Title

Gender of Siblings and Choice of College Major

Journal

CESifo Economic Studies, 61(1)

ISSN

1610-241X

Authors

Anelli, Massimo

Peri, Giovanni

Publication Date

2015-03-01

DOI

10.1093/cesifo/ifu028

Peer reviewed

Gender of Siblings and Choice of College Major

Massimo Anelli (University of California, Davis)

Giovanni Peri (University of California, Davis, CESifo)*

May 23, 2014

Abstract

In this study we analyze whether the gender composition of siblings within a family affects the choice of College Major. The question is whether a family environment that is more gender-homogeneous encourages academic choices that are less gender stereotyped. We use the last name and the exact family address contained in a unique dataset covering 30,000 Italian students graduated from high school between 1985 and 2005 to identify siblings. We follow the academic career of these individuals from high school to college graduation. We find that mixed gender siblings within a family tend to choose college majors following a stereotypical gender specialization. Namely, males have higher probability of choosing "male dominated" majors such as Engineering and women higher probability of choosing "female dominated" majors such as Humanities. Same-gender siblings, on the other hand, have higher probability of making non-gender stereotyped choices. This college major choice is not driven by the choice of high school academic curriculum, which appears to be mainly function of geographical proximity to schools.

Key Words: Gender, Choice of College Major, Family Structure

JEL Codes: I21, J12, J16, Z18.

*Addresses: Department of Economics, UC Davis, One Shields Avenue, Davis, CA 95616, USA. manelli@ucdavis.edu; gperi@ucdavis.edu.

1 Introduction

Gender bias in the choice of college major is a very important determinant of the gender wage gap in many OECD countries (Flabbi, 2011). In particular women choose majors as Engineering, Physical Sciences and Economics-Business at a much lower rate than similar men. As those majors are associated with high earnings, women's choice of major account for an important part of the earning wage gap with men. This difference is pervasive to most rich countries and has persisted over time. This time-persistence is remarkable as women have experienced improvements in their academic performance during the last forty years and they have overtaken men in college enrollment and graduation rates. This bias against choosing math-science intensive curricula appears, for women, already in high school. For those who enroll into college the bias becomes even more evident. In the U.S., for instance, only 18% of recent graduates in Engineering, but 64% of graduates in the Humanities, were women [1]. This difference in major choice is associated with large discrepancies in expected wages for female and male college graduates. Altonji et al. (2011) find that after adjusting for basic demographics, potential labor market experience, and graduate education, the gap in wages rates between men who graduated with a major in electrical engineering and men who graduated in a major in general education was 0.561 logarithmic points, almost 75% of their wage. This is nearly as large as the 0.577 logarithmic points difference between college graduates and high school graduates. Furthermore, they find that the standard deviation of the return to various majors is 0.177 for men, about double the typical estimate of the value of a year of school. Choosing different majors produces differences in wage outcomes as large as achieving substantially different levels of schooling. This evidence suggests that studying and understanding the determinants of the gender bias in the choice of major is an important component to implement education policies that can effectively reduce the gender wage gap.

In this paper we use a newly collected dataset covering 30,000 Italian students graduated from high school between 1985 and 2005 to analyze whether the gender composition of the set of siblings within a family affects the choice of academic track in High School and then the choice of College

[1] See table 301 in the U.S. Digest of Educational Statistics (2001) at Digest of Education Statistics 2011 website.

Major. In Italy, as in many European countries, students choose different academic tracks already in high school. The individuals included in our study have all chosen to attend college-preparatory high schools, but they may have chosen one of two different types: one focusing in Classic Studies (Liceo Classico) or another more grounded in Scientific Studies (liceo Scientifico).

In a recent paper using these data (Anelli and Peri, 2013), we found that classmates in high school affect the choice of major of male students. We found evidence that a larger share of same-sex high-school classmates increased the probability of choosing majors associated to high earning jobs (Economics/Business, Medicine, Engineering) for males. Complementarily to that analysis, in this paper we exploit the unique feature of the same data set, presented in Anelli and Peri (2013) to explore how the gender composition of siblings within a family affects the choice of high school type and college major. If a more gender-uniform environment reduces the tendency to form gender-stereotyped identities and choices, then families with one-gender siblings may have a similar effect as same-gender classes in stimulating individuals to make less "typical" choices. We analyze the effect of siblings' gender on the choice of high school academic curriculum and college major. We also analyze whether the order of birth within the family has an impact on those choices.

Our main findings are the following two. First, mixed-gender siblings groups encourage a more gender-stereotypical choice of major. Females with at least one brother are less likely to choose (typically "male") High Earning College Majors (HEM), while males with at least one sister are significantly more likely to choose HEM. Second, the choice of high school track, instead, does not reveal a similar effect of gender of siblings, but it seems to be driven by a common family choice (often of the closest high school) in which the younger siblings tend to follow the high school choice of the older sibling. These different results are consistent with the idea that parents have a very strong influence on the high school choice of all children (made when they are 12 years old), while students' preference and their gender identity are more relevant in the choice of college major (made at 18 years old).

