee determines the council size based on population, administrative districts, topography, transportation, and other conditions. The committee cannot split the smallest administrative district and make it a part of another ward. In sum, there are rules preventing the membership of interested individuals in the committee and also rules circumscribing how the election constituencies are drawn up.

2.5 Main Results

2.5.1 The numbers of candidates and councilors by gender

The results of regressing equation (2.2) are reported in Table 2.5. The most interesting result is captured by columns (1) and (2). In response to the treatment at cycle 4, parties initially put up more male ward candidates but gradually decrease the number of male ward candidates. Eventually, at election cycle 7, the parties in the treated municipal councils put up fewer male candidates than those in untreated councils. As for female ward candidates, the opposite pattern holds: the coefficient sign changes from negative (albeit statistically insignificant) to positive. Thus, the way parties select candidates in reaction to the gender quota is changing over time.

Focusing next on the columns for the councilors, we can see that similarly, the number of female ward councilors in the treated municipalities is lower in the beginning but is higher at the end.[12] Moreover, the higher number of female PR councilors in the treated municipalities at election cycle 4 more than compensates for the lower number of female ward councilors. Consequently, column (10) shows that there are statistically insignificantly more female councilors as a whole at election cycle 4 in the treated municipalities. Then, the coefficients for the later cycles grow in magnitude and become statistically significant.

What is driving the changing reaction to the gender quota? The parties in municipalities that got the initial treatment are changing their behaviors, becoming more female-friendly in their endorsement of candidates over time. A possible explanation is that that although parties countered the gender quota initially, the quota was not completely undone, as signified by the

[11] No bunching is rejected for randomly selected cutoffs of council size.

[12] Due to the addition of the running variable in the regression, mechanically the coefficients of columns (5) and (6), as well as those of columns (9) and (10), are of opposite signs. Also mechanically, the coefficients of columns (7) and (8) add up to 1.

positive coefficient for cycle 4 in column (10). Then, the consequent experience of female councilors induced parties to become more favorable towards female councilors. Section 2.6 explores deeper into this learning story.

2.5.2 Focusing on candidates likely to get elected

Because the analysis is at the municipality level, it is not straightforward to pin down where the effects are coming from. Many parties operate in a municipality, and each party puts forth a large number of ward candidates per ward. The changes in the composition of the electoral body may not mean much if the change in the candidate selection pattern is driven by parties or candidates in positions that have no hope in getting elected. To explain the source of the changing candidate selection with greater clarity, we point to Table 2.6.

Table 2.6 shows that even when we restrict our attention to candidates for whom election is probable, we see the same patterns of (a) the initial preference for males, and (b) the shift in the preference for females. In this table, the sample is restricted to main parties only: the Conservative party and Progressive party.[13] Columns (1) and (2) reproduce Columns (1) and (2) of Table 2.5 but for the main parties only. Columns (5) and (6) are even more selective; these are the candidates of the main parties, running as the candidates in the high-up positions on the ballot for the party in a ward (position 1 if the ward elects 1-2 councilors, and positions 1 and 2 if the ward elects 3-4 councilors). These candidates have a great chance of getting elected.

Because political parties can choose how many candidates to champion, it is difficult to interpret the results in columns (1) and (2) together.[14] Column (4) reports how the number of female ward candidates compares between treatment and control municipalities, when the total number of ward candidates is no different.[15] Hence, column (4) looks at how the *substitution* between the two genders occurs, when the total number of candidates is pre-set. Again, there are fewer females in cycle 4, and gradually more females afterwards. The pattern is mirrored when we restrict the candidates to those in useful positions.

Table 2.7 uses a different way of gauging the election probability of candidates. It shows how parties select candidates for different types of wards. We categorize wards based on whether the party had a stronghold in the previous election, in which case we call the ward "safe."[16] These are wards where we can assume the party candidates have a very high probability of being elected. Since whether a ward is safe is dependent on the party at hand, the regressions in Table 2.7 are at the ward×party level. We can see in column (5) and (7) that parties had a strong preference for placing male candidates in safe wards in cycle 4, especially in useful seats. The preference for men, however, disappears from the cycle 5 onward. Furthermore, we can see in columns (6) and (8) that,

[13]As it can be seen in Appendix Table 2.A2, the majority of candidates and elected councilors is affiliated with these two main parties – 54% and 83% on average, respectively – and their importance increases over time.

[14]The maximum number of ward candidates for a party is the total number of ward seats in the council, but there is no minimum.

[15]One may argue that the total number of ward candidates is a "bad control" in the regression, because it is an outcome of the treatment (Angrist and Pischke, 2009). However, as the total number of ward candidates equals the sum of the number of male and female ward candidates, the controlled regression in column (4) can be simply interpreted as a "summarization" of columns (1) and (2) together; if we were to regress the number of male ward candidates while controlling for the total number of candidates, the regression coefficients would equal exactly the negative of the coefficients in column (4).

[16]A party is considered to have a stronghold in a ward if the party wins the greatest vote share in the PR arm in the ward, and it got over 10 percentage points more vote share than the next popular party.

from cycle 5, parties in treatment municipalities started placing more women among candidates in unsafe wards. This is what is driving the overall increase the number of female candidates we observe at the municipality level. Therefore, although the number of female ward candidates increases faster over time in treatment than in control municipalities, the gains for women still remain bounded to the less-preferred wards with lower likelihood of election.

2.5.3 Placebo test

In the last sections, we found that parties reacted to the introduction of quotas in the PR arm by reducing the number of women among ward candidates immediately after the reform. The substitution away from women was stronger in ballot positions and wards where party candidates had higher chances at getting elected. However, from cycle 5, parties in municipalities above the threshold started placing more women among ward candidates.

Before going into the potential mechanisms that can explain the effects we observe, we report the result of a placebo test where we check that the thresholds are meaningful only after and not before the reform to the election system. This test provides supportive evidence that we are estimating the effect of the introduction of the quotas, and the results are not driven by treated municipalities being different from control municipalities ex-ante.

We test whether the number of male and female ward candidates changed at the threshold before and after the reform. We want to make sure that the probability of getting an additional PR seat upon the reform is not correlated with other factors that affect the number of male and female candidates, prior to the reform. Table 2.8 shows that up to election cycle 3, the effect of being past the threshold is not statistically significantly distinguishable from zero. It is at election cycle 4 that the treatment induces an effect, as expected.

2.6 Learning about women's competence

There are two broad mechanisms that can explain the initial substitution towards men and the following gradual increase in women among the candidates of parties in treatment municipalities. In the first mechanism, parties dynamically change their attitudes towards women. The initial observed substitution towards men could be due to parties preferring men as candidates and politicians, while the gradual rise in the number of women could be explained by parties' attitudes towards female candidates changing as they learn about women's competence through the exposure to female councilors. However, the dynamic treatment effects can also occur with absolutely no change in parties' perception of women. In the second mechanism, the effects are driven by mechanical changes in the pool of female candidates available. There might have been a shortage of women among the pool of potential candidates when quotas were introduced. If very few women were willing or were qualified enough to become councilors upon the introduction of the quotas, parties that were obliged to have more women among PR candidates would have had fewer women left to place among ward candidates. The overall increase in women in the following cycles then could be explained by the pool of experienced women increasing more in treated municipalities with respect to control ones thanks to the quotas.

We argue in the following sections that the first mechanism on learning appears to be more important.

2.6.1 Alternative mechanism: no learning, but mere change in the size of the pool of qualified women

Even if there is absolutely no learning by parties, the patterns that we see in the number of female candidates and councilors could be consistent with the initial shortage and gradual growth of qualified or experienced women. We provide evidence that this mechanism cannot be the whole story.

First of all, for the treatment effect of fewer female ward candidates in cycle 4 to be explained by a shortage of women, we should find that parties in treatment municipalities have greater trouble finding female candidates. We test this hypothesis. A party is defined as unconstrained if the number of female candidates in its party list is greater to the number of women it needs to list as candidates due to quotas.[17] Table 2.9 shows that parties below the threshold are not more unconstrained in the number of women they can list as candidates, particularly in cycle 4.

Secondly, if the only reason for the treatment effect of rising number of female ward candidates over cycles 5 through 7 was the greater availability of experienced female candidates, then there should be no shift in the gender preference among candidates who have zero councilor experience. However, Table 2.10 shows that the initial withdrawal of female candidates and the gradual reversal is present even for ward candidates who have never been elected before. Hence, the greater availability of experienced women in treated municipalities after the introduction of the quotas cannot be the only reason driving the changing reaction. In addition, even if every municipality had women in its parliament in cycle 4 thanks to the introduction of the quota, the number of female councilors was still very small and so it cannot account for all the rise in the female candidates over time.

2.6.2 Is there learning about women's competence?

Then, are parties really learning about women's competence? If parties are learning about the competency of females, then the learning can be expected to take place at a faster rate if parties are exposed to more competent females. We next examine how the treatment effects are different by the education level of the first female councilors of cycle 4. Panels A and B of Table 2.11 reproduce Table 2.5 for councils in which the female PR councilors of cycle 4 have below-median and above-median education levels, respectively. It is clear to see that the shift towards female candidates is more apparent and stronger when the first women are more educated, pointing towards parties learning about the competency of females over time.

[17]For example, the number of women a party needs to include in the party list is 1 if the number of PR seats for the municipality is 1 or 2, and 2 if the number of PR seats is 3 or 4.

2.6.3 Which parties are learning and how?

The fact that the treatment status at cycle 4 has lasting effects through to cycle 7 shows that the treatment and control municipalities evolve on different paths from cycle 4, through some dynamic linkages across election cycles. A dynamic linkage of particular interest is whether a party learns about the competency of women as a result of the previous election outcome, and then changes its strategy in the current election.

We compare the strategies of parties that marginally won a PR councilor to those that marginally lost a PR councilor in the previous election. This comparison gives us the causal effect of having previously won a PR seat. In close electoral races in which the outcome of the election is uncertain, the winner is typically determined by factors that are beyond the control of parties and candidates, so which party wins the seat can be considered as random (Lee, 2008).

We take marginal parties to be the *two* parties that either marginally won or lost the *last* PR seat for the municipality. In order to differentiate marginal winners from losers, we measure how far off the vote share received by a party was, from the share it needed to win that seat. For party p in municipal council c at election cycle t, this value is given by $v_{cpt} \equiv (\text{vote share})_{cpt} - \bar{v}_{cpt}$, where $\bar{v}$ denotes the verdict-determining vote share. As the simplest example of $\bar{v}$, when two parties are competing for one PR seat, $\bar{v} = 0.5$ for both parties. The precise way we compute $\bar{v}$ for all possible contest scenarios is detailed in Appendix Section 2.10.3.

Figure 2.7 shows the distribution of the vote shares received by the two marginal parties competing for the last PR seat in a municipality. The histogram shows that there are plenty of parties that received a vote share close to the share needed to win that seat. We employ a regression discontinuity design of the following form:

$$Y_{cpt} = \beta \times Winner_{cp,t-1} + f(v_{cp,t-1}) + X_{cpt} + \delta_n + \gamma_t + \epsilon_{cpt} \tag{2.3}$$

where $Winner_{cp,t-1} \equiv 1(v_{cp,t-1} \geq 0)$. We denote by $n \in \{1, 2\}$, whether the marginal candidate that won – or nearly won – the last PR seat corresponds to the first or second PR candidate in a party's list.[18] $f(v_{cp,t-1})$ is linear and allows for different slopes to the left and right of the cutoff $v_{cp,t-1} = 0$. X_{cpt} represents the control variables, including the number of ward seats and the total council size for the *contemporaneous* election, i.e. election cycle t. A further factor to note is that the sample includes only the two major parties in South Korea, in order to track the parties over time. Due to frequent changes to party names, as well as frequent dissolutions and merges of small parties, parties other than the two major ones are difficult to follow over time. Therefore, X_{cpt} also includes a dummy that indicates which of the two major parties party p is.

When we consider marginal victories in the PR arm, we need to differentiate between the cases when the marginal candidate is in the first or second position in the party list.[19] If the marginal candidate is first on the list, then the candidate is necessarily a female, as enforced by the quota. However, if the marginal candidate is second on the list, the candidate might be male or

[18] There are only five parties that won or nearly won a third PR councilor, so we exclude these parties. There is no party that won four PR councilors.

[19] We do not consider the cases of the marginal candidate being third or lower on the list. It is extremely rare that a single party wins that many PR candidates, as can be seen in Appendix Table 2.A1.

female. Thus, when we consider close elections around a number-1 PR candidate, we are comparing parties that marginally won a *woman* in the previous election cycle with parties that didn't win any PR candidate. On the other hand, for close elections around a number-2 PR candidate, we are comparing parties that marginally won a second PR councilor to parties that only won one PR councilor.

We also distinguish between two different types of parties: those that placed men as the number-2 candidate in the party list and those that did not, in the previous election. As parties can place a candidate of any gender in even positions in their lists, the parties that place women in even positions are expected to have more gender-equal attitudes ex-ante.

Table 2.12 reports the result of regressing equation (2.3), when the marginal candidate is in position 1 or position 2 of the party list, and when the party lists a male or female second. We find positive coefficients on $Winner$ in columns (1)-(4) of Panel A. Marginal winners put forth a higher female share among ward candidates in the following election cycle than marginal losers, when the marginal candidate is in the first position and when the party has a male as the number-2 PR candidate. Therefore, parties that had a prior preference towards men that update their beliefs on the competency of women, upon experiencing a female councilor. In addition, the positive coefficients imply that a party's strategy is affected by having a female councilor get elected from within the party. For example, we cannot say from the results that the Conservative party do not learn from experiencing a Progressive female councilor, but we can say that the degree of learning is greater for the Progressive party. In contrast, the null effect in columns (5)-(8) of Panel B – marginally winning a female number-2 candidate – implies that once a party wins a female councilor, an additional female councilor does not further impact the party's strategy in the following election. Moreover, the null effect in columns (5)-(8) of Panel A – marginally winning a female number-1 candidate for parties with a more gender-equal prior – implies that winning the first woman does not affect a party's future gender preference if the party had a more gender-equal attitude already.

2.7 Conclusion

This paper highlights that with time, affirmative action policies can still be effective despite an initial backlash, as long as the policies are not completely undone. Moreover, such is the case even in settings where the target group consists of a very small minority among the incumbents. Through exposure to the minority group, the policies provide incumbents with an opportunity to learn about the competency of the minority group. Once the learning takes off, the policy itself might be unneeded.

Although gender quotas in parliaments have been adopted broadly worldwide, there are still many countries that have none in place, such as Egypt, India, Liberia, Mauritius, Sao Tome and Principe, Sierra Leone, and Sri Lanka. Unsurprisingly, these countries also suffer from low levels of female representation in national parliaments. The South Korean setting of this paper is unique in that it studies the effect of a gender quota in the legislative body from a starting point of practically zero women. Therefore, this paper is informative about the effect of gender quotas where they are most needed.

What remains to be crystallized is exactly which aspect of women's competence parties are learning about. As of yet, we do not know whether it is their election campaign skills, their loyalty to the party, their keenness as legislators, or their ability to meet the demands of the electorate, that the parties update their beliefs on. Further evidence is needed in this direction.

This paper is a part of a bigger agenda that attempts to study how a gender quota might trigger a gradual process of learning in favor of women. To tackle the precise mechanisms through which the learning takes place, we plan to study in future work the specific interactions among councilors recorded in the transcripts of council meetings.

References

Ahern, K. R. and A. K. Dittmar (2012). The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation *. *The Quarterly Journal of Economics 127*(1), 137–197.

Angrist, J. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion* (1 ed.). Princeton University Press.

Bagues, M. and P. Campa (2021). Can gender quotas in candidate lists empower women? evidence from a regression discontinuity design. *Journal of Public Economics 194*, 104315.

Baltrunaite, A., P. Bello, A. Casarico, and P. Profeta (2014). Gender quotas and the quality of politicians. *Journal of Public Economics 118*(C), 62–74.

Besley, T., O. Folke, T. Persson, and J. Rickne (2017). Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. *American Economic Review 107*(8), 2204–42.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Chattopadhyay, R. and E. Duflo (2004). Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica 72*(5), 1409–1443.

Cho, H.-O. and E. H. Kim (2010). A Study of Historical Context and Institutionalization Process of Gender Quota System and Women's Political Empowerment Movement in Korea 2000-2010. *Journal of Korean Social Trend and Perspective* (79), 110–139.

Clayton, A. (2015). Women's political engagement under quota-mandated female representation: Evidence from a randomized policy experiment. *Comparative Political Studies 48*(3), 333–369.

Dahlerup, D. and L. Freidenvall (2013). Electoral gender quota systems and their implementation in europe. Technical report, The European Parliament, the Committee on Women's Rights and Gender Equality.

De Paola, M., V. Scoppa, and R. Lombardo (2010). Can gender quotas break down negative stereotypes? evidence from changes in electoral rules. *Journal of Public Economics 94*(5), 344 – 353.

Esteve-Volart, B. and M. Bagues (2012). Are women pawns in the political game? evidence from elections to the spanish senate. *Journal of Public Economics 96*(3), 387–399.

Inter-Parliamentary Union (2021). Global and regional averages of women in national parliaments. Monthly data, 1997-2021.

International IDEA (2021). Gender quotas database. Global Database of Gender Quotas in parliaments worldwide by International Institute for Democracy and Electoral Assistance, Inter-Parliamentary Union and Stockholm University.

Jang, G. S. (2008). Busan Provincial Council gets "Poor Report Card". *"No-cut" News*.

Jeon, J. Y. (2013). Political Dynamics of Adopting Gender Quota in Korea. *Journal of Korean Politics 22*(1), 29–52.

Jin, J. W. (2018). Female nomination subsidy monopolized by two major parties. *Women News*.

Kim, S. H. (2005). Local Council Elections - "Reformed" or "Deformed"? *Pressian News*.

