Sibling Gender Composition and Participation in STEM Education

Anne Ardila Brenøe*
University of Copenhagen and IZA

September 22, 2017

Abstract

This paper studies how sibling gender composition affects participation in Science, Technology, Engineering, and Mathematics (STEM) education. To identify the causal effect of sibling gender, I focus on a sample of first-born children who all have a younger biological sibling, using rich administrative data on the total Danish population. The randomness of the second-born siblings' gender allows me to estimate the causal effect of having an opposite sex sibling relative to a same sex sibling. The results suggest that having a younger opposite sex sibling increases the probability of enrolling in a gender-stereotypical field of education, with the largest relative effects for women. Although sibling gender composition has no impact on men's probability of actually completing a STEM education, it has a powerful effect on women's success within these fields: first-born women with a younger brother are eleven percent less likely to complete any field-specific STEM education relative to women with a younger sister. I provide evidence that parents with mixed sex children gender-specialize their parenting more and invest more time in their first-born same sex child than parents with same sex children. These findings indicate that the family environment plays in important role for shaping interests in STEM fields.

JEL classification: I2, J1, J3
Keywords: Sibling gender, gender-stereotype, STEM, education, field of study.

*University of Copenhagen, Department of Economics, Øster Farimagsgade 5, Building 26, 1353 Copenhagen K, Denmark. aab@econ.ku.dk. I thank Marianne Bitler, David Card, Ilka Gerhardts, Jennifer Graves, Mette Gørtz, Shelly Lundberg, Heather Royer, Jenna Stearns, Melanie Wasserman, Ulf Zölitz, and seminar participants at the University of Copenhagen (Department of Economics and Department of Sociology), University of California-Santa Barbara, University of California-Berkeley, University of California-Davis, the IZA Summer School in Labor Economics 2017, the Workshop: Education, Skills, and Labor Market Outcomes 2017, the Copenhagen Education Network Workshop 2017, the International Workshop on Applied Economics of Education 2017, and Lund University for helpful discussions and comments.

1 Introduction

Although women today, on average, attain more education than men across most OECD countries, large gender differences persist in the choice of field of study (OECD, 2016). Only 28 percent of students enrolled in tertiary education are female within Science, Technology, Engineering, and Mathematics (STEM), while women represent 54 percent of all students. Meanwhile, the returns to field of study vary as much as the returns to level of education, with the greatest returns to STEM fields (Altonji et al., 2012; Kirkebøen et al., 2016). The gender segregation in field of study persists into occupational choice in the labor market and thereby contributes to the gender wage gap (Blau and Kahn, 2016; Gallen et al., 2017). At the same time, the STEM workforce is the main contributor to technological innovations, representing the main source of long-run economic growth.[1] Yet, many countries face a shortage of STEM graduates. Given the larger returns to investing in STEM education for both the individual and society, we need to better understand how the social environment interacts with the decision to participate within STEM fields —and in particular women's decision given their current underrepresentation.

Why are so few women in STEM fields compared to men? While boys and girls enter school with same levels of math ability, girls lose interest in math and science throughout elementary school with the consequence that boys have a math test score advantage by middle school (Kahn and Ginther, 2017). Several studies document that different aspects of the social environment during childhood affect gender differences in math test scores. Fewer studies, however, trace effects into the actual choice of studying and working within STEM fields. This is, in part, due to limited data availability, as one needs to link childhood exposure to later educational and, preferably, adult outcomes.

In this paper, I focus on one possible causal factor critical for the development of girls and boys' interests in STEM fields during childhood: sibling gender composition. I use high-quality administrative data for the total population in Denmark from 1980 through 2015 to provide causal estimates of the impact of sibling gender composition on participation in STEM education. In particular, I exploit the random assignment of the second-born child's gender, conditional on the sex of the first-born child. The crux of my identification strategy is to compare STEM participation for first-born children with a second-born opposite sex sibling to those with a same sex sibling. I do this separately for men and women, born between 1962 and 1986, who all have a younger biological sibling (same mother and father). This approach stands in contrast to previous studies on sibling gender composition, predominantly studying educational attainment, as they generally include all siblings both in the measure of sibling gender

[1] See, for instance, Atkinson and Mayo (2010); Peri et al. (2015) and references therein.

composition and in the estimation sample (Amin, 2009; Anelli and Peri, 2014; Bauer and Gang, 2001; Butcher and Case, 1994; Conley, 2000; Cools and Patacchini, 2017; Cyron et al., 2017; Hauser and Kuo, 1998; Kaestner, 1997; Oguzoglu and Ozbeklik, 2016).[2] Considering the effects of older siblings' gender on younger siblings' outcomes is challenging due to selection bias. As explained in greater detail in Section 2, selection bias arises, for instance, if parents decide to have a second child depending on their first child's gender. This, in turn, leads to biased estimates if parents with different gender preferences raise their children differently.