The rest of the paper is organized as follows. Section 2 describes the data and shows its summary statistics. Section 3 presents and discusses our empirical approach. Section 4 shows the results of our estimation. Section 5 concludes the paper.

2 Descriptive Statistics

Our data cover around 30,000 individuals who graduated from College-Preparatory High Schools ("Licei" in Italian), hereafter referred to as CPHS, between 1985 and 2005 in the city of Milan, Italy. Data have been collected from hard copy records of documents from 15 high schools in Milan that represent the vast majority of all CPHS in the city[2]. They contain information about the type of high school attended, the class and year of graduation, the high school exit exam score, the identity of classmates during the last year of high school and the exact location within Milan of the home where they lived during high school. We have then linked these data to the student's records from all Universities in Milan. About 77.6% of those high school graduates enrolled in one of the University within Milan, the city that offers the widest and the highest quality choice of tertiary education in Italy. We are able, therefore, to track the choice of major for those high school graduates who enrolled in a university in Milan[3]. Relevant characteristic of the Italian educational CPHS system is that each institution is either a Scientific High School ("Liceo Scientifico") offering to students a more math and science intensive curriculum or a Classical High School ("Liceo Classico") which offers a classic-humanities intensive course of studies. No institution in our sample offers both academic curricula. For what concerns higher education, Italian perspective college students apply to a specific college major in a specific University. Therefore, the choice of major takes place during the last year of High School, when student apply to college. Since the cost of switching major later on in their academic career is substantial (it entails dropping out of the current academic program and re-applying to a different program and essentially re-starting college), the major enrolled at the time of college application is also the major of graduation for the large majority of college students.

In order to identify siblings in our dataset, we use the last name and the exact family address contained in our dataset (street, and apartment number). To maximize the precision of address matching we have transformed each address into geographic coordinates using Google geo-coding

[2]We focus on CPHS graduates because they are a very homogenous group of students for which selection into university is not an issue: in our sample 84% of the students graduating from a CPHS has completed a college degree and 13% has enrolled university at least once without completing any degree.

[3]19% of the unmatched students have attended college abroad. 21% did not go to college. The remaining 60% either attended another minor tertiary education institution in Milan for which we do not have data or universities in different cities in Italy.

system, rather than relying on the name of the "street" that sometimes may have small misspelling or differences[4]. We thus define as siblings, students with the same last name and living at the identical geographic coordinates (latitude and longitude). Out of 29,370 individuals we are able to identify 5443 siblings in 2627 families. Every identified family with more than one child in our dataset, has on average 2.11 siblings and the maximum number of siblings in one family is 5. Out of these 2627 identified families, and importantly for our identification, 663 have only female siblings, 710 only male siblings and the remaining 1373 families have mixed gender siblings. For the subsample of surveyed individuals we know the exact number of siblings that we can compare with those we have identified. Our address-based match gives the exact number of siblings for 35% individuals, while underestimates of 1 sibling 49% of the cases. Out of all individual for which we do not observe any other sibling graduating from CPHS, 32% are actual single-children, while the rest have at least one sibling who did not graduate from one of the CPHS under analysis.

The characteristics of the dataset collected allow us to identify only siblings in each family who graduated from a CPHS. We are unable to identify other siblings who were attending other types of High School (non college-preparatory). Hence, when interpreting our results we should keep in mind that we are estimating the effect of siblings gender composition on college major choice, conditional on those siblings attending CPHS. This does not affect the internal validity of our work, however it should be taken into considerations when interpreting results. We are restricting our analysis to those families who have at least two children attending a CPHS. This essentially excludes families with a single child and families that send only one child to a CPHS. Let us emphasize that individuals going to other types of high schools are far less likely to go to college afterwards. Our universe of CPHS graduate siblings is thus the more relevant group to analyze college major choice. To give a measure of the specific characteristics of the families we are focusing on, we exploit the precise family structure information collected for a sub-sample of individuals. For a stratified 10% random sub-sample of the initial 30,000 individuals we collected more detailed information from telephone interviews conducted in June 2011 by the professional company "Carlo Erminero & Co.". The additional information covers several variables regarding the family background, parental

[4]We thank Mehrez Ben Salem e Zied Bouyahia for crucial technical assistance in geo-coding addresses.