Kim, W. H., M. J. Kim, H. C. Lee, and H. K. Kim (2003). *Research on measures to expand female candidate nomination by political parties.* Korea Women's Development Institute.

Lee, D. S. (2008). Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics 142*(2), 675 – 697. The regression discontinuity design: Theory and applications.

Lee, J. J. (2019). *The Legislative Impact Analysis of Gender Quota in Local Council District Elections*. Number 44. National Assembly Research Service.

Lee, M. S. (2003). Measures to expand female political representation in preparation for the 17$^{\text{th}}$ general election. Panel discussion in the Meeting of Female Members of the National Assembly and Municipal Councils To Increase the Participation of Women in Politics.

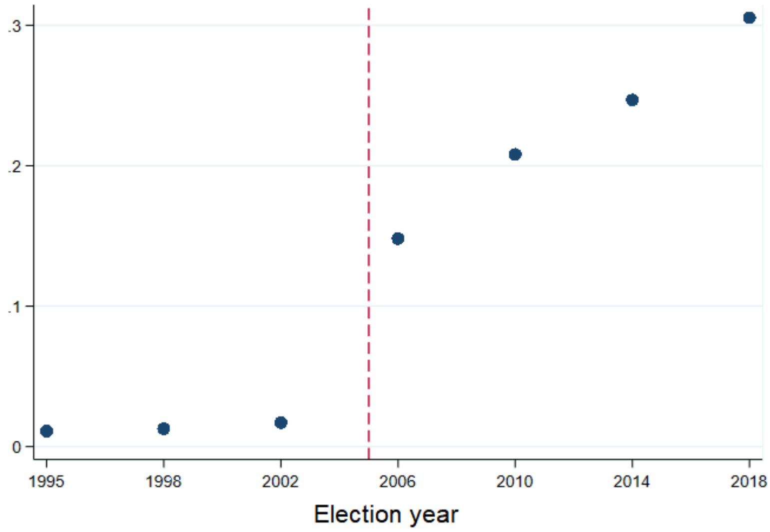
Local Decentralization Bureau – Election and Local Council Division (2019). Ministry of the Interior and Safety to Block "Self-Assessment" of Overseas Training for Local Councils and to Redeem Unjust Spending. Ministry of the Interior and Safety.

Ministry of the Interior and Safety (2018). Municipal councilor's activities guide. Publication Registration Number: 11-1741000-000101-15.

National Election Commission (2018). Election subsidy payments by year (2001-2018). National Election Commission Data Resource Archive.

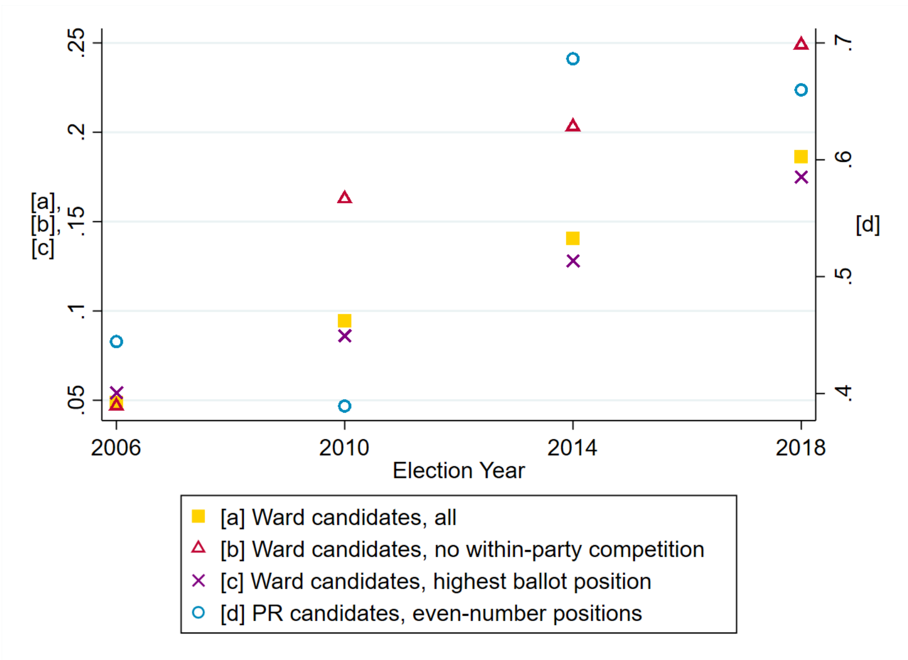
2.8 Figures

Figure 2.1: Proportion of females in municipal councils, nationwide average



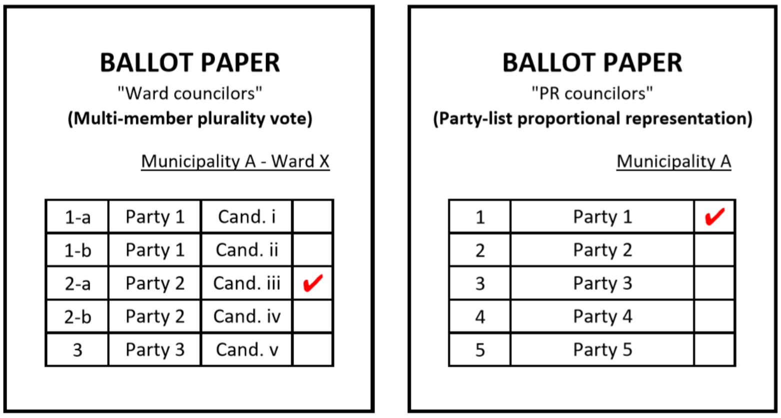
Notes: This figure illustrates the nationwide average of the gender ratio in municipal councils, for every election cycle since their emergence. The red dotted line indicates the year of the major reform that instituted the gender quota.

Figure 2.2: Share of females among non-quota candidates



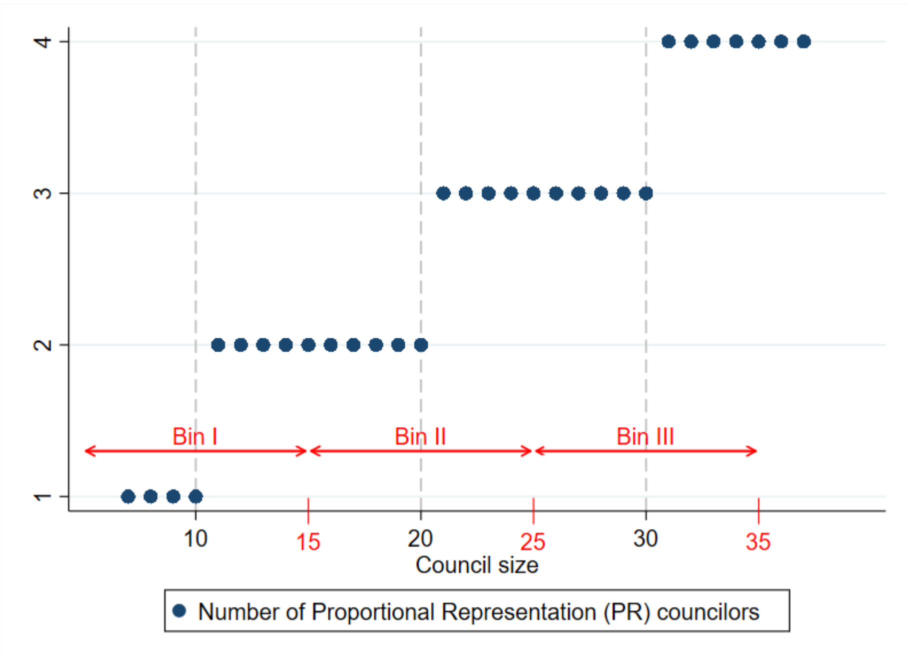
Notes: This figure plots the share of females among [a] all ward candidates, [b] ward candidates with no within-party competition, [c] ward candidates that have within-party competition but is ranked the highest, and [d] PR candidates in even-number party list positions. The left-hand vertical axis corresponds to [a], [b], and [c], whereas the right-hand one corresponds to [d].

Figure 2.3: Ballot papers in municipal council elections



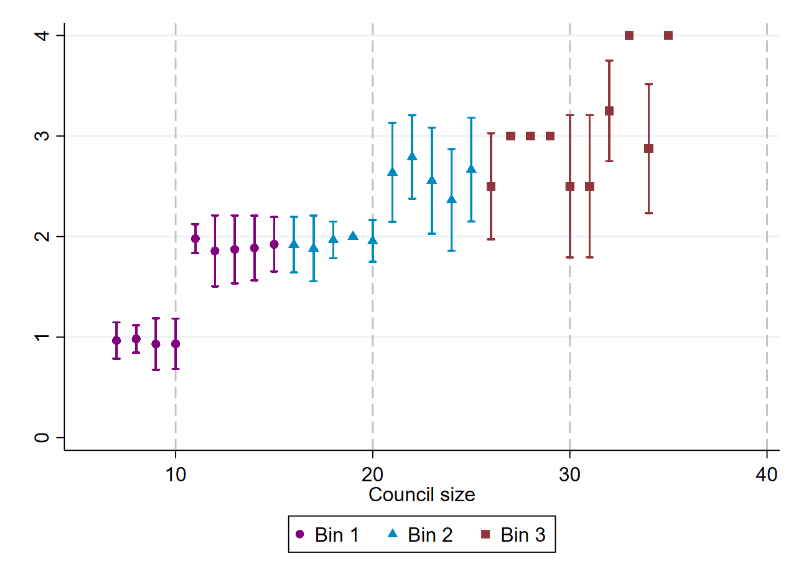
Notes: This figure illustrates the ballot papers that a voter residing in ward X of municipality A receives for the municipal council elections. The one on the left is used to vote for ward councilors and the one on the right is used to vote for PR councilors. The red ticks indicate how the voter might vote.

Figure 2.4: Councils by bins around each threshold



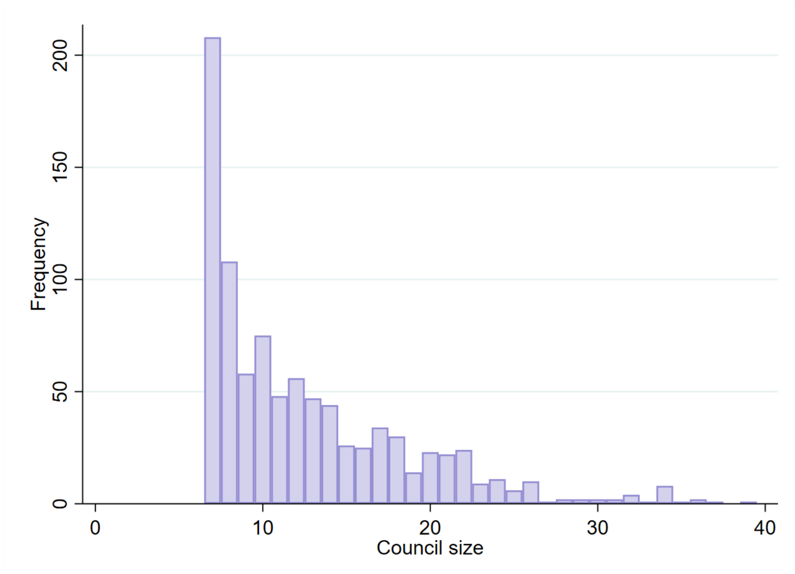
Notes: This figure depicts how the number of seats reserved for the proportional representation arm increases as a step function of the total number of councilors in a municipality. There are municipalities that do not correspond to the step function, because they are formed by the union of multiple municipalities *after* the election took place. The municipalities pre- and post-union are all excluded from the sample as outliers and are not shown in this figure. Moreover, the figure depicts how a council is categorized into a bin based on its most proximate threshold.

Figure 2.5: Average number of female PR councilors by council size



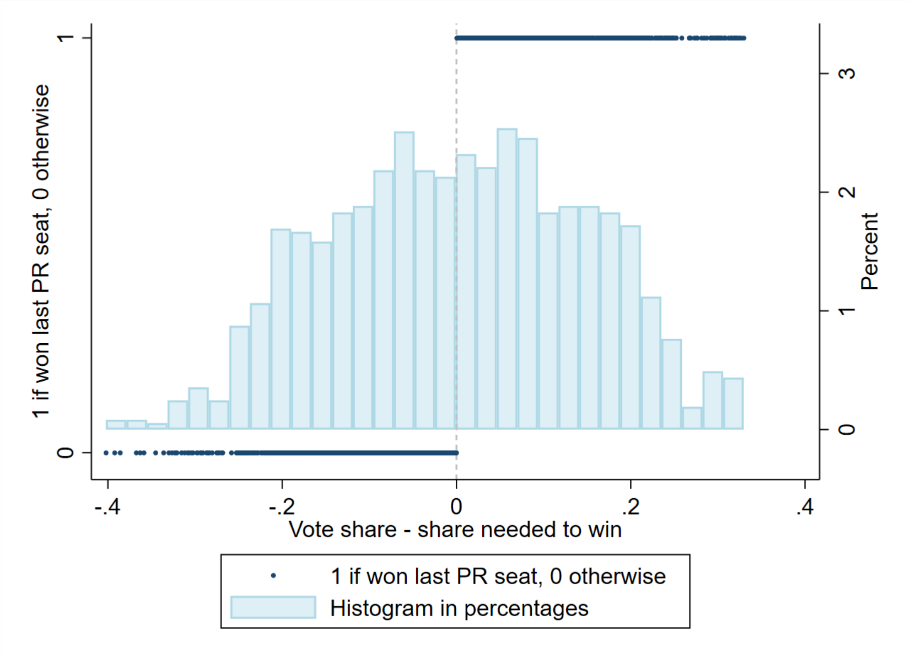
Notes: The error bars indicate standard deviation of the number of female PR councilors by council size. Where the erroe bars are missing, there is only one municipality for that council size. Therefore, we can tell that there are only a small number of municipal councils belonging to bin 3.

Figure 2.6: Histogram of council size



Notes: The sample includes all municipal councils of election cycles 4, 5, 6, and 7.

Figure 2.7: Marginal winners and losers of the last PR seat



Notes: This figure shows the distribution of the vote shares received by the two marginal parties competing for the last PR seat in a municipality. The vote share is computed to be the share of votes received among qualifying parties, i.e. parties that received more than 5% of the raw votes in the PR election arm.

2.9 Tables

Table 2.1: Amendments to legislation on municipal council elections

First applicable election year	Amendment
2006	[PR] Proportional representation introduced
	[W] Single-member plurality vote → Multi-member plurality vote
	[PR] Odd-number candidates in party lists must be female (not enforced)
	[W] Subsidies to parties for nominating female candidates
2010	[PR] Odd-number candidates in party lists must be female (enforced)
	[W] At least one female candidate per general election district

Notes: Adapted from Lim (2018). [PR] indicates rules relating to proportional representation councilors and [W] to ward councilors.

Table 2.2: Descriptive statistics: Gender composition of municipal councils

	Election cycle (year)						
	1 (1995)	2 (1998)	3 (2002)	4 (2006)	5 (2010)	6 (2014)	7 (2018)
Total number of councilors							
Min.	7	7	7	7	7	7	7
Mean	19.9	15.0	15.0	12.6	12.6	12.8	12.8
Max.	50	40	41	36	34	43	44
Number of PR councilors							
Min.	-	-	-	1	1	1	1
Mean	-	-	-	1.63	1.63	1.67	1.70
Max.	-	-	-	4	4	5	5
Gender ratio							
Min.	0	0	0	0	0.06	0.08	0.10
Mean	0.01	0.01	0.02	0.15	0.21	0.25	0.29
Max.	0.43	0.22	0.28	0.46	0.57	0.86	0.64
Gender ratio among PR councilors							
Min.	-	-	-	0*	0*	0.50	0.50
Mean	-	-	-	0.87	0.96	0.97	0.98
Max.	-	-	-	1	1	1	1
Minimum number of women required							
Min.	-	-	-	0	1	1	1
Mean	-	-	-	1.12	1.12	1.13	1.13
Max.	-	-	-	2	2	3	3

Notes: *gender quotas were introduced in 2005. However, during the election of 2006, they remained merely a strong recommendation, so it was still legal to place a male in slot 1 of party lists. Most municipalities complied, but 14 of them had no female PR councilors. In election year 2010, the minimum of the gender ratio among PR councilors is 0 because in one council the elected woman was invalidated for being a member of multiple parties.

