Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Stereotypical Selection

Martina Zanella

TEP Working Paper No. 0224

January 2024

Revision date: August 2024

Trinity Economics Papers
Department of Economics

Stereotypical Selection

Martina Zanella*

Abstract

Women continue to be underrepresented and face significant challenges in establishing careers within traditionally male-dominated fields. Does minority status in and of itself create a barrier to women's success? Conventional wisdom posits that it does, as under-representation is often associated with negative stereotypes about the group (e.g. women in math), which are shown to be detrimental to performance. However, the decision to enter stereotypically male fields is endogenous and may, in part, be motivated by a desire to challenge these stereotypes. This paper aims to assess how minority status affects performance when selection is endogenous. I study the performance of 14,000 students at an elite university across 16 departments, in a real-world setting that combines a choice with well-defined stereotypes - university major - with exogenous variation in peer identity - quasi-random allocation of students across class groups within the same course. I find that being in the minority negatively affects only the students who choose majors in line with societal expectations (e.g. men in math). These are the students who belong to the majority group, and those who are found to hold beliefs that more strongly align with these expectations. In line with social identity considerations being incorporated into educational choices, the evidence suggests that ex-ante different "sensitivity" to stereotypes and the composition of the environment induces students to select different majors and subsequently react to the composition of the environment in a self-fulfilling way.

JEL codes: D91, I24, J15, J16, J24

*Department of Economics, Trinity College Dublin; email: zanellam@tcd.ie. I acknowledge the collaboration with the London School of Economics, that I thank for the data and the possibility to administer the survey. I am grateful for the advice and support from Nava Ashraf, Oriana Bandiera, and Robin Burgess throughout the project. I also thank advice from Manuel Bagues, Tim Besley, Gharad Bryan, Katherine Casey, Jeremiah Dittmar, Martina Kirchberger, Camille Landais, Alan Manning, Guy Michaels, Gaia Narciso, Steve Pischke, Anna Raute, Daniel Reck, Mark Schankerman, Daniel Sturm, and numerous seminar and conference participants, for their extremely helpful comments. A special acknowledgment goes to Judith Shapiro for priceless guidance in navigating the LSE system, feedback, and support. The use of the data and the survey were approved by the LSE Ethics Committee in October 2017 and April 2021.

I. INTRODUCTION

Despite significant progress in reducing inequality in the labour market over recent years, women remain under-represented in traditionally male occupations, giving rise to an economically consequential gap in labor market outcomes (Blau and Kahn, 2017; Goldin, 2014b). This under-representation can be traced back to initial differences in educational choices[1] that are further amplified in the labor market as women struggle to establish a career in these fields[2]. Does minority status in and of itself create a barrier to women's success in traditionally male fields? Can this be one of the reasons why we see fewer and fewer women moving up the ladder?

Conventional wisdom suggests this is the case. Under-representation usually goes hand in hand with the existence of negative stereotypes about the group (e.g. women in math), which experiments in Psychology and Economics show are detrimental to performance (Bordalo et al., 2019; Spencer et al., 2016; Hoff and Pandey, 2006; Steele and Aronson, 1995), particularly when in a minority (Karpowitz et al., 2024; Born et al., 2020; Chen and Houser, 2019).

However, the answer is not so straightforward. While experiments shut down the selection channel altogether, the choice to enter a traditionally male field, despite the negative stereotype and the under-representation of women, is *endogenous*, and may even be motivated by challenging societal expectations. As such, women in traditionally male occupations may not be as influenced by stereotypes and the gender of their peers as the average woman in the population.

This is what this paper sheds light on – how minority status, and more generally, peers' identity, affects performance when *selection* is *endogenous*. This margin has mostly remained unexplored due to the challenge of finding a setting where choices and minority status could be separated. In this paper, I overcome this challenge by studying a setting with high-stakes real-world decisions that combines a choice characterized by well-defined stereotypes with exogenous variation in peer identity. I study the performance of >14,000 students enrolled in undergraduate programs at a selective UK university. I look at whether students' choice of major – in line or against gender stereotypes across 16 different subjects – interacts with the effect of peers' gender on performance once students are in the classroom.

I show that considering selection is crucial. I find that being surrounded by different-gender peers negatively affects only those students who made a choice in line with societal expectations. These are the students who belong to the majority group (e.g. men in math). Hence, this indicates that while peers' identity *can* explain gender gaps in performance, the effect is due to the *majority group* benefiting from being *over-represented*, rather than the *minority group* suffering from being *under-represented*.

The paper proceeds in two steps. I start by laying down a theoretical framework that explains how minority status is expected to affect students' performance in the absence of selection. I formalize the two channels indicated by the Economics and Social Psychology literature as the main reasons behind the influence of peers' identity on performance: (i) homophily - a preference to engage with same-gender

[1] OECD (2020); Bertrand (2020); McNally (2020)
[2] Lundberg and Stearns (2019); Bertrand (2018); Beede et al. (2011)

peers (Inzlicht and Good, 2006) that limits opportunities to benefit from peer-to-peer interactions and support when in the minority (Tinto, 1975); and (ii) stereotypes, which distort beliefs about ability when group belonging is made salient by minority status (Karpowitz et al., 2024; Bordalo et al., 2019; Spencer et al., 2016).

I then complement this framework with a simple Roy model of major choice to show how selection into majors can alter these predictions. The proposed mechanism suggests that students incorporate stereotypes and gender composition when choosing their majors. As a consequence, the students who are more likely to select counter-stereotypical majors may be those with ex-ante lower "sensitivity" to stereotypes and peers' gender, who are then more resilient to minority status once in the classroom.[3]

After having explained how the mechanisms work in theory, I test them empirically by exploiting a novel dataset on the universe of students enrolled in undergraduate programs at the London School of Economics from 2008 to 2017, that I assembled by combining rich administrative data on students[4] with information regarding their attitudes and social networks collected through a survey. I focus on students during their first year at the university. In the first year, interactions within the classroom are important as students have yet to form their social networks outside the classroom.[5] Furthermore, students do not have experience regarding the academic system or their performance in this new system, and this is the first time that they interact with each other. Hence, stereotypes and social interactions can play a strong role in informing priors and expectations regarding students' performance and the performance of their peers.

In terms of empirical strategy, two ingredients are necessary – well-defined stereotypical/counter-stereotypical choices and exogenous variation in the gender composition of classmates. I define stereotypical and counter-stereotypical choices by leveraging the variety of majors offered by the university and the heterogeneity in their gender composition. While 49% of the students enrolled in undergraduate programs at the university are women, women represent the minority of students enrolled in math-intensive disciplines and the majority of students enrolled in humanistic disciplines. This closely reflects the stereotype that "women are worse than men in math and science while better at reading and humanities" (Carlana, 2019; Reuben et al., 2014). Hence, I indicate as "stereotypical" the choice of students who select in departments characterized by a high share of students of their gender, while a "counter-stereotypical" choice is that of a student who selects in a department where their gender is under-represented.[6]

[3]This is in line with the recent growing literature on selection into occupations that suggest that negative stereotypes regarding group-specific skills, identity considerations, and social norms affect beliefs regarding expected returns and the prestige of occupations, driving individuals out of occupations that do not align with societal expectations (Del Carpio and Guadalupe, 2022; Kugler et al., 2021; Cortes and Pan, 2018; Pan, 2015; Oxoby, 2014; Goldin, 2014a; Bertrand, 2011; Akerlof and Kranton, 2000)

[4]Application information, academic history, and performance in each course, allocation into courses and tutorials, background information, instructors' characteristics

[5]These have been shown to have an effect on performance on their own (for example, roommates - Corno et al., 2022; Jain and Kapoor, 2015; Stinebrickner and Stinebrickner, 2006; Sacerdote, 2001; or study groups - Oosterbeek and van Ewijk, 2014; Russell, 2017)

[6]As a proxy of stereotypical selection, I use the average share of men and women enrolled in undergraduate programs in

The exogenous variation in peers' identity comes from three key features of the setting. Firstly, students are quasi-randomly allocated to *classes* - small groups of maximum 15-20 people - to attend tutorials. Secondly, first-year courses are big modules characterized by multiple classes, as students attend courses common to different departments during their first year. This allows to control for *course fixed effects*, accounting for potential differences in courses common across students. Lastly, students attend on average 4 courses during their first academic year. Hence, each student can be observed across multiple classes, allowing to include *student fixed effects*. This takes care of student-specific characteristics that affect students' performance and choices. Hence, the causal effect of class gender composition is identified by comparing the performance of the *same student* across all the courses that they attend during their first year, where they are allocated to classes with an exogenous composition of classmates, net of *course* and *student* fixed effects.

This identification strategy relies on four assumptions, which I validate empirically. (i) Class allocation is as good as randomly assigned, conditional on scheduling constraints. (ii) Scheduling constraints are orthogonal to students' gender. (iii) The groups of classmates experienced in different courses are only partially overlapping, allowing to address the reflection problem that usually characterizes the estimation of peer effects (Bramoullé et al., 2020; De Giorgi et al., 2010). (iv) The variation in class composition is characterized by enough common support across departments to compare the effect for students of the same gender who made different choices of major.[7]

The empirical evidence confirms the model predictions. First of all, I find that students who make choices of majors confirming societal expectations are affected significantly differently by the identity of their peers when they are in the classroom compared to students who made a choice of major that challenges societal expectations. For students who made a choice of major in line with gender stereotypes, e.g. men in math, a 10% increase in the share of same-gender classmates *increases* course grades by 0.3314 points (2.03% of a standard deviation) on average. On the other hand, for students who chose a major stereotypically associated with the opposite gender, e.g. women in math, a 10% increase in the share of same-gender classmates *decreases* final course grades by 0.2839 points (1.74% of a standard deviation) on average. Lastly, class composition does not affect students who choose a major with a balanced gender ratio.

The effect is present for both men and women, albeit it is stronger for men.[8] However, the estimated patterns are not gender-specific. I replicate the analysis defining class composition and stereotypical

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, however, robust to other proxies of stereotypical selection.

[7]Numerous first-year courses are common across majors and class allocation does not depend on the major of enrolment. As such, two students who chose the same major and are in the same class in a given course are not necessarily allocated to the same class in other courses (the share of same-program peers within the classroom is on average 0.48).

[8]This is not surprising in this setting as being in the minority in a male-dominated department is the result of stronger selection as male-dominated departments have higher rejection rates. Secondly, the stark segregation across departments is driven by men choosing to apply significantly more to male-dominated departments.

selection along ethnic lines, exploiting the diversity of the environment (48% of students are Asian, while 36% are White) and the variety of majors offered by the institution. The findings for gender also apply to ethnicity, suggesting that the estimated patterns are related to selection rather than group-specific dynamics.

Moreover, the results are not driven by other competing explanations. I provide an array of robustness checks – testing for linearity, controlling for teaching assistants' or other peers' characteristics (including ability), addressing the eventuality of spillovers and mechanical effects, placebo checks, and different definitions of stereotypical choices – to address potential concerns.

Secondly, the evidence is consistent with a story of ex-ante different "sensitivity" to social identity considerations inducing students to select different majors and then react to the composition of the environment in a self-fulfilling way. Three key pieces of evidence support this statement. The same stronger response for students who made stereotypical choices of major is found when replicating the analysis considering another characteristic of the environment that might reinforce stereotypical associations – the gender of the class teacher. Moreover, I find that the students who made stereotypical choices that suffer the most from being surrounded by opposite-gender classmates are those who come from countries with more conservative gender norms (proxied by the Global Gender Gap Index of the student's country of origin), as we would expect if ex-ante sensitivity to stereotypes was the mechanism driving the results. Lastly, the results of a Gender-Scientific Implicit Association Test (Greenwald et al., 1998), confirm that the students who uphold the strongest stereotypical associations are indeed those who made a stereotypical choice of major, i.e. men enrolled in math-intensive fields and women enrolled in programs related to humanities. The answers to questions regarding students' propensity to engage with same-gender peers display the same pattern.[9]

Taken together, these findings have important policy implications. They indicate that if belonging to the minority group is the result of choices that challenge societal expectations and strong selection, the members of the minority group are a selected group of people who do not suffer from being under-represented. On the contrary, those who believe more strongly in the stereotype and are more sensitive to the composition of the environment are those who belong to the majority group. This suggests that, in these settings, popular measures targeting members of the minority to level the playing field – such as reallocation policies, networking groups, or mentoring programs – might focus the attention on the "wrong" group, risking perpetuating negative stereotypes, and taking resources away from tackling the structural issue at place.

The paper contributes to the recent literature that theorizes that individuals incorporate stereotypes regarding group-specific skills, identity considerations, and social norms when making occupational choices (e.g. Del Carpio and Guadalupe, 2022; Kugler et al., 2021; Cortes and Pan, 2018; Pan, 2015; Oxoby, 2014; Goldin, 2014a). Providing empirical proof that social identity considerations are inter-

[9]A survey was administered to all the students enrolled in undergraduate programs at the university in 2021. 500 students (10% of the student body) answered the survey. The response rate is similar to that of other surveys performed on undergraduate students at the university.

nalized in choices is challenging due to the difficulty of isolating preferences from other correlated characteristics of occupations. This paper contributes to this literature not only by adding a piece of evidence in support of this hypothesis but also by displaying its cruciality when designing policies to tackle inequality in performance.

Moreover, my work contributes to the literature that aims at understanding the importance of stereotypes and social identity considerations in explaining the under-confidence and under-performance of minorities (e.g. Coffman et al., 2021; Bonomi et al., 2021; Karpowitz et al., 2024; Born et al., 2020; Bordalo et al., 2019; Chen and Houser, 2019; Bordalo et al., 2016; Coffman, 2014) by showing that acknowledging selection into minority status is crucial. Indeed, it demonstrates that selection moderates the extent to which biased environments, the identity of peers, or in general social identity considerations, can impact performance.

Lastly, this paper also relates to the significant body of literature that studies the effect of the identity of peers on performance and choices in real-world environments in higher education[10] (Shan, 2020; Zölitz and Feld, 2021; Griffith and Main, 2019; Booth et al., 2018; Huntington-Klein and Rose, 2018; Hill, 2017; Oosterbeek and van Ewijk, 2014; Giorgi et al., 2012; Griffith, 2010). Both the methodology and the analysis of my paper build on this work. However, the majority of these studies focus on a single field or estimate average effects, overlooking the interplay between selection and the effect of peers' identity and providing only a partial picture. 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, this paper provides evidence of an additional unexplored margin that might be able to reconcile apparently diverging results – endogenous selection into minority status.[11]

The remainder of the paper is organized as follows. Section II. describes the setting and the data. Section III. presents the theoretical framework. Section IV. introduces the empirical part of the paper by describing the empirical strategy and the tests in support of the identifying assumptions. Sections V. and VI. illustrate the results. Finally, Section VII. concludes.

II. Institutional Setting and Data

The setting of this paper is the London School of Economics and Political Science (LSE), one of the most selective and competitive universities in the UK. In the UK, students can only apply to 5 programs across

[10]The focus of this paper's contribution is restricted to higher education as the mechanisms at play in this setting might be different from those affecting students at different points in their career. Indeed, students have different needs, interests, and objectives before they enter this high-stakes environment, where their performance will likely determine their lifetime occupation.

[11]To the best of my knowledge, the only other paper in Economics that assesses the heterogeneity of peer effects based on selection is Pregaldini et al. (2020), which compares students who have self-selected into a STEM or language specialization in a Swiss High School. In line with the results of this paper, they find heterogeneous effects depending on students' choices, which they attribute to differences in willingness to compete. A few other studies assess heterogeneous treatment effects based on the program of enrollment (Hill, 2017) or course type (Zölitz and Feld, 2021 and Oosterbeek and van Ewijk, 2014), finding, once again, that aggregate results hide significant heterogeneity. None of them, however, explicitly explores selection in line or against societal expectations (and into minority status) as a potential reason behind these patterns.

all UK universities. Hence, applying to undergraduate programs is a strategic decision that presents high opportunity costs. This is particularly true at the LSE as securing an offer of admission from this institution is very challenging. Students apply from every part of the world to attend undergraduate programs at the University, competing for a fixed number of available places. This process results in a 75% rejection rate every year.[12]

Four unique features of the Institution make the LSE a particularly apt setting to study this research question. First, the University is characterized by an exceptionally multicultural and diverse student body: 49% of undergraduate students are women, roughly 36% of students are White, 48% are Asian (21% of which Chinese, and 12% Indian), and 4% are Black.

Secondly, it offers a very diverse portfolio of courses – 44 undergraduate degree programs across 16 academic departments. These programs span different disciplines – from Mathematics, Finance, and Economics to Anthropology and Sociology – and are characterized by stark gender segregation. In departments such as Mathematics, Finance, and Economics men represent $>65\%$ of the students enrolled, while departments such as Anthropology and Sociology are characterized by $>75\%$ of female students (Figure I).

Thirdly, the academic system is homogeneous across programs belonging to different disciplines. Students attend on average 4 courses each academic year during two 10-week-long teaching terms – Michaelmas term (September- December) and Lent term (January- April). During the first year, which is the focus of this paper, most courses are compulsory.[13] Moreover, most of those courses are common across programs.[14]

Lastly, for each course, students attend lectures and *classes* (tutorials). While 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 into small groups of maximum 15-20 students.[15]

The empirical strategy of the paper exploits students' assignment to these class groups and their gender composition. The reason is twofold. The students officially allocated to the same class group represent the peers each student engages with during the whole year. Students attend 10 classes each term, one per week (generally one hour long). They are allocated to a class group before the beginning of the academic year, and can only change under exceptional circumstances via an official request that has to be approved. This official allocation reflects very closely actual attendance since classes are compulsory and teaching assistants are required to register students' attendance.[16]

[12]Each year the University admission office receives approximately 26,000 applications for roughly 1,700 places.

[13]Students are generally allowed to choose at most one course.

[14]See Table A.VI in Appendix A. for more detailed information.

[15]The university regulations fix a cap of 17 students per class group. Exceptions are however frequent. Indeed, the average class size in the sample is 13.47, but 2.19% of class groups have more than 17 students

[16]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 to class groups 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.

Moreover, class group 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, students cannot be in two classes at the same time. Hence, the gender composition of the pool of students officially allocated to the same class group can be considered as good as random conditional on course clashes. This assumption is tested empirically in Section IV.E..

II.A. Data

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 the university, alongside some individual background characteristics, such as gender, age at enrolment, country of origin, ethnicity, term time accommodation, and the previous schools attended. This array of information is available for the academic years 2008/09 - 2017/18. The university also collects and stores students' application information – the total number of students who applied, were accepted, or rejected to each program, divided by gender, ethnicity, and country of birth, alongside a complete list of all the qualifications that students submit when filing their applications[17], which I use to construct a measure of ex-ante ability.

Class Register The class register contains information on lecture and class 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 also 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. [18]

Human Resources Records Background information (gender and ethnicity) from Human Resources records were matched with the class register for a subsample of teachers who were employed at the university with an official contract during the academic years considered.

Survey A survey was administered to all the students enrolled in undergraduate programs at the LSE during the academic year 2020/2021. The questionnaire was administered via email. 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-

[17] This information is only available for the cohorts of students enrolled in undergraduate programs between 2011 and 2017.

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

specific skills, and an Implicit Association Test eliciting the association between male-Scientific, and female-Humanistic. More information on the survey can be found in Appendix G.

Additional Sources I integrated the data about students and courses with information on the institutional system and undergraduate programs gathered from the university course guides and programme regulations[19]. Moreover, the paper makes use of data on the World Economic Forum Overall Global Gender Gap Index between 2006 and 2018.[20]. Lastly, information regarding enrollment in undergraduate programs and STAFF composition across fields of study for the overall UK higher education system is collected from HESA, the Higher Education Statistics Agency website.[21]

II.B. Sample selection

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

The focus on first-year students is motivated by three factors. First, students' choices are strongly restricted during the first year, allowing to minimize the problem of selection into courses. Second, during the first year, students enrolled in different programs have several courses in common. This provides a variation in peers' identity in first-year courses that goes beyond the variation within each program. Moreover, it makes sure that first-year courses are large and characterized by numerous classes, allowing to observe significant variation in the composition of classes within each course. Third, this is the best moment to study how behavior and performance are affected by the environment. Students have empty information sets regarding each other as it is the first time they meet. Thus, their expectations regarding their performance and the performance of their peers mostly rely on priors rather than the information they collected by interacting with each other.

The sample is restricted to modules that students attend during the first term, class groups with a reasonable size[22], and students who never changed class groups during the term.[23] After this first selection, the sample is further restricted to courses with at least two class groups and students for which I can observe a test score in at least two courses. The final sample is composed of 54,554 course-year-class group-level observations, corresponding to information on 14,297 students (99.36% of the original sample). As it can be seen in Table I, these students attend on average 3.82 courses in their first year, for a total of 509 courses with an average course size of 138.68 students, split into 9.40 class groups of 13.48 students each.[24]

[19]Link: https://www.lse.ac.uk/resources/calendar2023-2024/PreviousAcademicSessions.htm

[20]Source: The World Bank data. More information can be found in the Appendix C.3.

[21]Link: https://www.hesa.ac.uk/

[22]The data on class allocation are not created for research purposes and have never been used to this aim. Hence, using the dataset required a substantial cleaning effort.

[23]Evidence of the independence of class group changes on class group composition is provided in Appendix C.2..

[24]More details on the sample selection can be found in Appendix C.1..

II.C. Course performance

The main outcome of the paper is course performance, which is measured using the final overall grade that each student obtains after attending a course.

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, the combination of final exams and essays, or take-home assignments during the year. A student is deemed to have failed a course if their 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 out of 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 overall 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.[25]. Second, the LSE is characterized by a blind marking system – lecturers do not know the identity of the author of the piece of work they are marking.[26] This allows to disregard potential concerns such as discrimination.

Table II displays some descriptive statistics. First-year final grades are characterized by substantial variation. The average grade is 60.32, with a standard deviation of 16.34 points. Moreover, students do not perform in the same way in all courses. The within-student standard deviation is 8.61, and 40% of students experience a gap of at least 20 points between the highest and the lowest obtained grade – Figure II. This is not unusual in this context. Not all first-year courses count for the final degree classification to the same extent. Indeed, a student registered in a BA or BSc program needs to have passed first-year courses to the value of at least three credits to be eligible to progress to the second year.[27] Moreover, only the "year-one average" counts for the final degree classification.[28]

[25]For reference: BSc-Programmes-Marking-Frame.

[26]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 marks, the two markers discuss and agree on the final mark. The process is completely blind: students do not indicate their names in the exams, but only a candidate number which is secret to examiners and only known to the student. For reference: Marking-and-moderation, Challenging-results-and-appeals.

[27]The only exception is the Bachelor of Laws in which students progress to the second year if they have passed all four credits.

[28]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 will be counted once to make a total of eight first-year grades.

III. THEORETICAL FRAMEWORK

How does the gender composition of the classroom influence students' performance in a course? Additionally, why might students' initial choice of major play a role? To answer these questions, I develop a simple theoretical framework that explains how minority status affects students' performance, drawing from key insights in Economics and Social Psychology literature. I derive predictions regarding the effect of minority status in the absence of selection and then extend the framework to illustrate how accounting for selection can alter these predictions. The proposed mechanism suggests that students incorporate social identity considerations when choosing their majors. Consequently, students who exhibit lower ex-ante sensitivity to these considerations are more likely to select counter-stereotypical majors and display greater resilience to minority status once in the classroom.

III.A. The effect of peers' identity

Students are risk-neutral and choose educational investments (e_i) to maximize their utility $U(e_i, a_i^b, g, s_g, \beta_i)$

$$\max_{e_i} \beta_i F(a_i^b, e_i) - C(e_i, \delta_i s_g) \tag{1}$$

Students benefit from investing in effort as they care about their course performance $F(.)$.[29] 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. Lastly, following Spence (1974), we assume that the cost function is convex and increasing in effort: $C_e > 0$ and $C_{ee} > 0$.

Peers' identities affect students' performance via two main channels.

Stereotypes Minority status impacts students' performance by affecting expectations and confidence about personal achievement. Building on Bordalo et al. (2019), I assume that students hold imperfect information regarding their abilities. Beliefs about their ability rely on stereotypes and are affected by stereotypical distortions, which exaggerate true differences in ability between groups. In this sense, stereotypical distortions contain a kernel of truth (Bordalo et al., 2016).