income, job and education, current employment and current family situation of the individual. For this representative sub-sample we are therefore able to identify the exact number of siblings for each individual and, among those, the number of siblings who graduated from a CPHS. From this sub-sample we learn that 25% of families in our sample have a single child, 53% have two children and the remaining 22% have 3 or more children. Of the families with two or more children the representative sub-sample shows that 16% of them had all kids attending a CPHS, 22% of them had at least two children attending a CPHS, while in 84% of them at least one child attended some other type of high school. In Table 1 we show descriptive statistics separately for families of two or more children in which only one child attended a CPHS and for those in which at least two children attended a CPHS. These statistics identify the differences between families in our analyzed siblings sample and other families not included in the sibling sample. Mean difference and the relative t-statistic are reported in the last two columns of Table 1. Families with only one child attending a CPHS are characterized by lower household income (especially lower income for their fathers) and substantial lower probability that parents went to college (almost 20 percentage points less). They also live in neighborhoods with lower average house prices. These numbers show that the group of families for which we observe two or more children attending a CPHS are at the high end of the wealth and education distribution. So our sample of reference for this study includes families with high commitment to the education of their children and to invest time and money in their education. The college major choice of individuals in these families is therefore a very interesting outcome: it is certainly the result of family environment and the role played by parents and siblings is relevant, as those families are likely to be academic aware and involved. Moreover children in these families are on average less likely to make stereotypical major choices. We have computed the average gender bias in college major choice (calculated as the ratio of the percentage of men enrolling High Earning Majors over the ratio of the percentage of women) for the two types of family. Families with only one child at CPHS are characterized by a higher gender bias in the probability of enrolling a HE Major[5]. This goes in the expected direction of lower stereotypical choices for wealthier and more educated families. We thus interpret our results on families with 2

[5]The percentage of men choosing HEM in families with only one child attending CPHS is 2.11 times higher of the percentage of women, while this ratio is 1.84 for families with two or more children in CPHS

or more children at CPHS as a lower bound effect for stereotypical choice of major.

3 Empirical Methodology

We first analyze the effect of gender composition of siblings in a family on the choice of college major and then on the choice of high school track. In order to focus on choice categories that exhibit the largest gender differences, we simplify the college major choice in two outcomes only, captured by dichotomous dummies. The first outcome is the probability of enrolling into one of the three majors associated with the highest expected earnings and also with a very strong male-bias in aggregate (Economics & Business, Engineering and Medicine). This variable is coded as equal to one $P(HEM = 1)$ when a student chooses a high earning major and it is equal to 0 otherwise[6]. The second outcome is the probability of choosing a Scientific high school and it is coded as a one if the individual chooses it, $P(SHS = 1)$ and a 0 if he/she does not.

The main empirical model is, therefore, specified as follows:

$$y_{i,h} = \alpha + \beta \left(SibChar\right)_{i,h} + \gamma \left(X_{i,h}\right) + \varepsilon \tag{1}$$

where $y_{i,h}$ is one of the two dycothomous outcomes (relative to high school or college choice) described above for individual i living in family h. We regress these outcomes, separately for boys and girls, on an indicator describing a specific characteristic of his/her siblings $SibChar_{i,h}$, while also controlling for individual and family characteristics $\gamma X_{i,h}$. The main variable that we consider is the gender of siblings in the family by looking at the effect of having *at least one sibling of opposite gender* attending a College-Preparatory High School (CPHS)[7]. The interpretation of the coefficient β is straightforward. It represents the association between the presence of other-sex siblings in the family and the probability of making a certain academic choice. While the gender of one's siblings

[6] In Anelli, Peri (2013) we show, using the same data, how positive labor market outcomes are more significantly associated to the choice of "high earning majors" (HEM) than to the choice of other major categorizations, such as STEM majors.

[7] It is important to recall the we are restricting the analysis to families in which at least 2 siblings graduated from one of the CPHS under analysis as described in section 2. We are thus estimating the effect of siblings gender composition on college major choice, conditional on those siblings attending CPHS

is certainly exogenous to individual's academic choice and hence a causal interpretation of the coefficient β is legitimate, an intriguing question is what is the channel through which the gender of a sibling affect one's academic choice. One may think, on one hand, that parents with children of different gender, may direct their "traditional" view of school choice on them. Hence mothers will encourage daughters to choose classical studies and humanities majors (that they are more likely to have attended) and father may channel their preferences for engineering, science and business on males. When children are all the same gender, such "specialization" may be less clear. On the other hand children may pattern their preferences in comparison with those of their siblings and observing their choices. Again female may differentiate themselves from male brother choices by choosing a humanities, classical study major, and similarly for men. Problems in our identification strategy might arise if parents' preferences about the gender of their children were to affect fertility decisions (e.g. stopping rule) and at the same time these preferences are correlated with gender stereotypes within the family. Although the families under analysis (wealthy, highly educated) are unlikely to have strong preferences over the gender of their children, we test the hypothesis that families with same gender siblings (all female or all male) and those with mixed-gender siblings are systematically different. In Table 2 a simple test for difference in means neatly show that families with same gender children are not statistically different from those with mixed-gender children for several observable characteristics (e.g. family income, mother's and father's education, etc.). If preferences about gender were to affect fertility decisions for this sample, we should in particular expect a systematic difference in the number of children per family. However, also this difference is not statistically significant.