Table 2.3: Identification: RDD causal effect on number of female PR councilors

	Number of female PR councilors		
	Bin 1	Bin 2	Bin 3
	(1)	(2)	(3)
Treat at cycle 4× Cycle 4	0.84***	0.52***	-0.23
	(9.35)	(2.82)	(-0.42)
Treat at cycle 4 × Cycle 5	0.84***	0.32*	0.44
	(12.28)	(1.71)	(1.21)
Treat at cycle 4 × Cycle 6	0.77***	0.58***	0.04
	(10.18)	(3.33)	(0.10)
Treat at cycle 4 × Cycle 7	0.77***	0.57***	0.16
	(9.61)	(3.68)	(0.40)
Running variable form	council	council	council
N	670	198	33

Notes: This table reports the results of regressing (2.2), separately for each bin, with the number of female PR councilors as the outcome variable. t statistics from standard errors clustered by municipality in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: Identification: Balance tests on pre-determined characteristics

***Panel A*: Population characteristics**

	Population		Voting age population			Households	
	Total	Foreign	Total	Male	Female	Total	Foreign
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treat at cycle 4	-23.97	0.01	-17.22	-7.84	-9.38	-5.59	0.01
	(-0.78)	(0.78)	(-0.76)	(-0.70)	(-0.82)	(-0.50)	(0.83)
Running variable	31.49***	-0.00	23.22***	11.29***	11.94***	10.66***	-0.00
	(5.50)	(-0.84)	(5.47)	(5.38)	(5.55)	(5.15)	(-0.86)
Running variable form	council	council	council	council	council	council	council
N	219	219	219	219	219	219	219

***Panel B*: Political leaning, economic, and ward division characteristics**

	Past vote share by party		Budget		Ward characteristics	
	Conservative	Progressive	Total	Council expenses	Num of wards	Seats per ward
	(8)	(9)	(10)	(11)	(12)	(13)
Treat at cycle 4	-0.02	-0.00	54.16	0.02	-0.23	0.17
	(-0.26)	(-0.12)	(0.63)	(0.36)	(-1.30)	(1.48)
Running variable	-0.00	0.00	19.04	0.06***	0.45***	-0.06**
	(-0.08)	(0.08)	(1.30)	(7.84)	(11.71)	(-2.53)
Running variable form	council	council	council	council	ward	ward
N	219	219	219	219	219	219

Notes: The regression specification follows equation (2.2), and the sample consists of bins 1 and 2 at election cycle 4. t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5: Main Results: RDD causal effect on number of candidates and councilors

	All political parties									
	Candidates				Councilors					
	Ward		PR		Ward		PR		All	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treat at cycle 4 × Cycle 4	3.70***	-0.24	0.94***	0.96***	0.36*	-0.36*	0.09	0.76***	-0.29	0.29
	(3.19)	(-0.69)	(4.27)	(4.39)	(1.65)	(-1.65)	(1.16)	(9.01)	(-1.03)	(1.03)
Treat at cycle 4 × Cycle 5	0.56	0.49	0.63***	1.20***	-0.28	0.28	0.10	0.71***	-0.88***	0.88***
	(0.61)	(1.37)	(3.77)	(5.36)	(-1.24)	(1.24)	(1.52)	(9.23)	(-3.17)	(3.17)
Treat at cycle 4 × Cycle 6	-1.39*	0.91**	0.25*	1.02***	-0.47*	0.47*	0.08	0.70***	-1.06***	1.06***
	(-1.66)	(2.18)	(1.70)	(4.66)	(-1.76)	(1.76)	(1.31)	(8.89)	(-3.45)	(3.45)
Treat at cycle 4 × Cycle 7	-2.23**	1.10**	0.21	1.27***	-0.73**	0.73**	0.04	0.69***	-1.33***	1.33***
	(-2.23)	(2.49)	(1.35)	(5.83)	(-2.50)	(2.50)	(0.61)	(9.04)	(-3.99)	(3.99)
Running variable form	ward	ward	council	council	ward	ward	council	council	council	council
N	868	868	868	868	868	868	868	868	868	868

Notes: The regression specification is given by equation (2.2). The sample includes only bins 1 and 2. t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: Main Results: RDD causal effect on number of ward candidates likely to get elected

	Main political parties only							
	All ward positions				Useful ward positions			
	Male	Female	All	Female	Male	Female	All	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treat at cycle 4 × Cycle 4	1.56*	-0.25	1.31	-0.43*	1.51***	-0.36*	1.15**	-0.54***
	(1.92)	(-1.09)	(1.52)	(-1.95)	(2.85)	(-1.91)	(2.11)	(-2.92)
Treat at cycle 4 × Cycle 5	0.59	0.52**	1.10	0.37*	0.38	0.43**	0.81	0.31*
	(0.79)	(2.09)	(1.31)	(1.73)	(0.75)	(2.32)	(1.49)	(1.81)
Treat at cycle 4 × Cycle 6	0.90	0.83***	1.72**	0.60**	0.50	0.60**	1.09**	0.43*
	(1.22)	(2.82)	(2.11)	(2.25)	(1.01)	(2.44)	(2.00)	(1.94)
Treat at cycle 4 × Cycle 7	0.06	1.29***	1.35*	1.11***	0.37	0.73***	1.10**	0.56**
	(0.08)	(4.00)	(1.74)	(3.66)	(0.74)	(2.90)	(2.11)	(2.37)
Total ward candidates				0.13***				0.15***
				(8.37)				(7.88)
Running variable form	ward	ward	ward	ward	ward	ward	ward	ward
N	867	867	867	867	867	867	867	867

Notes: The regression specification is given by equation (2.2). The sample includes only bins 1 and 2 and is restricted to the two main parties. The number of observations is 867 instead of 868 since in one municipality main parties only have proportional candidates. Useful positions refer to candidates in the high-up positions on the ballot for the party in a ward (position 1 if the ward elects 1-2 councilors, and positions 1 and 2 if the ward elects 3-4 councilors). t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: Main Results: RDD causal effect on number of ward candidates in safe and unsafe wards

	Main political parties only, (ward×party)-level regressions							
	All ward positions						Useful ward positions	
	Male	Female	All	Female	Female Safe	Female Unsafe	Female Safe	Female Unsafe
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treat at cycle 4 × Cycle 4	0.06	0.00	0.06	-0.00	-0.01	0.00	-0.05**	-0.01
	(0.68)	(0.11)	(0.74)	(-0.13)	(-0.40)	(0.09)	(-2.02)	(-0.21)
Treat at cycle 4 × Cycle 5	-0.05	0.07**	0.01	0.06**	0.04	0.10***	0.01	0.10***
	(-0.62)	(2.34)	(0.18)	(2.26)	(1.12)	(2.60)	(0.33)	(3.15)
Treat at cycle 4 × Cycle 6	-0.03	0.06*	0.03	0.06*	-0.04	0.10***	-0.01	0.06**
	(-0.33)	(1.95)	(0.39)	(1.76)	(-0.74)	(2.64)	(-0.18)	(2.05)
Treat at cycle 4 × Cycle 7	-0.08	0.07**	-0.01	0.07**	0.04	0.09**	-0.01	0.05
	(-0.96)	(2.18)	(-0.15)	(2.16)	(0.70)	(2.55)	(-0.21)	(1.55)
Total ward candidates				0.11***	0.10***	0.13***	0.09***	0.10***
				(10.17)	(7.05)	(9.10)	(4.55)	(4.22)
Running variable form	ward	ward	ward	ward	ward	ward	ward	ward
N	6035	6035	6035	6035	2337	3698	2337	3698

Notes: The regression specification is given by equation (2.2). The sample includes only bins 1 and 2 and is restricted to the two main parties. The level of observation is party ward. A ward is considered as safe for a party if the party wins the greatest vote share in the PR arm in the ward, and it got over 10 percentage points more vote share than the next popular party. Unsafe wards are all the others. t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: Identification: Placebo check - Pre-quotas period

	Number of ward candidates			
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
Treat × Cycle 1	0.92	0.01	0.03	0.04
	(1.32)	(0.26)	(1.33)	(0.31)
Treat × Cycle 2	0.40	-0.05	-0.61	0.03
	(1.12)	(0.24)	(1.01)	(0.30)
Treat × Cycle 3	1.06	0.06	0.01	0.15
	(1.02)	(0.27)	(0.90)	(0.32)
Treat × Cycle 4	3.22***	0.72**	2.93**	0.59**
	(1.21)	(0.28)	(1.15)	(0.27)
Treat × Cycle 5			-0.71	1.29***
			(0.83)	(0.31)
Treat × Cycle 6			-2.29***	1.58***
			(0.79)	(0.42)
Treat × Cycle 7			-2.79***	1.58***
			(0.98)	(0.44)
Running variable form	ward	ward	ward	ward
N	899	899	1577	1577

Notes: The running variable and treatment status are defined contemporaneously. While municipality size and divisions remained almost unchanged from cycle 4 onward, they changed dramatically during the first three election cycles. Thus, it would be inaccurate to define treatment for the first three cycles using cycle 4 municipality characteristics. The regression specification is given by equation (2.1). t-statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9: Mechanisms: Probability of being unconstrained

	Pr(unconstrained)		
	(1)	(2)	(3)
Treat at cycle 4 × Cycle 4	-0.02	0.09	0.09
	(-0.46)	(1.19)	(1.27)
Treat at cycle 4 × Cycle 5	-0.02	-0.09	-0.12*
	(-0.49)	(-1.19)	(-1.65)
Treat at cycle 4 × Cycle 6	-0.04	-0.02	-0.04
	(-0.89)	(-0.33)	(-0.69)
Treat at cycle 4 × Cycle 7	-0.00	-0.08	-0.10
	(-0.05)	(-1.18)	(-1.38)
Running variable from	council	council	council
Sample	all parties	main parties	main parties participating in ward arm
N	3795	1557	1520

Notes: The regression specification is given by equation (2.2). The sample includes only bins 1 and 2. A party is defined as unconstrained if the number of female candidates in the party's list is greater than the number of women the party is obliged to place in its list due to quotas. In column (1), all parties are included; column (2) includes only the two main parties; in column (3) the sample is restricted to only the main parties in municipalities where the parties have at least one ward candidate. t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.10: Mechanisms: Effect on number of rookie ward candidates

	Rookie ward candidates	
	Male	Female
	(9)	(10)
Treat at cycle 4 × Cycle 4	0.82	-0.27*
	(1.33)	(-1.66)
Treat at cycle 4 × Cycle 5	-0.63	0.01
	(-1.05)	(0.05)
Treat at cycle 4 × Cycle 6	-0.29	0.15
	(-0.57)	(0.84)
Treat at cycle 4 × Cycle 7	-0.45	0.38*
	(-0.86)	(1.82)
Running variable form	ward	ward
N	868	868

Notes: The regression specification is given by equation (2.2). The sample includes only bins 1 and 2.t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.11: Mechanisms: Heterogeneity by education level of first PR female councilors

	***Panel A*: Below-median education level**									
	Candidates				Councilors					
	Ward		PR		Ward		PR		All	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treat at cycle 4 × Cycle 4	2.46	-0.55	1.33***	0.67**	0.53	-0.53	0.21*	0.61***	0.06	-0.06
	(1.28)	(-0.94)	(4.36)	(2.06)	(1.33)	(-1.33)	(1.66)	(4.39)	(0.12)	(-0.12)
Treat at cycle 4 × Cycle 5	0.33	0.47	0.94***	1.22***	0.07	-0.07	0.11	0.67***	-0.46	0.46
	(0.22)	(0.75)	(3.49)	(3.34)	(0.16)	(-0.16)	(0.96)	(4.84)	(-0.84)	(0.84)
Treat at cycle 4 × Cycle 6	-1.04	0.63	0.54**	0.44	-0.17	0.17	0.18	0.56***	-0.59	0.59
	(-0.75)	(0.83)	(2.35)	(1.47)	(-0.36)	(0.36)	(1.51)	(3.77)	(-1.03)	(1.03)
Treat at cycle 4 × Cycle 7	-1.90	1.26	0.58*	1.08***	-0.57	0.57	0.09	0.61***	-1.05	1.05
	(-0.99)	(1.58)	(1.96)	(3.32)	(-0.94)	(0.94)	(0.80)	(4.33)	(-1.50)	(1.50)
Running variable form	ward	ward	council	council	ward	ward	council	council	council	council
N	387	387	387	387	387	387	387	387	387	387
	***Panel B*: Above-median education level**									[()]35pt
Treat at cycle 4 × Cycle 4	4.03***	0.16	0.75**	1.12***	0.11	-0.11	0.01	0.88***	-0.71**	0.71**
	(2.67)	(0.35)	(2.55)	(3.80)	(0.41)	(-0.41)	(0.06)	(8.06)	(-2.08)	(2.08)
Treat at cycle 4 × Cycle 5	0.49	0.72	0.56***	1.17***	-0.64**	0.64**	0.13*	0.71***	-1.30***	1.30***
	(0.39)	(1.56)	(2.72)	(4.09)	(-2.26)	(2.26)	(1.69)	(7.83)	(-3.87)	(3.87)
Treat at cycle 4 × Cycle 6	-1.75	1.27**	0.22	1.38***	-0.77**	0.77**	0.06	0.77***	-1.49***	1.49***
	(-1.48)	(2.40)	(1.17)	(4.55)	(-2.24)	(2.24)	(0.78)	(8.66)	(-3.87)	(3.87)
Treat at cycle 4 × Cycle 7	-2.44*	1.07*	0.14	1.28***	-0.88***	0.88***	0.04	0.71***	-1.57***	1.57***
	(-1.93)	(1.97)	(0.79)	(4.38)	(-2.66)	(2.66)	(0.55)	(7.68)	(-4.20)	(4.20)
Running variable form	ward	ward	council	council	ward	ward	council	council	council	council
N	481	481	481	481	481	481	481	481	481	481

Notes: The regression specification is given by equation (2.2). The sample includes only bins 1 and 2. t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.12: Mechanisms: Effect of marginally winning a PR councilor in the previous election

***Panel A*: Position of marginal candidate: 1**

	Female share among party's ward candidates							
	Gender of second position: male				Gender of second position: female			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Winner	0.07*	0.08*	0.10*	0.21***	0.00	0.03	-0.02	-0.04
	(1.73)	(1.73)	(1.81)	(3.09)	(0.04)	(0.33)	(-0.20)	(-0.23)
Bandwidth ($\lvert v_{cpt} \rvert$)	0.20	0.15	0.10	0.05	0.20	0.15	0.10	0.05
N	297	231	157	84	104	70	49	24

***Panel B*: Position of marginal candidate: 2**

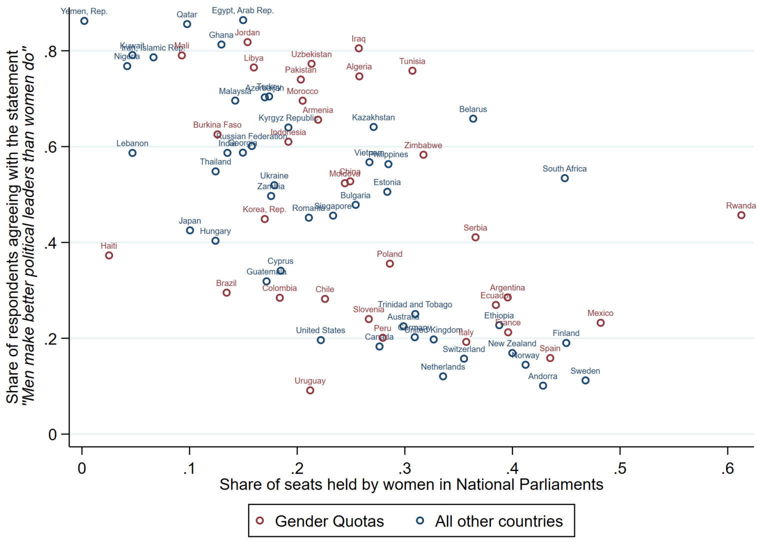
	Female share among party's ward candidates							
	Gender of second position: male				Gender of second position: female			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Winner	0.01	0.03	0.04	-0.05	-0.01	-0.01	-0.04	-0.05
	(0.26)	(0.68)	(0.92)	(-1.17)	(-0.31)	(-0.21)	(-0.84)	(-0.72)
Bandwidth ($\lvert v_{cpt} \rvert$)	0.20	0.15	0.10	0.05	0.20	0.15	0.10	0.05
N	113	84	69	33	98	75	62	33

Notes: The regression specification is given by equation (2.3). The Mean Square Error-optimal bandwidth is selected (Calonico et al., 2014). The standard errors remain very similar when they are clustered at the municipality level. t statistics from standard errors clustered by municipality×party in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.10 Appendix

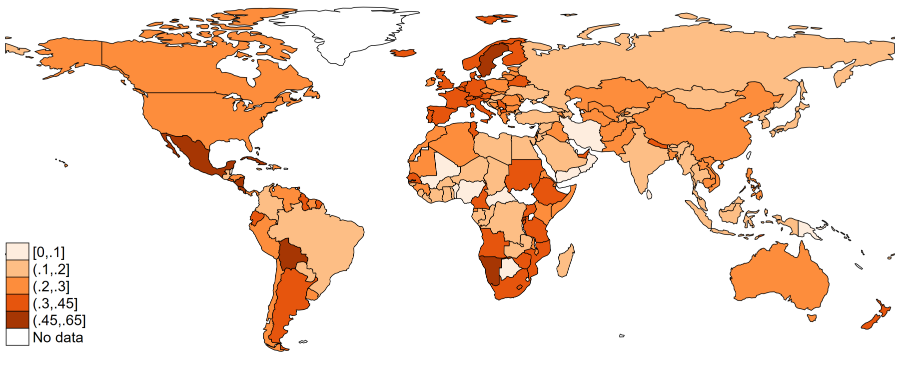
2.10.1 Additional Tables and Figures

Figure 2.A1: Female share in national parliaments and attitudes towards women



Notes: Sources: Attitudes towards women: World Values Survey waves 5 (2005-2009) and 6 (2010-2014); Share of seats held by women in national parliaments: Inter-Parliamentary Union, average of years 2018, 2019, 2020; Gender quotas in national parliaments: International IDEA Institute for Democratic and Electoral Assistance Gender Quotas Database, 2020

Figure 2.A2: Proportion of seats held by women in national parliaments (%)



Notes: Source: Inter-Parliamentary Union, average of years 2018, 2019, 2020

2.10.2 Appendix to Identification

Confirming that the number of female PR councilors changes only at the thresholds

In order to buttress the regression discontinuity design, we test whether there is a change in the number of female PR councilors as council size increases, at points *other* than the thresholds. We regress, for each value of $x \in \{-4, -3, ..., 3, 4\}$ i.e. distance from the threshold,

Table 2.A1: Allocation of proportional representation seats across parties

	Councils by the number of PR seats					
	1 PR seat		**2 PR seats**		**3 PR seats**	
	N.	Percent.	N.	Percent.	N.	Percent.
Election Cycle 4						
1 Party	117	100%	15	17.86%	0	0
2 Parties	0	0	69	82.14%	15	83.33%
3 Parties	0	0	0	0	3	16.67%
Election Cycle 5						
1 Party	117	100%	5	6.02%	0	0
2 Parties	0	0	78	93.98%	13	72.22%
3 Parties	0	0	0	0	7	27.78%
Election Cycle 6						
1 Party	110	100%	18	20.22%	0	0
2 Parties	0	0	71	79.78%	17	100%
3 Parties	0	0	0	0	0	0%
Election Cycle 7						
1 Party	105	100%	9	9.89%	1	4.35%
2 Parties	0	0	82	90.11%	18	94.74%
3 Parties	0	0	0	0	1	5.26%
Total	**449**		**347**		**72**	

Notes: The sample is restricted to bins 1 and 2, i.e. to municipal councils with up to 25 councilors.