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

A_g is the average ability of group g, and μ_i is the 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 $\theta_i \sigma(s_g)(A_g - A_{-g})$, where $(A_g - A_{-g})$ are actual differences in performance between the two groups. The size of the bias depends on how much stereotypes are top of mind ($\sigma(s_g)$) for the student. This is where minority status plays a role. Under-representation

[29] This is a simplification since recent papers show that students also care about their image (e.g. Bursztyn et al., 2019 and Ashraf et al., 2014).

affects the extent to which stereotypes are salient in individuals' minds by reminding them about their identity and their belonging to the group that makes them distinct (Karpowitz et al., 2024; Born et al., 2020; Chen and Houser, 2019; Hoff and Pandey, 2006; Inzlicht et al., 2006) Hence, we assume that stereotypical distortions decrease in the size of their group s_g, $\sigma_{s_g} \leq 0$.[30]

$$\frac{\partial a_i^b}{\partial s_g} = \theta_i \sigma_{s_g}(A_g - A_{-g}) \tag{3}$$

Whether minority status induces students to overestimate or underestimate their ability depends on the nature of the stereotype about their group (Bordalo et al., 2019; Coffman, 2014). If students choose a major in line with stereotypes, students belong to the "better" performing group ($A_g - A_{-g} > 0$). Hence, being under-represented, through stereotypical distortions, leads them to overestimate their ability – "stereotype lift". On the contrary, if students are enrolled in a counter-stereotypical major, they belong to the "worse" performing group ($A_g - A_{-g} < 0$). As a consequence, being under-represented depresses their beliefs regarding their ability – "stereotype threat" (Spencer et al., 2016).

Lab and field experiments in the academic setting document a widespread belief that women are worse than men in mathematics and science while better at reading (Carlana, 2019; Ellemers, 2018; Reuben et al., 2014; Lane, 2012). Hence, to make an illustrative example, being surrounded by women in a math-intensive major induces men to believe they are better compared to when surrounded by men. On the contrary, being surrounded by men in a math-intensive degree induces women to believe they are worse compared to when surrounded by women.

We are assuming that men perform better on average ($A_M - A_W > 0$) in math-intensive majors (stereotypical male), while women perform better on average ($A_W - A_M > 0$) in majors related to humanities (stereotypical female). This is consistent with what we observe and with the definition of stereotypical choices used in the empirical part of the paper.[31]

Preferences for same-gender peers The second channel via which class composition affects students' performance is through the cost of effort. Insights from Social Psychology suggest that 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). Hence, we assume that students prefer to interact with those who share their demographic traits. This is formalized by assuming that minority status affects performance by reducing opportunities for social interactions and mutual academic assistance, which increases the marginal cost of effort, $C_{s_g} \leq 0$ and $C_{es_g} \leq 0$.

Individuals are heterogeneous along three key dimensions – the strength of stereotypical associations (θ_i) and their cost of interacting with opposite-gender students (δ_i). The former determines the strength

[30] More details on the definition and derivations of beliefs on ability can be found in Appendix D.

[31] Figure B.VIII 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.

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. Lastly, students are heterogeneous in ability (μ_i).

III.B. Predictions in the absence of selection

Students' performance in the course is a direct function of their effort choice and their believed ability $F(a_i^b, e_i^*)$. Given the assumptions on the concavity of F and convexity of the cost function C in effort, there exists at least an interior solution such that

$$e_i^* : \beta_i F_e - C_e = 0 \tag{4}$$

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

$$\frac{\partial F(a_i^b, e_i^*)}{\partial s_g} = \theta_i \sigma_{s_g}(A_g - A_{-g})[F_a - \frac{\beta_i F_e F_{ea}}{\beta_i F_{ee} - C_{ee}}] + \delta_i \frac{F_e C_{es_g}}{\beta_i F_{ee} - C_{ee}}\Bigg|_{(a_i^b, e_i^*)} \tag{5}$$

This leads to the following predictions.[32]

Prediction 1 ***The effect of minority status in counter-stereotypical departments***
If $F_{ea} > 0$ then $\frac{\partial F(a_i^b, e_i^*)}{\partial s_g} > 0$ when $A_g - A_{-g} < 0$

Prediction 1 states that by reinforcing stereotype threat and reducing opportunities for emotional and academic support, minority status reduces performance for students enrolled in counter-stereotypical departments when effort and ability are complements in learning.[33]

Prediction 2 ***Comparative statics***
$|\frac{\partial F(a_i^b, e_i^*)}{\partial s_g}|$ is increasing in θ_i and δ_i

According to Prediction 2, the effect of being in the minority is stronger the stronger stereotypical associations (θ_i) and the higher the cost of interacting with the opposite gender (δ_i) for the student.

III.C. The role of selection into majors

A recent growing literature on selection into occupations suggests that negative stereotypes regarding group-specific skills, identity considerations, and social norms affect beliefs regarding expected returns and the prestige of occupations, driving individuals out of occupations that do not align with societal

[32] Detailed derivations can be found in Appendix D.

[33] This is true also when ability and effort are substitutes in learning but the positive effect associated with the higher perceived ability and the improved interactions are stronger than the substitution effect.

expectations (e.g. Del Carpio and Guadalupe, 2022; Kugler et al., 2021; Cortes and Pan, 2018; Pan, 2015; Oxoby, 2014; Goldin, 2014a).

If this is true, this has important implications for our investigation. Indeed, the students who decide to enroll in a major dominated by the opposite gender, despite the negative stereotypes regarding their group (women in math-intensive fields and men in humanities), might hold beliefs that are less in line with stereotypes and might face lower costs of interacting with students of the opposite gender, being, as a consequence, less affected by minority status.

I illustrate this by using a simple Roy model of major choice. Building on Del Carpio and Guadalupe (2022) theoretical framework, I assume that student i of gender g chooses between two options, a stereotypical and a counter-stereotypical major, which provide the following returns:

$$\begin{aligned} \text{Stereotypical major (S) - Returns: } & R^S = W^S A_i^S \\ \text{Counter-stereotypical major (C) - Returns: } & R^C = \frac{W^C A_i^C}{(B\Theta_i)(\Delta_i S_{-g})} \end{aligned} \tag{6}$$

The returns depend on the future streams of earnings of the jobs the major offers access to, which depend on the returns to skill in each sector (W^C and W^S). Stereotypes and norms prescribe what people are expected to do based on observable identity traits. Hence, choosing a counter-stereotypical major implies a cost of deviating from norms, $B\Theta_i$. This is characterized by two components: an identity cost B, which is higher the more counter-stereotypical is the major, and an individual specific component Θ_i, which indicates how sensitive individual i is to norms and stereotypes. A second component enters the returns for choosing the counter-stereotypical major, $\Delta_i S_{-g}$. As before, I assume that students prefer to engage with same-gender peers. As a consequence, they sustain a cost when choosing a counter-stereotypical major, as the share of same-gender peers is lower. This cost is higher the higher the share of opposite gender peers in the major S_{-g} and the higher the strength of i's preferences Δ_i.

For simplicity, let's abstract from considering preferences for same-gender students and let's only focus on the identity cost associated with choosing a major that challenges societal expectations.[34] If we log-linearize and assume that a_i^S, a_i^C, δ_i, and θ_i are distributed following a normal distribution $N(0, \sigma_x^2)$, with $x = a^S, a^C, \theta$ respectively, a student chooses the counter-stereotypical major if and only if the returns from it are at least as great as the returns from choosing the stereotypical major:[35]

$$Pr(\text{Choosing C}) = Pr(ln(R^C) > ln(R^S)) = Pr(a_i^C - a_i^S - \theta_i > w^S - w^C + \beta) = 1 - \Phi(z) \tag{7}$$

where $z = \frac{w^S - w^C + \beta}{\sigma_h}$.

Having characterized the probability of choosing a counter-stereotypical major, we can now derive predictions regarding the conditions under which this probability is higher. Keeping everything else constant, the probability of choosing the counter-stereotypical major decreases in the identity cost (β).

[34] The discussion can be easily generalized to preferences for same-gender students.

[35] Derivations can be found in Appendix D.

$$\frac{\partial Pr(\text{Choosing C})}{\partial \beta} < 0 \tag{8}$$

This is intuitive. As we are assuming that stereotypes prescribe what people are expected to do based on individuals' observable identity, the higher is the identity cost, the lower are the perceived returns from enrolling into a counter-stereotypical major.

Moreover, assuming that students internalize stereotypes when making their choice of major implies that students applying to stereotypical and counter-stereotypical majors are heterogeneous in terms of their sensitivity to stereotypes (θ_i). The average sensitivity to stereotypes (θ_i) of students who apply to counter-stereotypical majors can be characterized as follows:

$$E(\theta_i|\text{Choosing C}) = \frac{\sigma_\theta \sigma_d}{\sigma_h}(\rho_{\theta d} - \frac{\sigma_\theta}{\sigma_d})\lambda(z) \tag{9}$$

where $h \equiv a_i^C - a_i^S - \theta_i$, $\lambda(.) = \frac{\phi(.)}{1-\Phi(.)}$ is the Inverse Mills Ratio, with $\lambda'(.) > 0$, and $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$. $d \equiv a_i^C - a_i^S$, is the comparative advantage of individual i in the counter-stereotypical major compared to the stereotypical major.

Prediction 3 ***The effect of major choices***

$\frac{\partial E(\theta_i|Choosing\,C)}{\partial \beta} > 0 \; if \; \rho_{\theta d} > \frac{\sigma_\theta}{\sigma_d}$

$\frac{\partial E(\theta_i|Choosing\,C)}{\partial \beta} < 0 \; if \; \rho_{\theta d} < \frac{\sigma_\theta}{\sigma_d}$

If a stereotype exists regarding choices of major and the identity cost associated with making a choice challenging this stereotype is greater than zero, the students applying to majors in line or against stereotypes are heterogeneous in terms of their sensitivity to this stereotype (θ_i).

In particular, if the correlation between sensitivity to the stereotype and the comparative advantage in the counter-stereotypical major in the population is strong ($> \frac{\sigma_\theta}{\sigma_d}$), the students who apply to counter-stereotypical majors are those who strongly believe in the stereotype. On the other hand, if the correlation between sensitivity to the stereotype and the comparative advantage in the counter-stereotypical major in the population is weak or negative ($< \frac{\sigma_\theta}{\sigma_d}$), the students who apply to counter-stereotypical majors are those who believe less in the stereotype.

IV. Empirical Strategy

The theoretical framework presented above explains the key channels through which minority status affects students' performance once they are in the classroom and how selection into majors plays a role. The main objective of the empirical analysis is to empirically test whether these predictions are met. Two ingredients are necessary to do so – define stereotypical and counter-stereotypical major choices and causally identify the effect of a change in peers' identity. I discuss each of them in this section.

IV.A. Endogenous selection into minority status: major choices and stereotypes

Following an approach widely used in the literature (e.g. Hebert, 2020; Bostwick and Weinberg, 2021; Kugler et al., 2021), I infer gender stereotypes from the gender composition of the field. Specifically, 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 graphical description of the distribution of students across departments can be found in Figure I.[36] A man's choice is considered in line with stereotypes if the student chooses a major that belongs to majority male departments. On the other hand, a male student makes a choice of major that goes against societal expectations if he enrols in a majority female department. In the same way, a woman's choice is considered more in conformity with stereotypes the higher the share of women in her department of enrollment between 2008-2017.

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. First, the gender distribution of students across departments barely changes between 2008 and 2017 (Figures B.I), and it closely mirrors the distribution of men and women across subjects in the UK Higher Education system (Figure B.II). Second, it is the result of men and women applying to systematically different programs rather than the university's selection process (Figure B.III). Third, 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 better at reading (Carlana, 2019; Reuben et al., 2014; Ellemers, 2018; Lane, 2012), majority male departments are primarily characterized by math-intensive programs, while majority female departments are characterized by programs related to humanities.[37] Lastly, the gender composition of the different departments is well-known to students, as confirmed by the survey evidence presented in Figure B.IV.

I exploit two definitions of stereotypical selection – a continuous measure and a categorical measure. In the first case, the measure of stereotypical selection is simply the average share of same-gender students in each department of enrolment across the academic years 2008-2017. This continuous measure relies on the assumption that the effect of the department share of same-gender peers is linear. Furthermore, it is sensitive to the presence of outliers. To relax these assumptions, I define a categorical measure of stereotypical selection, where I categorize choices using a definition that becomes gradually more stringent. I start by considering, as students making stereotypical and counter-stereotypical choices, students who are enrolled in the top and bottom five departments in terms of the average share of same-gender students, respectively. I then replicate the same analysis by considering the top and

[36] In order to count each student only once, I consider each cohort in the first year of undergraduate degree. Moreover, the definition is based on the department of initial enrollment of the student. For 1.66% of students, the department of enrollment at the end of the year does not coincide with the initial department of choice. The analysis replicated with the department of enrollment at the end of the year delivers the same results (E.).

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

bottom 3 and 2 departments in terms of the share of same-gender students.[38]

IV.B. The causal effect of peers' identity

The second necessary step to test the model predictions is to estimate the causal effect of a variation in peers' identity. The identification of the effect of peers' identity is generally problematic because of two well-known issues: (i) the problem of correlated effects, which stems from correlated unobserved characteristics due, for instance, to endogenous choices of peers; (ii) the reflection problem, i.e. the difficulty of distinguishing between the impact of peers' outcomes (endogenous peer effects) and peers' characteristics (contextual peer effects) due to simultaneity in the behavior of interacting agents (Manski, 1993).

These two concerns are alleviated in this setting thanks to three features of the environment. First, class allocation can be considered as good as random conditional on scheduling constraints, addressing the problem of endogenous peer choices. Second, since students attend multiple courses during their first academic year, the same student is observed in different courses. Hence, I can control for student fixed effects, alleviating concerns of correlated effects (Bramoullé et al., 2020). Third, numerous first-year courses are common across programs and class allocation does not depend on the program of enrolment. Hence, students who are in the same class in a course are not necessarily allocated to the same class in other courses. This allows to alleviate the reflection problem by exploiting the presence of excluded peers, i.e. the sets of peers of peers do not perfectly coincide as classmates of classmates are not necessarily classmates (De Giorgi et al., 2010).

I estimate the causal effect of peers' identity by comparing the performance of the same student i across the courses c that they attend during their first academic year a, where they have been assigned to classes g with exogenous peers' characteristics, net of course and student fixed effects. The specification is given by the equation below:

$$y_{iacg} = \alpha_{ac} + \alpha_i + \beta \times \mathit{SLM}_{iacg} + \epsilon_{iacg} \tag{1}$$

The independent variable y_{iacg} is student i's grade in course c in academic year a. SLM_{iacg} is the *share of students like me*, i.e. the share of same-gender classmates in class group g of course c: the share of females for women and the share of males for men.[39] β captures the effect of an increase in the share of same-gender students in the class. Lastly, α_{ac} and α_i are course $\times$ year and student fixed effects respectively. The standard errors are clustered at the class level, which is the level of randomization.

Course $\times$ year fixed effects (α_{ac}) are essential to obtain exogenous variation in peers' characteristics. Students are randomly allocated to classes only within courses. Hence, course fixed effects are essential

[38] 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 is 16.

[39] In Appendix E.6., I provide evidence that defining class composition as the share of same-gender peers using the leave-out mean (Angrist, 2014) does not affect the results.

to capture differences in the 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.[40] In addition, course fixed effects control for other observable and unobservable individual traits and course characteristics that determine differences in students' performance across courses.

In addition to course $\times$ year fixed effects, I control for individual fixed effects α_i, which allows to address several 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. 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, individual fixed effects allow to control for observable and unobservable individual characteristics that are student-specific, common across courses, and determinants of program (course) selection and average performance.[41] Lastly, including individual fixed effects allows to estimate the effect of class composition by exploiting a *within-student* variation, rather than a *within-course* variation in class composition. While the former compares the performance of the same student across courses, the latter compares the performance of students allocated to different classes within the same course, assuming that classes in different courses are independent of each other. 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. Indeed, students' performance in a course might not only be affected by their classmates in the course, but also by their experience in other courses. I discuss the extent to which spillovers affect the estimates in Section V.B., where I show that the estimates controlling for individual fixed effects are more conservative than the estimates without individual fixed effects.

Two final potential concerns are addressed by the characteristics of the grading system at the university. First, discrimination, or differential treatment of minority students during the examination process, does not represent a concern in this setting, since grading is characterized by blind marking.[42] Second, the effect of peers' identity on the performance of men and women can be separately identified only under an assumption of no zero-sum game. This assumption does not seem unreasonable in the LSE

[40]For example, let's consider John. John is enrolled in the BSc in Economics and attends a Mathematics course, MA100, and an Economics course, EC100. EC100 is compulsory also for students enrolled in other undergraduate programs, such as BSc in Government, which is characterized by a higher share of women. Hence, EC100 is attended by a higher number of women compared to MA100. Thus, John will be more likely to have a higher share of female classmates in EC100 compared to MA100 overall. Once we control for the overall course composition, the identifying variation we are left with is the deviation of class composition from the overall course gender composition. As class allocation is orthogonal to individual characteristics, this is as good as random. Hence, the share of female classmates John is exposed to is exogenous once we control for course fixed effects.

[41]As an example, let's go back to John. John chooses BSc in Economics. He attends the compulsory courses 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. However, he struggles in MA100 and ST102, where he has been randomly allocated to male-majority classes. 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.

[42]A detailed description of what this implies can be found in Section II.C.

setting as marking is based on absolute grading. Hence, grades are based on the performance of each student, without any ranking of students.[43]

IV.C. The main specification

The identification strategy explained in Section IV.B. and the definition of stereotypical (counter-stereotypical) choices detailed in Section IV.A. are then combined to estimate the interplay between selection into minority status and the effect of class composition. I estimate the following two specifications.

Continuous Measure

$$y_{iacg} = \alpha_{ac} + \alpha_i + \beta_1 \times \mathit{SLM}_{iacg} + \beta_2 \times \mathit{SLM}_{iacg} \times \mathit{STS}_i + \epsilon_{iacg} \tag{2}$$

where the independent variable y_{iacg} is student i's grade in course c in academic year a. SLM_{iacg} is the *share of students like me*, i.e. the share of same-gender classmates in class group g of course c: the share of females for women and the share of males for men. 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. Lastly, α_{ac} and α_i are course $\times$ year and student fixed effects respectively. The standard errors are clustered at the class level, which is the level of randomization.

If being surrounded by same-gender classmates improves performance ($\beta_1 > 0$), a positive β_2 coefficient implies that the performance of students who made a stereotypical choice of major (women in humanities and men in mathematics) is more sensitive to the composition of the class compared to the performance of students who made a choice of major that challenges societal expectations. A negative coefficient, on the other hand, implies that the students whose performance is more affected by being in the minority are those who are enrolled in departments traditionally associated with the opposite gender (women in mathematics and men in humanities).

Categorical Measure

$$y_{iacg} = \alpha_{ac} + \alpha_i + \beta_c \times \mathit{SLM}_{iacg} + \beta_n \times \mathit{SLM}_{iacg} \times \mathit{NS}_i + \beta_s \times \mathit{SLM}_{iacg} \times \mathit{SS}_i + \epsilon_{iacg} \tag{3}$$

where the independent variable y_{iacg} is student i's grade in course c in academic year a. SLM_{iacg} is the *share of students like me*, i.e. the share of same-gender classmates in class group g of course c. *stereotypical selection* (SS_i) is a dummy equal to one if the student is enrolled in the 2,3, and 5 departments with the highest share of same-gender students. The omitted category are students enrolled in the 2,3, and 5 departments with the lowest share of same-gender students. *neutral selection* (NS_i) is a dummy equal to one for students who are not enrolled in top or bottom 2,3, and 5 departments. Lastly, α_{ac} and α_i are course $\times$ year and student fixed effects respectively. The standard errors are clustered at the class level, which is the level of randomization.

[43]More information is provided in Section II.C., while a graphical demonstration can be found in Bandiera et al. (2010).

The coefficients of interest are β_c, β_n, and β_s. β_c provides 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 a measure of the extent to which the effect of a change in the gender composition of classmates is different if the student is enrolled in a "neutral" or stereotypical department.

IV.D. Variation in the share of same gender classmates

We might be worried that the random allocation of students into classrooms creates a problem of small variation in classroom composition that could potentially lead to overestimating the effect in an analogous way as a problem of weak instrument in instrumental variable regressions (Angrist, 2014).

This is not a concern in this setting for a number of reasons. First, women represent a substantial share of the undergraduate population - 49%, as shown in Table I. Second, first-year courses are big - the average course size is 138.68 students and the average course has 9.40 classes. Third, the average share of same-program classmates is 0.48, as several first-year courses are common across departments and the allocation of students to classes is independent of the program of enrolment. Taken together these characteristics of the setting imply that students who are in the same class in a course are not necessarily allocated to the same class in other courses, allowing to exploit a variation in the share of same-gender classmates that goes beyond the variation within the program of enrollment.

The variation in class gender composition is substantial, as shown in Table II. The average share of same-gender classmates is 0.56, with a standard deviation of 0.16. Furthermore, most of the variation in the share of same-gender classmates is individual-specific – between classes for the same individual, and it cannot be explained by the course chosen or the program of enrollment. Indeed, as shown in Table II, controlling for course and student fixed effects, the residual variation represents 68.71% of the overall variation.[44] Moreover, if we plot the within-student (across-courses) variation in the share of same-gender classmates, obtained by taking the difference between the highest and lowest share of same-gender classmates for each student – Figure III, we can see that 50% of the students in the sample experience a difference between the highest and the lowest share of same gender classmates between classes of at least 30%.

Looking at Figure IV and Figure V, we can also 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 stereotypical and counter-stereotypical departments share significant common support – a necessary condition to be able to assess if the effect of changes in class composition differs for students who chose different majors.

IV.E. Validity of the identification strategy

The identification strategy relies on the following assumptions: (i) students are assigned to classes in a way that is orthogonal to their characteristics, conditional on scheduling constraints; (ii) there are no

[44] Following Olivetti et al. (2020)

factors that are correlated with class gender composition that differentially affect men and women; (iii) scheduling constraints are exogenous: men and women do not systematically select different courses. In this section, I discuss the evidence concerning the validity of these assumptions.

Test 1: 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 women, conditional on a given share of women in the course and scheduling constraints. Details on the specification used can be found in Appendix E.1..

The results confirm that students are not systematically assigned to classes based on their sex. Indeed, Table III shows that women do not have a higher probability of being in class groups with same-gender peers than men do.

Test 2: Class group allocation does not predict students' characteristics Following Feld and Zölitz (2017) and Braga et al. (2016), I test whether class group allocation predicts students' characteristics. I regress students' individual characteristics on class group dummies and scheduling constraints, running one regression for each combination of course and academic year included in the sample of analysis. I then test whether the class dummies in each regression are jointly significantly different from zero. According to Murdoch et al. (2008), the p-values of this test should be uniformly distributed with a mean of 0.5 if class allocation is random. More details on the test can be found in Appendix E.2., where I also show the results of a Monte Carlo simulation that confirms that, indeed, the more tests we perform, the more the p-values converge to a uniform distribution with mean 0.5.

The results of the test are consistent with conditional random allocation. Figure VI shows the p-values obtained from the tests of joint significance of the class group dummies. The probability that p-values are smaller or equal to 5% is roughly 5%, while the probability of p-values being smaller or equal to 10% is roughly equal to 10% for all the outcomes.

Test 3: Balance checks The third piece of evidence provides support for assumption (ii) – there are no factors that are correlated with class gender composition that differentially affect men and women – by showing that the share of female classmates is not correlated to the variation in other 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, once we control for gender. In addition, following the procedure suggested by Guryan et al. (2009), I test whether the share of female peers and same-gender peers in the class is not systematically correlated with students' characteristics once we control for the respective course leave-out means.

The results in Table IV corroborate the assumption that the variation in class gender composition is not systematically related to the variation in other peers' characteristics. More details on the interpretation of the results and the specification used can be found in Appendix E.3..

Test 4: Exogeneity of scheduling constraints The tests presented above provide evidence that students are as good as randomly allocated to class groups, conditional on scheduling constraints. With this test, I provide evidence that the limited course choices that first-year students are allowed to make after choosing a major are not gender specific.

Following Lavy and Schlosser (2011), I perform a permutation test where I simulate 1000 random unconstrained allocations of students to class groups. For each course, I randomly allocate students to class groups, under the sole assumption that course size, number of class groups, and class group size should reflect the observed ones. For each course and academic year, I then compute the difference between each class group share of female students and the course share of females for both the observed and the simulated allocations. Finally, I perform a Two-sample Kolmogorov-Smirnov test to verify that the observed and the simulated distributions of this statistics are not significantly different. More details can be found in Appendix E.4..

Comparing the actual allocation to this simulated unconstrained allocation represents a test for the assumption that students of different genders do not select systematically different courses among the first-year optional courses. If men and women were to select systematically different courses, the observed allocation would reflect systematically different course clashes. As such, it would not be consistent with an allocation that assigns students to random class groups without taking into account the constraints deriving from scheduling clashes with other courses.

The results of the test are shown in Figure VII. We fail to reject the null hypothesis that the distribution of the observed statistics and the simulated statistics are different at every significance level (1%, 5%, and 10%).[45]

V. Selection and the effect of peers' identity

V.A. The effect of minority status

Table V displays the results of Specification (2) in (Column 1) and Specification (3) in Columns (2)-(4), which make use of the continuous and categorical measures of stereotypical selection respectively. Two interesting patterns emerge. First, class gender composition affects students' performance in the course. The magnitude of the effect is small. However, it is in line with what other studies estimate for other higher education settings (e.g. Zölitz and Feld, 2021 for the School of Business and Economics at Maastricht University, and Booth et al., 2018 for the University of Essex). This is surprising and it confirms that even highly motivated and selected students in a strongly competitive environment such as the LSE are subject to social biases, and their performance is influenced by the identity of their peers.[46]

[45] Appendix E.4. also presents the result of the same test performed using a different statistics: the course standard deviation in the class group share of females.