In order to understand whether it is simply the gender composition of a family to determine the "major specialization" of siblings or whether observing and reacting to the other sibling's choice is important, we also estimate a second specification. In this, we look at how the gender of a sibling interacts with the order of birth. We analyze the effect of gender on the choice of later-born siblings conditional on the gender of the first-born sibling. Namely we estimate the following specification,

again separately for men and women

$$y_{i,o>1,h} = \alpha + \beta_1 \left(SibChar_{i,o=1,h}\right) + \gamma \left(X_{i,o>1,h}\right) + \varepsilon \tag{2}$$

where $y_{i,o>1,h}$ is the outcome/choice of student i, who is the second- or higher-order born sibling (of order $o > 1$) in family h. The term $(SibChar_{i,o=1,h})$ is a dummy equal to one if the gender of the first-born sibling ($o = 1$) is different from the gender of individual i. The term $(X_{i,o>1,h})$ captures the individual and family characteristics for the second or higher order born sibling.

These specifications will allow us to answer the two following questions. (i) Does having siblings of the opposite sex increase the probability of an individual choosing more "gender-stereotyped" high school track and college major? In particular for women such a hypothesis would be associated with an estimate of $\beta < 0$, because the presence of brothers would decrease the probability of choosing a HEM. To the contrary for men such hypothesis will correspond to $\beta > 0$. (ii) Is this gender-stereotyped choice in the presence of opposite sex siblings stronger for second and higher-order children? In this case $\beta_1 > \beta$ for males and $\beta_1 < \beta$ for females. This would imply that the gender identity is particularly patterned by an individual in comparison with an older sibling of opposite gender.

For the choice of major outcome we test the sibling effect both for the unconditional probability of attending one of the HEM majors $P(HEM = 1)$ and for the probability conditional on the high school track choice $P(HEM = 1|SHS)$. Comparing the estimates for these two outcomes allows us to test if the gender effect of siblings is stronger on the choice of high school, a decision largely affected by parents as it is made when children are 12/13 years old, or if it is stronger on the choice of college major, even conditionally on high school. This second choice is likely to be more of a choice of the individual as it is made when he/she is 18 years of age.

4 Results

We first look at the effect of siblings' gender directly on the probability of attending a High Earning Majors $P(HEM = 1)$. In a second set of results we focus on the choice of high school track $P(SHS = 1)$. Finally, we present results for the probability of attending one of the HEM majors conditional on the high school track choice $P(HEM = 1|SHS)$.

In Table 3 columns 1 and 2 we analyze the entire sample of CPHS graduates who enrolled in college. For these two columns we look at the effect of having at least another sibling observed attending a CPHS, separately for men and women and independently of their siblings' gender. We control for a family wealth proxy based on the relative house value of the neighborhood where the students lived at the time of high school attendance and for a dummy equal to one if the individual is a twin. Results show that having at least one sibling attending a CPHS increases the probability of attending a High Earning Major (Economics & Business, Engineering and Medicine) by almost 5 percentage points for women on a baseline of 20.8% and by 6.6 percentage points on a baseline of 45.6% for men. These effects should be interpreted as a family effect. Having at least one sibling attending a CPHS can be interpreted as belonging to a family that sent at least two or all siblings to a CPHS. We are thus comparing families with stronger preferences for investing in their kids' human capital to families that have only one kid at CPHS. As showed in Table 1 these latter families are characterized by lower household income, substantially lower probability that parents went to college and they live in neighborhoods with lower average house prices. Students belonging to these families may be less likely to go to competitive (and demanding) majors. It is also important to notice that in our sample, 25% of the families have a single child. Hence among those students for which we do not observe another sibling, some are single children. However for the sub-sample that was randomly selected for a phone survey in 2011 we are able to exactly identify the number of children in the family. In columns 3 and 4 of Table 3 we replicate specification of column 1 and 2 after restricting the estimation to this surveyed sub-sample and after excluding families with a single child. Effects are still large in magnitude and close to those of columns 1 and 2, although the effect is now larger for women (7.4 percentage points increase in the probability of attending a

HEM). Overall, in families with more than one child graduating from CPHS, there is thus a higher probability of these children to choose a HEM. This is probably the result of selection of these families among those committed to higher education and possibly also with more economic means.