This paper makes three important contributions to the existing literature. First, I study the effect of sibling gender composition on educational STEM choice from first place of enrollment after compulsory schooling (grade 9) through highest completed education by age 30; this is to observe the emergence and persistence of the effect and is only possible due to the unique dataset. Second, I use a new strategy to estimate the causal effect of sibling gender compared to previous studies, which reduces concerns about selection bias. Third, to the best of my knowledge, I am the first to conduct a large quantitative analysis of how sibling gender composition affects child-parent interactions, thereby providing a detailed picture of likely channels through which the effects on STEM participation operate.

My results suggest that having an opposite sex sibling increases the probability of choosing a gender-stereotypical education. Although sibling gender has only a limited effect on men, it has a significant impact on women's participation in STEM education. Having a second-born brother decreases first-born women's likelihood of ever enrolling in and completing any field-specific STEM education by respectively 5.5 and 10.5 percent. The reduced probability of choosing a program with STEM focus is already present at the first place of enrollment after compulsory schooling and persists into STEM college completion and occupational choice through age 40. Thus, women with a younger brother are more likely to opt out of STEM already at the time of high school application with important consequences for their further educational specialization, field of occupation, and labor market earnings. Meanwhile, men with a younger sister relative to men with a brother are only more likely to enroll in a STEM program, but not consistently more likely to ever complete a field-specific STEM degree or to work within a STEM occupation.

[2] The only exception from such strategy is Peter et al. (2015), which investigates the effect of a co-twin's gender on educational attainment and earnings. Moreover, Gielen et al. (2016) employs a difference-in-differences strategy to estimate the effect of having a male twin on earnings; yet, their interest is whether exposure to prenatal testosterone (rather than sibling gender composition per se) has an effect on earnings. The literature on sibling gender composition and educational attainment provides inconsistent findings, though with an overweight of studies reaching statistically insignificant associations. A general problem for the studies is, however, small sample sizes of typically 1,000–10,000 observations, often making a rejection of a tight zero finding impossible.

Why does sibling gender alter women and men's likelihood of choosing STEM fields? The impact on field choice could be due to changes in aspects like preferences, interests, identity, and ability. However, I rule out the latter, ability, as sibling gender composition does not affect school performance or attainment. Sibling gender could affect identity, and thereby preferences and interests, through child-parent and/or child-sibling interactions. I provide compelling evidence that changes in child-parent interactions —and, in particular, increased gender-specialized parenting in families with mixed sex children —play an important role for the changes in STEM participation. Drawing on time use data, I show that parents of mixed sex children invest more gender-specifically in their first-born child, especially in families with first-born daughters, than parents of same sex children. This translates into a substantially worse relationship between fathers and their first-born daughters when the second-born child is male relative to female. Moreover, I find the effects on STEM choice to be strongest for individuals with a more "gender-stereotypical" same sex parent. In line with the same sex education argument (Booth et al., 2013; Schneeweis and Zweimüller, 2012), I further show that young boys with a younger sister are more exposed to gender-stereotypical behavior within the family than boys with a younger brother. Consequently, my findings emphasize that if policy makers want to increase the number of people —and particularly women —within STEM fields, they need to focus on early exposure to gender-stereotypes in the social environment, including the family.

My focus on the social environment is consonant with recent studies that trace gender gaps in educational outcomes to factors such as teacher stereotypes (Lavy and Sand, 2015), the gender of school peers and teachers (Anelli and Peri, 2014, 2016; Bottia et al., 2015; Brenøe and Zölitz, 2017; Carrell et al., 2010; Favara, 2012; Oguzoglu and Ozbeklik, 2016; Zölitz and Feld, 2017), and parental role models (Brenøe and Lundberg, 2017; Cheng et al., 2017; Humlum et al., 2017). Only two existing studies investigate associations between sibling gender composition and field of college enrollment (Anelli and Peri, 2016; Oguzoglu and Ozbeklik, 2016). These studies, however, face challenges in terms of selection bias and data availability. At the same time, no previous study has examined how the effects develop from childhood through adulthood or provided a comprehensive analysis of possible mechanisms.