Table 2.A2: Candidates and councilors' party affiliation

		Candidates				Councilors			
		Direct		Proportional		Direct		Proportional	
	N	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Election Cycle 1									
Independent	226	1	0			1	0		
Election Cycle 2									
Independent	228	1	0			1	0		
Election Cycle 3									
Independent	228	1	0			1	0		
Election Cycle 4									
Independent	230	0.41	0.15	0	0	0.11	0.14	0	0
Progressive party	230	0.16	0.08	0.29	0.17	0.20	0.18	0.18	0.25
Conservative party	230	0.25	0.13	0.47	0.30	0.54	0.31	0.64	0.38
Election Cycle 5									
Independent	228	0.32	017	0	0	0.14	0.16	0	0
Progressive party	228	0.21	0.16	0.31	0.29	0.33	0.27	0.41	0.39
Conservative party	228	0.33	0.18	0.49	0.31	0.43	0.25	0.44	0.40
Election Cycle 6									
Independent	227	0.34	0.16	0	0	0.13	0.15	0	0
Progressive party	227	0.24	0.16	0.38	0.27	0.37	0.25	0.40	0.37
Conservative party	227	0.36	0.18	0.55	0.30	0.49	0.26	0.59	0.38
Election Cycle 7									
Independent	226	0.20	0.16	0	0	0.09	0.14	0	0
Progressive party	226	0.32	0.12	0.43	0.20	0.54	0.21	0.66	0.33
Conservative party	226	0.29	0.16	0.38	0.25	0.35	0.20	0.33	0.33

Table 2.A3: Identification: Effect of council size on number of female PR councilors

	x value								
	-4	-3	-2	-1	0	1	2	3	4
Coefficient ($\hat{\beta}$)	-0.03	0.03	-0.03	-0.01	0.92***	-0.01	-0.03	-0.03	0.09
Standard error	(-0.36)	(1.32)	(-0.99)	(-0.20)	(15.00)	(-0.14)	(-0.36)	(-0.35)	(1.04)
N	267	380	210	170	168	150	136	111	87

Notes: This table reports the results of regression equation (2.4). The sample includes only bins 1 and 2. t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$$\text{(number of female PR councilors)}_{cbt} = \beta \times TreatOne_{cbt} + \delta_b + \gamma_t + \epsilon_{cbt} \tag{2.4}$$

$$\text{where } TreatOne_{cbt} = \begin{cases} 1, & \text{if (council size)}_{cbt} = x \\ 0, & \text{if (council size)}_{cbt} = x - 1 \end{cases}$$

Equation (2.4), therefore, estimates the change in the number of female PR councilors when the council size increases by 1, for all points around the threshold. Table 2.A3 reports the results. It confirms that there is a positive effect only at the threshold.

Robustness to bandwidth choice

Table 2.A4: Identification: Alternative bandwidths

	Candidates				Councilors					
	Ward		PR		Ward		PR		All	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
***Panel A*: distance**† ≤ 4										
Treat	0.37	0.31	0.58***	1.23***	-0.15	0.15	0.09**	0.91***	-0.06	1.06***
	(0.46)	(0.94)	(4.64)	(8.29)	(-0.74)	(0.74)	(1.97)	(19.58)	(-0.26)	(4.78)
N	868	868	868	868	868	868	868	868	868	868
***Panel B*: distance** ≤ 3										
Treat	0.14	0.33	0.51***	1.24***	-0.13	0.13	0.08	0.92***	-0.04	1.04***
	(0.16)	(0.99)	(3.77)	(7.95)	(-0.59)	(0.59)	(1.57)	(17.77)	(-0.19)	(4.48)
N	811	811	811	811	811	811	811	811	811	811
***Panel C*: distance** ≤ 2										
Treat	0.89	0.37	0.54***	1.29***	-0.13	0.13	0.08	0.92***	-0.06	1.06***
	(0.99)	(1.07)	(3.59)	(7.50)	(-0.59)	(0.59)	(1.33)	(16.07)	(-0.23)	(4.16)
N	514	514	514	514	514	514	514	514	514	514
***Panel D*: distance** ≤ 1										
Treat	0.34	0.44	0.61***	1.25***	-0.24	0.24	0.09	0.91***	-0.15	1.15***
	(0.34)	(1.17)	(3.55)	(6.66)	(-0.94)	(0.94)	(1.36)	(14.03)	(-0.53)	(4.02)
N	320	320	320	320	320	320	320	320	320	320
***Panel E*: distance** $= 0$										
Treat	0.38	0.42	0.58***	1.23***	-0.24	0.24	0.08	0.92***	-0.16	1.16***
	(0.37)	(1.10)	(3.26)	(6.53)	(-0.91)	(0.91)	(1.25)	(15.00)	(-0.55)	(3.99)
N	168	168	168	168	168	168	168	168	168	168

Notes: This table reports the results of regression equation (2.1). The sample includes only bins 1 and 2.
† Distance refers to the distance to the threshold. To illustrate, the council sizes for which distance equals 0 are 10, 11, 20, and 21, while the council sizes for which distance equals 1 are 9, 12, 19, and 22. t statistics from standard errors clustered by municipality in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

2.10.3 Appendix to Learning about women's competence

Computing the running variable in the regression discontinuity design of section 2.6.3

The purpose of the regression discontinuity design of section 2.6.3 is to compare the strategies of parties that marginally won a PR councilor to those that marginally lost a PR councilor in the previous election. Thus, we are interested in the causal effect of having won a female PR councilor. We take marginal parties to be the *two* parties that either marginally won or lost the *last* PR seat. In order to differentiate marginal winners from losers, we measure how far off the vote share received by a party was, from the share it needed to win that seat. The running variable for party

p in municipal council c at election cycle t equals $v_{cpt} = voteshare_{cp,t-1} - \bar{v}_{cp,t-1}$, where $\bar{v}$ denotes the verdict-determining vote share.

To compute $\bar{v}$, we first need to describe the rules by which PR seats get allocated:

Rules for allocating PR seats

1. Among parties running for prop rep in a locality, only the parties getting ¿=5% of votes qualify.
2. Of the qualifying parties, first compute X =(number of prop MP seats in the locality)×(vote share of each qualifying party).
3. Allocate to each qualifying party the number of seats equal to the integer part of X.
4. Allocate the remaining seats by the ranking of the decimal part of X.

E.g. Municipal council A has 3 PR seats. There are 3 parties (1, 2, and 3) running for proportional representation. The vote shares of the parties are: party 1: 60%, 2: 38%, and 3: 2%. Party 3 got less than 5%, so it does not qualify. Among the qualifying parties, the vote shares are then party 1: $60/(60+38) \approx 61.22\%$, and 2: $38/(60+38) \approx 38.77\%$. The values of X's are party 1: $3 \times 0.6122 \approx 1.83$, and 2: $3 \times 0.3877 \approx 1.16$. Parties 1 and 2 both have 1 in the integer part of X, so they first get one PR councilor each. The last PR seat goes to party 1, because $0.83 > 0.16$.

Below, we compute $\bar{v}$ for all possible contest scenarios.[20] While doing so, we distinguish whether the marginal candidate that won – or nearly won – the last PR seat corresponds to the first, second, or third PR candidate in a party's list. For notational convenience, we call s the position in the party list of the marginal candidate of a party, and V the sum of the vote shares (among qualifying parties) received by the two marginal parties.

$\boxed{1}$ When there is one PR seat in the municipality

i) The two most popular parties contest over the only PR seat. Marginal parties: ranks 1 and 2

- Rank 1: $\bar{v} = \frac{V}{2}$, $s = 1$
- Rank 2: $\bar{v} = \frac{V}{2}$, $s = 1$

$\boxed{2}$ When there are two PR seats in the municipality

i) The contest is over whether the rank-2 party wins the second PR seat. Marginal parties: ranks 1 and 2

- Rank 1: $\bar{v} = \frac{2V+1}{4}$, $s = 2$

[20] An example is when there are three PR seats in a municipality, and the rank-1 and rank-2 parties contest over the last seat. Let v_n denote the vote share (among qualifying parties) received by the rank-n party. Rank 1 wins if $3v_1 - 2 > 3v_2 \iff v_1 > v_2 + \frac{2}{3}$. Therefore, $\bar{v}$ for the rank-1 party equals $v_2 + \frac{2}{3}$. On the other hand, $\bar{v}$ for the rank-2 party equals $v_1 - \frac{2}{3}$.

– Rank 2: $\bar{v} = \frac{2V-1}{4}$, $s = 1$

$\boxed{3}$ When there are three PR seats in the municipality

i) The contest is over whether the third PR seat goes to the rank-1 party or the rank-2 party. Marginal parties: ranks 1 and 2

– Rank 1: $\bar{v} = \frac{3V+2}{6}$, $s = 3$

– Rank 2: $\bar{v} = \frac{3V-2}{6}$, $s = 1$

ii) Where the rank-2 party wins a seat for sure, the contest is over whether the third PR seat goes to the rank-1 party or the rank-3 party. Marginal parties: ranks 1 and 3

– Rank 1: $\bar{v} = \frac{3V+1}{6}$, $s = 2$

– Rank 3: $\bar{v} = \frac{3V-1}{6}$, $s = 1$

$\boxed{4}$ When there are four PR seats in the municipality

i) The contest is over whether the fourth PR seat goes to the rank-1 party or the rank-2 party. Marginal parties: ranks 1 and 2

– Rank 1: $\bar{v} = \frac{4V+3}{8}$, $s = 4$

– Rank 2: $\bar{v} = \frac{4V-3}{8}$, $s = 1$

ii) Where the rank-1 party wins two seats for sure and the rank-2 party wins a seat for sure, the contest is over whether the fourth PR seat goes to the rank-1 party or the rank-2 party. Marginal parties: ranks 1 and 2

– Rank 1: $\bar{v} = \frac{4V+1}{8}$, $s = 3$

– Rank 3: $\bar{v} = \frac{4V-1}{8}$, $s = 2$

iii) Where the rank-1 party wins two seats for sure and the rank-2 party wins a seat for sure, the contest is over whether the fourth PR seat goes to the rank-2 party or the rank-3 party. Marginal parties: ranks 2 and 3

– Rank 2: $\bar{v} = \frac{4V+1}{8}$, $s = 2$

– Rank 3: $\bar{v} = \frac{4V-1}{8}$, $s = 1$

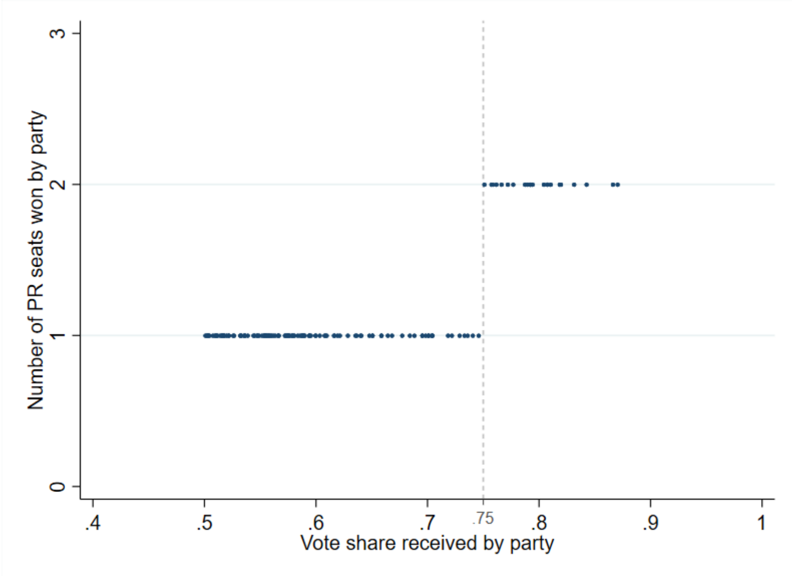
iv) Where the rank-2 and rank-3 parties win a seat each for sure, the contest is over whether the fourth PR seat goes to the rank-1 party or the rank-4 party. Marginal parties: ranks 1 and 4

– Rank 1: $\bar{v} = \frac{4V+1}{8}$, $s = 2$

– Rank 4: $\bar{v} = \frac{4V-1}{8}$, $s = 1$

As an example, take the case of the rank-1 party in a municipality with two PR seats and two qualifying parties. The party's $\bar{v} = 0.75$, according to the computation given above. Indeed, Figure 2.A3 shows that among such rank-1 parties, those receiving a vote share greater than 0.75 win two PR councilors whereas those receiving a vote share below 0.75 win one PR councilor.

Figure 2.A3: Marginal winners and losers of the last PR seat, among rank-1 parties in municipalities with two PR seats and two qualifying parties



Notes: This figure shows that in municipalities with two PR seats and two qualifying parties, the rank-1 parties must receive a vote share greater or equal to 0.75 in order to win both PR seats. The reason the vote share received is always greater than 0.5 is because these parties are the rank-1 parties. Note that the vote share is the share of votes among qualifying parties only.

Chapter 3

Gender Ratio and Group Interactions: Evidence from Municipal Council Meetings in South Korea

3.1 Introduction

In recent years, the under-representation of women in decision-making bodies, either in politics or business, has attracted increasing attention. Measures such as gender quotas have been widely employed[1], and are one of the main policy recommendations used to fight gender imbalance.[2] However, little is known about the decision-making process in the mixed-sex groups that these policies foster, mainly due to the challenge of getting detailed information on group interactions outside of experimental settings. Nonetheless, shedding light on this margin is crucial if we want to improve the status quo; equality in numbers might not translate into equality in voice in decision-making, especially in countries with conservative gender norms. Women might face barriers against fully contributing to the discussion despite having a seat at the table, if they are introduced into male-dominated bodies before attitudes have changed sufficiently to accommodate them.

In this paper, we aim at documenting to what extent and how group dynamics, the decision-making process, and decision outcomes changed after the introduction of the gender quota in local political bodies in South Korea. We do so by analyzing the word-for-word transcripts of over 150,000 meetings: over 60 meetings per year across 26 years, 7 election cycles, and 226 municipal councils.

The South Korean setting is particularly well-suited to answer this research question. First,

[1]130 countries around the world employed gender quotas in the national parliament according to International IDEA (2021)

[2]The 2021 World Economic Forum, for example, recommended setting targets for women in leadership on a government and business level to advance more women into management and leadership positions (World Economic Forum, 2021).

the gender quota of 2006 instigated a sizable and abrupt change in the composition of municipal councils. In the first election with the quota, the female share of councilors jumped from an average of 3% to an average of 15%. The arrival of women therefore had the potential to shake up group interactions in council meetings.

Second, we have access to unusually rich and detailed information regarding these local political bodies' decision-making processes and operations. Municipal councils are legally required to publish information regarding their elected members (such as gender, age, education, work experience, party affiliation), and detailed minutes of council meetings, with precise information regarding the identity of attendees, people speaking, and transcription of every speech. Every municipality has a website where all this information is available to the public.

Third, the setting not only allows us to study how group dynamics change immediately after the introduction of the quota, but also how they evolve. By exploiting eleven years (three election cycles) before and fifteen years (four election cycles) after the introduction of the policy, we have access to data about 21,470 councilors, each of them attending on average of 320 meetings per election cycle. Hence, we can shed light on the evolution of group interactions for the same group of councilors during the four years of their term, as well as for different groups of councilors across multiple election cycles.

We perform three types of analyses and present three sets of evidence. The first part of the paper focuses on documenting the effect of the quota on female representation. We perform an analysis at the meeting level to show how the gender composition of meeting attendees changed within the same municipality after the introduction of the quota. As expected, the quota successfully increased female representation at meetings. Attendance increased on average by 1.11 women in the first election cycle (cycle four) after the policy implementation and kept growing in subsequent cycles. As the number of elected female councilors kept increasing above quota requirements in subsequent cycles, the number of women attending meetings increased steadily. By the fourth election cycle after the introduction of the policy, 2.2 additional women attended meetings on average in the same municipality relative to the pre-quota numbers.

The second part of the paper discusses the extent to which the increased representation gave rise to an increased in participation and influence in discussions by analyzing the transcripts of each meeting. First of all, we exploit detailed meeting records to compare attendees and speakers and study the evolution in the propensity to speak during meetings ("extensive margin"). Second, following the political science and psychology literature (e.g. Karpowitz and Shaker, 2012; Newman et al., 2008; Mehl et al., 2007), we analyze meeting transcripts and define three measures of group dynamics: the number of interventions, the total length of speech, and the average length of speech for each speaker. We use these measures to shed light on the role that different parties play in the conversation and on the nature of the interactions ("intensive margin"). The total length of speech - defined as the total number of words that each person says in a meeting - is an indication of how much "space" a person takes up during a meeting. The average length of speech - defined as the average number of words that each person says in an intervention, together with the number of interventions, instead provide us with an indication of the nature of the interactions. We perform a within-meeting individual-level analysis to estimate the gender gap in the propensity to speak and the nature of interactions between men and women attending the same meeting. By controlling for meeting fixed effects, we can isolate group dynamics from changes in meeting characteristics over time that would confound the estimate of the gender gap.