In columns 5 and 6 we focus on the main question of this paper. We restrict our analysis to families for which we observe at least two siblings attending a CPHS (i.e. the same individuals that had at least 1 sibling in CPHS in columns 1-4) and we include a dummy equal to one if the student has at least a sibling of opposite gender attending a CPHS. We run specification 1 separately for women and men and we control for the order of birth. Results show that male siblings graduating from a CPHS who have at least a sister also graduating from a CPHS are 5 percentage points more likely to enroll in a High Earning Major with respect to similar male siblings who have only brothers graduating from a CPHS. The effect is symmetric for female siblings, although weaker and not significant. They are 2.5 percentage points less likely to enroll in a HEM if they have at least a brother at CPHS with respect to female siblings with only sisters. These effects thus show stronger gender-stereotypical specialization in the choice of major in mix-gender sibling groups, especially for males.

Males who grew up in families with sisters seem to reinforce their "male role" in their choice of major. Females growing up in families with brothers, also reinforce, but less significantly, their "female role" in the choice of major. This gender specialization appears to enhance stereotypical major choices. This result is in line with experimental evidence (Booth and Nolen (2011) and Booth et al. (2012)) according to which mixed-gender environments tend to reinforce stereotypes. One possible mechanism through which this effect takes place is the competition between mixed-gender siblings. As found by relevant experimental research (Gneezy et al., 2003; Niederle and Vesterlund, 2007, 2010) women tend to under-performance with respect to their potentials when competing with men. Hence sisters comparing themselves with brothers may, over time, learn to reduce their performance in what are considered as "typically male" activities. However, parents might also play a crucial role in directing their children towards this gender specialization. In the presence of mixed-gender offspring they themselves (father and mother) may assume more stereotyped roles, with mothers advising daughters and father advising sons. To the contrary, in families with one-

gender siblings they might be more neutral and more symmetrically involved when it comes to influence their kids' major choice. In columns 7 and 8 we test wether the order of birth plays an important role in determining this gender specialization effect on choice of major. With respect to columns 5 and 6, we restrict the estimation to second or higher order siblings. In this case, similarly to what has been done in column 5 and 6, we include a dummy equal to one if the second or higher order born has a first order born sibling of opposite gender. We also control for the age difference with first born sibling. According to these estimates, second order born females with an elder brother are not less likely to enrol in a HEM than those with an elder sister. The difference between this estimate and the one of column 5 (-2.5 percentage points) is, however, not very large and not significant[8]. Column 8 shows that second order born males are 6.6 percentage points more likely to enroll in a HEM if their eldest sibling is female. This result is quite significant and slightly larger than the effect estimated in column 6. Overall, therefore, we find that the gender stereotypical specialization of male siblings in their choice of major is particularly strong if they grow up in a mixed-gender sibling group with at least one sister going to high school. Female siblings also have a symmetric tendency to more gender stereotyped major choices in mixed siblings groups, but the effect is not as strong nor as significant.

In Table 4 we step back in the academic decision process of students and replicate the same specifications of Table 3 on the probability $P(SHS = 1)$ of attending a Scientific High School. The first two columns cover again the entire sample of students graduating from a CPHS in Milan between 1985 and 2005. Results show that having at least one sibling attending a CPHS increases the probability of attending a SHS (scientific high school) by 4.2 percentage points for women on a baseline of 43.3% while the effect for men is zero. It appears that families with stronger preferences for investing in the education of their children are more likely to send their daughters to a SHS while they make similar choices to all other families for their sons. In columns 3 and 4 we restrict the estimation to the sub-sample for which we know the full family composition and we exclude single child families. Estimates are identical to those of the first two columns, although the coefficient for

[8]In a symmetric regression of first order born siblings' probability of enrolling in a HEM conditional on gender of later born that we do not report here, we find that first order born women with second or higher order brothers are 5.2 percentage points less likely to choose a HEM

women is now not significant (as a consequence of a smaller sample size).

Columns 5 and 6 are the focus of our analysis. We restrict again the sample to families for which we observe at least two siblings attending a CPHS. Interestingly, we find estimates of opposite sign relative to those for the probability of enrolling a HEM (see Table 3 columns 5 and 6) for both men and women. In this case female students having at least one brother are 6.5 percentage points *more* likely to attend a SHS , while men with at least one sister have a symmetric negative not significant effect (-3.4). Recall that the choice of high-school, takes place when students are 12/13 years of age and therefore the role of the parents may be much stronger than in the choice of college major. Results show that for what concerns the choice of high school and differently from results on the major choice, mix gender siblings reduces (especially for women) gender stereotypical choices. The results in columns 7 and 8 are enlightening in terms of the mechanism driving these effects. They show that the probability of attending a SHS is just a function of first born sibling's (random) gender. If the first born sibling is male, his unconditional probability to go to a SHS is higher[9] and no matter the gender of later born siblings, families seem to prefer sending them to the same school of the first born children, increasing as a consequence the probability that a later daughter will attend a SHS. Symmetrically if the first born sibling is female, her unconditional probability to go to a SHS is lower and consequently their later born siblings will be less likely to attend a SHS, no matter their gender. This pattern is very consistent with families choosing the high school of their first child according to a stereotypical pattern of boys in SHS and girls in classic high schools. Then, probably because they get to know the school, or because of convenience, they encourage their younger kids to attend the same school. Also the last row of Table 3 shows that geographical proximity of the high school is a crucial determinant of the probability of attending. If a SHS is the closest to one's home this increases the probability of attending a SHS by 20 percentage points, independently of the gender of the kid.