[46] One caveat is that first-year courses are not high-stakes courses as the "year-one average" counts for only one-ninth of the final classification grade.

Secondly, the effect is not the same for every student. Selection into majors plays a significant role. In particular, the students who perform better when surrounded by same-gender peers in the classroom are students who enrolled in majors in line with stereotypes. Indeed, Column (1) shows that the coefficient of "share of students like me" is negative and significant, while the coefficient of the interaction "share of students like me x stereotypical selection" is positive and significant in all specifications. Moreover, these patterns are confirmed when we use the categorical definition of stereotypical selection. The impact of a change in the share of same-gender classmates decreases (almost linearly) when we use a less strict definition of stereotypical and counter-stereotypical choices, indicating that this is not the result of outliers or students enrolled in particular departments.

This result significantly diverges from what we would have predicted if we had generalized the findings of experiments where selection has been shut down by design. According to the literature, stereotypes exercise a detrimental effect when they are made salient in the mind of a person who belongs to a group that is negatively stereotyped[47] – in this setting, students who are enrolled in counter-stereotypical majors (men in humanities and women in math-intensive majors). Hence, we would have expected these students to perform worse when surrounded by different-gender classmates. The evidence points to the opposite direction. The performance of students who made a counter-stereotypical choice of major is not hindered (if anything is boosted) when they are in the minority in the classroom.

What this exactly implies in terms of direction, magnitude, and significance can be inferred from Panel B of Table V. For students who are enrolled in stereotypical departments, a 10% increase in the share of same-gender classmates increases course grades by 0.3314 points, equivalent to 2.03% of a standard deviation (Column 3). On the other hand, for students enrolled in counter-stereotypical departments, a 10% increase in the share of same-gender classmates decreases final course grades by 0.2839 points, i.e. 1.74% of a standard deviation (Column 3), albeit the effect is less robust. Lastly, the performance of students who are enrolled in departments with a balanced gender ratio is not affected by class composition.

If we consider that stereotypical choices of major imply selecting a program where the student's gender is majoritarian, these results indicate that students who chose to enroll in departments where being in the majority is the norm (e.g. men in math) suffer from being under-represented once in the classroom. Students who chose to enroll in departments where being in the minority is the norm (e.g. women in math) perform better when allocated to classes where their gender is under-represented. Those students who chose to enroll in departments with a balanced gender composition are not affected by the composition of the class.

What part of the distribution does the effect come from? Table VI provides evidence of the effect at different points of the grade distribution. Course grades range from zero to one hundred. 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 their

[47] For reference Ellemers (2018); Spencer et al. (2016); Steele and Aronson (1995).

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.

Class composition does not affect the probability of dropping out of the course, as displayed in Column (1). This is different from what has been estimated by other papers, that found in different settings that the composition of the environment affects persistence within courses and majors (Shan, 2020; Zölitz and Feld, 2021; Hill, 2017; Griffith, 2010). It is however not surprising in this context. The drop-out rate from first-year courses is low (2.9%). As mentioned in Section II., students can only apply to 5 programs across all UK universities, and the LSE rejection rate is as high as 75% every year. Hence, students who apply at the LSE are likely highly motivated. Furthermore, as only the "year-one average" counts for the final classification grade, students might decide to take the final exam even if they do not expect to perform well.

Regarding the decomposition of the average effect estimated in Table V, Columns (2) - (5) show that having same-gender students in the classroom lifts the performance of students who made stereotypical choices of major at every point of the grade distribution, with the effect increasing toward the top of the distribution.[48] On the other hand, a higher share of same-gender students in the classroom reduces the performance of the students who made a counter-stereotypical choice of major who are in the low-middle part of the grade distribution. Indeed, it has a small but not robust effect on the probability of passing the course and it prevents students from achieving first and upper-second-class honors (60 or more). However, it does not have an impact on the probability of getting a distinction.

V.B. Selection into minority status or other alternative explanations?

The main challenges to interpreting the results presented in the previous section as the interplay between selection and the effect of peers' gender are two. First, the variation in peers' gender might be correlated with differences in other characteristics of peers, programs, teachers, or other potential confounders. Second, choices of majors may be important for reasons other than the fact that they carry information regarding stereotypes. In this section, I attempt to address these concerns.

Major choices as informative of gender stereotypes I provide two pieces of evidence to support the hypothesis that the results are related to patterns of selection in line or against stereotypes regarding *gender* skills. First, I perform a placebo test where I estimate Specifications (2) and (3) by adding the interaction between the measure of stereotypical selection and the share of same ethnicity classmates, same program classmates, same background classmates, and peers' average ability at entry. Table A.I

[48]To illustrate this, let's consider the estimates in Panel C. A 10% increase in the share of same-gender students in the classroom increases the probability of getting a distinction in the course by 0.68 percentage points. Considering that 23% of students on average get a distinction, this represents a 2.96% increase in the probability of getting a distinction ($= 0.68/0.296$). The same calculation delivers an increase of 0.44% in the probability of passing the course, 0.94% increase in the probability of getting at least a lower second pass honor, and 1.39% increase in the probability of getting at least a first-class honor.

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. Second, Table A.III shows that the results are robust to other proxies of stereotypical selection: the share of women and men among applicants to undergraduate programs at the LSE, the share of men and women enrolled in different subjects in higher education in the UK, and the share of men and women among the staff working in each subject in higher education in UK.[49]

Difference in teachers or other peers' characteristics correlated with gender Students enrolled in different departments might be exposed to teachers with different characteristics. In particular, there might be more female teachers in majority-female departments. This can be problematic if female teachers treat women and men differently. Moreover, academic regulations are different across programs. Or else, math-intensive departments might attract students with different characteristics compared to disciplines that are related to humanities. Lastly, peers' gender might be related to other peers' characteristics, which might induce us to erroneously attribute the estimated effects to the gender composition of the class when it is instead due to other peers' characteristics (Bramoullé et al., 2020; Angrist, 2014; Manski, 1993).

To explore if this affects the results, I estimate Specifications (2) and (3) controlling for teaching assistant fixed effects, the share of the same ethnicity, same background (defined as students coming from State-funded or Independent schools), and same program students in the classroom. The results, in Table A.IV, are robust to the introduction of these controls. Despite I cannot directly test whether unobserved individual characteristics bias the estimated effects, the evidence of this exercise, in addition to the results of the balance checks, provides a consistent picture that this is unlikely the case.

Ability peer effects Another source of heterogeneity for the peers of students enrolled in male compared to female-dominated departments is ability. Selection and admission processes are different across programs. Thus, students who are enrolled in different departments might be exposed to peers with different abilities. Interpreting the estimated effects as the interaction between choices of major and the effect of belonging to the minority becomes problematic if ability is correlated with gender.

To address this potential concern, I estimate Specifications (2) and (3) controlling for average peers' ability, defined based on qualification scores at entry.[50] Results can be found in Column (5) of Table A.IV. They indicate that ability peer effects and the ability of peers changing along gender lines cannot explain the estimated patterns.

Differences in the strength of minority status across departments Despite the significant common support (evidence provided in Section IV.A.), the variation in the share of same-gender classmates is higher for students who are enrolled in stereotypical departments. This can explain the different effects of class

[49]Details regarding the measures used can be found in Table A.II.

[50]The sample is restricted to students with admission information for whom I was able to reconstruct a qualification score at entry, as explained in Appendix C.4..

composition estimated for students enrolled in different departments if the effect of a change in peers' gender composition is non-linear. Figure B.VII displays the estimates of a local polynomial regression for each value of the share of same-gender classmates. There is no evidence of non-monotonicity in the effect of a change in the share of same-gender classmates.

Spillovers across courses and mechanical effects The estimated effects using the within-student variation (including individual fixed effects) could be overestimated in the presence of spillovers across courses[51], which could rationalize the stronger effect found for the students who made a stereotypical choice of major given the higher within-student variation in the share of same-gender classmates. Moreover, including individual fixed effects might lead to overestimating the effect of an increase in the share of same-gender classmates due to a mechanical relation between individual observed and unobserved characteristics and class composition (Angrist, 2014). To test if this is the case, in Table A.V I display the results of the effect of same-gender classmates estimated by comparing the performance of students who are enrolled in the same course, but who have been assigned to classes with a different share of same-gender classmates - "within-course variation" – Column (3).[52] In Column (4), the sample is further restricted to one observation per student, which controls for the fact that observations for the same student in different courses are not independent. The results in Table A.V show that the estimates exploiting the within-student variation are smaller in magnitude compared to the estimated effects obtained exploiting the within-course variation.

On the other hand, controlling for course fixed effects might mechanically induce us to estimate an opposite effect for students belonging to "opposite" groups (e.g. men and women) within the classroom as it forces the average grades in the course to be equal to 0. Column (2) of Table A.V displays the results of a specification that does not include course $\times$ year fixed effects. They confirm that mechanical effects are unlikely to be the reason why we observe the estimated heterogeneity in the effect of same-gender classmates for students who made different choices of major.

V.C. Gender differences

Is the effect of stereotypical selection different for men and women? To answer this question I estimate Specifications (2) and (3) by 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 VII displays the results.

[51]For instance, being in a minority in a course might induce students to perform worse in the course and, at the same time, perform better in other courses where the share of same-gender classmates is higher.

[52]Under the assumption that students are assigned to classes independently of their assignment in other courses (after controlling for scheduling constraints), 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. One caveat has to be kept in mind. Spillovers across courses might also induce to overestimate the effect when exploiting the within-course variation. For instance, if being in a minority in one course has such a demotivating effect that students perform worse also in the other courses they attend, not controlling for student fixed effects would lead to overestimating the effect.

Both men and women are affected by the composition of the classroom, but men display a stronger effect. Column (1) shows that the men who chose majors in line with stereotypes benefit significantly more from having classmates of their gender compared to men who made choices against stereotypes, while the effect for women is smaller in magnitude and not statistically significant. Moving to Columns (2), (3), and (4), however, we notice 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 and not statistically significantly different from the effect estimated for men using the same definition of stereotypical selection. This indicates that both female and male performance is affected by the identity of their peers, but men's performance more strongly, as the effect is significant and strong also for students who didn't enroll in extreme departments in terms of gender composition.

This result is not surprising in this setting for two reasons. First, being in the minority in a male-dominated department is the result of stronger selection. Indeed, male-dominated departments are the most competitive, as they are characterized by higher rejection rates. Secondly, the stark gender segregation observed across departments seems to be driven mostly by men's application choices. Figure B.V shows the number of applicants for each offer made in the same year by each department by gender. While we don't see a strong difference in applications for women, for the same number of available places, a disproportionately higher number of men apply to departments that end up being male-dominated.[53]

These results are in line with what has been found by other studies that explored the heterogeneity of the effect of peers' gender across courses and departments (e.g. Zölitz and Feld, 2021; Hill, 2017; Oosterbeek and van Ewijk, 2014).[54] However, to the best of my knowledge, none of the existing studies explores the role that selection in line with or against societal expectations plays in explaining the heterogeneity in the evidence, which is the novel contribution of this paper.

V.D. Counterfactual scenario: reallocation policy in male-dominated fields

The evidence presented thus far shows that selection into minority status plays a role. To illustrate the importance of this result for policy, I use the estimated effects for men and women to perform a back-of-the-envelope calculation. I estimate what would happen to students' performance in the counterfactual

[53] This is in line with findings from Card and Payne (2021), that the gender gap in STEM readiness can be explained by the lower rate of university entry by non-STEM-oriented males.

[54] Oosterbeek and van Ewijk (2014) exploits the random allocation to class groups of first-year undergraduate students in economics and business at the University of Amsterdam and finds that males in groups with a high share of females do worse in courses with high math content. Moreover, Zölitz and Feld (2021) finds that an increase in female classmates reduces women's probability and increases men's probability of choosing a male-dominated major in the School of Business and Economics at Maastricht University. In line with the results of this paper, they also find that the students whose grades are affected by a change in the gender composition of the classroom are women in non-mathematical courses and men in mathematical courses (albeit men in mathematical courses perform better when surrounded by more women). Lastly, Hill (2017) finds that an increase in female peers is associated with a reduced likelihood of graduating in Business for males and suggestive evidence that females are less likely to complete STEM degrees (conditional on graduating) when exposed to cohorts with greater female shares.

scenario of a reallocation policy that enforces a more *balanced gender ratio* in *male-dominated departments*.

In male-dominated departments a negative gender gap exists.[55] The average share of women enrolled in undergraduate programs is 35.6%, and male students perform better than female students enrolled in the same course by 1.304 points on average (Figure B.VIII).

I find that enforcing a more balanced gender ratio in male-dominated fields reduces the gender gap performance. However, it lowers average performance at the same time, as neither men nor women benefit from a higher share of female classmates. To illustrate this, let us consider Column (3) of Table VII, for example. A 10% increase in the share of women decreases male students' performance by 0.3977 points, i.e. 2.4% of a standard deviation. At the same time, a 10% increase in the share of women decreases course grades for female students by 0.2263 points, i.e. 1.38% of a standard deviation. Hence, a 10% increase in the share of women in the class, holding the characteristics of students enrolled in these departments constant, reduces the gender gap by 13.15%.[56] However, this effect is the combination of two negative effects: a reduction in performance for men (equivalent to 30.5% of the gap), and a reduction in performance for women (equivalent to 17.35% of the gap). Hence, as the only students who benefit from being surrounded by same-gender peers are the students who made choices of major in line with stereotypes and belong to the majority group (men), a more balanced gender ratio negatively affects performance for both men and women.

This simple exercise illustrates that the gender composition of the environment can explain the gender gaps in performance – namely women performing worse than men in male-dominated fields and better than men in female-dominated fields. However, the effect is due to the *majority group* benefiting from being *over-represented*, rather than the *minority group* suffering from being *under-represented*. This indicates that, in these environments, also policies such as classroom segregation would not be beneficial for the members of the minority group.

V.E. 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. A description of the empirical specification and the evidence of its validity can be found in Appendix F..

The institution is characterized by an exceptionally diverse environment. As it can be seen in Table I, only 36% of students are White, while 48% of the population of undergraduates are Asian, primarily Chinese and Indian.

Furthermore, ethnic groups are distributed unequally across departments. The pattern along ethnic lines is exactly the opposite with respect to the distribution of traits that characterize departments along gender lines. Figure B.X shows that the departments where ethnic minorities – Asian students – represent

[55] For this exercise let's consider as male-dominated departments the 3 departments with the highest share of men.

[56] This is obtained by adding the effect for men (0.3977/1.304*100=30.5% of the gap) and women (0.2263/1.304*100=17.35% of the gap).

the majority of enrolled students are the math and science-intensive ones, which are also the departments where women are under-represented.

As for gender, this reflects stereotypes regarding group-specific skills and roles. Indeed, 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).

The results displayed in Table VIII confirm the generalizability of the patterns estimated for gender. Both White and Asian students benefit from being surrounded by more classmates of their ethnicity when they make a stereotypical selection of major. This is robust to controlling for the share of same-gender classmates (Column 2). These findings confirm that the estimated patterns concern the interaction between being in a minority and selection into fields, and are not only related to gender or gender-specific dynamics.

VI. Mechanisms: why selection matters

The last prediction of the model, Prediction 4, indicates that if social identity considerations are internalized in degree choices, the students enrolled in different departments are heterogeneous along traits that determine the effect of peers' identity on performance. In this section, I provide evidence that this is indeed the case. In particular, the evidence points towards students who made choices of majors not in line with societal expectations having weaker stereotypical associations, and facing lower costs of interacting with students of the opposite gender – exactly those characteristics that make them more resilient to the effect of changes in peers' identity.

VI.A. Direct evidence: the survey

Direct evidence on the strength of stereotypical associations and the cost of interacting with opposite-gender students is provided by leveraging the results of a survey administered to a subsample of students at the university.[57]

Figure VIII displays the 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. A positive score indicates that the student unconsciously associates women with humanities and men with science and math. A negative score indicates the opposite association. We can see that students on average associate men with math and science and women with humanities. However, this implicit association is significantly stronger for students who are enrolled in departments in line with stereotypes.[58]

Regarding the cost of interacting with opposite-gender students, the evidence is provided in Figure IX. The students who are enrolled in counter-stereotypical fields nominate significantly more opposite-

[57]Details on the survey can be found in Appendix G.

[58]These patterns are not specific to the sample of students that took part in the survey. Evidence in this direction is discussed at length in Appendix G.1..

gender students when asked about their friends, study-mates, and peers they ask questions on the material. This might be partially explained by the different availability of students of the opposite gender across majors. However, if this was the only reason why these patterns emerged, we would observe the same differences for all the categories. On the contrary, a significantly bigger difference is observed when students are asked to think about the peers they ask questions on the material. This suggests that students enrolled in counter-stereotypical majors might have a lower cost of engaging with opposite-gender peers, allowing them to benefit from academic support from their peers independent of their gender.

VI.B. Additional evidence

Individuals might possess implicit stereotypical associations in a self-favorable form only because of the tendency to associate self with desirable traits – in this case, own gender with the subject they study (Rudman et al., 2001). Moreover, once students decide to select a major, they might be induced to act in a way that confirms their decision to reduce cognitive dissonance (Akerlof and Dickens, 1982; Festinger, 1957). Or else, the same fact of being enrolled in a major might affect students' implicit stereotypical associations or preferences for same-gender peers. Three pieces of evidence suggest that these mechanisms cannot be the sole explanations.

First, the elicited implicit stereotypical associations and preferences for same-gender peers are not significantly stronger for students who have been exposed more to the environment. Figure B.XI shows that differences between students enrolled in stereotypical and counter-stereotypical majors are significant also when the sample is restricted to first-year students.[59]

Secondly, similar results are obtained when the analysis is replicated using another characteristic of the environment that might trigger social identity considerations – the gender of the class teacher. Table IX shows that the performance of students who made a counter-stereotypical choice of major 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 towards the mechanism being related to heterogeneity in sensitivity to social identity considerations.

Lastly, not every student who chooses a stereotypical major is affected by peers' gender to the same extent. In Table X, I estimate Specification (1) interacting the share of same-gender classmates with the Global Gender Gap Index (GGI) of each student's country of origin – a proxy for the ex-ante strength of gender stereotypes in students' minds.[60] I split students into terciles based on the GGI of students' countries of origin, where a higher tercile indicates that the student comes from a country with more progressive gender norms. Focusing on students who made a stereotypical choice of major – Column

[59]The results, however, are less precise due to the limited sample size.

[60]Guiso et al. (2008) provides evidence that the country's GGI is significantly correlated with gender gaps in math and reading. Countries with lower GGI have bigger math gender gaps in favor of men, while countries with high GGI have bigger reading gender gaps in favor of women.

1, we see that those who suffer the most from being surrounded by opposite-gender classmates are the students coming from countries with low GGI, i.e. countries with more conservative gender norms. On the contrary, students coming from countries where gender norms are more progressive do not suffer as much.[61] This is exactly what we would have expected if ex-ante sensitivity to stereotypes was the mechanism driving the results.

Taken together, these pieces of evidence suggest that, while the decision taken, the desire to avoid cognitive dissonance, or the tendency to associate self with desirable traits might play a role, these cannot be the only reasons for the estimated patterns to appear. Ex-ante "sensitivity" to stereotypes and social norms inducing students to select different majors and then react to the composition of the environment in a self-fulfilling way seems to be a more plausible explanation.

As a final remark, differences in ability, and in particular, minorities being positively selected, could be an alternative potential explanation for why the performance of students who made a counter-stereotypical choice of major is less affected by the composition of the environment. Using as a proxy for ability students' qualification scores at entry, I can provide some, albeit limited, evidence in this concern. This hypothesis does not seem to be supported by the evidence. Indeed, Figure B.XIII shows that the students who are enrolled in counter-stereotypical departments do not display higher qualification scores at entry.

VII. Discussion and Implications for Policy

This paper provides evidence that when belonging to a minority group is the result of choices that challenge societal expectations and strong selection, the members of the minority group are a selected group of people who do not suffer from being under-represented. On the contrary, those who uphold preferences to engage with same-gender peers and beliefs more strongly in line with societal expectations are those who belong to the majority group. They are the ones who are sensitive to the composition of the environment and benefit from being over-represented in the field.

In these environments, initiatives targeting minorities and bringing them together to level the playing field and equalize performance might not be the most effective and might even risk contributing to perpetuating stereotypes. Indeed, the back-of-the-envelope calculation presented in Section V.C. shows that bringing minorities together might deliver unintended effects in the short-run, by depressing performance for the members of the minority group. Moreover, it might reduce the possibility for the majority group to learn about the competence of minorities by engaging with them. The findings suggest that it is indeed the majority group who more strongly believes that stereotypes are true and would benefit from being exposed to a diverse environment. Lastly, policies targeting minorities could contribute in making group

[61] A visual representation of the correlation between the GGI index of the country of origin of the students and the share of females in their department can be found in Figure B.XII in Appendix. It shows that departments with a lower share of women are more likely chosen by students coming from countries with conservative gender norms (Panel A). Moreover, this relationship is not driven by a particular ethnic group (Panel B).

belonging more distinct, or even create potential backlash against minority individuals. Policies that aim at normalizing entering into certain occupations, by for example acting "down the ladder" rather than "up the ladder" (e.g. counter-stereotypical role models in lower education) or on the general public (e.g. ban of harmful gender stereotypes in ads) might be more effective in delivering a long-lasting change.

The findings of the paper hint towards interesting implications also for the effectiveness of measures such as affirmative action policies in equalizing *participation*. Indeed, while this discussion goes beyond the direct implications of this paper, two considerations directly follow from its findings. By introducing individuals who might have otherwise selected more stereotypical majors in environments where their group is severely under-represented, affirmative action policies might negatively affect the performance of the minority group in the short term. However, through their effect on the composition of the environment, they could potentially affect selection in subsequent periods, generating a trade-off between short and long-term effects. This represents another important avenue for research.

The setting of this paper, the LSE, is a very selective and competitive environment. While this makes it particularly suitable to study this research question, it renders its findings not easily generalizable to other educational environments. The findings might be more informative for the discussion around policies targeting selective and competitive working environments – for example, decision-making bodies or leadership positions.

REFERENCES

Akerlof, G. and W. T. Dickens (1982). The economic consequences of cognitive dissonance. *American Economic Review 72*(3), 307–19.

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

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics 30*, 98–108. Special Section articles on "What determined the dynamics of labour economics research in the past 25 years? edited by Joop Hartog and and European Association of Labour Economists 25th Annual Conference, Turin, Italy, 19-21 September 2013 Edited by Michele Pellizzari.

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.

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. (2018). Coase lecture – the glass ceiling. *Economica 85*(338), 205–231.

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

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature 55*(3), 789–865.

Bonomi, G., N. Gennaioli, and G. Tabellini (2021, 09). Identity, Beliefs, and Political Conflict*. *The Quarterly Journal of Economics 136*(4), 2371–2411.

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.

Bramoullé, Y., H. Djebbari, and B. Fortin (2020). Peer effects in networks: A survey. *Annual Review of Economics 12*(1), 603–629.

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.

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. and A. A. Payne (2021). High school choices and the gender gap in stem. *Economic Inquiry 59*(1), 9–28.

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, K., C. B. Flikkema, and O. Shurchkov (2021). Gender stereotypes in deliberation and team decisions. *Games and Economic Behavior 129*, 329–349.

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

Corno, L., E. La Ferrara, and J. Burns (2022, December). Interaction, stereotypes, and performance: Evidence from south africa. *American Economic Review 112*(12), 3848–75.

Cortes, P. and J. Pan (2018, 07). 425Occupation and Gender. In *The Oxford Handbook of Women and the Economy*. Oxford University Press.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010, April). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics 2*(2), 241–75.

Del Carpio, L. and M. Guadalupe (2022). More women in tech? evidence from a field experiment addressing social identity. *Management Science 68*(5), 3196–3218.

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.

Festinger, L. (1957, June). *A Theory of Cognitive Dissonance*. Stanford University Press.

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. (2014a). A Pollution Theory of Discrimination: Male and Female Differences in Occupations and Earnings. In *Human Capital in History: The American Record*, NBER Chapters, pp. 313–348. National Bureau of Economic Research, Inc.

Goldin, C. (2014b, 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. (2010). Persistence of women and minorities in stem field majors: Is it the school that matters? *Economics of Education Review 29*(6), 911–922.

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.

Jain, T. and M. Kapoor (2015). The impact of study groups and roommates on academic performance. *The Review of Economics and Statistics 97*(1), 44–54.

Karpowitz, C., S. D. O'Connell, J. Preece, and O. Stoddard (2024). Strength in numbers? gender composition, leadership, and women's influence in teams. *Journal of Political Economy 0*(ja), null.

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.

Lundberg, S. and J. Stearns (2019, February). Women in economics: Stalled progress. *Journal of Economic Perspectives 33*(1), 3–22.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

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.