Two patterns emerge. First of all, group dynamics change significantly before and after the introduction of the quota. Before the quota, women's propensity to speak is actually higher than men's, and women take up a bigger part of the conversation. This suggests that the very few female councilors pre-quota were a very selected group of women. However, after the introduction of the quota, the gender gap in participation becomes negative as attendance in meetings does not immediately translate into an equal voice for the women introduced by the quota. The participation gap is large in magnitude; women have a 6.1% lower probability of speaking than men attending the same meeting. Furthermore, the women who speak are significantly less vocal than men (as measured by total length of speech), driven by a lower number of speech instances compared to men.

Secondly, the composition of municipal council meeting interventions evolves over time after the quotas. We find that the gap in participation both at the extensive and intensive margin decreases every election cycle and almost disappears by the final cycle (four cycles after the introduction of the quota). Furthermore, group interactions are not "static", but evolve for the same group of councilors during the four years of appointment. Female councilors participate more, become more vocal, and speak up more frequently over time, significantly reducing the gender gap by the end of the term.

In the last part of the paper, we exploit the depth of the information available to shed light on the mechanisms underlying the evolution in group interactions observed in the data after the quota, in terms of experience, learning, and selection.

A part of the evolution in the gender gap over time is related to women having lower job experience than men. By comparing rookies to incumbents within the same meeting, we can show that rookies of both genders participate significantly less and are significantly less vocal than incumbent councilors.

Nevertheless, experience alone cannot explain the overall effect. Female rookies display a negative gap in participation, number, and length of interventions relative to male rookies, who equally lack councilor experience, within the same meeting. This gap, however, almost closes by the end of the four years of the mandate. Moreover, the very same women start off as vocal as any other incumbent men when they become re-elected in the following election cycle. This suggests that women have the same potential as men but are not able to realize it in the beginning. The finding also resonates with lab and field experiments which document that token women participate less in group discussions and are perceived as less influential by their peers than women on majority-women teams (Karpowitz and Stoddard, 2020; Bordalo et al., 2019; Chen and Houser, 2019; Coffman, 2014). The closing of the gap is consistent with learning as women and men learn about each others' authority and influence.

Lastly, despite several papers in the economic literature provide evidence that the introduction of the quota affects competence and quality of candidates and elected politicians (Bagues and Campa, 2021; Bagues and Campa, 2020; Besley et al., 2017a; Baltrunaite et al., 2014), the change in the quality of councillors across cycles does not seem to be able to explain the evolution in female participation observed in this setting. We see no change across election cycles in the gender gap for rookies (and incumbents) in the first year of appointment, indicating that "new" women start off from the same position every election cycle.

We contribute to a steadily growing literature studying how group composition affects individuals' behavior in male-dominated settings (e.g. Gloor et al., 2020; Coffman and Shurchkov, 2019; Chen and Houser, 2017; Born et al., 2020) and mixed-gender settings (Dupas et al., 2021; Miller and Sutherland, 2021; Karpowitz and Stoddard, 2020). These papers find that women's perceived expertise, influence, and willingness to take on leadership roles are lower in male-dominated settings. Our paper expands on this literature by being one of the first papers to study how group composition affects the decision-making process and its outcomes and speak to its evolution over time in a setting outside the lab with real consequences.

This study also sheds light on the effects of the introduction of gender quotas. A sizable body of papers study the impact of quotas - and more generally female representation - on women's political presence (Beaman et al., 2009; Bhalotra et al., 2018; Casas-Arce and Saiz, 2015; Brollo and Troiano, 2016), competence/quality of future candidates and elected politicians (Lee and Zanella, 2021; Bagues and Campa, 2021; Bagues and Campa, 2020; Besley et al., 2017b; Baltrunaite et al., 2014), allocation of public expenditure (Braga and Scervini, 2017; Rigon and Tanzi, 2012; Ferreira and Gyourko, 2014; Chattopadhyay and Duflo, 2004; Clots-Figueras, 2012; etc.), and decision-making of hiring committees (Bagues et al., 2017). However, while several studies look at final outcomes, only a few can shed light on the mechanisms that lead to these results. Thanks to the requirement to publish transcripts of meetings, we can not only look at the effect of the quota on political outcomes, but we can speak to the nature and evolution of the decision-making process.

The remainder of the chapter is organized as follows. Section 2 describes the institutional setting and political environment in which the quota is introduced, and provides some information on the functioning and roles of municipal councils. In Section 3 we describe the data. Section 4 displays the evolution of attendees' composition and nature of interactions of meetings before and after the introduction of the quota. Section 5 describes the mechanisms through which the observed effects might occur, and Section 6 concludes.

3.2 Institutional setting

There are 226 municipal councils in South Korea. They were established in the mid-1990s to constitute the legislative arm and work with municipal governments, the executive branch, to oversee local matters. Councils have several legally defined responsibilities, including reviewing and approving the spending of municipal governments, adopting and revising local bills, monitoring the municipal governments' administrative functions, and examining petitions submitted by residents. Municipal governments administer around a third of South Korea's total public expenditure (Ministry of the Interior and Safety, 2018).

3.2.1 The electoral rule and gender the quota

Councilors are elected every four years. Seven elections have been held so far, with 2018 being the latest election year. Up to the third election in 2002, all councilors were directly elected through plurality vote in single-member constituent wards. In 2006 a parallel voting system was introduced. At least 10% of the councilors needed to be elected through party-list proportional representation. The remaining seats were reserved for plurality voting in multi-member constituent wards, where each constituency elected between 2 and 4 councilors [3].

In 2006, a gender quota was also put in place: all odd-number candidates in the proportional party-list needed to be female. The reform was first adopted in the general election for the National Assembly in 2004 and was subsequently expanded to municipal elections in 2006. The motivations behind the adoption of the quota seemed to be only partially related to the importance of increasing female representation. On the one hand, women's organizations strongly demanded increased female representation in politics, which was dramatically behind the international average at the time.[4] On the other hand, the adoption of the quota seemed to be motivated by political tactics (Jeon, 2013; Kim, 2005). Some argue that political parties wanted to increase the size of the National Assembly back to what it was before the size cut during the Asian Financial Crisis. Hence, the fact that the majority of the seats added with the quota would go to females made it a good excuse to expand the Assembly. Some others argue instead that political parties were interested in the introduction of a party nomination system and used the quota to justify it, as the gender quota was embedded in the proportional representation arm where party nomination was essential.

Independently of the reasons that led to their introduction, quotas resulted effective in increasing female representation. As illustrated in Figure 3.1, the female ratio sharply rose after the reform. Municipal councils were severely male-dominated prior to the reform, with only 3% of councilors being female. After the introduction of the quota in 2006, female representation in municipal councils increased steadily, reaching more than 30% in the last election in 2018. As the number of seats reserved for proportional representation is small, most elected councilors turned out to be the first candidates in the lists and, therefore, female. Furthermore, women's share increased also among the councilors directly elected through plurality vote.

3.2.2 Municipal councils' functioning

The activity of municipal councils is organized in sessions, where councilors meet to discuss their legislative matters and are composed of a series of meetings with different duration. Each session is made of plenary meetings and committee meetings. Plenary meetings are assemblies where every councilor is present. In contrast, committee meetings are attended only by committee members, and councillors discuss issues pertinent to the committee's expertise.

Each session starts with a plenary meeting where councilors approve the agenda to be discussed (e.g. proposals for new bills or amendments for existing bills, spending proposals, budget approval).

[3] The maximum number of candidates a party could nominate for a ward equals the preset number of seats for that ward.

[4] See Cho and Kim (2010) for a summary of the major activities of women's organizations.

Local councils discuss a wide range of topics, spanning from transport, infrastructure, energy, education, welfare, health to more administrative issues such as budget revision or approval. Once the agenda is set, the council chair delegates items to specific committees to be reviewed in detail. Items not relevant to a specific committee or particularly urgent are discussed in the plenary sessions. After each committee has discussed items relative to its expertise in separate committee meetings, they report their findings, discussion outcomes, or official proposals to the whole council in a plenary session where all proposals are voted on.

3.3 Data

We assembled a new dataset that combines rich background information on elected councilors and detailed information on council meetings. Municipal councils are legally required to publish information regarding their elected members and transcripts of each meeting. However, the publication of information is not motivated by research purposes but by the necessity to improve accountability and transparency. Hence, data are not stored in an organized, easy to collect and analyze format. The data collection has entailed webscraping these information directly from the website run by the National Election Commission and the website of each individual council.

3.3.1 Background information on elected councillors

By web scraping the website of the National Election Commission, we obtained information regarding various background characteristics of elected councilors such as name, gender, date of birth, age, occupation, education, relevant work experience, and information on electoral outcomes such as party affiliation, vote outcomes, councilors' roles, and committee membership.

Table 3.1 provides descriptive statistics on the gender composition of councils by election cycle. A total of 21,470 people held a seat in local municipal councils in South Korea across the 226 municipalities and 7 election cycles. Municipalities are very heterogeneous in size. The smallest municipality has 7 councilors, while the biggest one has around 40 elected members. This heterogeneity in size implies significant variation in the number of councilors required to be elected through the proportional arm (at least 10%), and as a consequence in the number of women that are required to hold a seat in the municipal council due to the quota system.

3.3.2 Information on council meetings

The second source of data is individual council websites. These contain detailed transcripts of council meetings, with precise information regarding meeting date and time, duration, agenda items discussed, the identity of attendees and people speaking, and the transcription of every speech with its duration. This information is made available to the public due to the legal requirement for municipal councils to publish transcripts of council meetings.

3.3.3 Sample of analysis

We collected data on 113 municipalities. Table 3.2 provides descriptive statistics on this subsample of municipalities. By comparing them to the overall South Korean municipal councils, it can be noticed that the municipal councils of the collected municipalities are slightly bigger with respect to the municipal council in the average municipality in South Korea. As a matter of fact, they display a higher number of women among elected councillors.[5]

The sample for the analysis includes municipalities and meetings for which we have information on attendees and speeches (96.4% of meetings). Furthermore, the analysis focuses on committee meetings, which represent 65% of the overall number of meetings in each locality. We exclude plenary meetings (28.67%) and extraordinary meetings (2.93%). This selection is made to ensure that we are comparing homogeneous meetings. While plenary meetings are events where everyone is present and that primarily consist of structured meetings (such as opening or closing ceremonies or meetings where councilors discuss the agenda for the session), committee meetings are the place where discussions happen and where proposals for new bills or amendments of existing bills, spending and budget proposals are revised and discussed in detail.

The unit of analysis is a meeting segment. In the same meeting, different issues (agenda items) are discussed. Each issue is the object of the discussion of a separate segment. Thus, each segment is considered as the reference unit of analysis of group interactions and decision-making.

The sample of analysis is further restricted to meeting segments where there is at least 1 interaction and that are not initial introductory segments where the chair is presenting the topic of discussion (1.41% of observations are dropped. In 90% of times, the person speaking is the chair, the committee chair or his substitute).

Table 3.3 displays some descriptive statistics for the final sample. It comprises information on a total of 7,579 councillors, each attending an average of 63 committee meetings per year, for a total of 252 committee meetings during the 4 years of their appointment. This gives us a total of 159,580 meetings where we can explore councillors interactions. Each meeting has on average 2.15 segments and each segment is attended by 7.67 councillors. Among the attending councillors 3.79 speak and the average number of interactions is 72.21.

3.4 Do women get a seat at the table?

Councils are required to publish information regarding the attendees at each meeting. We exploit this information to identify the number of men and women participating at meetings in different election cycles. This allows us to study the evolution of the gender composition of meetings over time. We use the following specification, separately for men and women:

$$y_{csml} = \alpha_l + \sum_{c=1,c\neq 3}^{7} \beta_c \times \text{Election Cycle}_c + \gamma X_{smlc} + \epsilon_{clms} \tag{3.1}$$

Where the dependent variable is the number of women or men attending segment s of meet-

[5]These municipalities were not selected at random, but they were the municipalities for which it was easier to collect data. We are currently collecting data for the remaining municipalities.

ing m, held in election cycle c, in municipality l. α_l represents municipality fixed effects, while $\sum_{t=1}^{7}$ Election Cycle$_t$ are dummies equal to one for each of the seven election cycles in the period of analysis. Cycle 3, the election cycle just before the introduction of the quota, is the reference cycle. Thus, β_c measures the within-municipality variation in the average number of women and men attending meetings in each election cycle compared to the status quo in the election cycle before quotas were introduced. X_{smlc} are segment level controls, such as the total number of interactions in the segment and a fixed effect for the type of meeting (budget, regular, or special meeting). They are introduced to control for potential changes in the length and type of meetings across election cycles. The standard errors are clustered at the municipality level.

Figure 3.2 displays the results of Specification 3.1. Each circle represents the β_c for men and women in each election cycle, in blue and red respectively. The gender composition of meetings drastically changes after the introduction of the quota in cycle 4 elections. As expected, the quota is successful in increasing the number of female elected councillors, and as a consequence, female representation at meetings. Attendance increases by 0.90 women in election cycle 4 compared to election cycle 3. The number of women at meetings more than doubled, as the number of women attending meetings before the introduction of the quota was on average 0.33. In line with the fact that the number of elected female councillors grows above the the quota requirement in election cycles 5, 6, and 7, the number of women attending meetings increases steadily. It reaches 2.04 additional women compared to cycle 3 in the last election cycle.

The increase in women is accompanied by a more than proportional decrease in the number of male councillors over time. As a matter of fact, the number of male councillors in the first cycle after the introduction of the quota (cycle 4) is 2.75 lower than the last cycle before the quota (cycle 3), and keeps going down reaching 3.81 fewer men in cycle 7.[6]

3.5 Does this translate into women having a voice?

3.5.1 Propensity to speak

Increased attendance at meetings does not necessarily translate into increased participation in discussions, as women might face barriers to actively contribute to the decision making process when introduced in male-dominated environments. Exploiting the requirement for local councils to publish detailed transcripts of each meeting, we compare attendance records and speakers records to determine who speaks during each session. We define a measure for the propensity to speak at meetings, i.e. the probability that a person makes at least one intervention during a meeting segment.

Evolution across election cycles

Do women and men have the same probability to speak at meetings? To give an answer to this question, we compare the extent to which men and women attending the same meeting segment

[6]The number of men in meetings in the first election cycle is significantly higher than the following election cycles before the introduction of the quota. This is due to the fact that municipal councils were slightly fewer and bigger when they were initially introduced.

take part to the conversation. Figure 3.3 shows the results of the following specification:

$$y_{csmli} = \alpha_{ms} + \sum_{c=1}^{7} \beta_c \times \text{Female}_i \times \text{Election Cycle}_c + \epsilon_{icsml} \tag{3.2}$$

The unit of analysis is individual i attending segment s of meeting m in cycle c and municipality l. The dependent variable is a dummy equal to one if a person speaks in the segment and zero otherwise. α_{ms} represent segment fixed effects, which are introduced to control for observed and unobserved factors that make segments different from each other, such as meetings duration, committee, topic, structure, type and number of attendees, day and month, etc. Segment fixed effects also capture election cycle and year fixed effects, controlling for observed and unobserved factors that are year specific but common across localities. Furthermore, they additionally capture municipality $\times$ year fixed effects, allowing us to control for potential confounding factors that are year and municipality specific. $\sum_{c=1}^{7} \text{Female}_i \times \text{Election Cycle}_c$ are a series of election cycle dummies interacted with a female dummy. Hence, β_c captures the gender gap in the probability to speak between men and women who attend the same segment s of the same meeting m in cycle c and municipality l, keeping common segment characteristics constant. Standard errors are clustered at the municipality level. The exact magnitude of the estimated effects can be found in Table 3.4.

We can notice two patterns. First, before the introduction of the quota there were very few women, but they were as vocal as men, or slightly more. As a matter of fact, there is no significant gender gap in propensity to speak in cycles 1 and 2, and a significant and positive gender gap in cycle 3 (women had a 3.4 percentage points higher probability to speak at meetings with respect to men).[7]. However, gender dynamics change when gender quotas are introduced. Conditional on attendance, women have a significantly lower probability of speaking during meetings in the first cycle after the introduction of the policy. In terms of magnitude, the participation gap is economically large: women have a 6.1 percentage points lower probability to speak, which corresponds to a gap of 12.2% (compared to the mean of the dependent variable in the same cycle, 50.2 percentage points).

Second, municipal council meeting dynamics evolve over time. The negative gender gap in propensity to speak decreases over time, becoming equal to 4.4 percentage points in cycle 5, 2.1 percentage points in cycle 6, and finally disappears in cycle 7 (0.005). Perhaps women acquire greater knowledge and confidence, group interactions evolve over time and women become better assimilated into the group, or the pattern is driven by a change in the type of female and male councillors that are elected over time.

Evolution within election cycles

The richness of the data allows us to explore an additional margin. Since we have access to meeting transcripts and local councils are required to provide information regarding the date of the meeting, we can identify when each meeting takes place. This allows us to explore how participation evolves within the same election cycle for the same group of councillors across the

[7]Standard errors are bigger before the introduction of the quota as there were very few women among MPs

years of the appointment. Figure 3.4 shows how the probability to participate at meetings changes over the years, by displaying the residuals, $\hat{\epsilon}_{igmt}$, averaged by gender g and year t of a regression of the probability to speak Y_{igmt} for councilor i of gender g in segment s of meeting m in year t on segment fixed effects α_s:

$$Y_{isgmt} = \alpha_s + \epsilon_{isgmt} \tag{3.3}$$

Segment fixed effects account for observable and unobservable factors that make segments within meetings different from each other, such as the agenda and the group of councilors, providing a comparable indication of the gender gap in the nature of interactions over time. We employ the logarithm of the length of speech (total and average), so that the coefficient of the female dummy will provide us information on the gender gap, expressed in percent difference in length of speech between men and women.