These two pieces of evidence together indicate that while the choice of high school is likely driven by the convenient location relative to home and usually it is common to all children in a family, the choice of college major involves more gender differentiation and in mixed-gender sibling groups

[9]In our sample 68% of boys choose a SHS with respect to 42% of girls.

more gender specialization.

Finally in Table 5 we replicate the estimations of Table 3 for the probability of enrolling in a HEM while controlling for the choice of high school. This allows us to test that the effect of siblings gender composition emerges at the time of the college-major choice and affects both individuals who have attended a SHS or a Classic High School. While the choice of a SHS increases the probability of choosing a HEM[10], we see that family gender composition plays an important role in affecting the choice of HEM so that the impact, especially on males, is clearly present even controlling for their high school choice.

In Table 4 we also include school/cohort fixed effects, within school/cohort ranking in the high school exit exam, female share in high school classroom[11] to control for high school performance and environmental factors that are likely to influence the probability of choosing a HEM. Estimates in Table 5 are very similar to those of Table 3 and for column (7) and (8) the effect is even reinforced for males. This confirms that gender of other siblings, especially older ones, affects the choice of major of students. In particular males with an older sister are much more likely to choose HEM independently from the high school choice that they made. In fact, those male individuals were more likely to choose a classical High School (Column 8 of Table 3), probably following the family choice driven by the oldest daughter. Nevertheless the tendency to choose HEM in response to having an older sister is strong enough for males that overall an older sister increases the male-stereotypical choices by men (Column 8 table 2).

5 Conclusions

Gender bias in the choice of college major is an important determinant of the gender wage gap and it is shown to be highly persistent in all advanced economies. Understanding the factors associated to the gender differential in the choice of major is thus crucial for identifying effective policies aimed at reducing the gender wage gap. In this work we focus on families that have at least two children

[10] 42.5% of all students graduating from a SHS enroll in a HEM versus only 20.5% of those graduating from a Classical Studies High School.

[11] As found in Anelli, Peri 2013 same-sex share in the high school class is an important determinant of the probability of attending one of the HEM

attending a College-Prep High School of Milan to analyze whether the gender composition of sibling groups affects the choice of academic track in High School and the choice of College Major. We find that mixed-gender siblings are associated with a stronger gender-stereotypical specialization across majors. Male siblings who have at least a sister attending a College-Prep High School are 5 percentage points more likely to enroll in a High Earning Major (Economics/Business, Medicine, Engineering), relative to male siblings with only brothers. The effect is symmetric (but weaker and not significant) for female, who are 2.5 percentage points less likely to enroll in a High Earning Major siblings if they at least one brother, relative to female siblings with only sisters. These effects suggest that gender-mixed groups reinforce gender-segmented college major decisions, especially for males within the family.

Interestingly, we also find that gender composition of siblings does not influence the choice of high school track. This decision, made much earlier in life, is highly influenced by parents and it appears to be mainly function of geographical proximity of schools. Parents have preferences for having all their kids in the same high school and they tend to choose the type of high school more stereotypical for the gender of their first child (e.g. Scientific High School for male first children) and then encourage younger siblings to attend the same school, independently of their gender.

Our findings on the choice of college major show that this decision reflects more the preference of the students (rather than of their families) relative to the choice of high school and, importantly, that a mixed-gender environment in the family reinforces gender stereotypical preferences. When all siblings are of the same sex their identity and preference for academic paths is less influenced by gender stereotypes. It is thus important to raise awareness in students, their teacher and their families about the existence of this gender bias. This paper emphasizes that the environment they live in, may crucially affect these perceptions especially at the moment of choosing college education.

References

[1] Altonji, J.G., Blom, E. and Meghir, C. 2012, "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers", Annual Review of Economics, vol. 4, no. 1, pp. 185-223.

[2] Anelli, M. and Peri, G. 2013, Peer Gender Composition and Choice of College Major, National Bureau of Economic Research, Inc, NBER Working Papers: 18744.

[3] Booth, Alison L. & Cardona Sosa, Lina & Nolen, Patrick J., 2011. "Gender Differences in Risk Aversion: Do Single-Sex Environments Affect their Development?", IZA Discussion Papers 6133, Institute for the Study of Labor (IZA).