Pregaldini, D., U. Backes-Gellner, and G. Eisenkopf (2020). Girls' preferences for stem and the effects of classroom gender composition: New evidence from a natural experiment. *Journal of Economic Behavior Organization 178*, 102–123.

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.

Rudman, L. A., A. G. Greenwald, and D. E. McGhee (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin 27*(9), 1164–1178.

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.

Sacerdote, B. (2001, 05). Peer Effects with Random Assignment: Results for Dartmouth Roommates*. *The Quarterly Journal of Economics 116*(2), 681–704.

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.

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.

Stinebrickner, R. and T. R. Stinebrickner (2006). What can be learned about peer effects using college roommates? evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics 90*(8), 1435–1454.

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 (2021). The effect of peer gender on major choice in business school. *Management Science 67*(11), 6963–6979.

VIII. TABLES

Table I
Descriptive Statistics: Students' Characteristics

	N.	Mean	SD	Min	Max
Students' Characteristics:					
Females	14297	0.49	0.50	0	1
Age at entry	14297	18.55	1.23	16	56
White	14297	0.36	0.48	0	1
Indian	14297	0.12	0.32	0	1
Chinese	14297	0.21	0.41	0	1
Black	14297	0.04	0.20	0	1
Other Asian	14297	0.15	0.36	0	1
Other	14297	0.04	0.19	0	1
Missing	14297	0.09	0.28	0	1
Qualification score at entry	9382	503.58	35.67	240	540
Within student var. in share of same gender classmates	14297	0.26	0.12	0	0.83
Classes and courses:					
Course size	509	138.68	163.73	12	1011
N. classes per course	509	9.40	10.79	1	69
Class size	4863	13.48	2.40	6	26
N. classes per student	14297	3.82	0.60	2	7
Class composition:					
Share of females	54554	0.49	0.17	0	1
Share of same gender classmates	54554	0.56	0.16	0.06	1
Share of co-ethnic classmates	54554	0.32	0.19	0.038	1
Share of same program classmates	54510	0.48	0.33	0.038	1
Average peers qualification score at entry	35598	500.69	16.31	420	540
Female TA	29736	0.37	0.48	0	1
Same gender TA	29736	0.52	0.50	0	1
Outcomes:					
Course Grade	54554	60.33	16.34	0	100
Pr(complete course)	54554	0.97	0.17	0	1

Notes: The descriptive statistics concern the sample of the analysis: 54554 course-year-class group level observations, corresponding to information on 14297 students. Before the sample selection, students attend 4.04 classes on average. This statistics decreases to 3.82 classes on average after data cleaning. 3.67% of students are observed only in 2 classes. Students for which we observe more than 4 classes are 5%, 86.2% of which attend 5 courses. These are students who attend language courses or half-unit courses during Michaelmas term.

Table II
Descriptive Statistics: Residual Variation

	N.	Mean	SD
Panel A: Course grades			
Raw	54554	60.33	16.34
Residual share after controlling for Course FEs	54554	0.00	15.84
Residual share after controlling for Student FEs	54554	0.00	8.61
Residual share after controlling for Student and Course FEs	54554	0.00	7.94
Panel B: Share of same-gender classmates			
Raw	54554	0.56	0.16
Residual share after controlling for Course FEs	54554	0.00	0.16
Residual share after controlling for Student FEs	54554	0.00	0.11
Residual share after controlling for Student and Course FEs	54554	0.00	0.11

Notes: Panel A displays the variation in course grades. The residual variation is obtained by taking the residuals of a regression of exam grades on course fixed effects, student fixed effects, and course and student fixed effects respectively. In Panel B, 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, student fixed effects, and course and student fixed effects respectively. In both panels, the sample consists of 54554 course-year-class group level observations, corresponding to information on 14297 students.

Table III
Identification: Effect of Gender On The Share of Same-gender Peers

	Share of same gender peers (1)
Panel A: Gender	
Female	0.003
	(0.004)
Course leave-out-mean	0.992***
	(0.015)
Observations	54554
Course fixed effects	Y
Parallel course dummies	Y

Notes: This Table displays the resuts of Specification (4). For each course c in academic year t I test that individual characteristics cannot predict the share of same gender peers in the class group g the student was assigned to. The dependent variable is the share of same gender peers. The Leave-out mean 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. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IV
Identification: Balance Check

Independent Variables (one at the time):	Share of females (1)	Share of same gender peers (2)	Share of female peers (3)
Asian Other	0.000	0.002	0.002
	(0.001)	(0.003)	(0.001)
Black	-0.002	0.002	-0.002
	(0.003)	(0.003)	(0.003)
Chinese	-0.000	-0.003*	-0.000
	(0.001)	(0.002)	(0.002)
Missing	-0.001	-0.002	-0.001
	(0.002)	(0.002)	(0.002)
Other	0.000	0.001	0.001
	(0.003)	(0.003)	(0.003)
White	-0.001	0.002	-0.000
	(0.001)	(0.001)	(0.001)
Independent School	-0.002*	-0.001	-0.003**
	(0.001)	(0.001)	(0.001)
Grammar School	-0.000	0.004	0.000
	(0.003)	(0.004)	(0.004)
Academy or Comprehensive School	0.001	0.002	0.001
	(0.001)	(0.001)	(0.001)
Colleges or Further Education	0.000	0.001	0.000
	(0.002)	(0.002)	(0.002)
Other	0.001	0.010	-0.000
	(0.009)	(0.011)	(0.010)
Not applicable	0.001	-0.001	0.001
	(0.001)	(0.001)	(0.001)
Age at entry	-0.000	-0.001	-0.000
	(0.000)	(0.000)	(0.000)
Qualification Score at entry	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Course Fixed Effects	Y	Y	Y
Parallel Course Dummies	Y	Y	Y
Female Dummy	Y	X	X
Course leave-out mean	X	Y	Y
N. Tests performed	14	14	14
N. Tests significant at 1%	0	0	0
N. Tests significant at 5%	0	0	1
N. Tests significant at 10%	1	1	1
Share Tests significant at 1%	0.000	0.000	0.000
Share Tests significant at 5%	0.000	0.000	0.071
Share Tests significant at 10%	0.071	0.071	0.071
Total N. Tests performed	42		
Total Share Tests significant at 1%	0		
Total Share Tests significant at 5%	0.048		
Total Share Tests significant at 10%	0.071		

Notes: This Table displays the results of Specification (6). For each course c in academic year t I test whether individual characteristics can predict the share of female – Column (1), the share of female peers – Column (2), the share of same gender peers – Column (3) – in the class g the student was assigned to. 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, and a dummy for each course the student attends during the academic year. In Column (1), I additionally control for a female dummy. In Columns (2) and (3), the female dummy is replaced with the course leave out mean. The number of observations is 54554 for all the tests with the exception of those that have the "Qualification Score at entry" as independent variable. In this case, the number of observations is 37625. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table V
Main Results: Stereotypical Selection

	Course grade			
	Continuous	Top and Bottom 2	Top and Bottom 3	Top and Bottom 5
	(1)	(2)	(3)	(4)
Panel A: Interaction				
Share of students like me	-5.946***	-3.556***	-2.839***	-1.215*
	(1.685)	(1.218)	(1.043)	(0.725)
Share of students like me × Neutral		3.741***	3.072***	1.060
		(1.289)	(1.139)	(0.960)
Share of students like me × Stereotypical Selection	12.410***	7.851***	6.153***	3.543***
	(3.137)	(1.543)	(1.312)	(0.942)
Panel B: Absolute Effect				
Share of students like me × Counter-stereotypical Selection		-3.556***	-2.839***	-1.215*
		(1.218)	(1.043)	(0.725)
Share of students like me × Neutral		0.185	0.233	-0.155
		(0.423)	(0.454)	(0.625)
Share of students like me × Stereotypical Selection		4.295***	3.314***	2.328***
		(0.902)	(0.773)	(0.585)
Observations	54554	54554	54554	54554
Course fixed effects	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y
Mean Dependent Variable	60.326	60.326	60.326	60.326
	(16.341)	(16.341)	(16.341)	(16.341)

Notes: This table provides evidence of the results of Specification (2) in Column (1) and Specification (3) in Columns (2)-(4) of Panel A. 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. If students didn't complete the course, the grade is coded as 0. Panel B presents the effect of a change in the share of same gender classmates for students who made different choices of major. For example, for students who made a stereotypical choice of major, the effect of a change in the share of same gender students (Panel B coefficient) is the sum of the coefficient of "Share of students like me" and "Share of students like me × Stereotypical Selection" in Panel A. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table VI
Main Results: Stereotypical Selection - Grades Distribution

	Pr(Drop-out) (1)	Grade ≥ 40 (2)	Grade ≥ 50 (3)	Grade ≥ 60 (4)	Grade ≥ 70 (5)
Panel A: Continuous Definition (Interaction)					
Share of students like me	0.018	-0.060*	-0.130***	-0.173***	-0.080
	(0.016)	(0.031)	(0.046)	(0.066)	(0.055)
Share of students like me × Stereotypical Selection	-0.028	0.125**	0.273***	0.339***	0.192*
	(0.030)	(0.057)	(0.086)	(0.124)	(0.102)
Panel B: Top & Bottom 2 (Absolute Effect)					
Share of students like me × Counter-stereotypical Selection	0.008	-0.037	-0.052	-0.086**	-0.044
	(0.011)	(0.024)	(0.035)	(0.042)	(0.040)
Share of students like me × Neutral	0.003	-0.001	0.001	-0.015	0.018
	(0.003)	(0.008)	(0.012)	(0.017)	(0.015)
Share of students like me × Stereotypical Selection	0.000	0.054***	0.097***	0.132***	0.066**
	(0.007)	(0.016)	(0.026)	(0.034)	(0.031)
Panel C: Top & Bottom 3 (Absolute Effect)					
Share of students like me × Counter-stereotypical Selection	0.010	-0.037*	-0.053*	-0.083**	-0.009
	(0.010)	(0.019)	(0.029)	(0.037)	(0.035)
Share of students like me × Neutral	0.003	0.001	0.003	-0.008	0.009
	(0.004)	(0.009)	(0.013)	(0.019)	(0.015)
Share of students like me × Stereotypical Selection	-0.000	0.041***	0.079***	0.087***	0.068**
	(0.006)	(0.014)	(0.022)	(0.029)	(0.027)
Panel D: Top & Bottom 5 (Absolute Effect)					
Share of students like me × Counter-stereotypical Selection	0.008	-0.013	-0.025	-0.036	-0.003
	(0.006)	(0.014)	(0.021)	(0.029)	(0.025)
Share of students like me × Neutral	0.005	-0.001	-0.006	-0.026	0.009
	(0.005)	(0.011)	(0.018)	(0.024)	(0.021)
Share of students like me × Stereotypical Selection	-0.002	0.023**	0.056***	0.059**	0.046**
	(0.005)	(0.011)	(0.017)	(0.023)	(0.020)
Observations	54554	54554	54554	54554	54554
Course fixed effects	Y	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y	Y
Mean Dependent Variable	0.029	0.931	0.843	0.628	0.232
	(0.167)	(0.254)	(0.364)	(0.483)	(0.422)

Notes: This table provides evidence of the results of Specification (2). In Panels B, C, and D, it presents the effect of a change in the share of same gender classmates for students who made different choices of major, as in Panel B of Table V. 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 in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table VII
Gender Differences: Stereotypical Selection

	Course grade			
	Continuous	Top and Bottom 2	Top and Bottom 3	Top and Bottom 5
	(1)	(2)	(3)	(4)
Panel A: Interaction				
Female x Share of students like me	-2.369	-3.012**	-2.263*	-0.532
	(2.162)	(1.326)	(1.220)	(1.026)
Female x Share of students like me × Neutral		3.733**	2.863**	0.874
		(1.456)	(1.388)	(1.378)
Female x Share of students like me × S. Selection	5.031	5.210**	4.155**	1.220
	(3.850)	(2.104)	(1.636)	(1.202)
Male x Share of students like me	-10.463***	-5.636*	-4.360**	-2.216**
	(2.499)	(2.999)	(1.995)	(0.956)
Male x Share of students like me × Neutral		5.258*	4.208**	1.555
		(3.064)	(2.099)	(1.300)
Male x Share of students like me × S. Selection	21.318***	10.613***	8.337***	5.745***
	(4.781)	(3.184)	(2.245)	(1.317)
Panel B: Absolute Effect				
Female x Share of students like me × C.S. Selection		-3.012**	-2.263*	-0.532
		(1.326)	(1.220)	(1.026)
Female x Share of students like me × Neutral		0.721	0.600	0.342
		(0.584)	(0.635)	(0.919)
Female x Share of students like me × S. Selection		2.199	1.892*	0.688
		(1.622)	(1.103)	(0.640)
Male x Share of students like me × C.S. Selection		-5.636*	-4.360**	-2.216**
		(2.999)	(1.995)	(0.956)
Male x Share of students like me × Neutral		-0.378	-0.152	-0.660
		(0.646)	(0.684)	(0.876)
Male x Share of students like me × S. Selection		4.977***	3.977***	3.530***
		(1.073)	(1.005)	(0.895)
Panel C: Gender differences (Female - Male)				
Counter-stereotypical Selection	8.093**	2.624	2.097	1.684
	(3.236)	(3.269)	(2.329)	(1.401)
Stereotypical Selection	-16.288***	-5.403	-4.182	-4.525**
	(6.005)	(3.637)	(2.714)	(1.769)
Observations	54554	54554	54554	54554
Course fixed effects	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y

Notes: This table provides evidence of the results of Specification (2) – Column (1) – and Specification (3) – Columns (2)-(4) – interacting the "Share of students like me" and "Share of students like me × Stereotypical Selection" with a dummy equal to one for females (Female) and a dummy equal to one for males (Male). In Column (1) stereotypical choices are defined based on 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 2 departments with the highest share of same gender students, while students in the 2 departments with the lowest share of same gender students are those who made a choice not in line with stereotypes. In Columns (3) and (4), I replicate the analysis by defining stereotypical and counter-stereotypical choices using the top and bottom 3 and 5 departments. The outcome variable is course grades, with incomplete courses coded as 0. Panel B displays the estimated effect of an increase in the share of same gender classmates for students who made different choices of major. Panel C displays the results of a test of the difference between the coefficient for men and women for each specification. "S. Selection" stands for Stereotypical Selection, while "C.S. Selection" stands for Counter-Stereotypical Selection. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table VIII
Beyond Gender: Stereotypical Selection and Ethnicity

	Course grade			
	(1)	(2)	(3)	(4)
Panel A: Continuous definition (Interaction)				
White $\times$ Share of students like me	-1.503	-1.503	-1.285	-1.282
	(1.999)	(1.997)	(2.012)	(2.010)
White $\times$ Share of students like me $\times$ S. Selection	8.206*	8.218*	8.357*	8.359*
	(4.434)	(4.431)	(4.453)	(4.449)
Asian $\times$ Share of students like me	-5.693***	-5.694***	-6.281***	-6.276***
	(1.857)	(1.857)	(1.862)	(1.862)
Asian $\times$ Share of students like me $\times$ S. Selection	14.147***	14.158***	14.549***	14.552***
	(3.645)	(3.644)	(3.642)	(3.642)
Other $\times$ Share of students like me	-12.359**	-12.304**		
	(4.836)	(4.833)		
Other $\times$ Share of students like me $\times$ S. Selection	65.700***	65.415***		
	(23.918)	(23.909)		
Share of same gender classmates		0.550		0.781*
		(0.367)		(0.399)
Observations	54554	54554	45447	45447
Course fixed effects	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y

Notes: This table provides evidence of the results of Specification (8). 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 Columns (2) and (4) I additionally control for the share of same gender classmates. Columns (3) and (4) replicate the same analysis excluding the "Other" ethnic group from the sample. “S. Selection” stands for Stereotypical Selection. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table IX
Mechanisms: Gender of Teaching Assistants

	Course Grade			
	Continuous	Top and Bottom 2	Top and Bottom 3	Top and Bottom 5
	(1)	(2)	(3)	(4)
Same Gender TA	-2.101**	-0.466	-0.285	0.000
	(0.843)	(0.556)	(0.461)	(0.372)
Same Gender TA $\times$ Neutral		0.181	-0.074	-0.438
		(0.594)	(0.512)	(0.455)
Same Gender TA $\times$ Stereotypical Selection	4.171***	2.150***	1.547**	0.669
	(1.611)	(0.759)	(0.625)	(0.535)
Observations	29736	29736	29736	29736
Course Fixed Effects	Y	Y	Y	Y
Student Fixed Effects	Y	Y	Y	Y

Notes: This table provides evidence of the results of Specification (2) – Column 1 – and Specification (3) – Columns (2)-(4), 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 (Same Gender TA). The sample is restricted to classes where I have information on the gender of the teaching assistant. 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. 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 in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

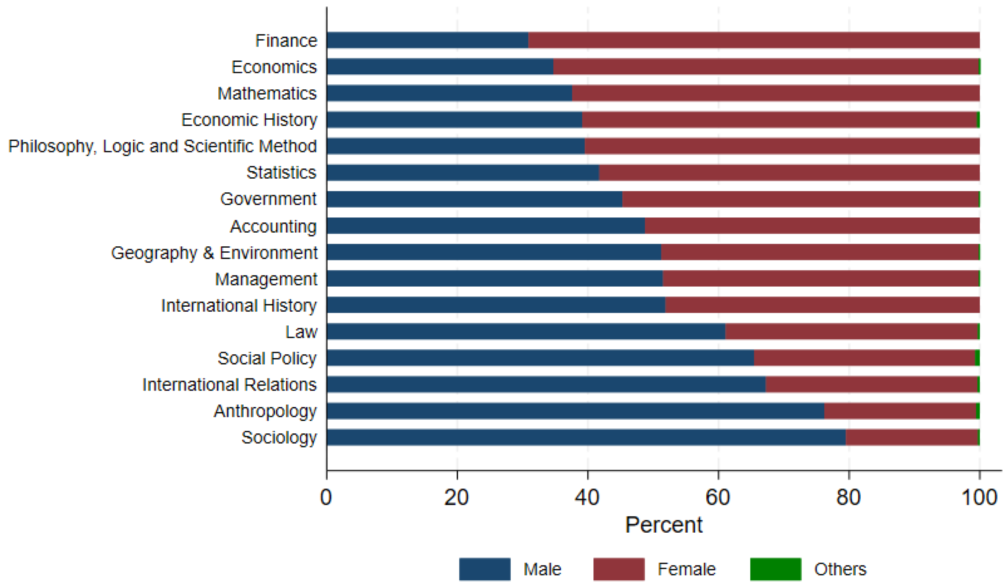
Table X
Mechanisms: Heterogeneity by Country of Origin - Global Gender Gap Index

	Course Result	
	Counter-Stereotypical	Stereotypical
	(1)	(2)
Share of students like me	-0.704	3.519***
	(1.367)	(1.102)
GGI Tercile=2 × Share of students like me	0.601	-0.426
	(1.856)	(1.473)
GGI Tercile=3 × Share of students like me	-0.011	-3.876**
	(1.956)	(1.526)
Observations	11373	20242
Course Fixed Effects	Y	Y
Student Fixed Effects	Y	Y

Notes: This table provides evidence of the results of Specification (1), interacting the "Share of students like me" with the Global Gender Gap Index (GGI) of each student's country of origin. I split students into terciles based on the GGI of students' countries of origin. A higher tercile indicates that the student comes from a country with more progressive gender norms. The reference category are students whose country of origin belongs to the first tercile – countries with more conservative gender norms. 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 in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

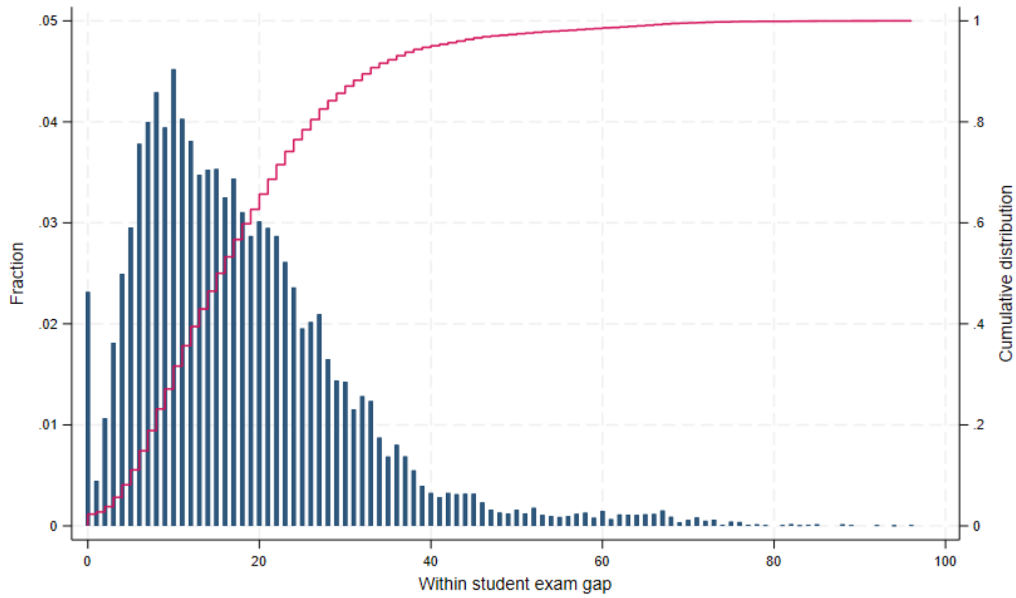
IX. FIGURES

Figure I
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.

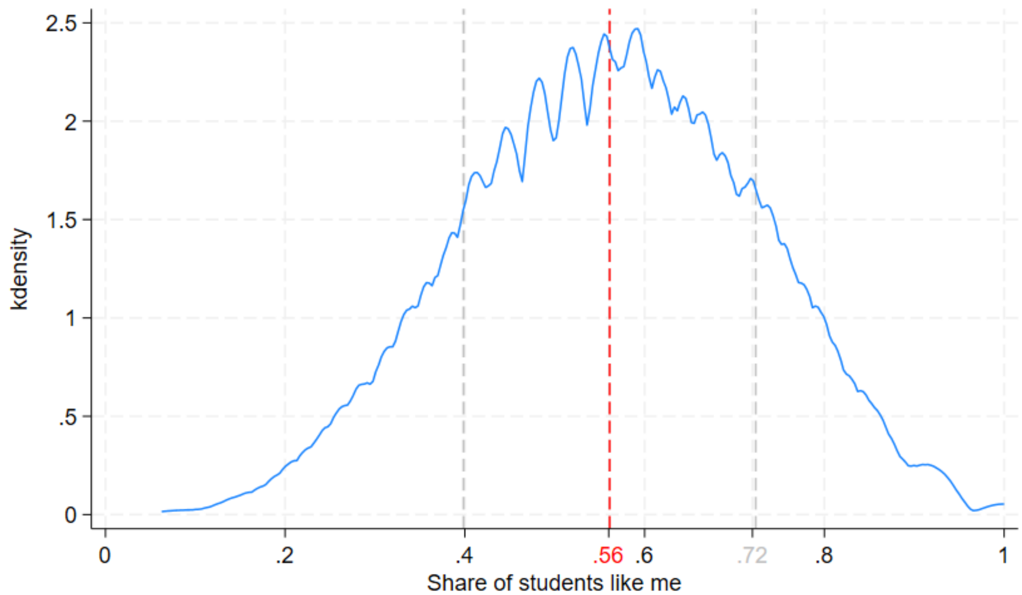
Figure II
Within-student Course Performance Gap



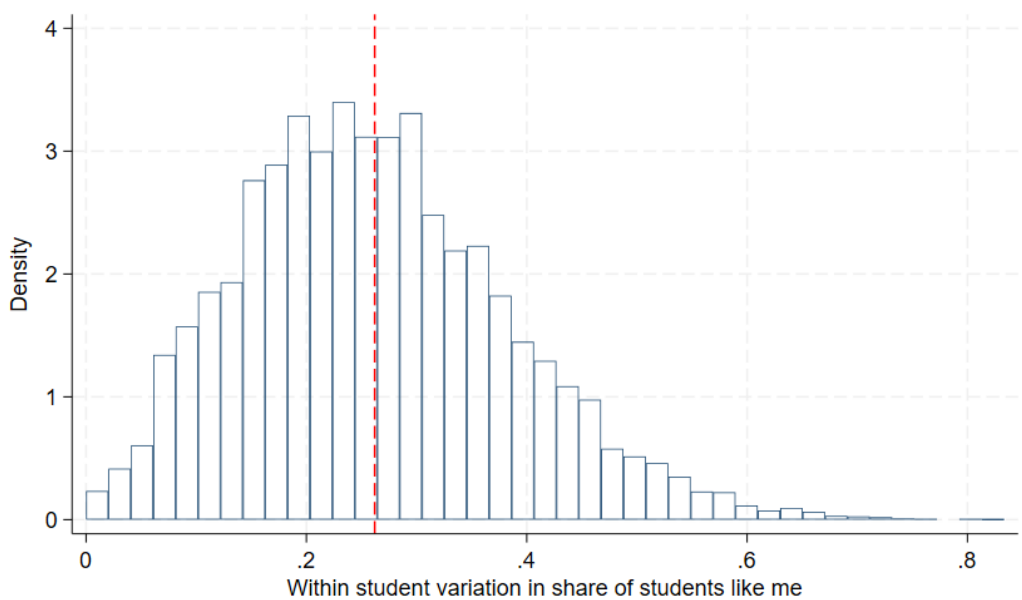
Notes: Following Bandiera et al. (2010), the figure displays the within student exam gap. This is defined as the difference between student i's highest and lowest exam mark across all the courses attended during Michaelmas term of the first year – the courses in analysis. The sample consists of 54554 course-year-class group level observations, corresponding to information on 14297 students.