Figure 3.4 shows that the propensity to speak is not "static", but evolves for the same group of elected councillors during the years of the appointment. Female councilors post-quotas start their appointment significantly less vocal than men. However, they gradually increase their propensity to speak over the years.[8] This pattern can be observed for all election cycles.

Table 3.5 and Figure 3.5 provide a formal test for the evolution of the gender gap in propensity to speak over the years of the appointment for the post-quota period.[9] It displays the estimates of the following specification:

$$y_{tmsli} = \alpha_s + \beta_1 \times \text{Female}_i + \sum_{t=2}^{4}[\beta_t \times \text{Female}_i \times 1(\text{Year} = t)] + \sum_{t=2}^{4}[\gamma_t \times 1(\text{Year} = t)] + \epsilon_{itsml} \tag{3.4}$$

where the unit of analysis is individual i attending segment s of meeting m in year t after the elections and municipality l. α_s represents segment fixed effects. $\sum_{t=2}^{4} \text{Female}_i \times 1(\text{Year} = t)$ are dummies equal to one for each year after the elections interacted with a female dummy. In this case, β_1 estimates the gender gap in participation in the first year after the elections, while β_2, β_3, and β_4 provide an indication of the extent to which the gender gap between men and women evolves across the years within the same election cycle. Standard errors are clustered at the municipality level.

3.5.2 Nature of interactions

The probability of speaking during meetings is not informative about the role of different parties in the conversation, or the nature of the interactions. In order to shed light on the evolution of group dynamics over time, we follow the political science and psychology literature, and we focus on the number of interventions during meetings and the length of speech (e.g. Karpowitz and Shaker, 2012; Newman et al., 2008; Mehl et al., 2007). The number of words in each speech is one of the linguistic tools most widely used as a proxy for voice and authority. Based on the count of words, we define two measures: the total length of speech and the average length of speech. The total length of speech, defined as the total number of words that each person says in a meeting, is an indication

[8] The estimates before the quota have to be taken with caution given the small number of women among councillors.

[9] In the period before the quota the number of women is too small to deliver meaningful estimates.

of how much "space" a person takes up during a meeting. Gender differences in total length of speech provide us with an indication of the extent to which the conversation is monopolized by a gender, leaving the other gender in a marginalized position during the meeting. On the other hand, the average length of speech, defined as the average number of words that each person says during their intervention, and the number of interventions provide us with an indication of the nature of the interactions. It makes a huge difference to group dynamics whether the speech is a monologue or part of a conversation with other individuals. Back-and-forth exchanges between individuals (e.g. Q&A or pro-con debate on bill proposals) are significantly shorter than monologues (e.g. audit report, bill proposal, 5-minute speech on a topic of choice). Thus, women displaying a higher number of interventions for a lower amount of time might indicate that they engage in more heated conversation with more back-and-forth, while a lower number of interventions for a longer amount of time might instead would be a sign of women and men taking up different roles and not really engaging in a debate.

Evolution within and across election cycles

Figure 3.6 displays the result of the same type of analysis (Specification 3.3) performed for the propensity to speak. It shows the residuals, $\hat{\epsilon}_{igmt}$, averaged by gender g and year t of a regression on segment fixed effects α_s of logarithm of total and average length of speech and number of interactions (Y_{igmt}) for councilor i of gender g who spoke at least once in segment s of meeting m in year t . Tables 3.5, 3.4, and Figures 3.7, 3.8 provide a formal test. Table 3.4 and Figure 3.7 display the estimates of Specification (3.2), providing evidence of the evolution of the gender gap in number of interactions, total and average length of speech across election cycles. Table 3.5 and Figure 3.8 display the estimates of Specification (3.4)

Exactly as the propensity to speak, group dynamics change significantly after the introduction of the quota. Focusing on the total length of speech (first graph), we can see that women not only were more likely to speak before the introduction of the policy, but they were also taking up a bigger part of the conversation. On the other hand, they speak significantly less and are significantly less vocal than men when they do after the introduction of the quota.[10] In terms of magnitude, women's average total length of speech is 25% lower with respect to men's speeches in cycle 4.

This is due to the fact that those women who take part to council meetings after the introduction of the quota engage in a lower number of interactions with respect to men (second graph). Women on average engage in discussions 3.14 times less than men attending the same meeting in cycle 4. The average length of speech (third graph) is not significantly different by gender in the period after the introduction of the quota, implying that it is not the case that women are excluded from conversations: men and women are likely to be involved in similar types of interactions with other meeting participants.

Furthermore, group dynamics appear to change across election cycles. The average gender gap in total length of speech - the area comprised between the red and the blue lines - gets smaller,

[10]The estimates before the quota have to be taken with caution given the small number of women among councillors.

indicating that women gradually become as vocal as men across election cycles, even though the gap does not close. Women speak 8.9% less than men by cycle 7. This seems to be driven by an increase in the average number of interactions rather than the average length of speech. The average gender gap in number of interventions reduces by as much as 50% by cycle 7.

Lastly, interactions evolve for the same group of elected councillors during the years of the appointment exactly as for the propensity to participate. Female councilors post-quotas become more vocal and speak up more frequently, significantly reducing the gender gap by the end of the appointment. Women in the first year of appointment take up 22.6% less space than men during meetings and speak on average 4.289 less times. These gaps decrease by 47.3% and 58.63% respectively by the last year of appointment.

3.6 Mechanisms

After the introduction of gender quotas in local councils in South Korea, the number of women participating into council meetings increases significantly. However, attendance does not immediately translate into full contribution into discussions. As a matter of fact, women initially display a lower propensity to speak, and also conditional on speaking are significantly less vocal than men. However, this negative gender gap in contribution decreases with time across election cycles, and within the same election cycle over the years of the appointment.

There exist a multiplicity of explanations that can be consistent with the observed evolution of group dynamics over time. For instance, several papers document the existence of a significant gender gap in the willingness to contribute to ideas (Bordalo et al., 2019; Chen and Houser, 2017; Coffman, 2014), and willingness to be leaders in group decision-making tasks (Born et al., 2018) in group settings in stereotypically male-typed domains in controlled laboratory experiments. These differences become stronger in mixed gender groups, particularly when the team is majority male. For example, Karpowitz and Stoddard (2020) in a multi-year field experiment find that token women participate less in group discussions. Furthermore, they are perceived as less influential by their peers and are less likely to be chosen to represent the group than women on majority-women teams. Figure 3.9 displays the gender composition of meeting attendees over the election cycles after the introduction of quotas. It appears very clearly that the share of women in meetings significantly increases over time along with the increase in the number of female councillors. Hence, a first potential explanation for the existence and evolution of this gender gap in participation can be differences in propensity to contribute to discussions between men and women, that are especially important in the first cycles, when elected councillors are still predominantly men.

Another strand of literature suggests that gender stereotypes may play an important role in understanding how teams discuss. For instance, Coffman and Shurchkov (2019) designs a controlled laboratory experiment where teams brainstorm answers to questions that vary according to the gender perception of the topic involved (the perceived “maleness” of the question). As the maleness of the question increases, women appear to be rewarded less for their ideas. This is partly due to the fact that they self-promote less, but partly also to the fact that they are less likely to be selected to answer on behalf of the group. In light of these findings, a second explanation for the evolution in group dynamics might consist in men and women gradually learning about each other's ability through experience and repeated interactions over time. In this direction, Lee and Zanella (2021) shows that parties in municipalities facing more stringent the quota initially tried to counteract the policy by putting forth fewer female candidates in the unconstrained arm. However, this pattern gradually overturns over time, with the shift towards female candidates being stronger when parties experienced more educated women in the previous election cycles, pointing towards learning about the competency of females being an important stepping point for reaching gender parity in opportunities.

Third, differences in contribution to discussions might just be related to experience. Before the introduction of the quota, women were almost absent from municipal councils. Hence, those women who got elected thanks to the introduction of the quota might have been new to the task with respect to men, and thus would need some time to "learn how to do the job".

Lastly, several papers in the economic literature provide evidence that the introduction of the quota has an effect on competence and quality of candidates and elected politicians (Bagues and Campa, 2021; Bagues and Campa, 2020; Besley et al., 2017a; Baltrunaite et al., 2014). The evolution of gender dynamics over time might be then related to a change in the quality of the elected local councillors, with the gender gap in average quality of councillors decreasing over time.

In the following sections we provide suggestive evidence to shed light on what channels are more likely to explain the observed patterns.

3.6.1 Experience

Besides beliefs updating via experiential learning, another observationally equivalent explanation for the change in group dynamics over time is an increase in candidates experience. Before the introduction of the quota, women were almost absent from municipal councils. Hence, those women who got elected due to the policy might have been new to the task with respect to men, and thus needing some time to "learn how to do the job". In order to explore this channel, we compare the evolution of propensity to speak, number of interventions and length of speech for incumbent and rookies. Rookies in election cycle t are all those councillors that have been elected in election cycle t for the first time in municipal councils. Any other councillor who has been elected at least once before election cycle t in municipal councils is considered an incumbent.

Figure 3.10 displays how group dynamics and the nature of interactions changed over time for rookies and incumbents of different gender. Exactly as in Figure 3.6, we follow Specification 3.3 and we display the residuals, averaged by gender and year, of a regression of each outcome (total length of speech, number of interventions and average length of speech respectively) on meeting segment fixed effects. Men are displayed in blue, while women are red. The dotted line corresponds to rookies, while the straight line describes incumbents' behavior.

Four patterns appear. (i) Part of the evolution of dynamics over time can be explained by women being new councilors. As a matter of fact, we can see that both female and male new councilors start their appointment speaking less and increase their participation over time. This is true also in terms of space they take into the discussions and number of interventions conditional on speaking.

However, the lack of experience does not affect men and women in the same way, as there exist a gender gap also among rookies. (ii) Rookie women have a lower propensity to speak, speak fewer times and for shorter amount of time compared to rookie men at the beginning of the election cycle.

Nonetheless, (iii) the gender gap among rookies does not remain constant over the years of the appointment, as female rookies gradually become as influential as men. Both male and female rookies' behavior converge towards incumbents' behavior over time within the same election cycle, but female rookies' "speed of convergence" is faster, as they gradually close the gap with respect to male rookies. Furthermore, the same rookie women become as (or more) participative and vocal, conditional on participation, than male incumbents in the following election cycle (additional evidence provided in Table 3.7). Taken together, these two pieces of evidence suggest that it is not differences in inherent talent that cause women to be less vocal and that rookie women gradually gain influence as councilors work together during the years. This hints towards the gender gap in behavior being related to initial differences in roles, perceptions of authority, influence or competence for men and women, that evolve over time with women and men working together.

Lastly, (iv) the gender gap of rookies and incumbents in the first year of appointment does not decrease across election cycles. This suggests that selection, and in particular the gender gap in the average quality of councillors decreasing over time, does not seem to be the main explanation for the evolution in group dynamics across cycles.

We provide a formal test for the evolution of the gender gap for rookies and incumbents in Tables 3.6 and 3.7. In Table 3.6 we present the estimates from Specification 3.5 to provide evidence regarding the evolution of the gender gap for rookies and incumbents within the election cycle.

$$\begin{aligned} y_{tmlsi} = {} & \alpha_{ms} + \beta_{1,1} \times \text{Rookie}_{it} + \beta_{2,1} \times \text{Rookie}_{it} \times \text{Female}_i + \beta_{3,1} \times \text{Female}_i + \\ & + \sum_{t=2}^{4} [\beta_{1,t} \times \text{Rookie}_{it} + \beta_{2,t} \times \text{Rookie}_{it} \times \text{Female}_i + \beta_{3,t} \times \text{Female}_i] \times 1(\text{Year} = t) + \epsilon_{iltms} \end{aligned} \tag{3.5}$$

where the unit of analysis is individual i attending segment s of meeting m in year t. This is analogous to the analysis performed in Specification 3.4. The only difference is that the effect is estimated for rookies and incumbents. The sample is restricted to councillors that we were able to trace over time (96.22% of the sample). Standard errors are clustered at the municipality level.

In order to shed light on the evolution of the gender gap in the first year of appointment for rookies and incumbents, we perform Specification 3.6, restricting the sample to the first year of appointment. Furthermore, we distinguish incumbents between post-quota incumbents and before-quota incumbents, depending on whether we observe the councilor for the first time before or after the introduction of the quota.

$$\begin{aligned} y_{csmli} = {} & \alpha_{ms} + \beta_{1,4} \times \text{Rookie}_{ic} \times \text{Female}_i + \beta_{2,4} \times \text{Incumbent}_{ic} \times \text{Female}_i + \\ & + \sum_{c=5}^{7} [\beta_{1,c} \times \text{Rookie}_{ic} + \beta_{2,c} \times \text{Incumbent}_{ic}] \times \text{Female}_i \times 1(\text{Election Cycle} = c) + \epsilon_{icsml} \end{aligned} \tag{3.6}$$

where the unit of analysis is individual i attending segment s of meeting m in election cycle c. Standard errors are clustered at the municipality level and the results are displayed in Table 3.7.

3.6.2 Environment characteristics: Share of women and critical mass

Another potential explanation for the change in the gender gap in participation across election cycles is the change in the composition of municipal councils over time. The number of women attending meetings might be crucial to foster women's participation and recognition for their work. In order to assess if meetings composition represents a crucial constraint, we exploit the variation in the gender composition of meetings within municipality and election cycle. Figure 3.11 displays the residual variation in gender composition of attendees after controlling for election cycle by municipality fixed effects. We perform the following analysis

$$y_{tlmsi} = \alpha_{sm} + \beta_1 \times \text{Female}_i + \beta_2 \times \text{Female}_i \times \text{Share of women}_{tlsm} + \epsilon_{itlms} \tag{3.7}$$

Where the unit of analysis is individual i in segment s of meeting m in municipality l and election cycle t. We control for segment fixed effects, which allow us to control for anything that makes meetings attended by more women different. This is particularly important given that women might have different areas of expertise and might be systematically assigned to particular committees. Furthermore, women or men with particular characteristics, for instance more experienced, might be assigned to different tasks. The coefficient of interest is the coefficient of the interaction between the share of women among attendees in the meeting and a female dummy, which will capture how the gap in participation between men and women changes with the share of women in the meeting. We also perform the same analysis by adding individual fixed effects. As we observe the same person across multiple meetings within the same election cycle, we can control for individual fixed effects and compare the behavior of the same person when surrounded by a different gender composition of attendees. The sample is restricted to the post-quota period and to meetings where the gender of all the attending councillors is known. This guarantees that the share of women is not measured with error.[11]. Standard errors are clustered at the municipality level.

In two additional specifications, we unpack the effect and we estimate a coefficient for each year of the appointment and each cycle.

$$y_{tsmli} = \alpha_{ms} + \sum_{c=1}^{4} [\beta_{1,c} \times \text{Female}_i + \beta_{2,c} \times \text{Share of women}_{mslt} \times \text{Female}_i] \times 1(\text{Election Cycle} = c) + \epsilon_{itmsl} \tag{3.8}$$

Where $1(\text{Time} = c)$ is a dummy equal to one for the year after the appointment or a dummy for the election cycle.

Table 3.8 displays the results of the analysis on gender composition. Panel A displays the results of Specification (3.8), while Panel B and Panel C provide the results of Specification (3.5) where the average result is decomposed by year of appointment and election cycle respectively.

[11] There are a few instances where we cannot identify the gender of the councillor as we do not have background information

Odd columns display the effect without controlling for individual fixed effects, that are instead present in the even columns.

In meetings with a higher share of women, women participate more and take up more space (total length of speech) conditional on speaking. In terms of magnitude, meetings with a 10% higher share of women are associated with a 0.06 percentage points lower gap in participation and 1.89% lower gap in total length of speech. Panel B shows that the effect is concentrated in the first year of appointment, where meetings with a 10% higher share of women are associated with a 0.010 percentage points lower gap in participation and 3.21% lower gap in total length of speech, but it fades away over the years. Furthermore, Panel C indicates that these patterns are stronger in the first election cycle, even though the coefficients are not significant.

However, the coefficients lose significance and also sensibly decrease in magnitude when we control for councilor fixed effects. On the one hand, the decrease in variation in the share of females when we control for individual fixed effects might be reducing our power to detect significant results. On the other hand, this could be consistent with the evidence we found on experience playing a role if meetings with a higher share of females are also meetings where more experienced women are present. This is likely to be the case, especially if we consider that in the first cycle after the introduction of the quota the women who got elected were still very few. So, a higher share of females is likely to be correlated with the presence of an experienced woman.

This evidence overall suggests that the increase in the share of women does not seem to be able to explain the whole pattern of increase in participation we observe across election cycles. As a matter of fact, we cannot completely exclude that a bigger presence of women might play a role in helping everybody to express their opinions when councillors start working together. However, consistent with learning, the importance of being surrounded by women decreases the more women and men work together and learn about each others' authority and influence.

3.7 Conclusion

This paper provides evidence on what happens to group interactions in very male dominated environments after the introduction of gender quotas by studying the evolution of participation into municipal council meetings in South Korea.

The more balanced gender ratio fostered by the introduction of the policy succeeds in changing the composition of attendees and speakers of South Korean local councils meetings. The increase in female representation however does not immediately result in equal voice for men and women. As a matter of fact, women display a lower propensity to speak during meetings at the beginning, and conditional on speaking, they intervene less often and talk for a lower amount of time. This gap gradually closes over the 3 election cycles and 15 years we observe after the introduction of the quota.

This change in group interactions can be partially explained by women gaining experience in politics. The evidence however also points towards experiential learning being a concurring factor, i.e. councillors gradually learning about each other's competence and influence by working together. These findings indicate that the effectiveness of affirmative action policies strongly depends on the progressiveness of the gender norms that characterize the environment they intervene to regulate. The introduction of gender the quota in local councils in South Korea was strongly advocated and occurred a few years after the introduction of the same policy in the national parliament. However, it took fifteen years for women to acquire an equal voice in decision-making.