[4] Booth, A. & Nolen, P. 2012, "Choosing to Compete: How Different Are Girls and Boys?", Journal of Economic Behavior and Organization, vol. 81, no. 2, pp. 542-555.

[5] Flabbi, Luca (2011). "Gender Differences in Education, Career Choices and Labor Market Outcomes on a Sample of OECD Countries." Organization for Economic Co-operation and Development Background Paper for the WDR 2012.

[6] Gneezy, Uri, Niederle, Muriel, Rustichini, Aldo (2003). "Performance in competitive environments: gender differences." Quarterly Journal of Economics. 118, 1049-1074.

[7] Niederle, Muriel, Vesterlund, Lise (2007). "Do women shy away from competition? Do men compete too much?" Quarterly Journal of Economics 122 (3), 1067-1101.

[8] Niederle, Muriel, Vesterlund, Lise (2010). "Explaining the gender gap in math test scores: the role of competition." Journal of Economic Perspectives 24 (2), 129-144.

Tables

Table 1: Descriptive Statistics

	Families with only one child at CPHS (1)		Families with 2 or more children at CPHS (2)		Diff.	
	Mean	S.D.	Mean	S.D.	(1)-(2)	t-stat
Log(House Value)	7.94	0.28	7.99	0.30	-0.0544*	(-2.35)
Father's Wage	1726.96	636.36	1883.01	580.28	-156.1***	(-3.41)
Mother's Wage	919.73	796.70	1016.08	818.21	-96.35	(-1.58)
Household Income	2596.50	1024.28	2858.25	1013.16	-261.8***	(-3.34)
Father went to college	0.45	0.50	0.64	0.48	-0.196***	(-5.54)
Mother went to college	0.29	0.45	0.48	0.50	-0.187***	(-5.21)
Both parents present in the family	0.94	0.24	0.95	0.21	-0.0183	(-1.15)

t-statistics for difference in mean in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Sample: Families of interviewed individuals, excluding families with a single child. In column 1 we report descriptive statistics for families with only child attending a College-Prep High School while in column 2 those for families with at least two children attending CPHS.
CPHS: College Preparatory High School

Table 2: Testing assumption of exogenous siblings' gender

	Families with same gender siblings only (1)		Families with mixed-gender siblings (2)			
	Mean	S.D.	Mean	S.D.	Diff. (1)-(2)	t-stat
Log(House Value)	7.98	0.31	8.01	0.29	-0.02	(-0.60)
Father's Wage	1874.51	590.75	1892.98	570.61	-18.46	(-0.23)
Mother's Wage	1089.04	806.64	935.02	827.05	154.00	(1.42)
Household Income	2949.02	1025.90	2751.89	992.62	197.10	(1.43)
Father went to college	0.62	0.49	0.67	0.47	-0.04	(-0.72)
Mother went to college	0.49	0.50	0.46	0.50	0.02	(0.36)
Both parents present in the family	0.97	0.18	0.94	0.24	0.03	(1.09)
# of children in the family	2.36	0.56	2.45	0.67	-0.09	(-1.07)

t-statistics for difference in mean in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Sample: Families of interviewed individuals with at least two children attending College Preparatory High School. In column 1 we report descriptive statistics for families with either all female or all male children. In column 2 those for families with mixed-gender children.

Table 3: Probability of choosing a High Earning Major

Dependent variable: High Earning Major (H.E.M.)=1								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Women	Men	Women	Men	Women	Men	Women later born	Men later born
(At least 1 Sib. in CPHS=1)	0.048***	0.066***	0.074**	0.052				
	(0.011)	(0.012)	(0.033)	(0.038)				
(At least 1 Broth in CPHS=1)					-0.025			
					(0.022)			
(At least 1 Sist in CPHS=1)						0.050**		
						(0.025)		
(1^{st} born sibling is male=1)							-0.009	
							(0.026)	
(1^{st} born sibling is female=1)								0.067*
								(0.036)
Birth Date Order					-0.000	-0.050**	0.017	-0.096*
					(0.013)	(0.019)	(0.056)	(0.052)
Observations	10,081	9,617	911	856	2,084	2,254	1,048	1,152
R-squared	0.003	0.004	0.010	0.006	0.002	0.007	0.006	0.011
Family wealth proxies	Y	Y	Y	Y	Y	Y	Y	Y
Twin dummy	Y	Y	Y	Y	Y	Y	Y	Y
Siblings Sample	N	N	N	N	Y	Y	Y	Y
Interviewed Sample	N	N	Y	Y	N	N	N	N
Excluding one child families	N	N	Y	Y	Y	Y	Y	Y
Age diff. with 1^{st} born							Y	Y

Method: OLS, Standard Errors clustered at the graduation neighborhood level. , *** p<0.01, ** p<0.05, * p<0.1.
Sample: Sample of all students graduating from one College-Prep High School (CPHS) between 1985 and 2005 in Milan enrolling in college.
Dependent variable: High Earning Major (HEM)=1
Control variables: Dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city. For columns 7 and 8 we include also the age distance from first born.