Figure III
Within-student Variation in Share of Same-gender Classmates

Panel A



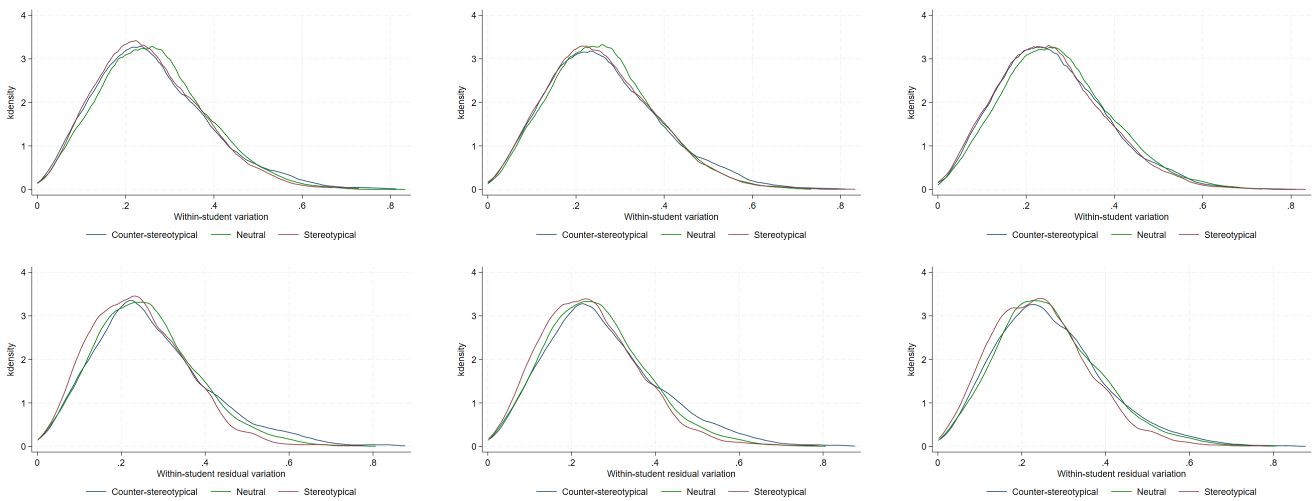
Panel B



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. The red line indicates the average variation."

Figure IV
Within-student Variation in Share of Same-gender Classmates Across Departments

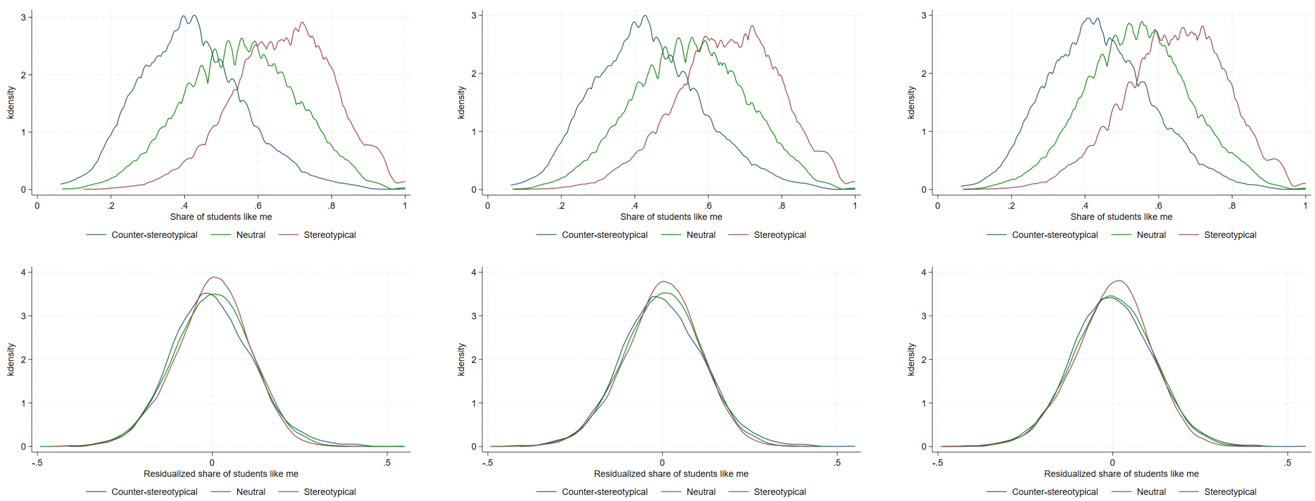
Panel A: Top & Bottom 2 **Panel B: Top & Bottom 3** **Panel C: Top & Bottom 5**



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. 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, obtained after regressing the share of same gender classmates on course $\times$ year fixed effects.

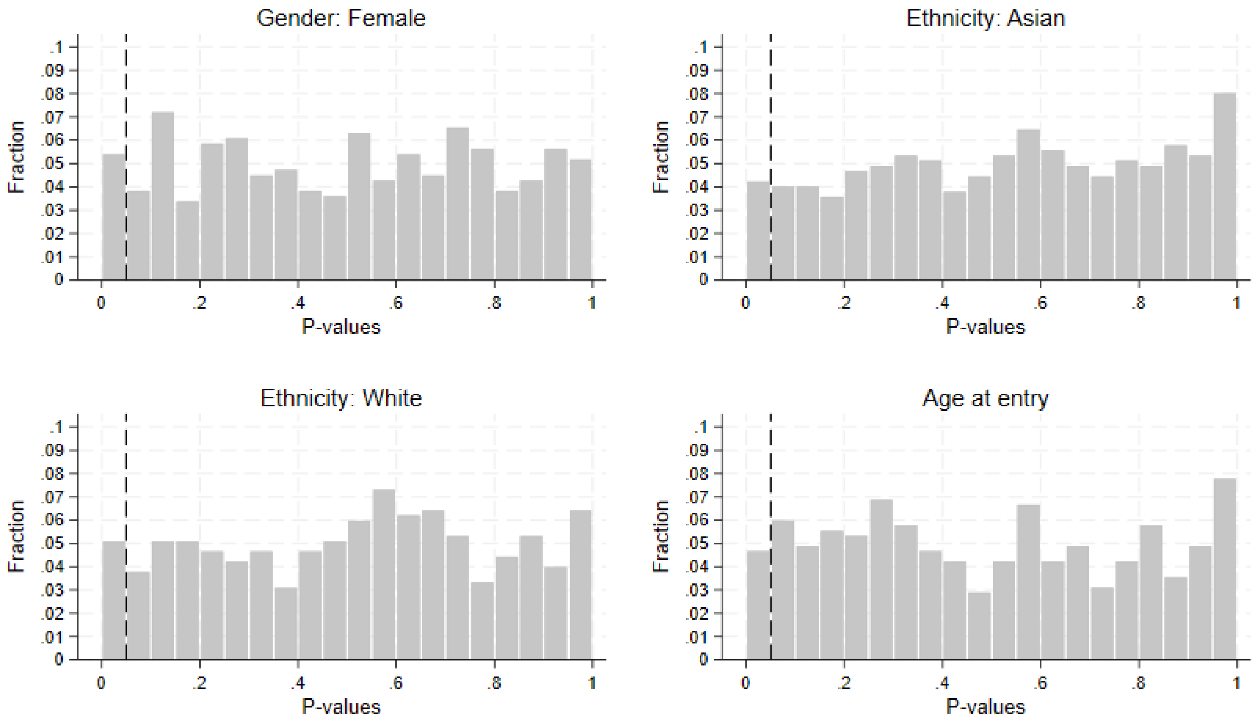
Figure V
Share of Same-gender Classmates Across Departments

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



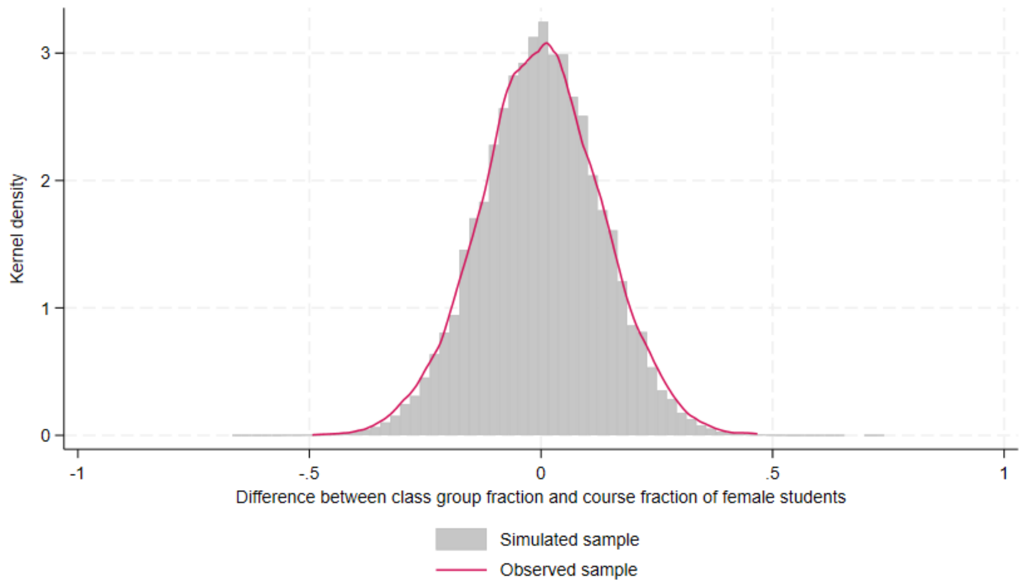
Notes: The figure displays the share of same gender classmates and the residual share of same gender classmates for students enrolled in stereotypical and counter-stereotypical departments, according to the three definitions used. The residual share of same gender classmates is obtained by taking the residuals of the regression of the same gender classmates on course × year fixed effects and student fixed effects.

Figure VI
Identification: Observed Sample P-values



Notes: This graph shows 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 clashes, for each one of the courses in analysis – Specification (5). 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. More information on the test can be found in Appendix E.3..

Figure VII
Identification: Allocation Simulation - Difference in Share of Female in Class and Course

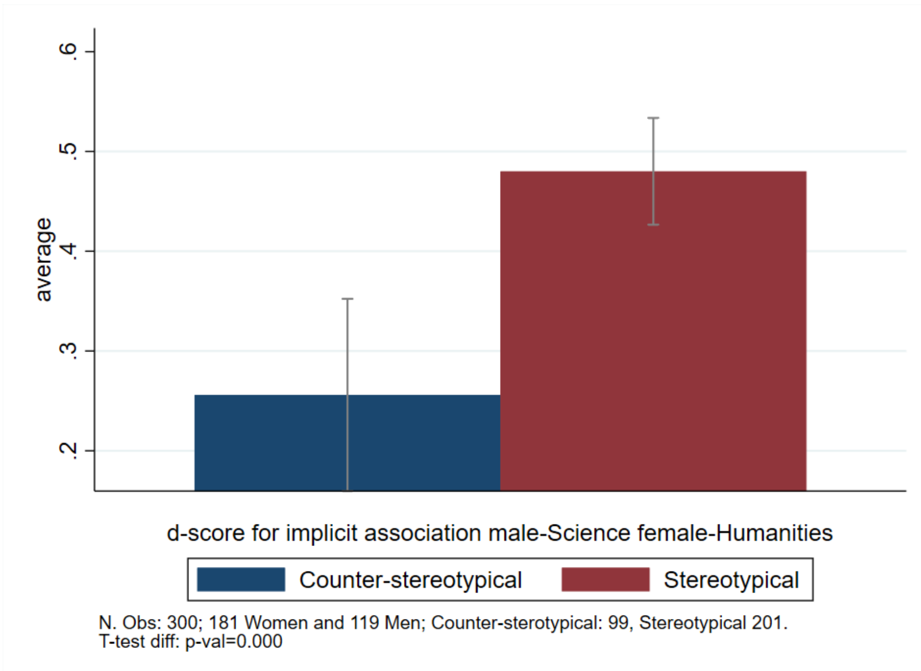


Two-sample Kolmogorov-Smirnov test

Smaller group	D	P-value
0:	0.0080	0.526
1:	-0.0083	0.500
Combined K-S:	0.0083	0.879

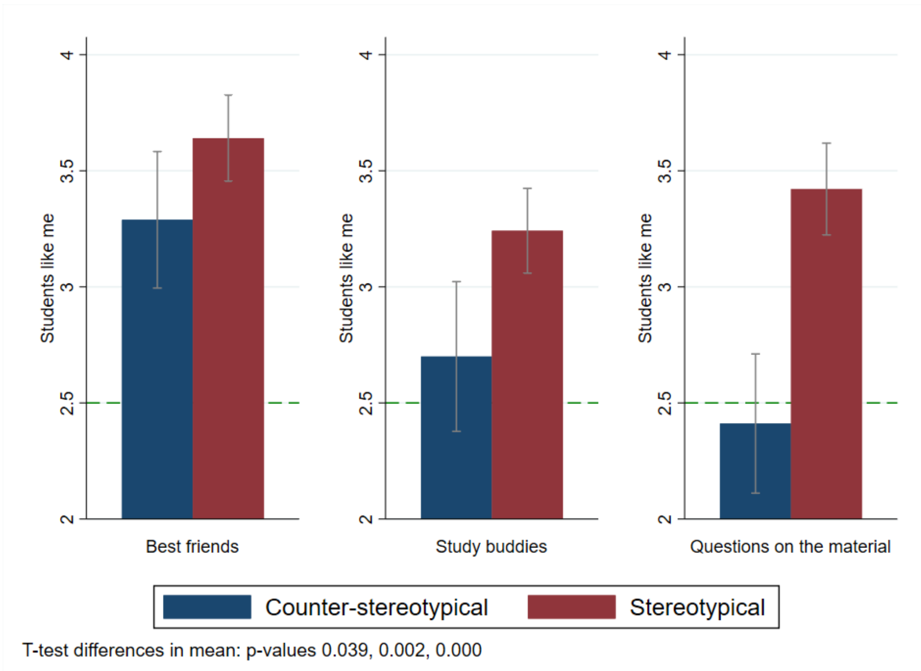
Notes: This Figure displays in dark grey the distribution of the difference between the share of women in each class and the share of women in the course for 1000 simulations of an unconstrained (not considering scheduling constraints) random allocation of students to classes. This is compared with the 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. The sample includes all the students that attend courses attended by the students in the sample of analysis and that did not change classes. More details on the test can be found in Appendix E.4..

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



Notes: The figure displays the 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. 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. The survey was administered to all the students enrolled in undergraduate programs in 2021. 500 students (10% of the student body) answered the survey. The response rate is similar to that of other surveys performed on undergraduate students at the university.

Figure IX
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 the Stereotypical and Counter-stereotypical groups 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. 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. The survey was administered to all the students enrolled in undergraduate programs in 2021. 500 students (10% of the student body) answered the survey. The response rate is similar to that of other surveys performed on undergraduate students at the university.

APPENDIX

Appendix A. ADDITIONAL TABLES

Table A.I
Robustness: Placebo Tests

	Course grade			
	(1)	(2)	(3)	(4)
Panel A: Continuous definition (Interaction)				
Share of students like me	-5.946***	-6.005***	-3.405**	-6.002***
	(1.685)	(1.685)	(1.683)	(2.067)
Share of students like me × Stereotypical Selection	12.410***	12.533***	7.564**	12.099***
	(3.137)	(3.138)	(3.140)	(3.846)
Share of co-ethnic classmates		2.772		
		(1.915)		
Share of co-ethnic classmates × Stereotypical Selection		-2.042		
		(3.530)		
Share of same program classmates			4.863***	
			(0.965)	
Share of same program classmates × Stereotypical Selection			-2.720	
			(1.764)	
Peers' average qualification score at entry				-0.022
				(0.020)
Peers' average qualification score at entry × Stereotypical Selection				0.056
				(0.036)
Observations	54554	54554	54554	35584
Course fixed effects	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y

Notes: This Table provides evidence of the results of Specification (2) in Column (1). Column (2) displays the results of Specification (2) with the addition of the measure of stereotypical selection interacted with the share of same ethnicity classmates – Column (2), the share of same program classmates – Column (3), and the average peers' qualification score at entry – Column (4). Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.II
Robustness: Departments Categorization

	LSE Undergraduate Enrolment (1)	LSE Undergraduate Applications (2)	UK HESA Undergraduate Enrolment (3)	UK HESA Staff (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^{+}	0.42^{-}
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: Column (1) displays the share of females among students enrolled in undergraduate programmes at the LSE. Column (2) displays the share of females among the students who applied to undergraduate programmes at the LSE. Column (3) shows the share of females among the students enrolled in undergraduate programmes in UK universities. Lastly, Column (4) displays the share of women among the staff working in UK universities. 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 is not present as a separate entry.

Table A.III
Robustness: Alternative Measures of Stereotypical Selection

	Course grade			
	LSE Undergraduates	LSE Applications	UK HESA Undergraduates	UK HESA Staff
	(1)	(2)	(3)	(4)
Panel A: Continuous definition (Interaction)				
Share of students like me	-5.946***	-4.953***	-7.483***	-4.796***
	(1.685)	(1.779)	(1.739)	(1.403)
Share of students like me × Stereotypical Selection	12.410***	10.531***	15.467***	10.275***
	(3.137)	(3.349)	(3.273)	(2.655)
Observations	54554	54554	54554	54554
Course fixed effects	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y

Notes: This table provides evidence of the results of Specification (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 (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 A.II. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.IV
Robustness: Peers' Characteristics and TA Fixed Effects

	Course grade				
	(1)	(2)	(3)	(4)	(5)
Panel A: Continuous definition (Interaction)					
Share of students like me	-5.946***	-3.782**	-5.983***	-3.500**	-5.994***
	(1.685)	(1.907)	(1.686)	(1.686)	(2.067)
Share of students like me × Stereotypical Selection	12.410***	8.311**	12.489***	7.605**	12.075***
	(3.137)	(3.571)	(3.139)	(3.142)	(3.846)
Panel B: Top & Bottom 2 (Interaction)					
Share of students like me	-3.556***	-2.062	-3.612***	-2.429**	-4.159**
	(1.218)	(1.336)	(1.218)	(1.216)	(1.638)
Share of students like me × Neutral	3.741***	2.284	3.813***	2.674**	4.102**
	(1.289)	(1.407)	(1.289)	(1.286)	(1.714)
Share of students like me × Stereotypical Selection	7.851***	5.491***	7.897***	5.452***	8.379***
	(1.543)	(1.747)	(1.543)	(1.543)	(2.053)
Panel C: Top & Bottom 3 (Interaction)					
Share of students like me	-2.921***	-2.116*	-2.955***	-1.973*	-3.723***
	(1.041)	(1.146)	(1.041)	(1.042)	(1.305)
Share of students like me × Neutral	3.185***	2.430**	3.256***	2.328**	3.740***
	(1.131)	(1.238)	(1.131)	(1.132)	(1.406)
Share of students like me × Stereotypical Selection	6.165***	4.879***	6.137***	4.145***	7.133***
	(1.310)	(1.474)	(1.309)	(1.319)	(1.662)
Panel D: Top & Bottom 5 (Interaction)					
Share of students like me	-2.839***	-1.914*	-2.878***	-1.887*	-3.208**
	(1.043)	(1.152)	(1.044)	(1.045)	(1.331)
Share of students like me × Neutral	3.072***	2.190*	3.130***	2.196*	3.088**
	(1.139)	(1.247)	(1.140)	(1.141)	(1.441)
Share of students like me × Stereotypical Selection	6.153***	4.680***	6.180***	4.146***	6.673***
	(1.312)	(1.479)	(1.312)	(1.320)	(1.686)
Observations	54554	51716	54554	54554	35584
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

Notes: This table provides evidence of the results of Specification 2 (Panel A) and Specification 3 (Panel B, C, D). 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 – Panels B, C, and D. The outcome variable is course grades, with incomplete courses coded as 0. 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. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.V
Robustness: Spillovers and Mechanical Effects

	Course Grade			
	(1)	(2)	(3)	(4)
Panel A: Continuous definition (Interaction)				
Share of students like me	-5.946***	-4.319**	-10.978***	-15.279***
	(1.685)	(1.990)	(2.368)	(4.723)
Stereotypical Selection			-7.974***	-10.872**
			(2.481)	(4.879)
Share of students like me × Stereotypical Selection	12.410***	9.421**	21.660***	29.566***
	(3.137)	(3.740)	(4.455)	(8.826)
Observations	54554	54554	54554	14297
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 (2) in Column (1). Column (2) displays the results of Specification (2) with no course fixed effects. Columns (3) and (4) display the results obtained by exploiting the within-course variation – Specification (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 in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

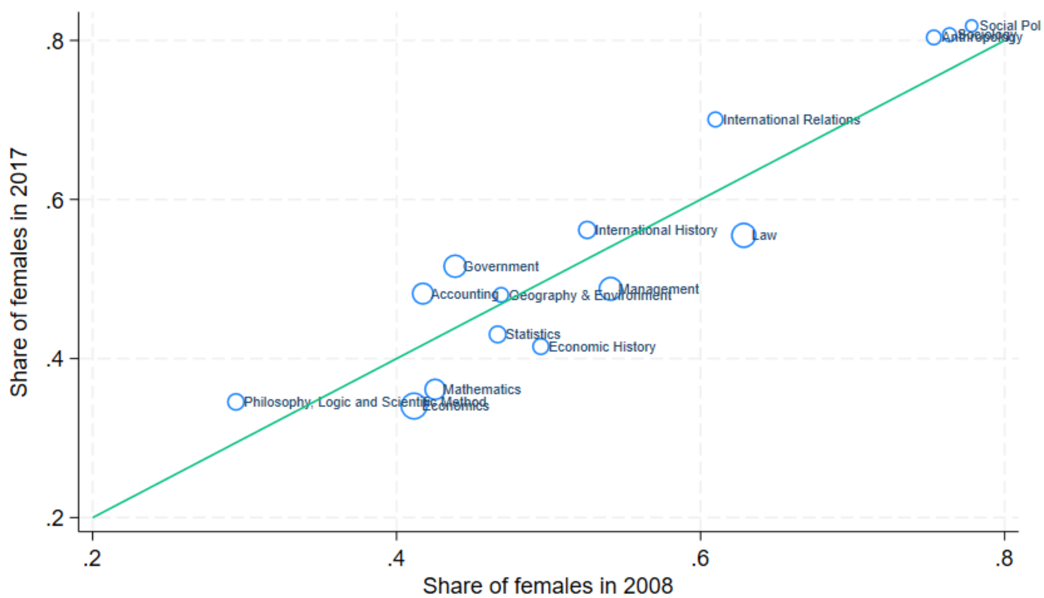
Table A.VI
Share of Courses Outside The Department of Enrolment

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*	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: This Table displays the share of the courses in program for first year undergraduate students outside the department of enrollment. For example, EC100 is a course taught by the Economics department. Hence, it is considered outside the department of enrollment for all the students who are enrolled in programs of study that do not belong to the Economics department. 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. The information is collected from the LSE Course Guides and Program Regulations for each academic year.
*Philosophy, Logic and Scientific Method