Further analysis is needed to study the nature of this change in the decision-making process and its effects on the allocation of public expenditure. This project is part of a bigger agenda. We are currently analyzing the transcripts of the meetings to shed light on the content of this voice that women gained over time. Furthermore, by exploiting yearly data on the allocation of the local governments' expenditure, we will be able to speak to the economic effects of these changes in the decision making process.

References

Bagues, M. and P. Campa (2020). Women and Power: Unpopular, Unwilling, or Held Back? A Comment. *Journal of Political Economy 128*(5), 2010–2016.

Bagues, M. and P. Campa (2021). Can gender quotas in candidate lists empower women? evidence from a regression discontinuity design. *Journal of Public Economics 194*, 104315.

Bagues, M., M. Sylos-Labini, and N. Zinovyeva (2017, April). Does the gender composition of scientific committees matter? *American Economic Review 107*(4), 1207–38.

Baltrunaite, A., P. Bello, A. Casarico, and P. Profeta (2014). Gender quotas and the quality of politicians. *Journal of Public Economics 118*(C), 62–74.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: Does exposure reduce bias?*. *The Quarterly Journal of Economics 124*(4), 1497–1540.

Besley, T., O. Folke, T. Persson, and J. Rickne (2017a). Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. *American Economic Review 107*(8), 2204–42.

Besley, T., O. Folke, T. Persson, and J. Rickne (2017b, August). Gender quotas and the crisis of the mediocre man: Theory and evidence from sweden. *American Economic Review 107*(8), 2204–42.

Bhalotra, S., I. Clots-Figueras, and L. Iyer (2018). Pathbreakers? women's electoral success and future political participation. *The Economic Journal 128*(613), 1844–1878.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019). Beliefs about gender. *American Economic Review 109*(3), 739–73.

Born, A., E. Ranehill, and A. Sandberg (2018, November). A man´s world? – The impact of a male dominated environment on female leadership. Working Papers in Economics 744, University of Gothenburg, Department of Economics.

Born, A., E. Ranehill, and A. Sandberg (2020). Gender and willingness to lead: Does the gender composition of teams matter? *The Review of Economics and Statistics 0*(ja), 1–46.

Braga, M. and F. Scervini (2017). The performance of politicians: The effect of gender quotas. *European Journal of Political Economy 46*, 1 – 14.

Brollo, F. and U. Troiano (2016). What happens when a woman wins an election? evidence from close races in brazil. *Journal of Development Economics 122*(C), 28–45.

Casas-Arce, P. and A. Saiz (2015). Women and power: Unpopular, unwilling, or held back? *Journal of Political Economy 123*(3), 641–669.

Chattopadhyay, R. and E. Duflo (2004). Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica 72*(5), 1409–1443.

Chen, J. and D. Houser (2017). Gender composition, stereotype and the contribution of ideas. Technical Report 17-26.

Chen, J. and D. Houser (2019). When are women willing to lead? the effect of team gender composition and gendered tasks. *The Leadership Quarterly 30*(6), 101340.

Cho, H.-O. and E. H. Kim (2010). A Study of Historical Context and Institutionalization Process of Gender Quota System and Women's Political Empowerment Movement in Korea 2000-2010. *Journal of Korean Social Trend and Perspective* (79), 110–139.

Clots-Figueras, I. (2012). Are female leaders good for education? evidence from india. *American Economic Journal: Applied Economics 4*(1), 212–244.

Coffman, Katherine Baldiga, C. B. F. and O. Shurchkov (2019). Gender stereotypes in deliberation and team decisions. Harvard Business School Working Paper 19-069. January 2019. (Revised January 2021.).

Coffman, K. B. (2014). Evidence on Self-Stereotyping and the Contribution of Ideas. *The Quarterly Journal of Economics 129*(4), 1625–1660.

Dupas, P., A. S. Modestino, M. Niederle, J. Wolfers, and T. S. D. Collective (2021, February). Gender and the dynamics of economics seminars. Working Paper 28494, National Bureau of Economic Research.

Ferreira, F. and J. Gyourko (2014). Does gender matter for political leadership? the case of u.s. mayors. *Journal of Public Economics 112*(C), 24–39.

Gloor, J., M. Morf, S. Paustian-Underdahl, and U. Backes-Gellner (2020, 01). Fix the game, not the dame: Restoring equity in leadership evaluations. *Journal of Business Ethics 161.*

International IDEA (2021). Gender quotas database. Global Database of Gender Quotas in parliaments worldwide by International Institute for Democracy and Electoral Assistance, Inter-Parliamentary Union and Stockholm University.

Jeon, J. Y. (2013). Political Dynamics of Adopting Gender Quota in Korea. *Journal of Korean Politics 22*(1), 29–52.

Karpowitz, Chris, J. P. and O. Stoddard (2020). Strength in numbers: A field experiment on gender, influence and group dynamics. IZA Discussion Paper No. 13741.

Karpowitz, Christopher F., M. T. and L. Shaker (2012). Gender inequality in deliberative participation. *American Political Science Review 106*(3), 533–547.

Kim, S. H. (2005). Local Council Elections - "Reformed" or "Deformed"? *Pressian News*.

Lee, J. E. and M. Zanella (2021). How do political parties respond to gender quotas? evidence from south korea. Working paper. https://jayeuijunglee.github.io/website/Lee_Zanella.pdf.

Mehl, M. R., S. Vazire, N. Ramírez-Esparza, S. Richard B, and J. W. Pennebaker (2007). Are women really more talkative than men? *Science 317*(5834), 82.

Miller, M. G. and J. L. Sutherland (2021). The effect of gender on interruptions at congressional hearings. Technical report.

Ministry of the Interior and Safety (2018). Municipal councilor's activities guide. Publication Registration Number: 11-1741000-000101-15.

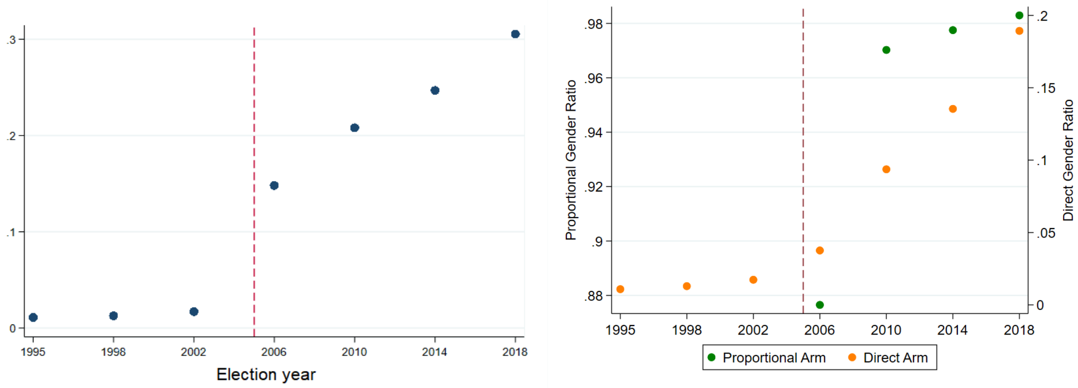
Newman, M. L., C. J. Groom, L. D. Handelman, and J. W. Pennebaker (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes 45*(3), 211–236.

Rigon, M. and G. M. Tanzi (2012, April). Does gender matter for public spending? Empirical evidence from Italian municipalities. Temi di discussione (Economic working papers) 862, Bank of Italy, Economic Research and International Relations Area.

World Economic Forum (2021). Global gender gap report.

3.8 Tables and Figures

Figure 3.1: Proportion of females in municipal councils, nationwide average



Notes: This figure illustrates the nationwide average of the gender ratio in municipal councils, for every election cycle since their emergence. The left panel illustrates the gender ratio among all councillors, while the right panel shows the gender ratio among proportional seats (green dots) and direct seats (orange dots). The vertical red dotted line indicates the year of the major reform that instituted the gender quota.

Table 3.1: Descriptive Statistics: Composition of municipal councils

	Overall Sample				
	N.	Mean	Std Dev	Min	Max
Total N. of Councillors					
Election cycle 1 (1995)	226	19.87	11.3447	7	60
Election cycle 2 (1998)	228	15.12	6.990416	7	40
Election cycle 3 (2002)	228	15.12	6.877084	7	41
Election cycle 4 (2006)	230	12.41	5.635811	7	32
Election cycle 5 (2010)	228	12.50	6.256977	7	55
Election cycle 6 (2014)	227	12.58	5.857225	7	38
Election cycle 7 (2018)	226	12.77	6.254611	7	39
N. of Female councillors					
Election cycle 1 (1995)	226	0.319	.8870975	0	7
Election cycle 2 (1998)	228	0.246	.721738	0	5
Election cycle 3 (2002)	228	0.338	.7936855	0	5
Election cycle 4 (2006)	230	1.787	1.219494	0	7
Election cycle 5 (2010)	228	2.618	1.824699	1	12
Election cycle 6 (2014)	227	3.057	1.982507	1	10
Election cycle 7 (2018)	226	3.814	2.492829	1	13

Notes: This table displays descriptive statistics on the number of councillors and female councillors for each election cycle for the 226 South Korean local municipal councils. Gender the quota were introduced in 2005. However, during the election of 2006, they remained merely a strong recommendation, so it was still legal to place a male in slot 1 of party lists. Most municipalities complied, but 14 of them had no female PR councilors. In election year 2010, the minimum of the gender ratio among PR councilors is 0 because in one council the elected woman was invalidated for being a member of multiple parties.

Table 3.2: Descriptive Statistics: Composition of collected sample of municipal councils

	Collected Sample				
	N.	Mean	Std Dev	Min	Max
Total N. of Councillors					
Election cycle 1 (1995)	108	23.82	11.74	7	50
Election cycle 2 (1998)	111	17.42	7.76	7	40
Election cycle 3 (2002)	111	17.57	7.72	7	41
Election cycle 4 (2006)	113	14.36	6.27	7	32
Election cycle 5 (2010)	113	14.69	7.24	7	55
Election cycle 6 (2014)	113	14.73	6.51	7	38
Election cycle 7 (2018)	113	15.16	7.01	7	39
N. of Female councillors					
Election cycle 1 (1995)	108	0.56	1.19	0	7
Election cycle 2 (1998)	111	0.40	.94	0	5
Election cycle 3 (2002)	111	0.55	1.01	0	5
Election cycle 4 (2006)	113	2.08	1.30	0	7
Election cycle 5 (2010)	113	3.25	2.01	1	12
Election cycle 6 (2014)	113	3.72	2.09	1	10
Election cycle 7 (2018)	113	4.68	2.65	1	13

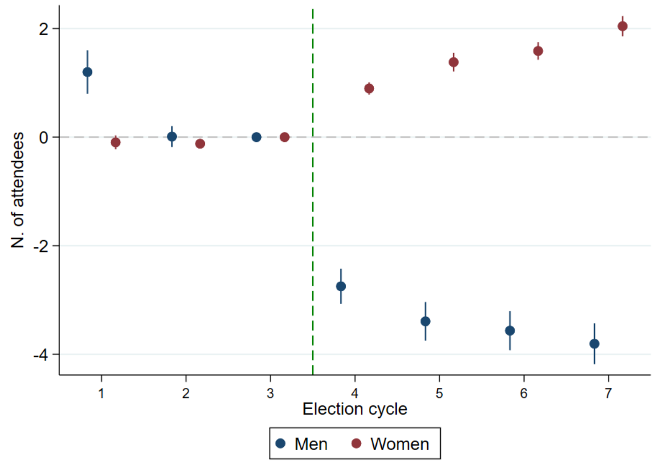
Notes: This table displays descriptive statistics on the number of councillors and female councillors for each election cycle for the 113 South Korean local municipal councils we collected data on.

Table 3.3: Descriptive Statistics: Meetings' characteristics

	N.	Mean	Std Dev
Analysis Sample			
N. election cycles	113	6.28	1.36
N. years	113	22.3	5.78
N. meetings	113	1,412.21	878.61
N. meetings per year	2,525	63.20	37.75
N. of segments per meeting	159,580	2.15	2.00
N. attendees per segment	342,450	7.67	2.54
N. speakers per segment	342,450	3.79	2.50
N. interactions per segment	342,450	72.21	132.84
Unique number of councillors	7,579		
Overall Sample			
N. election cycles	113	6.62	0.97
N. years	113	24.42	3.92
N. meetings	113	2,705.04	1,028.97
N. meetings per year	2,759	110.41	43.62
Committee meeting	2,759	0.65	0.18

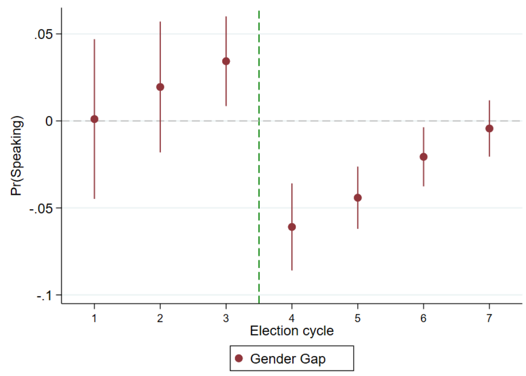
Notes: This table displays descriptive statistics on the analysis sample and the overall sample of meetings held in the 113 South Korean municipal councils we have data on.

Figure 3.2: N. attendees in meetings - across cycles



Notes: Evolution across election cycles of female and male attendees in meetings. Each dot represents the coefficients of the election cycle dummies in Specification 3.1. Each line represents 95% confidence intervals. Standard errors clustered at the municipality level. Standard errors clustered at the municipality level.

Figure 3.3: Gender gap in propensity to speak across cycles



Notes: Gender gap in probability to speak by election cycle. Each dot represents the coefficients of the interaction of election cycle dummies and female dummy as in Specification 3.2. Each line represents 95% confidence intervals. Standard errors clustered at the municipality level.

Table 3.4: Evolution of gender gap across election cycles

	Extensive margin	Intensive margin		
	Pr (Speaking)	N. interventions	Ln (Total length)	Ln (Average length)
	(1)	(2)	(3)	(4)
Election cycle 1 × Female	-0.033	-2.958	-0.079	0.052
	(0.025)	(1.922)	(0.115)	(0.051)
Election cycle 2 × Female	-0.015	-2.384	-0.073	-0.014
	(0.024)	(1.672)	(0.072)	(0.040)
Female	0.034***	2.543	0.088	0.013
	(0.013)	(1.634)	(0.067)	(0.039)
Election cycle 4 × Female	-0.095***	-5.680***	-0.330***	-0.051
	(0.017)	(1.969)	(0.074)	(0.045)
Election cycle 5 × Female	-0.078***	-5.638***	-0.264***	-0.024
	(0.017)	(1.605)	(0.073)	(0.045)
Election cycle 6 × Female	-0.055***	-6.146***	-0.255***	-0.020
	(0.015)	(1.780)	(0.077)	(0.042)
Election cycle 7 × Female	-0.039**	-4.096**	-0.177**	-0.004
	(0.016)	(1.719)	(0.073)	(0.042)
Observations	2581410	1285419	1285419	1285419
Male in election cycle 3 average	0.502***	20.638***	6.971***	4.733***
	(0.001)	(0.094)	(0.003)	(0.002)

Notes: The regression specification is given by Equation (3.2). Standard errors in parenthesis, clustered at the municipality level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3.4: Gender gap in propensity to speak across and within cycle

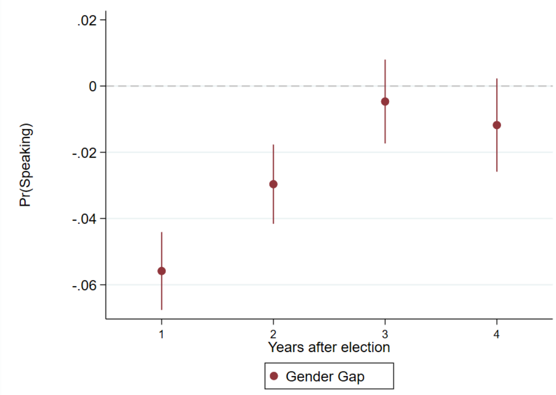
Notes: This figure displays the evolution of the probability to speak at meetings for men and women over time. The blue line represents men, while the red line represents women. Each dot represents the residuals, averaged by gender and year, of a regression of the outcome for councillor i in year t in segment s of meeting m on meeting fixed effects. The regression specification is given by Equation (3.3).