Table 4: Probability of choosing the scientific school track

Dependent variable: Scientific High School (S.H.S.)=1								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Women	Men	Women	Men	Women	Men	Women later born	Men later born
(At least 1 Sib. in CPHS=1)	0.042***	0.002	0.041	-0.003				
	(0.016)	(0.013)	(0.035)	(0.036)				
(At least 1 Broth in CPHS=1)					0.064***			
					(0.023)			
(At least 1 Sist in CPHS=1)						-0.034		
						(0.022)		
(1^{st} born sibling is male=1)							0.060*	
							(0.034)	
(1^{st} born sibling is female=1)								-0.063**
								(0.031)
Birth Date Order					0.024	-0.001	-0.033	-0.091*
					(0.018)	(0.012)	(0.060)	(0.050)
(Closest School S.H.S.=1)	0.201***	0.189***	0.272***	0.194***	0.179***	0.219***	0.200***	0.174***
	(0.042)	(0.038)	(0.046)	(0.052)	(0.049)	(0.041)	(0.046)	(0.044)
Observations	13,086	12,239	1,041	973	2,615	2,744	1,325	1,392
R-squared	0.047	0.050	0.068	0.069	0.048	0.077	0.058	0.072
Family wealth proxies	Y	Y	Y	Y	Y	Y	Y	Y
Twin dummy	Y	Y	Y	Y	Y	Y	Y	Y
Siblings Sample	N	N	N	N	Y	Y	Y	Y
Interviewed Sample	N	N	Y	Y	N	N	N	N
Excluding one child families	N	N	Y	Y	Y	Y	Y	Y
Age diff. with 1^{st} born							Y	Y

Method: OLS, Standard Errors clustered at the neighborhood level. *** p<0.01, ** p<0.05, * p<0.1.
Sample: Sample of all students graduating from one College-Prep High School (CPHS) between 1985 and 2005 in Milan
Dependent variable: Scientific High School (SHS)=1
Control variables: Dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city. For columns 7 and 8 we include also the age distance from first born.

Table 5: Probability of choosing a High Earning Major Conditional on High School Attendance

	Dependent variable: High Earning Major (H.E.M.)=1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Women	Men	Women	Men	Women	Men	Women later born	Men later born
(At least 1 Sib. in CPHS=1)	0.028***	0.045***	0.076*	0.037				
	(0.010)	(0.012)	(0.039)	(0.057)				
(At least 1 Broth in CPHS=1)					-0.030			
					(0.019)			
(At least 1 Sist in CPHS=1)						0.050**		
						(0.020)		
(1^{st} born sibling is male=1)							-0.013	
							(0.028)	
(1^{st} born sibling is female=1)								0.099***
								(0.031)
Birth Date Order					0.004	-0.051***	0.015	-0.041
					(0.021)	(0.019)	(0.063)	(0.058)
Constant	0.085***	0.281***	-0.040	0.504***	0.033	0.383***	-0.077	0.006
	(0.024)	(0.029)	(0.080)	(0.133)	(0.056)	(0.072)	(0.171)	(0.158)
Observations	10,324	9,843	936	876	2,114	2,290	1,062	1,173
R-squared	0.093	0.140	0.232	0.275	0.135	0.197	0.201	0.238
Graduation High School/Year FE	Y	Y	Y	Y	Y	Y	Y	Y
School/Cohort Ranking	Y	Y	Y	Y	Y	Y	Y	Y
Family wealth proxies	Y	Y	Y	Y	Y	Y	Y	Y
Class Female Share	Y	Y	Y	Y	Y	Y	Y	Y
Twin Dummy	Y	Y	Y	Y	Y	Y	Y	Y
Siblings Sample	N	N	N	N	Y	Y	Y	Y
Interviewed Sample	N	N	Y	Y	N	N	N	N
Excluding one child families	N	N	Y	Y	Y	Y	Y	Y
Age distance from 1^{st} born dummy							Y	Y

Method: OLS, Standard Errors clustered at the graduation high school/year level. , *** p<0.01, ** p<0.05, * p<0.1.
Sample: Sample of all students graduating from one College-Prep High School (CPHS) between 1985 and 2005 in Milan enrolling in college.
Dependent variable: High Earning Major (HEM)=1
Control variables: School/cohort fixed effects. Within school/cohort ranking in the high school exit exam, dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city, Female share in high school classroom, Birth date order, dummy=1 if the individual is a twin. For columns 7 and 8 we include also the age distance from first born.