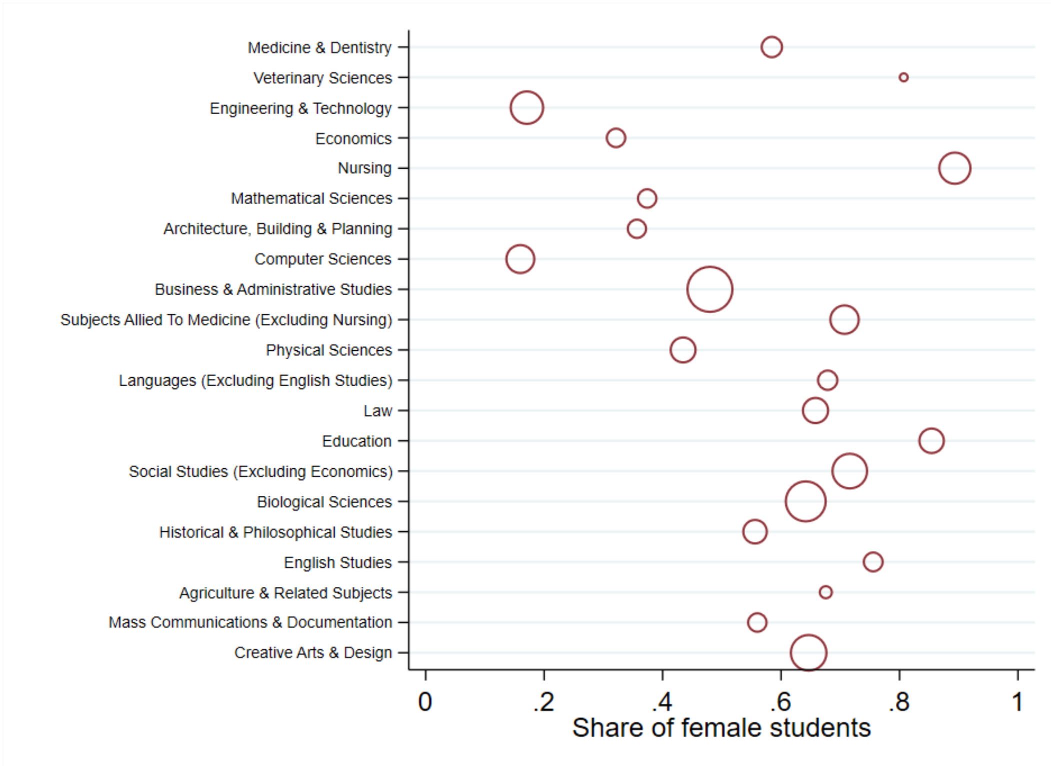
Appendix B. Additional Figures

Figure B.I
Departments Gender Composition Over The Years



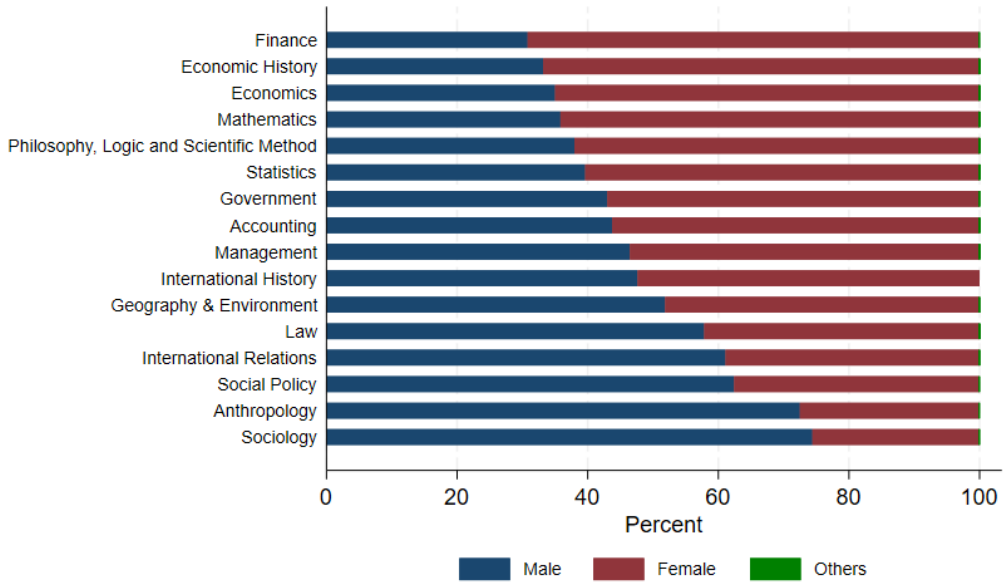
Notes: This figure illustrates the change in gender composition of LSE departments between 2008/09 and 2017/18. It is constructed based on the number of first year undergraduate students enrolled in bachelor programmes offered by each department in the two academic years. The size of the circle reflects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

Figure B.II
Enrollment UK Higher Education
(By Gender And Subject)



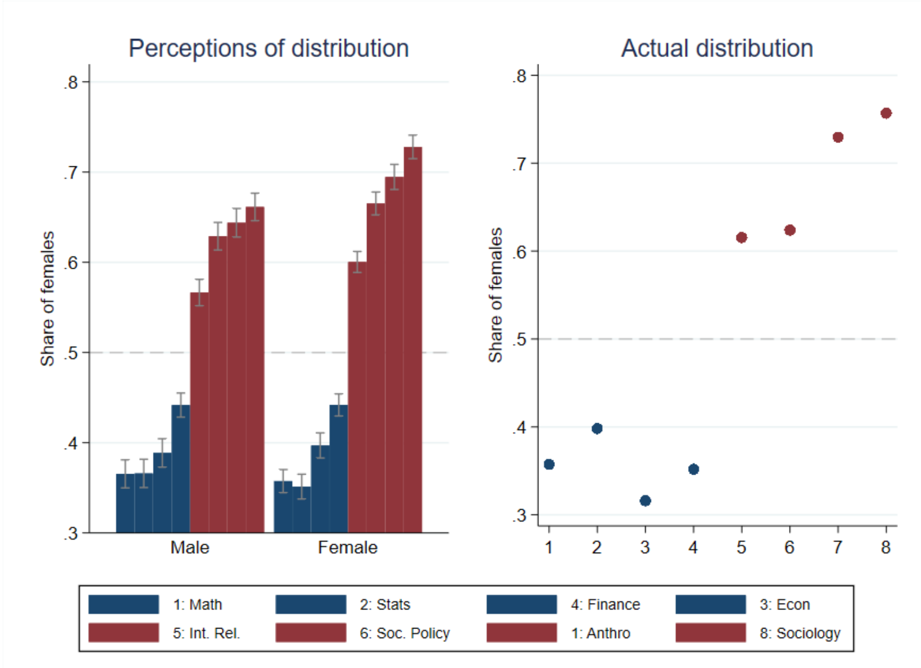
Notes: Information on the share of women among the overall population of undergraduate students enrolled in UK Universities. The size of the circle reflects the total number of students enrolled in each subject. Source: HESA Higher Education Student Data, academic year 2018-2019.

Figure B.III
Gender Composition of Applicants



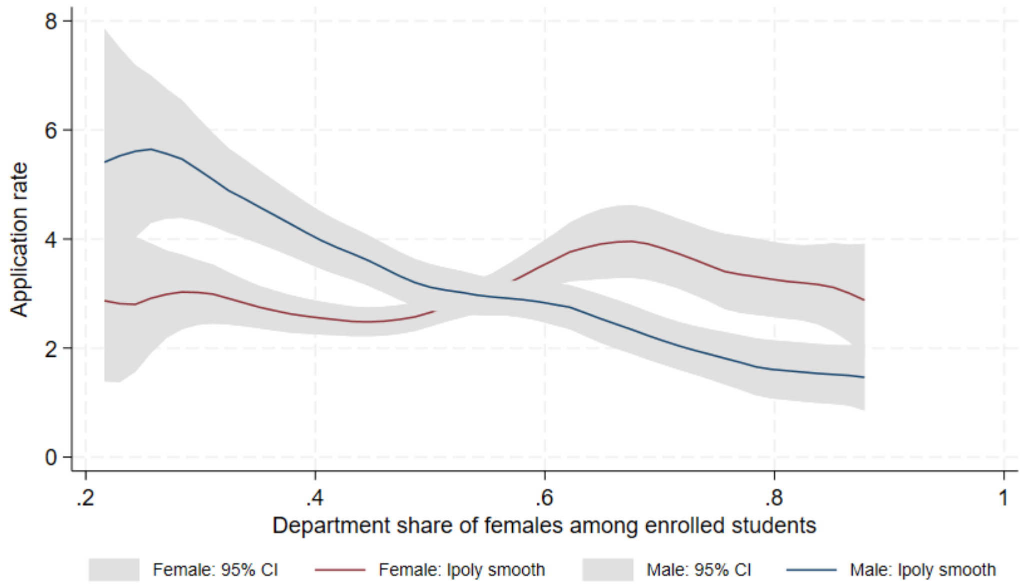
Notes: This Figure illustrates the gender composition of applicants for each department. It is constructed based on the number of men and women that apply to undergraduate programs at the university every year. The figure shows the average across academic years 2007/08 to 2019/20.

Figure B.IV
Explicit Beliefs: Share of Females Among Applicants



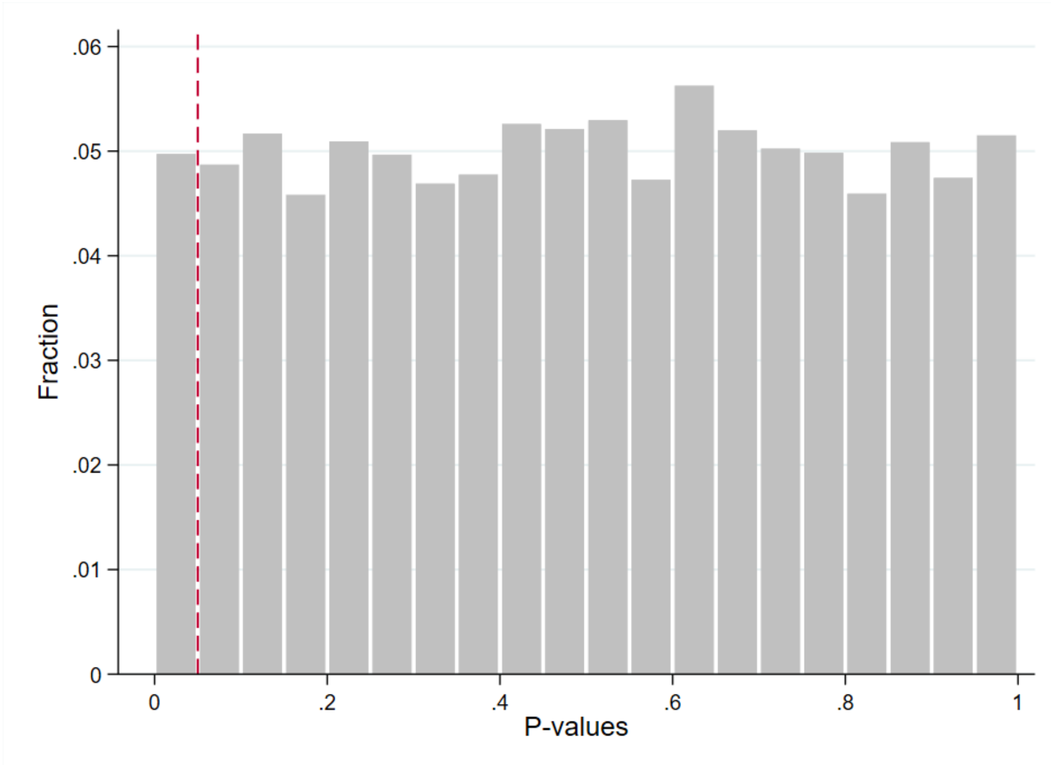
Notes: Panel A displays the answers to a question regarding the application of men and women across departments at the LSE. The question asks "Consider 10 people who apply for an undergraduate program in one of the following departments at the LSE. How many of them do you think are women?". Departments appear in a random order to students in the survey. The survey was administered to all the students enrolled in undergraduate programs in 2021. 500 students (10% of the student body) answered the survey. The response rate is similar to that of other surveys performed on undergraduate students at the university. Panel B displays the actual share of women among applicants to programs in the same departments (average across academic years 2007/08 to 2017/18).

Figure B.V
Application Rates by Gender



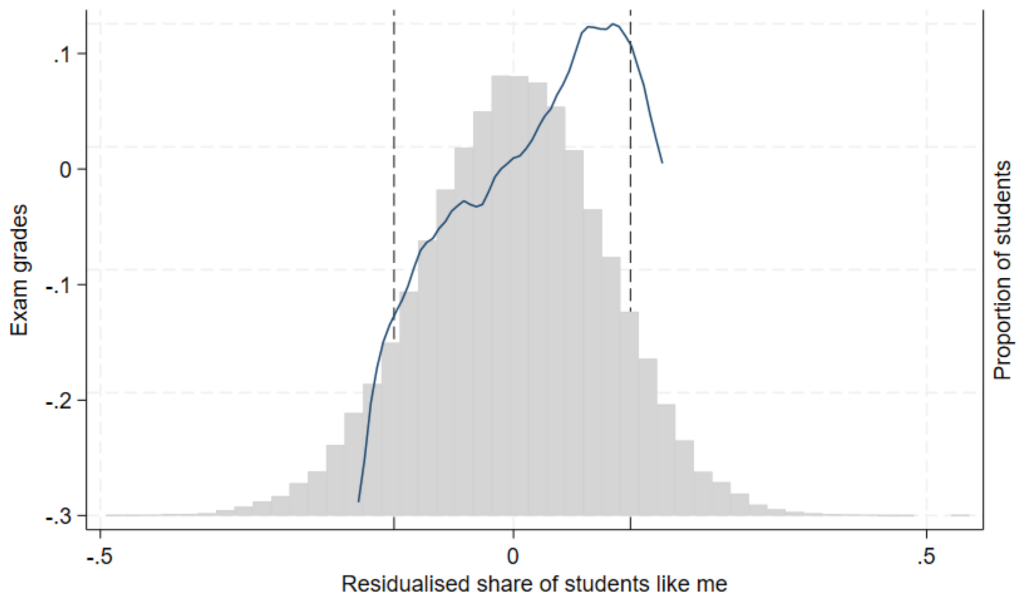
Notes: This figure displays a local polynomial plot of the relationship between the application rates for each program and the share of females among all the undergraduate students enrolled in the department the program belongs to. Application rates are calculated as the number of students that apply to each program divided by the number of available places (proxied by the total number of offers made). The sample includes all the years and all the undergraduate programs offered by the university. Male students are plotted in blue, while female students in red. The gray area represents 95% confidence intervals.

Figure B.VI
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 courses attended by the students in the sample of analysis and that did not change classes.

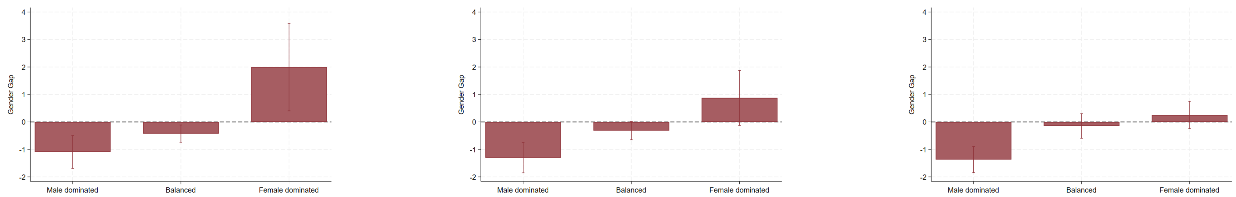
Figure B.VII
Robustness: Minority Effect - Linearity



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 as in Specification (1). 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.

Figure B.VIII
Gender gap: Course Performance in First Year Courses

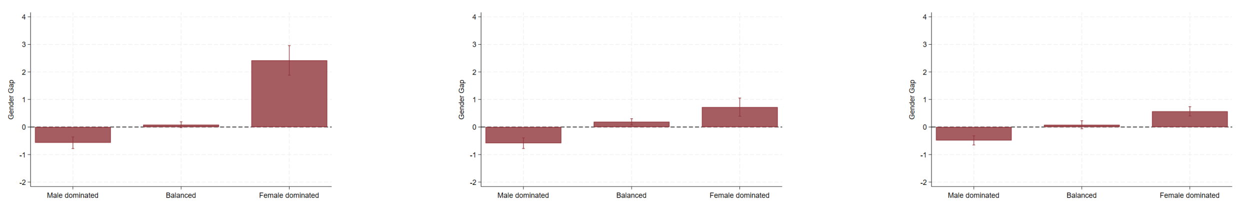
Panel A: Top & Bottom 2 | **Panel B: Top & Bottom 3** | **Panel C: Top & Bottom 5**



Notes: This figure displays the gender gap in course grades for first year students. The gender gap is calculated as the difference in course performance between women and men enrolled in the same course in the same year. Departments are categorized based on the top and bottom 2, 3, and 5 definition of stereotypical selection respectively. The sample includes the courses in analysis, however the picture is similar if all courses are included (Figure B.IX).

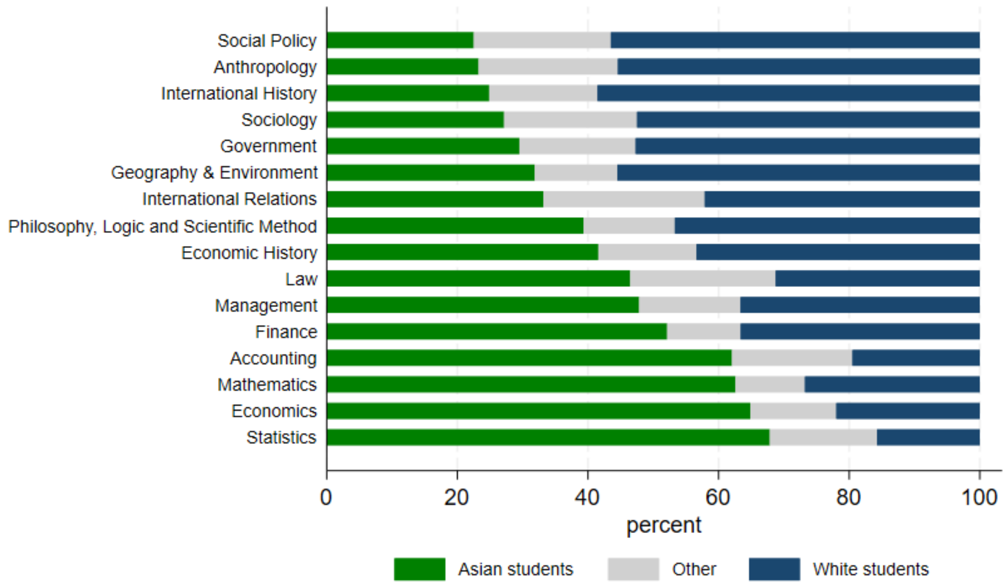
Figure B.IX
Gender gap: Course Performance in All Courses

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



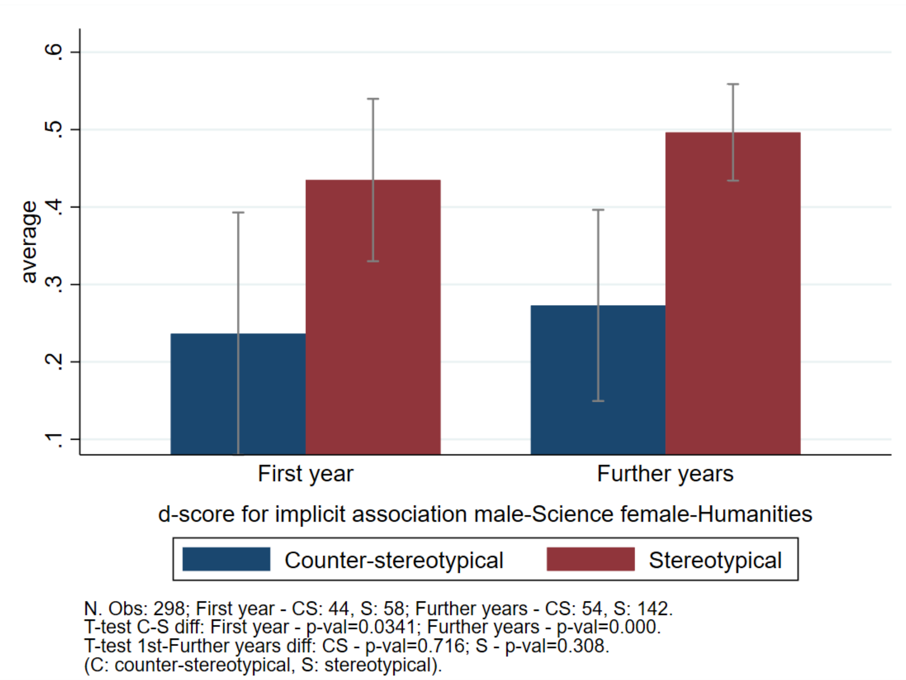
Notes: This figure displays the gender gap in course grades for undergraduate students. The gender gap is calculated as the difference in course performance between women and men enrolled in the same course in the same year. Departments are categorized based on the top and bottom 2, 3, and 5 definition of stereotypical selection respectively. The sample includes all years of degree.

Figure B.X
Departments Ethnic Composition



Note: The figure displays the distribution of Asian, White, and other ethnicity students enrolled in undergraduate programs across departments. 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. Other students is a residual category that includes students of other ethnic groups, but also students who provided no information regarding their ethnicity.

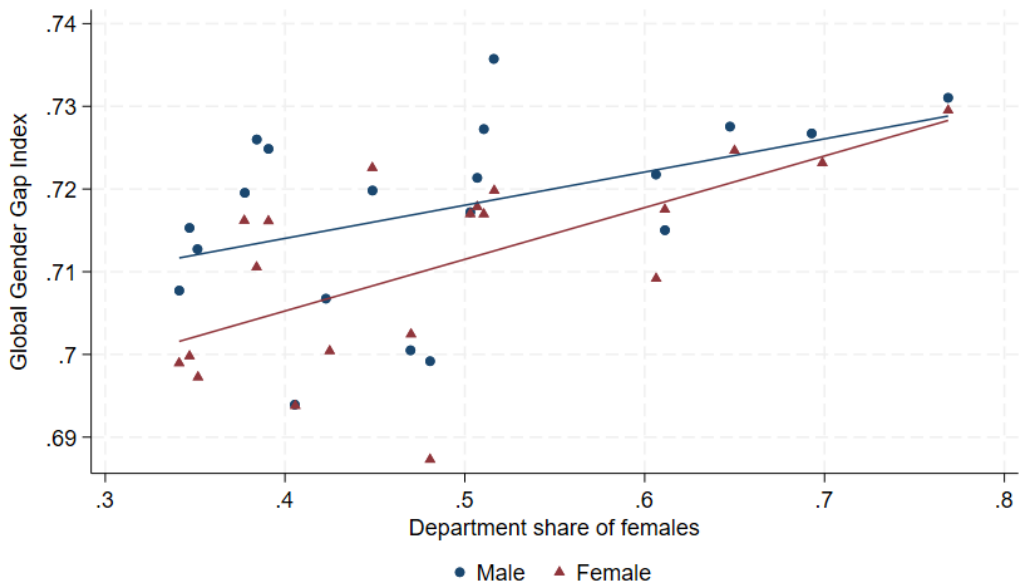
Figure B.XI
Mechanisms: Scientific-Male, Humanistic-Female IAT
(By Year of Enrollment)



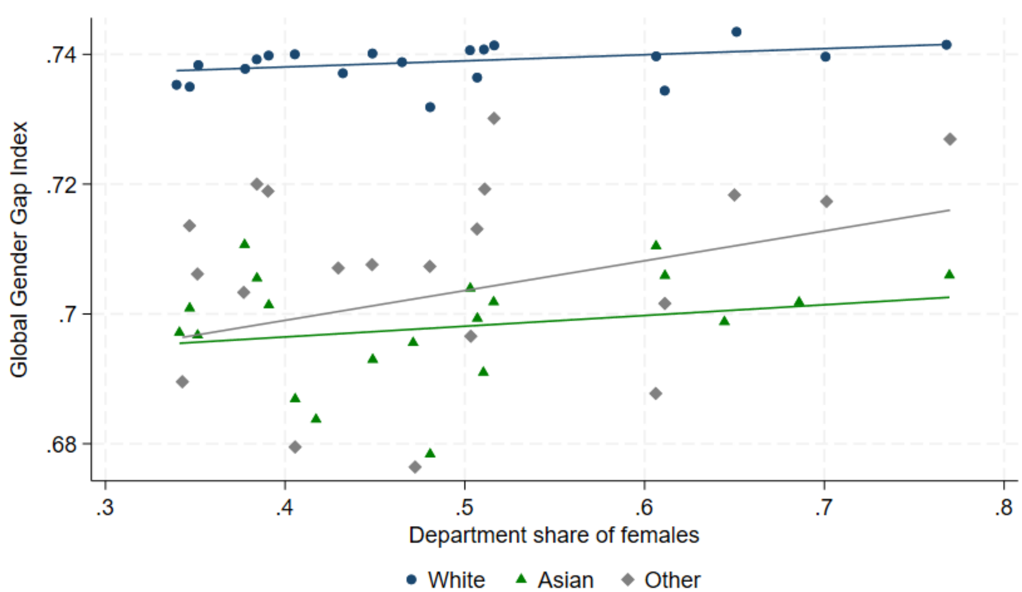
Notes: This figure displays the results of an implicit association test (Greenwald et al. 1998) to elicit students' associations between female-humanistic and male-scientific. A score of 0 indicates no association between male -scientific and female-humanistic; a positive score indicates that the student 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. 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. In the left panel, the sample is restricted to respondents who are enrolled in first year undergraduate programs. In the right panel, the sample includes respondents enrolled in their second, and third years undergraduate programs. The survey was administered to all the students enrolled in undergraduate programs in 2021. 500 students (10% of the student body) answered the survey. The response rate is similar to that of other surveys performed on undergraduate students at the university.