Table 3.5: **Evolution of gender gap within election cycle**

	Extensive margin		Intensive margin	
	Pr (Speaking)	N. interventions	Ln (Total length)	Ln (Average length)
	(1)	(2)	(3)	(4)
Female	-0.056***	-4.289***	-0.226***	-0.014
	(0.006)	(0.547)	(0.018)	(0.014)
Second year × Female	0.026***	1.776***	0.070***	0.000
	(0.003)	(0.323)	(0.013)	(0.007)
Third year × Female	0.051***	2.226***	0.115***	0.012
	(0.007)	(0.471)	(0.020)	(0.010)
Fourth year × Female	0.044***	2.515***	0.107***	0.015
	(0.008)	(0.585)	(0.025)	(0.013)
N	1565789	825326	825326	825326

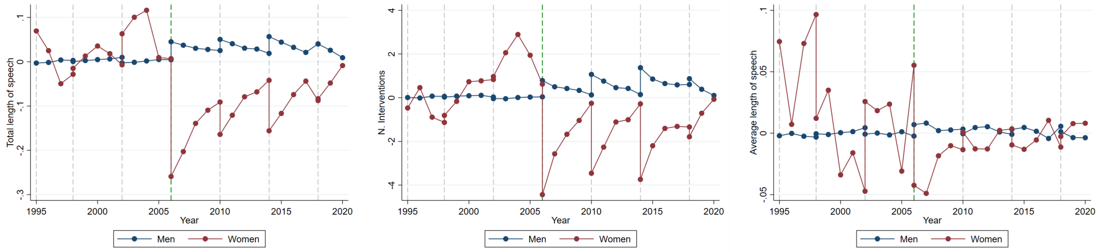
Notes: The regression specification is given by equation (3.4). The sample is restricted to the post-quota period. Standard errors in parenthesis, clustered at the municipality level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3.5: Gender gap in propensity to speak within election cycle



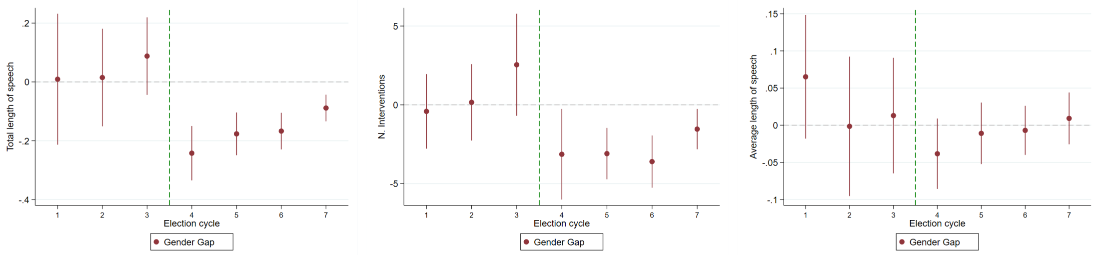
Notes: This figure displays the evolution of the gender gap in propensity to speak over the years of the appointment. The sample is restricted to the post-quota period. Each dot represents the coefficients of the interaction of election cycle dummies and female dummy as in Specification 3.4. Each line represents 95% confidence intervals. Standard errors clustered at the municipality level.

Figure 3.6: Gender gap in nature of interactions, conditional on speaking



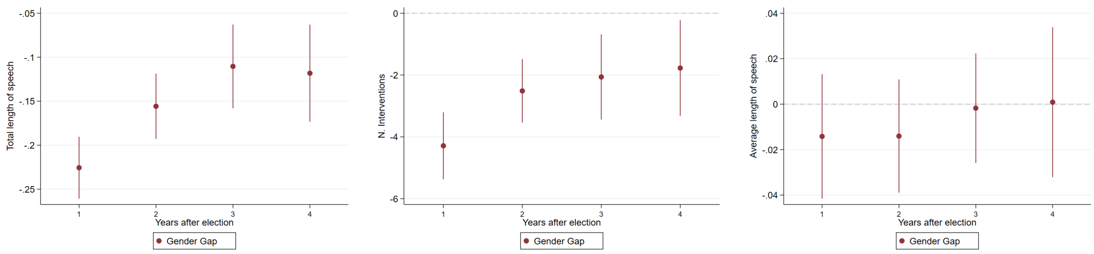
Notes: This figure displays the evolution of logarithm of total length of speech, number of interventions and logarithm of average length of speech respectively, conditional on speaking, for men and women over time. The blue line represents men, while the red line represents women. Each dot represents the residuals, averaged by gender and year, of a regression of each outcome for councillor i in year t in segment s of meeting m on meeting fixed effects. The regression specification is given by Equation (3.3)

Figure 3.7: Gender gap in nature of interactions, conditional on speaking across election cycles



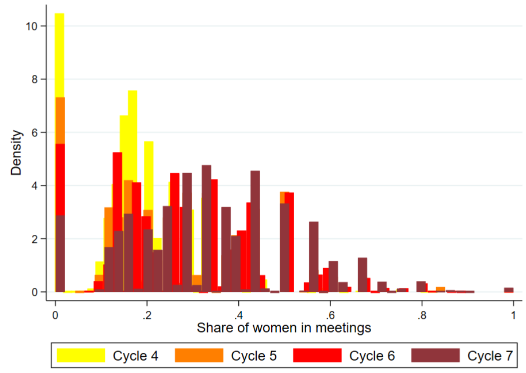
Notes: This figure displays the evolution of the gender gap in logarithm of total length of speech, number of interventions and logarithm of average length of speech respectively, conditional on speaking by election cycle. Each dot represents the coefficients of the interaction of election cycle dummies and female dummy as in Specification 3.2. Each line represents 95% confidence intervals. Standard errors clustered at the municipality level.

Figure 3.8: Gender gap in nature of interactions, conditional on speaking within election cycles



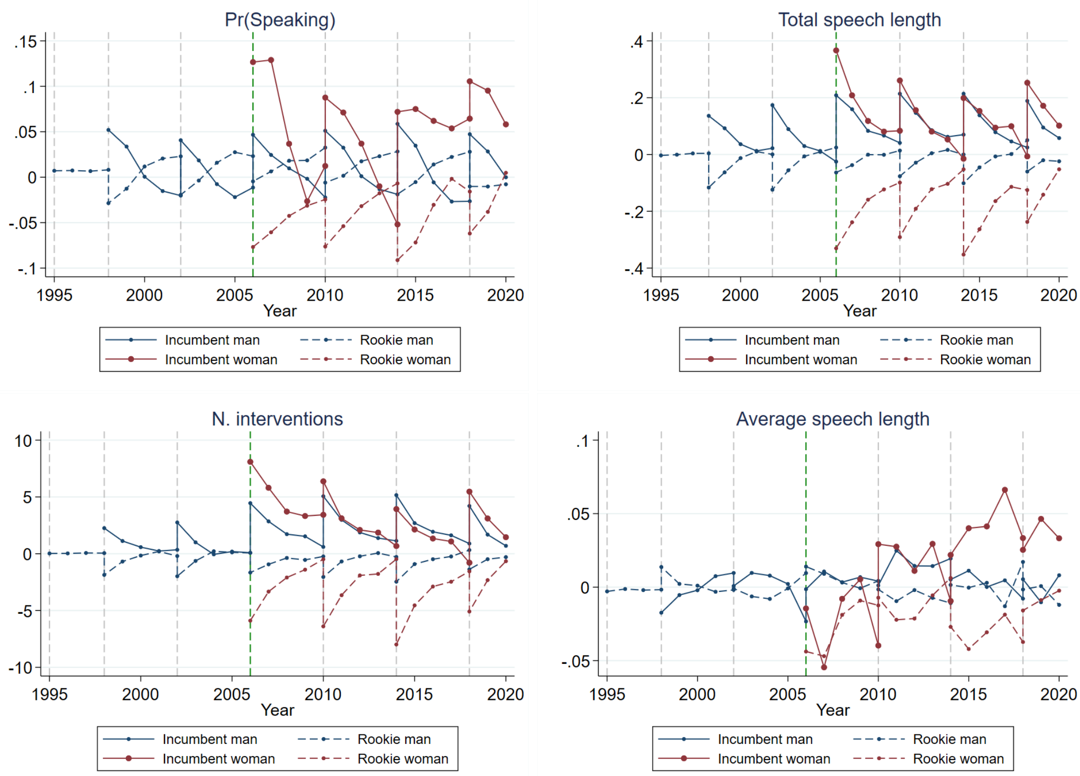
Notes: This figure displays the evolution of the gender gap in logarithm of total length of speech, number of interventions and logarithm of average length of speech respectively, conditional on speaking over the years of the appointment. The sample is restricted to the post-quota period. Each dot represents the coefficients of the interaction of election cycle dummies and female dummy as in Specification 3.4. Each line represents 95% confidence intervals. Standard errors clustered at the municipality level.

Figure 3.9: Mechanisms: Gender composition of meeting attendees



Notes: Histogram displaying the distribution in the share of women attending meetings across election cycles. Each color corresponds to a different election cycle. The unit of analysis is the meeting segment.

Figure 3.10: Experience: Incumbents vs Rookies



Notes: This figure displays the evolution of logarithm of total length of speech, number of interventions and logarithm of average length of speech respectively, conditional on speaking, for men and women over time. A distinction is made between rookies and incumbents. Rookies in election cycle t are all those councillors that have been elected in election cycle t for the first time in municipal councils. Hence, any other councillor, who has been elected at least once before election cycle t in municipal councils is considered an incumbent. The blue line represents men, while the red line represents women. The straight line represents incumbents, while the dotted line rookies. Each dot represents the residuals, averaged by gender and year, of a regression of each outcome for councillor i in year t in segment s of meeting m on meeting fixed effects. The regression specification is given by Equation (3.3)

Table 3.6: Experience: Incumbents vs Rookies, post the quota period

	Extensive margin	Intensive margin		
	Pr (Speaking)	N. interventions	Ln (Total length)	Ln (Average length)
	(1)	(2)	(3)	(4)
Panel A: Interactions				
Female	0.052***	0.221	0.024	0.029
	(0.012)	(1.014)	(0.034)	(0.018)
Second year × Female	0.012**	-0.095	0.011	0.013
	(0.006)	(0.546)	(0.017)	(0.010)
Third year × Female	-0.001	-0.652	-0.009	0.008
	(0.020)	(0.947)	(0.037)	(0.021)
Fourth year × Female	-0.003	-0.099	-0.006	0.022
	(0.024)	(1.115)	(0.043)	(0.028)
Rookie	-0.064***	-7.342***	-0.332***	-0.003
	(0.011)	(0.623)	(0.022)	(0.013)
Second year × Rookie	0.031***	3.799***	0.144***	-0.015*
	(0.005)	(0.385)	(0.012)	(0.008)
Third year × Rookie	0.107***	4.763***	0.251***	-0.009
	(0.012)	(0.647)	(0.023)	(0.014)
Fourth year × Rookie	0.113***	5.289***	0.287***	-0.006
	(0.014)	(0.792)	(0.026)	(0.015)
Female × Rookie	-0.123***	-4.639***	-0.272***	-0.059***
	(0.015)	(1.036)	(0.038)	(0.020)
Second year × Female × Rookie	0.010	1.824***	0.055**	-0.012
	(0.007)	(0.640)	(0.024)	(0.014)
Third year × Female × Rookie	0.038*	2.993**	0.124***	0.010
	(0.022)	(1.174)	(0.044)	(0.029)
Fourth year × Female × Rookie	0.031	2.711**	0.106**	-0.005
	(0.027)	(1.359)	(0.050)	(0.035)
Panel B: Gender Gap				
Incumbents:				
First year × Female	0.052***	0.221	0.024	0.029
	(0.012)	(1.014)	(0.034)	(0.018)
Second year × Female	0.065***	0.126	0.035	0.042**
	(0.014)	(0.821)	(0.031)	(0.019)
Third year × Female	0.052***	-0.431	0.015	0.037
	(0.016)	(1.286)	(0.042)	(0.024)
Fourth year × Female	0.049**	0.122	0.018	0.051
	(0.019)	(1.496)	(0.039)	(0.032)
Rookies:				
First year × Female	-0.071***	-4.418***	-0.248***	-0.030*
	(0.007)	(0.530)	(0.019)	(0.016)
Second year × Female	-0.048***	-2.689***	-0.182***	-0.029**
	(0.007)	(0.537)	(0.023)	(0.014)
Third year × Female	-0.033***	-2.077***	-0.133***	-0.011
	(0.008)	(0.601)	(0.025)	(0.015)
Fourth year × Female	-0.043***	-1.806**	-0.148***	-0.013
	(0.009)	(0.774)	(0.032)	(0.019)
Observations	1544408	814772	814772	814772

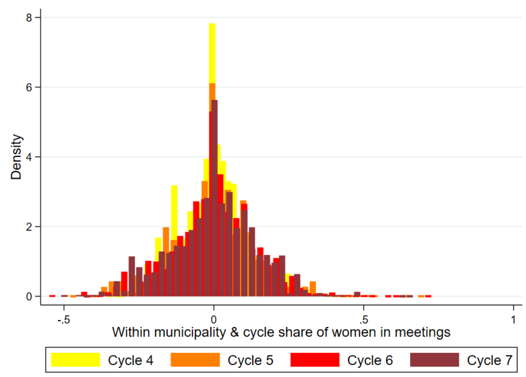
Notes: The table presents the results of Specification (3.5). The sample is restricted to councillors that we were able to trace over time (96.22% of the sample). Standard errors are clustered at the municipality level.

Table 3.7: Experience: First year - Incumbents vs Rookies, post the quota period

	Extensive margin	Intensive margin		
	Pr (Speaking)	N. interventions	Ln (Total length)	Ln (Average length)
	(1)	(2)	(3)	(4)
Rookie × Female	-0.098***	-6.566***	-0.411***	-0.057**
	(0.014)	(1.647)	(0.059)	(0.026)
Rookie × Female × Election cycle 5	-0.006	-1.159	0.029	0.043
	(0.016)	(1.798)	(0.071)	(0.037)
Rookie × Female × Election cycle 6	-0.023	-3.590*	-0.090	0.009
	(0.017)	(1.960)	(0.070)	(0.033)
Rookie × Female × Election cycle 7	0.022	0.530	0.101	0.037
	(0.018)	(1.908)	(0.068)	(0.034)
Incumbent pre-quota × Female	0.142***	7.722***	0.314***	-0.055
	(0.051)	(2.042)	(0.083)	(0.078)
Incumbent pre-quota × Female × Election cycle 5	-0.178***	-6.372	-0.265*	0.049
	(0.056)	(4.415)	(0.137)	(0.129)
Incumbent pre-quota × Female × Election cycle 6	-0.195**	-6.229**	-0.285**	0.049
	(0.077)	(2.892)	(0.134)	(0.131)
Incumbent pre-quota × Female × Election cycle 7	-0.056	-3.323	-0.168	0.011
	(0.096)	(6.142)	(0.116)	(0.074)
Incumbent post-quota × Female	0.113***	5.055***	0.249***	0.049*
	(0.018)	(1.257)	(0.041)	(0.026)
Incumbent post-quota × Female × Election cycle 5	-0.033	0.707	-0.010	-0.016
	(0.036)	(2.700)	(0.078)	(0.041)
Incumbent post-quota × Female × Election cycle 6	-0.043	-2.543	-0.092	-0.026
	(0.027)	(1.818)	(0.071)	(0.034)
Observations	477077	254237	254237	254237

Notes: The table presents the results of Specification (3.6). The sample is restricted to councillors that we were able to trace over time (96.22% of the sample) in their first year after the appointment. Standard errors are clustered at the municipality level.

Figure 3.11: Environment composition: Within municipality and election cycle variation in gender composition of meeting attendees



Notes: The histogram displays the distribution of the share of women attending meetings within each municipality and cycle across election cycles. The within-municipality share is obtained by taking the residuals of a regression of the share of women on municipality × election cycle fixed effects. Each color corresponds to a different election cycle. The unit of analysis is the meeting segment.

Table 3.8: Environment composition: Effect of meeting attendees' composition

	Extensive margin		Intensive margin					
	Pr (Speaking)		N. interventions		Ln (Total length)		Ln (Average length)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Overall								
Female × Share of women	0.061*	-0.024	2.832	0.714	0.189**	-0.023	0.037	0.011
	(0.033)	(0.048)	(2.385)	(1.591)	(0.091)	(0.087)	(0.058)	(0.045)
Panel B: Heterogeneity by election year								
Female × First year × Share of women	0.103**	-0.025	4.766	2.654	0.321**	0.079	0.089	0.045
	(0.040)	(0.052)	(3.224)	(2.314)	(0.126)	(0.124)	(0.070)	(0.060)
Female × Second year × Share of women	0.056	-0.024	1.496	-2.066	0.153	-0.102	0.063	0.003
	(0.048)	(0.022)	(3.041)	(1.783)	(0.120)	(0.072)	(0.072)	(0.036)
Female × Third year × Share of women	0.066	0.022	3.477	-2.857	0.150	-0.174	-0.018	-0.067
	(0.049)	(0.054)	(3.442)	(2.525)	(0.109)	(0.126)	(0.075)	(0.073)
Female × Fourth year × Share of women	0.013	-0.025	2.860	-4.999	0.158	-0.233	-0.014	-0.107
	(0.054)	(0.062)	(3.454)	(3.299)	(0.140)	(0.161)	(0.090)	(0.091)
Panel C: Heterogeneity by election cycle								
Female × Share of women	0.059	0.038	0.215	0.014	0.262	-0.132	0.047	-0.146
	(0.099)	(0.116)	(0.305)	(0.218)	(0.355)	(0.255)	(0.158)	(0.139)
Female × Election cycle 5 × Share of women	0.055	-0.095	0.231	-0.056	0.321	-0.014	0.090	0.042
	(0.062)	(0.063)	(0.144)	(0.140)	(0.208)	(0.151)	(0.134)	(0.085)
Female × Election cycle 6 × Share of women	0.005	-0.033	0.091	-0.018	0.104	-0.000	0.013	0.018
	(0.068)	(0.091)	(0.134)	(0.155)	(0.177)	(0.205)	(0.110)	(0.100)
Female × Election cycle 7 × Share of women	-0.008	0.024	-0.091	-0.058	-0.150	-0.012	-0.059	0.046
	(0.076)	(0.094)	(0.156)	(0.151)	(0.162)	(0.174)	(0.097)	(0.100)
Observations	1472683	1472683	772315	772315	772315	772315	772315	772315
Segment FEs	Y	Y	Y	Y	Y	Y	Y	Y
Individual x Election cycle FEs	X	Y	X	Y	X	Y	X	Y
Female	Y	X	Y	X	Y	X	Y	X
Female × Year within cycle	Y	Y	Y	Y	Y	Y	Y	Y
Female × Election cycle	Y	Y	Y	Y	Y	Y	Y	Y

Notes: The table displays the results of Specification (3.8). In Panel B and C the effects are decomposed to display heterogeneity by election year and election cycle respectively. The sample is restricted to meetings where the gender of all the attending councillors is known. This guarantees that the share of women is not measured with error. Standard errors are clustered at the municipality level.