Figure B.XII
Mechanisms: Global Gender Gap Index

Panel A: By Students' Gender

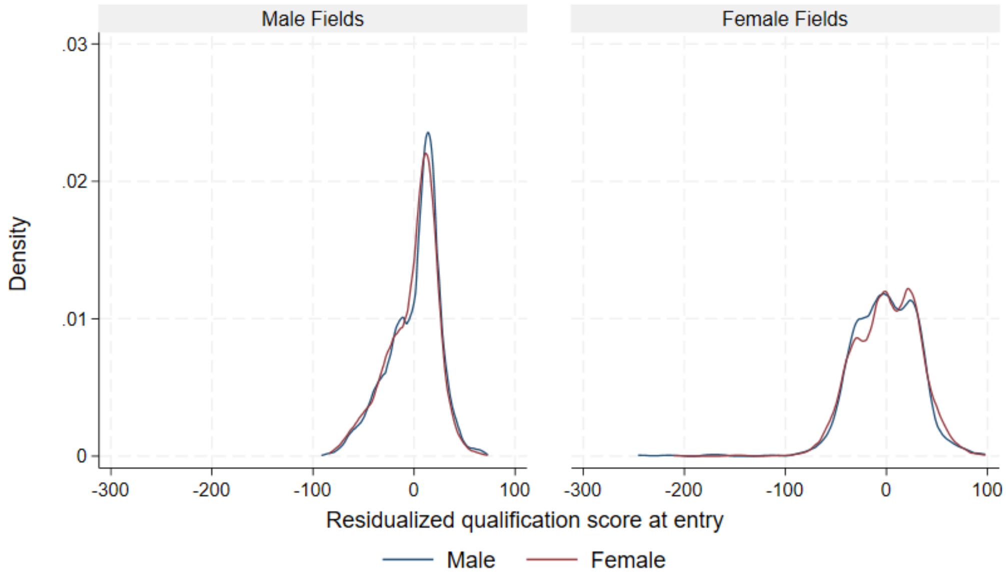


Panel A: By Students' Ehnicity



Notes: 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 in analysis for which I have information on the GGI of the country of origin.

Figure B.XIII
Mechanisms: Qualification Scores at Entry By Department and Gender



Notes: The graph displays the residualized qualification score at entry, residuals of regression of qualification score on country of origin 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: 9,382 students, 90.19% of the students enrolled in undergraduate programs in LSE between academic years 2011/12 and 2017/18. Following Campbell et al. (2019), I construct qualifications scores at entry based on the best three exam results among the A-level qualification scores students declared when applying to the 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 - 0.000, Female Fields - 0.311. P-values of T-test of equality in means: Male Fields - 0.040, Female Fields - 0.167.

Appendix C. APPENDIX TO THE DATA SECTION

C.1. Sample selection

The full sample of students who are enrolled in undergraduate programs at the 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 who are attending courses for the first time.

The sample is further restricted to courses that students attend during the first term. A few programs are characterized by half-year courses that are scheduled only 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 courses that first-year 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 26 students (excluding the smallest and largest 1% of observations) to exclude unreasonable and unlikely class sizes.[62] Furthermore, I exclude from the sample all the students who changed class groups during the first term, as in this case, I cannot identify the initial exogenous class allocation. This happens 5.08% of the time.[63] Lastly, I exclude class groups where I observe more than 50% of students changing groups. These occurrences correspond to class group "restructuring" - for instance when a class group gets canceled and students are re-shuffled to other class groups (5.08% 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: 54,554 course-year-class group level observations, corresponding to information on 14,297 students (99.36% of the original sample), who attend on average 3.82 courses in their first year.[64]

The final sample is made of 509 courses, with an average course size of 138.68 students, split into 9.40 class groups of 13.48 students each.

C.2. Students changing class groups

Students have the possibility to change class group during the term. 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 arise during the term, or external circumstances. In order to be able to change

[62]The university regulations fix a cap of 17 students per class group. There are frequent exceptions. 2.19% of classes have more than 17 students. However, class groups with more than 26 students seem very unlikely and thus are excluded from the sample.

[63]In Appendix C.2. I present the results of tests that show that the decision to change class group is not driven by the composition of the group.

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

class group, students are required to submit an official request. Considering the sample of first year undergraduate students who attended the courses for the first time, this happens 5.08% 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.

Firstly, Table C.I shows 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 class group fixed effects, while Columns (3) and (5) include a control for parallel courses, a dummy for each course the student attends in Michaelmas term. The table is divided into two panels. Panel A considers all the class groups the student has been allocated to during the term. If a student changes class group once during Michaelmas term, the student is present in the sample twice for the same course. Panel B consider only one class group per student. If a student changes class group, I randomly select one of the groups to which he has been allocated.

Secondly, 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 change 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 87.66% 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 C.II 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. I regress 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, a dummy for whether the student is a female, and the interaction of the two.

We can see from Panel B of Table C.I that women have a slightly higher probability of changing class group. However, this probability is small, between 0.4 and 0.7 percentage points higher with respect to a man in the same class group[65]. Moreover, we can see from Table C.II that the decision to change class group is independent of the gender composition of the class group.

These results indicate that when the students who changed class group are excluded from the analysis, the result is an unbalanced sample. However, this shouldn't generate a bias in the estimated effects since attrition is orthogonal to class group composition. Furthermore, since the students who change class group are excluded from the measure of class group composition, the share of female peers will be underestimated, but in an way that is independent of class group composition.

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

Table C.I
Decision To Change Class Group

	Dummy=1 if student changed class group				
	(1)	(2)	(3)	(4)	(5)
Panel A: All class groups					
Female	0.008**	0.008***	0.008***	0.015***	0.015***
	(0.003)	(0.002)	(0.002)	(0.004)	(0.004)
Observations	61632	61632	61632	33268	33268
Panel B: Random class groups					
Female	0.004**	0.004*	0.004*	0.007*	0.007*
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)
Observations	58533	58533	58533	30169	30169
Class group Fixed Effects	X	Y	Y	Y	Y
Parallel Courses	X	X	Y	X	Y

Notes: This table shows the results of a regression where the independent variable is a dummy equal to one if the student changes class group. The sample is restricted to Michaelmas term, first year undergraduate students who attend 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 changes class group, the student is present in the sample twice for the same course. Panel B consider only one class group per student. If a student changes 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 changes class group. Standard errors in parenthesis are clustered at class group level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.II
Decision To Change Class Group - Initial Allocation

	class group share of females		
	(1)	(2)	(3)
Initial allocation	0.007	0.007	0.007
	(0.009)	(0.008)	(0.008)
Initial allocation × Female	0.005	0.005	0.005
	(0.011)	(0.008)	(0.008)
Female	0.127***	0.072***	0.072***
	(0.008)	(0.006)	(0.006)
Observations	5214	5214	5214
Course Fixed Effects	X	Y	Y
Parallel courses	X	X	Y

Notes: This 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 change class group once in Michaelmas term, for whom I am 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. Standard errors in parenthesis are clustered at class group level in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C.3. 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.[66] 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 the 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 country's Global Gender Gap Index for 92.54% of students in the sample. I use this information as a proxy for the ex-ante strength of stereotypes regarding gender roles in the student's mind.

C.4. Ex-ante ability measure

I construct a measure of ability at entry based on the information that students provide during the admission process. I am able to construct a qualification score for a total of 9,382 students, 90.19% of the students in sample between academic years 2011/12 and 2017/18.[67]

The university bases its admission decisions on personal statements, academic achievements, 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 the LSE. Some students take courses that are equivalents to A-Levels. In these cases I calculate their A-Level-equivalent scores based on the university conversion tables for foreign students.

Figure C.I 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: 503.58 QCA points[68] with a standard deviation of 35.67[69].

Despite the limited variation, qualification scores are highly correlated with course grades. This can be seen in Panel B of Figure C.I, where year one course grades are plotted against qualification scores

[66] Source: The World Bank data

[67] Admission information are not available for students who enrolled at the LSE before 2011.

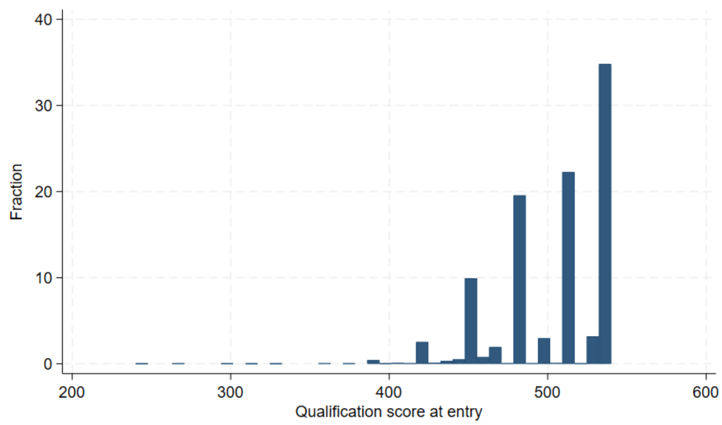
[68] This corresponds to a score in between a person that got two A-levels with A* and one A-level with A, and a person that has one A-level with A* and two A-levels with A.

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

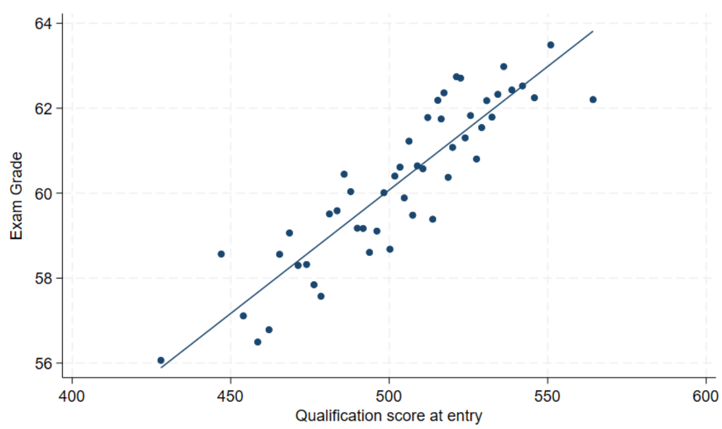
for all the students in analysis, controlling for students' country of origin (to control for differences in qualifications), program of enrolment, and course $\times$ year fixed effects. We can conclude that qualifications score at entry, albeit not a perfect proxy, contain some information regarding the students' ability at entry.

Figure C.I
Qualification Scores at Entry

Panel A



Panel B



Notes: Panel A displays qualification scores at entry for 9382 students, 90.19% 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 declare when applying to the 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 a binned scatterplot of the correlation between qualification scores at entry and first-year course grades controlling for the student's country of origin (to control for differences in qualifications), program of enrolment, and course $\times$ year fixed effects

Appendix D. Appendix to the theoretical framework

D.1. Derivation of the effect of changes in class composition on beliefs on ability

By assumption $\sigma_{s_g} < 0$, hence the effect of a change in class group composition on beliefs on ability depends on the nature of the stereotype in the field, which determines whether $(A_g - A_{-g})$ is smaller or grater than zero.

$$\frac{\partial a_i^b}{\partial s_g} = \overbrace{\theta_i \sigma_{s_g}}^{\leq 0}(A_g - A_{-g}); \tag{10}$$

If students chose a major in line with stereotypes ($A_g - A_{-g} > 0$) – e.g. men in math, an increase in the share of students of their gender depresses their beliefs regarding their ability. On the other hand, if students are enrolled in counter-stereotypical majors ($A_g - A_{-g} < 0$) – e.g. women in math, an increase in the share of same-gender classmates improves their beliefs regarding their ability.

D.2. Derivation of the effect of changes in class composition on effort choices

Let's define $G = \beta_i F_e - C_e = 0$, then

$$\frac{\partial e_i^*}{\partial s_g} = -\frac{\frac{\partial G}{\partial s_g}}{\frac{\partial G}{\partial e}} = -\frac{\beta_i F_{ea} \frac{\partial a_i^b}{\partial s_g} - \delta_i C_{es_g}}{\beta_i F_{ee} - C_{ee}} \tag{11}$$

By assumption, $\beta_i > 0$, $C_{ee} > 0$, $F_{ee} < 0$, and $C_{es_g} < 0$. Hence,

$$\frac{\partial e_i^*}{\partial s_g} = -\frac{\beta_i F_{ea} \frac{\partial a_i^b}{\partial s_g} - \overbrace{\delta_i C_{es_g}}^{\leq 0}}{\underbrace{\beta_i F_{ee} - C_{ee}}_{\leq 0}} \tag{12}$$

D.3. Derivation of the effect of changes in class composition on course grades

$$\begin{aligned}
\frac{\partial F(a_i^b, e_i^*)}{\partial s_g} =& F_a(a_i^b, e_i^*)\frac{\partial a_i^b}{\partial s_g} + F_e(a_i^b, e_i^*)\frac{\partial e_i^*}{\partial s_g} \\
=& F_a \theta_i \sigma_{s_g}(A_g - A_{-g}) - F_e \frac{\beta_i F_{ea} \frac{\partial a_i^b}{\partial s_g} - \delta_i C_{es_g}}{\beta_i F_{ee} - C_{ee}} \\
=& F_a \theta_i \sigma_{s_g}(A_g - A_{-g}) - F_e \frac{\beta_i F_{ea} \theta_i \sigma_{s_g}(A_g - A_{-g}) - \delta_i C_{es_g}}{\beta_i F_{ee} - C_{ee}} \\
=& \theta_i \sigma_{s_g}(A_g - A_{-g})[F_a - \frac{\beta_i F_e F_{ea}}{\beta_i F_{ee} - C_{ee}}] + \delta_i \frac{F_e C_{es_g}}{\beta_i F_{ee} - C_{ee}}\Bigg|_{(a_i^b, e_i^*)}
\end{aligned} \tag{13}$$

By assumption, $F_e > 0$, and $F_a > 0$. Hence,

$$\frac{\partial F(a_i^b, e_i^*)}{\partial s_g} = \overbrace{\theta_i \sigma_{s_g}}^{\leq 0}(A_g - A_{-g})[\overbrace{F_a}^{\geq 0} - \frac{\overbrace{\beta_i F_e}^{\geq 0} F_{ea}}{\underbrace{\beta_i F_{ee} - C_{ee}}_{\leq 0}}] + \delta_i \frac{\overbrace{F_e C_{es_g}}^{\leq 0}}{\underbrace{\beta_i F_{ee} - C_{ee}}_{\leq 0}}\Bigg|_{(a_i^b, e_i^*)} \tag{14}$$

If the student is enrolled in a counter-stereotypical department, $(A_g - A_{-g}) < 0$. As a consequence, if $F_{ea} > 0$, i.e. effort and ability are complement in the course performance function, an increase in the share of same gender classmates increases performance. An increase in the share of same gender classmates increases performance also if $F_{ea} < 0$, i.e. effort and ability are substitute, but the increased marginal returns from the higher perceived ability and the lower effort cost are stronger than this substitution effect.

If the student is enrolled in a stereotypical department, $(A_g - A_{-g}) > 0$, the overall effect of an increase in the share of same gender classmates depends on whether the effect on the marginal cost of effort (improved interactions) prevails on the effect on beliefs on ability (reduction in perceived ability).

D.4. Derivation of the probability of choosing a counter-stereotypical major

Since for simplification, we assumed that a_i^C, a_i^S, and θ_i follow a normal distribution $N(0, \sigma_x^2)$, with $x = a^S, a^C, \theta$ respectively, the probability of choosing a counter-stereotypical major is equal to

$$\begin{aligned} Pr(\text{Choosing C}) &= Pr(ln(R^C) > ln(R^S)) \\ &= Pr(w^C + a_i^C - \beta - \theta_i > w^S + a_i^S) \\ &= Pr(\underbrace{a_i^C - a_i^S - \theta_i}_{h} > w^S - w^C + \beta) \\ &= Pr\left(\frac{h_i}{\sigma_h} > \frac{w^S - w^C + \beta}{\sigma_h}\right) \\ &= 1 - \Phi\left(\frac{w^S - w^C + \beta}{\sigma_h}\right) \\ &= 1 - \Phi(z) \end{aligned} \tag{15}$$

where $z = \frac{w^S - w^C + \beta}{\sigma_h}$ and $h \equiv a_i^C - a_i^S - \theta_i$ is normally distributed with mean 0 and variance σ_h^2.

$\Phi(z)$ is increasing in z, hence, from this, it is easy to see that an increase in β reduces the probability of applying to the major everything else equal.

D.5. Derivation of the average strength of stereotypical associations for students choosing a counter-stereotypical major

Let's define $d \equiv a_i^C - a_i^S$, the relative comparative advantage of individual i in the counter-stereotypical major compared to the stereotypical major. It follows that

$$\begin{pmatrix} \theta \\ h = d - \theta \end{pmatrix} \sim N \begin{pmatrix} 0, & \sigma_\theta^2 & \sigma_{\theta d} - \sigma_\theta^2 \\ & \sigma_{\theta d} - \sigma_\theta^2 & \sigma_\theta^2 + \sigma_d^2 - 2\sigma_{\theta d} \end{pmatrix} \tag{16}$$

Hence, the average strength of stereotypical associations for students who decide to apply to a counter-stereotypical major is equal to

$$\begin{aligned} E(\theta_i|ChoosingC) &= E(\theta_i|\frac{h_i}{\sigma_h} > z) \\ &= \sigma_\theta E(\frac{\theta_i}{\sigma_\theta}|\frac{h_i}{\sigma_h} > z) \\ &= \sigma_\theta E[E(\frac{\theta_i}{\sigma_\theta}|\frac{h_i}{\sigma_h})|\frac{h_i}{\sigma_h} > z] \\ &= \sigma_\theta E[\rho_{\theta h}\frac{h_i}{\sigma_h}|\frac{h_i}{\sigma_h} > z] \\ &= \sigma_\theta \rho_{\theta h} E[\frac{h_i}{\sigma_h}|\frac{h_i}{\sigma_h} > z] \\ &= \sigma_\theta \rho_{\theta h} \frac{\phi(z)}{1 - \Phi(z)} \\ &= \sigma_\theta \rho_{\theta h} \lambda(z) \\ &= \frac{\sigma_\theta \sigma_d}{\sigma_h}(\rho_{\theta d} - \frac{\sigma_\theta}{\sigma_d})\lambda(z) \end{aligned} \tag{17}$$

where $\lambda(.) = \frac{\phi(.)}{1-\Phi(.)}$ is the Inverse Mills Ratio, with $\lambda'(.) > 0$, and $\rho_{xy} = \frac{\sigma xy}{\sigma_x \sigma_y}$.

Given the equation above, we can characterize the effect of an increase in β on the strength of stereotypical associations in the mind of the students who decide to apply to it.

$$\frac{dE(\theta_i|Choosing\,C)}{d\beta} = \frac{\sigma_\theta \sigma_d}{\sigma_h}(\rho_{\theta d} - \frac{\sigma_\theta}{\sigma_d})\frac{d\lambda(z)}{d\beta} \tag{18}$$

Since z is increasing in β and $\lambda'(.) > 0$, then

$$\uparrow \beta \Rightarrow \begin{cases} \uparrow E(\theta_i|Choosing\,C) & \iff \quad \rho_{\theta d} > \frac{\sigma_\theta}{\sigma_d} \\ \downarrow E(\theta_i|Choosing\,C) & \iff \quad \rho_{\theta d} < \frac{\sigma_\theta}{\sigma_d} \end{cases} \tag{19}$$

Appendix E. APPENDIX TO THE EMPIRICAL STRATEGY

E.1. Test 1: 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 class groups where there are more same-gender peers, conditional on a given share of same-gender peers in the course and scheduling constraints. The specification used is the following:

$$SGP_{-i,acg} = \alpha_{ca} + \beta \times F_i + \gamma \times SGP_{-i,ca} + \sum_{a=1}^{n_a} \sum_{p=1}^{n_{p,a}} \delta_{p,a} \times SC_{i,ap} + \epsilon_{iacg} \tag{4}$$

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 group 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 students of the same gender is higher for female students than for male students. Lastly, the allocation of students is constrained by the fact that students who attend more than one course in the same term cannot attend two classes at the same time. To take care of this, I control for a dummy for each course that each student takes during each academic year ($\sum_{a=1}^{n_a} \sum_{p=1}^{n_{p,a}} SC_{i,ap}$). I perform this test for all the courses where there are at least two classes and standard errors are clustered at the class level.

The results are presented in Table III.

E.2. Test 2: Class group allocation does not predict students' characteristics

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

$$y_{ig} = \sum_{g=1}^{n_g} \alpha_g \times G_{i,g} + \sum_{p=1}^{n_a} \gamma_p \times SC_{i,p} + \epsilon_{ig},\ \forall a, c \tag{5}$$

where the dependent variables are pre-determined individual characteristics - gender, ethnicity, age at enrollment - of student i enrolled in course c in year t, who has been allocated to class g. The explanatory variables are dummies for each class 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 ($G_{i,g} = \mathbb{1}(\textit{i's group}{=}g)$). The allocation of students is constrained by the fact that students who attend more than one course in the same term cannot attend two classes at the same time. To take care of this, I control for a dummy for every other course that students take during the academic year ($\sum_{p=1}^{n_a} SC_{i,p}$). I run one regression for each combination of course $c \times$ academic year a that is included in the sample of analysis.[70] The sample for each regression consists of all the students enrolled in course c in the same academic year a that didn't change class group during Michaelmas term. Furthermore, the sample is restricted to all the courses that have at least 2 class groups. For this test, Chinese, Indian, and other Asian students have been aggregated in a unique category called Asian to have a big enough sample to be able to perform meaningful tests.[71]

I test that class group dummies are jointly significantly different from zero, i.e. students are allocated to classes independently of their characteristics, conditional on scheduling constraints. According to Murdoch et al. (2008), the p-values of this test should be uniformly distributed with a mean of 0.5 if class allocation is random. In this case, courses have different sizes, the number of classes is not fixed, and classes belonging to the same course might be of different sizes. To check that also in this case the p-values converge to a uniform distribution with a mean of 0.5, I Perform a Monte-Carlo simulation where I randomly allocate to class groups the students enrolled in the first-year of degree, that didn't change class groups during the term. I perform this 1000 times, under the assumption that course size, number of seminars, and class size are equal to the observed one. Figure B.VI shows the result of the simulation. Indeed p-values converge to a uniform distribution with mean 0.5.

Table E.I shows the results of the tests performed. 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) - (3), (4) - (6), (7) - (9), and (10) - (12), show the results of Specification (5) with the female dummy, Asian dummy, White dummy, and age at entry as dependent variables, respectively.

Table E.I
Identification: Observed Sample P-values - Other Characteristics

	Clashes: Mandatory + Elective Courses											
	Female			Asian			White			Age at Entry		
	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)
Courses in sample	444	0.054	0.092	449	0.042	0.082	452	0.051	0.085	451	0.047	0.106
All courses	1415	0.049	0.090	1383	0.047	0.087	1406	0.041	0.082	1460	0.043	0.094

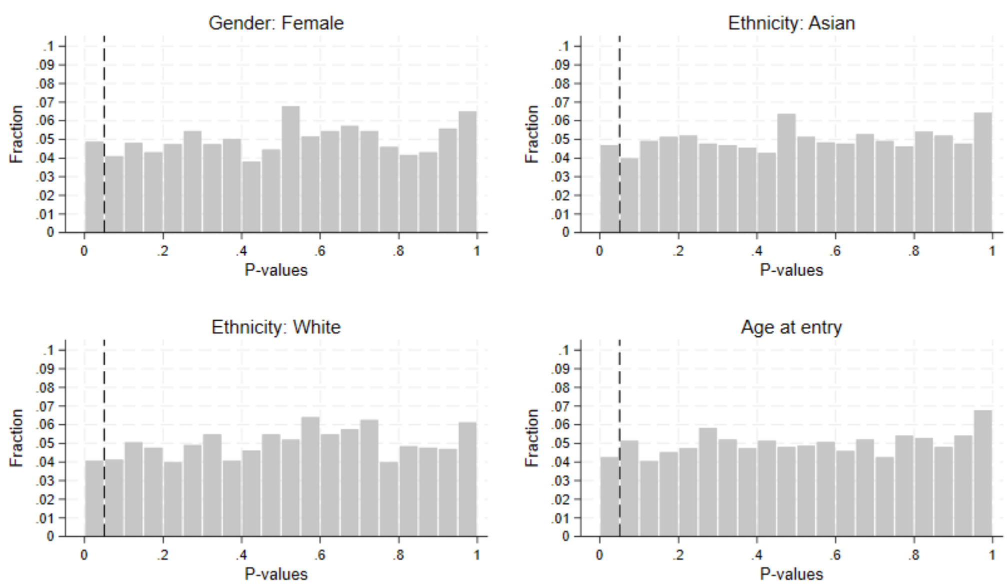
Notes: This table shows the results of the tests of joint significance of the class group dummies in Specification (5). The sample consists of the courses in analysis in the first row and all the undergraduate courses in the second row. Column "N." displays the number of performed regressions, column "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.

[70] The tests performed are 444 for gender, 449 and 452 when the Asian dummy and White dummy are the dependent variables respectively, and 451 for age at entry.

[71] 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 students are allocated randomly, class allocation might still predict students' characteristics. To illustrate this, let's 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.

The proportion of tests with a p-value smaller than 0.05 and 0.10 is consistent with random allocation of students into class groups. As a matter of fact, the fraction of p-values smaller or equal to 5% is roughly 5% and the fraction of p-values smaller or equal than 10% is roughly 10%. The higher is the number of tests performed, the more the p-values converge to a uniform distribution with mean 0.5, as it can be seen in the second row of Table E.I. A graphical illustration of the results of the test is presented in Figure VI for the courses in sample, and in Figure E.I for all undergraduate courses. The results confirm that conditional on course clashes, students are allocated to class groups independently from their individual characteristics.

Figure E.I
Identification: Observed Sample P-values, all courses



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 all undergraduate courses - Specification (5). The test includes all undergraduate courses with at least 2 class groups and all the students that attend courses attended by the students in the sample of analysis and that did not change classes.

E.3. Test 3: Balance checks

Table IV shows the results of the following two specifications:

$$SF_{acg} = \alpha \times T_i + \beta \times F_i + \sum_{a=1}^{n_a} \sum_{p=1}^{n_{p,a}} \delta_{p,a} \times SC_{i,ap} + \epsilon_{iacg} \tag{6}$$

where SF_{acg} is the share of females in the class group g, course c, and academic year a, T_i are individual traits such as ethnicity, previous school characteristics, age at entry, and qualification score at entry, and F_i is a dummy equal to one if the student is a woman. I control for a series of dummies equal to one for for every other course that students take during the academic year, and zero otherwise ($\sum_{a=1}^{n_a} \sum_{p=1}^{n_{p,a}} SC_{i,ap}$), as the allocation of students to classes is random conditional on scheduling constraints. Standard errors are clustered at the class group level.

$$\textit{SFP}_{-i,acg} = \alpha \times T_i + \beta \times \textit{SFP}_{-i,ca} + \sum_{a=1}^{n_a} \sum_{p=1}^{n_{p,a}} \delta_{p,a} \times SC_{i,ap} + \epsilon_{iacg} \tag{7}$$

Following the procedure suggested by Guryan et al. (2009), I also test whether the share of female peers $\textit{SFP}_{-i,acg}$ in the class group is not systematically correlated with individual characteristics T_i, once we control for the respective course leave out mean $\textit{SFP}_{-i,ca}$ and a series of dummies equal to one for all the courses students attend in the same year, and zero otherwise ($\sum_{a=1}^{n_a} \sum_{p=1}^{n_{p,a}} SC_{i,ap}$). Standard errors are clustered at the class level. I replicate this also using the share of the same gender peers rather than the share of female peers.

Only one of the estimated correlations appears to be significantly different from zero at 10% level of significance for the sample of analysis in Columns (1) and (2), and one coefficient appears to be significantly different from zero at a 5% level of significance in Column (3). 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 the results found when we aggregate all the columns as 4.8% of the test performed is significant at 5%, while 7.1% of the tests performed is significant at 10%.[72]

E.4. Test 4: Exogeneity of scheduling constraints

Following Lavy and Schlosser (2011), I perform a permutation test where I simulate 1000 random allocations of students to class groups. For each first-year course, I randomly allocate to class groups the students enrolled in the course that don't change class group during the first term, under the assumption that course size, number of seminars, and class group size are equal to the observed one.

For each course c in academic year a, I compute the difference between the share of female students in each class group and the share of females in the course ($\textit{Diff}_{gca}$) for both the observed sample allocation and the 1000 simulated random allocations:

$$\textit{Diff}_{gca} = \textit{Class share of female students}_{gca} - \textit{Course share of female students}_{ca}, \ \forall g, c, a$$

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:

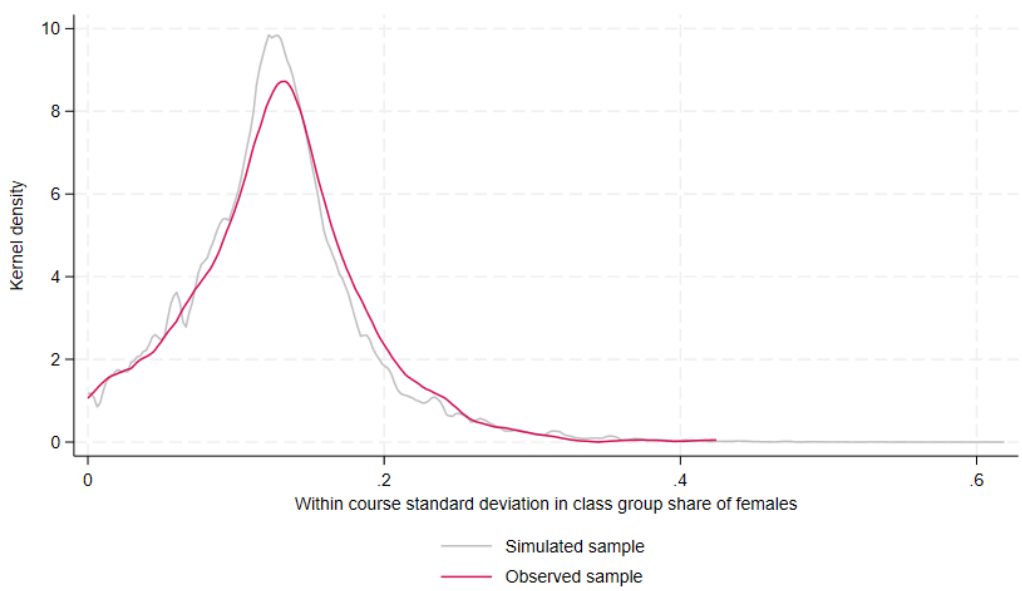
[72]Within column, this would be consistent only with Columns (1) and (2). However, the results within column should be taken with caution as the number of regressions is low, and the property described above holds only when the number of regressions is large. This is why looking at the three columns together is more indicative in this case.

$$Ho : Diff_o = Diff_s$$

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

In addition, these results are confirmed also when I perform a Two-sample Kolmogorov-Smirnov test to test the null hypothesis that the observed distribution of the course standard deviation in the share of females and the randomly simulated distribution are not significantly different. Figure E.II shows that the distribution of the observed course standard deviation in the proportion of females resembles the distribution that would result if the gender composition of each class group was randomly generated.

Figure E.II
Identification: Allocation Simulation - Within-course Std. Dev. in Class Groups Share of Females



Two-sample Kolmogorov-Smirnov test

Smaller group	D	P-value
0:	0.0154	0.774
1:	-0.0445	0.118
Combined K-S:	0.0445	0.235

Notes: This figure displays in grey the distribution of the 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. The sample includes all the students that attend courses attended by the students in the sample of analysis and that did not change classes. More details on the test can be found in Appendix E.4..

E.5. Department of choice rather than department of enrollment

The definition of selection used in the analysis is based on the department of initial choice. However, for a small percentage of students (1.66%), the department of enrollment at the end of the year does not coincide with the department of initial choice, as they change department over the year. Table E.II below, shows the results of Specification (2) and (3) when I replicate the analysis using the department of enrollment. We can see that the patterns are unchanged.

Table E.II
Robustness: Stereotypical Selection, Department of Enrollment at The End of The Year

	Course grade			
	Continuous	Top and Bottom 2	Top and Bottom 3	Top and Bottom 5
	(1)	(2)	(3)	(4)
Panel A: Interaction				
Share of students like me	-5.776***	-3.581***	-2.965***	-1.151
	(1.704)	(1.234)	(1.044)	(0.728)
Share of students like me × Neutral		3.795***	3.276***	1.032
		(1.303)	(1.136)	(0.957)
Share of students like me × Stereotypical Selection	12.078***	7.750***	6.149***	3.396***
	(3.169)	(1.558)	(1.316)	(0.947)
Panel B: Absolute Effect				
Share of students like me × Counter-stereotypical Selection		-3.581***	-2.965***	-1.151
		(1.234)	(1.044)	(0.728)
Share of students like me × Neutral		0.214	0.311	-0.120
		(0.423)	(0.453)	(0.622)
Share of students like me × Stereotypical Selection		4.169***	3.184***	2.245***
		(0.902)	(0.776)	(0.587)
Course fixed effects	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y
Observations	54554	54554	54554	54554

Notes: This table provides evidence of the results of Specification (2) in Column (1) and Specification (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 choice 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 in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

E.6. Same gender peers

We might be worried that class group gender composition is defined using the full mean rather than the leave-out mean (Angrist, 2014). Here I replicate the results defining class gender composition as the "share of peers like me", i.e. the share of same gender students among classmates excluding myself. We can see that this matters little as the estimated coefficients become slightly smaller in magnitude, but the results are unchanged. This is due to the fact that the identification strategy is exploiting a within-student variation and the independent variable is defined as share of same gender classmates rather than share of female classmates, hence it is constructed in the same way for men and women.

Table E.III
Robustness: Share of Same-gender Peers

	Course grade			
	Continuous	Top and Bottom 2	Top and Bottom 3	Top and Bottom 5
	(1)	(2)	(3)	(4)
Panel A: Full mean (Interaction)				
Share of students like me	-5.946***	-3.556***	-2.839***	-1.215*
	(1.685)	(1.218)	(1.043)	(0.725)
Share of students like me × Neutral		3.741***	3.072***	1.060
		(1.289)	(1.139)	(0.960)
Share of students like me × Stereotypical Selection	12.410***	7.851***	6.153***	3.543***
	(3.137)	(1.543)	(1.312)	(0.942)
Panel B: Leave-out mean (Interaction)				
Share of peers like me	-5.396***	-3.271***	-2.634***	-1.130*
	(1.572)	(1.137)	(0.970)	(0.674)
Share of peers like me × Neutral		3.432***	2.849***	1.011
		(1.202)	(1.057)	(0.890)
Share of peers like me × Stereotypical Selection	11.267***	7.251***	5.688***	3.246***
	(2.924)	(1.438)	(1.218)	(0.875)
Course fixed effects	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y
Observations	54554	54554	54554	54554

Notes: This table provides evidence of the results of Specification (2) in Column (1) and Specification (3) in Columns (2)-(4). In Panel A, class group composition is defined as the share of same gender classmates, while in Panel B as the share of same gender peers, except me. The outcome variable is course grades, with incomplete courses coded as 0. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix F. APPENDIX TO THE ANALYSIS ALONG ETHNIC LINES

To replicate the analysis along ethnic lines, I estimate the following specification:

$$y_{iacg} = \alpha_{ac} + \alpha_i + \sum_{e=1}^{3} E_{i,e} \times [\beta_{1,e} \times SLM_{iacg} + \beta_{2,e} \times SLM_{iacg} \times STS_i] + \epsilon_{iacg} \quad (8)$$

where students are divided into three groups $E_{i,e} = \mathbb{1}(\text{i's ethnic group} = e)$, where e is equal to Asian, White, or Other, which is a residual category that includes other ethnicity students and students who provided no information on their ethnicity. This categorization is motivated by the necessity of having large enough groups to perform the analysis. The *share of students like me* (STS_{iacg}) is the share of same-ethnicity classmates that student i experiences in class group 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. 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.[73]

$\beta_{2,e}$ 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 in line with stereotypes regarding ethnic-specific skills compared 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.

F.1. 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.

Firstly, 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 equivalent of Specification 4 for ethnicity.

Table F.IV 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.

[73]The results are robust to restricting the sample to only Asian and White students as it can be seen in Columns (3) and (4) of Table VIII.

Table F.IV
Identification: Effect of Ethnicity On The Share of Same-ethnicity Peers

	Share of same ethnicity peers		
	(1)	(2)	(3)
Panel A: 3 groups			
Asian	-0.002		
	(0.004)		
Other	-0.007		
	(0.004)		
Panel B: 6 groups			
Black		-0.005	
		(0.004)	
Chinese		0.005	
		(0.004)	
Missing		-0.002	
		(0.005)	
Other		-0.003	
		(0.004)	
White		0.003	
		(0.004)	
Panel C: 7 groups			
Indian			-0.001
			(0.004)
Chinese			0.004
			(0.004)
Black			-0.004
			(0.005)
Other Asian			-0.002
			(0.004)
Other			-0.004
			(0.005)
Missing			0.000
			(0.005)
Course leave-out mean	0.990***	0.992***	1.003***
	(0.011)	(0.012)	(0.012)
N	54554	54554	54554
Course fixed effects	Y	Y	Y
Parallel course dummies	Y	Y	Y

Notes: This table displays the result of a test that students belonging to different ethnic groups are not assigned to class groups with a different share of same-ethnicity peers (y_{itcg}). In each regression, I control for the leave-out mean of the share of same-ethnicity peers at the course level. The definition becomes more disaggregated across the different columns, following the aggregation of ethic groups in the different panels. Parallel course dummies are a series of dummies that are equal to one for all the courses students attend in the same year, and zero otherwise. Students that changed class group are excluded since their initial allocation cannot be observed. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The second test I perform is discussed in Section E.2.. 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 VI and Table E.I show 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 (5.1%) display a p-value smaller than 0.05 and 8.5% display a p-value smaller than 0.10. Regarding the Asian dummy, 4.2% of tests display a p-value smaller than 0.05, and 8.2% of tests display a p-value smaller than 0.10.

Lastly, I produce an array of "balance tests" to study whether the variation in the share of different ethnicity 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. Since the variation used in the analysis along ethnic lines is the variation in the share of White, Asian, and other students, this is the categorization I am going to use to perform this test.

It seems that the variation in the share of same ethnicity classmates is related to other individual characteristics,in particular, the characteristic of the previous school attended (Table F.V). However, controlling for program of enrollment fixed effects reduces the problem, as it can be seen in Table F.VI. The percentage of coefficients significant at 5% level or more are 2.2% and the share of coefficients significant at the 10% level or more are 6.67%. This indicates that once we control for clashes and the program of enrollment, which we flexibly do in the main empirical strategies as we control for individual fixed effects, the share of different ethnicity peers in the classroom is unrelated with peers' individual characteristics.

Table F.V
Identification: Effect of Other Characteristics On The Share of Same-ethnicity Peers

Independent Variables:	Share of White (1)	Share of Asian (2)	Share of White peers (3)	Share of Asian peers (4)	Share of same ethnicity peers (5)
Female	0.001	-0.000	0.001	0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Independent School	-0.002	0.001	-0.002	0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Grammar School	-0.004	0.008**	-0.004	0.008**	-0.009**
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Academy or Comprehensive School	0.003***	-0.001	0.004***	-0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
College or Further Education	0.001	0.001	0.001	0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Other	0.017*	-0.010	0.018*	-0.010	0.005
	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)
Not applicable	-0.001	-0.002	-0.002	-0.002	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age at entry	-0.001*	0.000	-0.001**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Qualification Score at entry	-0.000*	0.000	-0.000*	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ethnicity dummies	Y	Y	X	X	X
Course Fixed Effects	Y	Y	Y	Y	Y
Parallel Course Dummies	Y	Y	Y	Y	Y
Leave-out class mean	X	X	Y	Y	Y

Notes: This table displays the results of balance checks. 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. Each row corresponds to a separate regression where the dependent variable is the share of Asian or White classmates (Columns (1) and (2)), the share of Asian and White peers (Columns (3) and (4)), and the share of same ethnicity peers (Column (5)), while the independent variable is the individual characteristic, controlling for course fixed effects, a dummy for each course the student attends during the academic year, and ethnic group dummies. The number of observations is 54554 for all the tests with the exception of those that have the "Qualification Score at entry" as independent variable. In this case, the number of observations is 37625. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table F.VI
Identification: Effect of Other Characteristics On The Share of Same-ethnicity Peers (B)

Independent Variables:	Share of White (1)	Share of Asian (2)	Share of White peers (3)	Share of Asian peers (4)	Share of same ethnicity peers (5)
Female	0.001	-0.000	0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Independent School	-0.002	0.001	-0.002	0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Grammar School	-0.005	0.006*	-0.005	0.007*	-0.010***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Academy or Comprehensive School	0.002	-0.001	0.002	-0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
College or Further Education	-0.001	0.001	-0.001	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Other	0.009	-0.003	0.010	-0.003	0.009
	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)
Not applicable	0.001	-0.002	0.001	-0.002	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age at entry	-0.000	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Qualification Score at entry	-0.000	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ethnicity dummies	Y	Y	X	X	X
Course Fixed Effects	Y	Y	Y	Y	Y
Parallel Course Dummies	Y	Y	Y	Y	Y
Leave-out class mean	X	X	Y	Y	Y
Program of Enrollment Fixed Effects	Y	Y	Y	Y	Y

Notes: This table displays the results of balance checks. 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. Each row corresponds to a separate regression where the dependent variable is the share of Asian or White classmates (Columns (1) and (2)), the share of Asian and White peers (Columns (3) and (4)), and the share of same ethnicity peers (Column (5)), while the independent 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 program of enrollment fixed effects. The number of observations is 54554 for all the tests with the exception of those that have the "Qualification Score at entry" as independent variable. In this case, the number of observations is 37625. Standard errors in parenthesis are clustered at the class group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix G. SURVEY

A complementary online survey is administered to students in academic year 2020/2021 (June-July 2021). The survey is designed using Qualtrics and is sent to all the students enrolled in undergraduate programs at the University through their institutional email address via the university system. Participants agree to take part in the survey and sign a consent form in which it is explained that the survey is part of a research project aimed at understanding the determinants of academic and labour market performance. The students are informed that if they agree to participate, they will be asked a few questions regarding demographics, their experience as a student, and their beliefs and attitudes. The time to complete the survey is around 10 minutes and students are entered in a draft for three Amazon vouchers.

The survey includes 4 sections. In the first section, students are asked some questions regarding their experience at the university, disregarding as much as they could the last year of remote teaching. Following Bursztyn et al. (2019), students are asked questions regarding the extent to which they think image is important, and what elements contribute to being popular. Moreover, following Bursztyn et al. (2017), the survey includes questions regarding how much students feel comfortable participating in class.

In the second section, students perform a double-target Implicit Association Test (Greenwald et al., 1998). I follow 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 follow the approached used in the Gender-Science IAT designed by Project Implicit [74]. I use 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 (Math, Physics, Engineering, Statistics, Chemistry, Science, Biology) and Humanistic disciplines (Literature, History, Arts, Languages, Humanities, Anthropology, Sociology).

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.

[74] implicit.harvard.edu. More details on the organization can be found below.

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 is incorporated in Qualtrics using the ad-hoc approach designed by Carpenter et al. (2021).[75]

The third section of the survey consists in questions regarding explicit associations. Following Delfino (2021) and Carlana (2019), I ask 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 contains questions regarding demographic information and the students' social network.

G.1. Gender-Scientific IAT - Evidence from Harvard Project Implicit

We might be concerned that the patterns observed are specific to the sample of students that took part in the survey.

Although the respondents might be a selected sample of students not representative of the overall population of undergraduate students at the university, this selection bias is problematic for the interpretation of the results only if it is systematically correlated with choices of major and implicit attitudes. In particular, for the selection bias to lead to an overestimation of the effects, there should be a positive systematic correlation between the probability of participating in the survey and implicit stereotypical associations among students who made a stereotypical choice of major, and a negative correlation among students who made a counter-stereotypical choice of major. In other words, for the selection bias to lead to an overestimation of the effects, people who uphold stronger stereotypical associations who selected into stereotypical fields (e.g. men in math) should be significantly more inclined to take the test than people of the same gender who uphold weaker stereotypical associations and who chose an occupation/education in the same fields. At the same time, it should be the case that the sample of respondents is significantly skewed towards people who uphold weaker stereotypical associations who selected into counter-stereotypical fields (e.g. women in math) compared to people of the same gender who chose an occupation/education in the same fields. This seems unlikely.

Moreover, the very same pattern can be observed when analyzing the data of a gender-science implicit association test carried out by Harvard University Project Implicit[76], which covers a significantly wider population.[77] The results can be found in Figure G.I below. The analysed sample pertains the

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

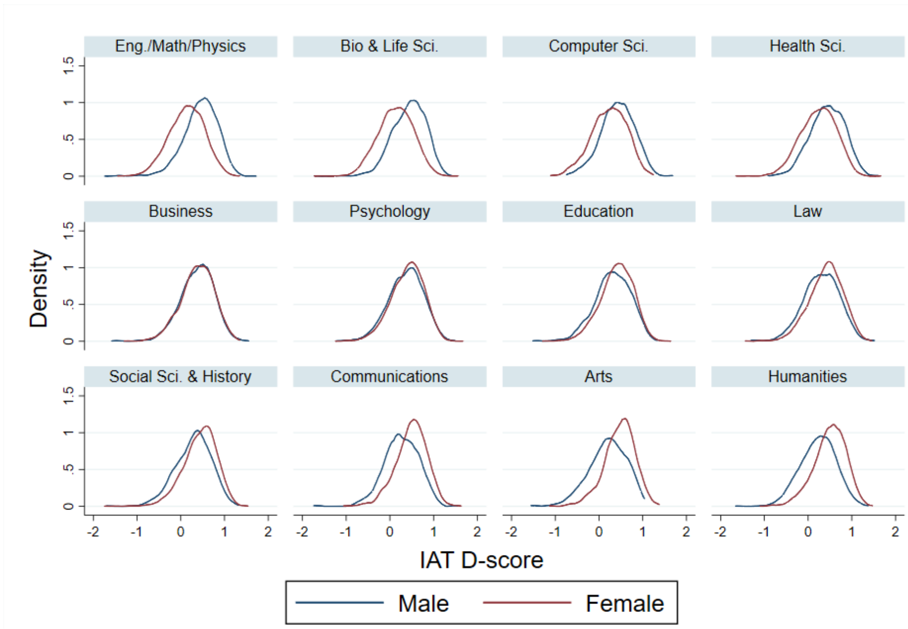
[76] Information can be found at this website https://implicit.harvard.edu/implicit/takeatest.html.

[77] Project Implicit is a non-profit organization and international collaborative of researchers who are interested in implicit

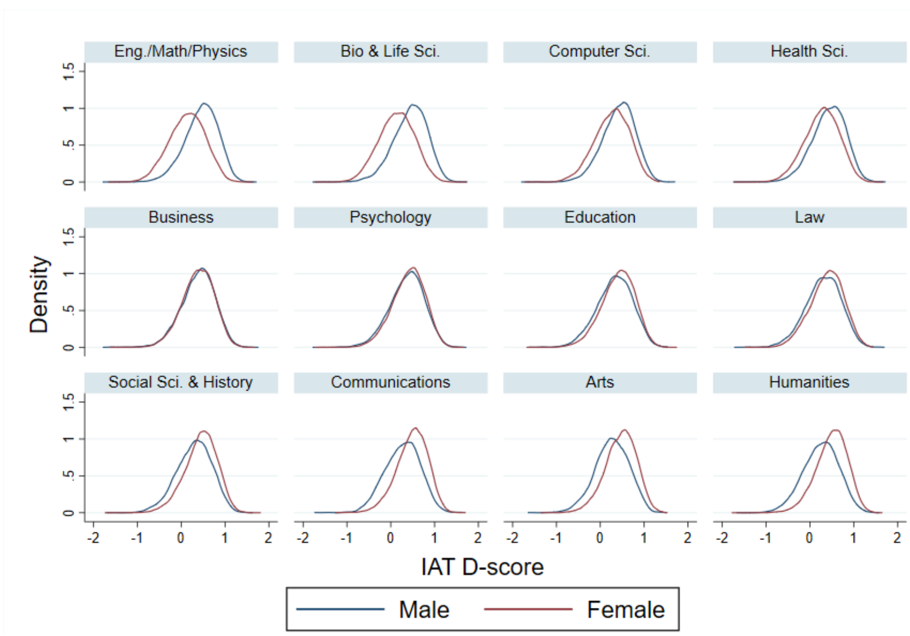
results of the Gender-Science IAT collected from 2003 to 2020. The sample includes (a) respondents who are currently students - 63.648 observations, and (b) all the respondents that declare their university major - 421.641 respondents. Whether we look at respondents who are current students or all the respondents that declare their university major, we can observe that men uphold stronger stereotypical associations compared to women when they attended a science-math related major, while weaker stereotypical associations compared to women when they attended a humanities-related major. This confirms that the observed patterns are not specific to LSE students.

social cognition. It was founded in 1998 by three scientists – Dr. Tony Greenwald (University of Washington), Dr. Mahzarin Banaji (Harvard University), and Dr. Brian Nosek (University of Virginia). The mission of the organization is to educate the public about bias and to provide a “virtual laboratory” for collecting data on the internet. Everybody is allowed to take a test and the results are collected by the organization and made publicly available for researchers (https://osf.io/y9hiq/).

Figure G.I
Project Implicit: Gender-Science IAT



(a) Currently studying



(b) All respondents

Notes: Results of an implicit association test (Greenwald et al. 1998) to elicit students' associations between female-humanistic and male-scientific carried out by Harvard University Project Implicit. 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. The analysed sample pertains the results of the Gender-Science IAT collected from 2003 to 2020 (https://osf.io/y9hiq/). The sample includes (a) respondents who are currently students - 63.648 observations, and (b) all the respondents that declare their university major - 421.641 respondents.