In contrast to my findings, Anelli and Peri (2014) do not find a relationship between sibling gender and women's probability of enrolling in a high earnings (male-dominated) college major. Meanwhile, they do find that men of any parity with any and especially those with an older sister are more likely to enroll in such major compared to men without any sister. Nevertheless, they do not examine whether these effects persist into actual degree completion or labor market outcomes, which seems important as Anelli and Peri (2016), using the same dataset, find that gender peer effects on college major choice for men are only short lived. Moreover, the authors face an important limitation for the study of sibling gender composition, as the dataset only includes siblings who complete an academic high school degree in one city in Italy. Oguzoglu and Ozbeklik (2016) ask a narrower question whether fathers employed in STEM fields differentially invest in their daughters depending on whether they also have at least one son. The study finds that women enrolled in college with a father within a STEM occupation are less likely to declare a STEM major if they have at least one brother compared to those without any brother. The authors suggest that this patterns is due to differences in fathers' transmission of occupation-specific tastes and preferences, although they are unable to test for such mechanisms.[3] While Anelli and Peri (2016) and Oguzoglu and Ozbeklik (2016) make an important contribution by studying how field of college major correlates with sibling gender composition, I add to this literature by examining, in a more complete way, how sibling gender composition affects participation in STEM from puberty well into adulthood and how it affects child-parent interactions.

2 Empirical Strategy

The aim is to estimate the causal effect of sibling gender composition on STEM participation. Simply comparing children from families with different gender compositions would, however, not provide valid estimates of the causal effects of sibling gender composition due to selection. An empirical fact in developed countries is that parents are more likely to have a third child if their first two children are of same compared to mixed gender (Angrist and Evans, 1998; Angrist et al., 2010; Black et al., 2005). Thus, the gender composition of current children affects parity progression, indicating that parents have preferences over the gender composition of their children. A parental preference for having at least one child of each gender could explain this fertility pattern. Alternatively, if some parents prefer boys, others prefer girls, and parents progress to the next parity until they have a child of their preferred gender, we would also expect the observed fertility pattern. Therefore, the final gender composition of all children in a given family is endogenous to the gender of the first child(ren).

Even if we compare children of the same parity and gender but with older siblings of different gender, we might have a selection problem. As an example, suppose we want to estimate the effect of the gender of the first-born child on the outcome of the second-born child. Assume that parents either prefer a girl, a boy, or are indifferent and assume, for simplicity, that they only have a second child if their first child is not of the

[3] The related literatures on the impact of single sex education (e.g. Booth et al. (2013); Eisenkopf et al. (2015); Jackson (2012); Lee et al. (2016)) and gender peer composition (e.g. Brenøe and Zölitz (2017); Hoxby (2000); Lavy and Sand (2015); Anelli and Peri (2016); Hill (2017)) also provide mixed results.

preferred gender.[4] However, estimating the effect of having an older brother compared to an older sister for the sample of second-born girls would give biased estimates if parents with different gender preferences raise their children differently. As Table 1 illustrates, among those second-born children who are actually born, girls with an older brother come from families who prefer girls, while girls with an older sister come from families who prefer boys. Thus, second-born girls who have an older brother do not come from similar families as those with an older sister. If parents with a certain gender preference raise their children of that gender more gender-stereotypically and having a sibling of the opposite sex makes educational choices more gender-typed, we would expect the bias to magnify the estimated effect of sibling gender. Put differently, the selection bias problem arises because we only observe the outcome for second-born children who are actually born.[5]

Table 1

Example: Parental Gender Preference and Fertility Choice

Gender of 1^{st} child	Girl						Boy					
Gender Preference	Girl		Indif.		Boy		Girl		Indif.		Boy	
Have 2^{nd} child	no	no	no	no	yes	yes	yes	yes	no	no	no	no
					↓	↓	↓	↓				
Gender of 2^{nd} child (if born)	G	B	G	B	G	B	G	B	G	B	G	B

Assumptions for example: 1) parents prefer either a girl, a boy, or are indifferent and 2) parents only have a second child if their first child is not of the preferred gender.

To reach the goal of estimating the causal effect of sibling gender composition, an ideal experiment would be to let parents decide before having their first child how many children they want and when they want their children to be born without any possibility to change this. The experimenters would then randomize the gender of all children, such that the gender of each child and the gender composition of each sibship would be completely random to the parents. In this case, we would compare children of same birth order and gender with the same number of siblings to each other and have variation in the number of sisters relative to brothers. Such experiment is, nevertheless, not ethically feasible.

The second-best experiment, in contrast, is viable and occurs as a natural exper-

[4] An extreme reason for not having a second child if the first child is not of the preferred gender could be divorce. Some U.S. studies find an increased divorce risk when having a first-born girl (Bedard and Deschenes, 2005; Dahl and Moretti, 2008), while Kabtek and Ribar (2017) do not find support for this for the Netherlands.

[5] Appendix A.1 shows the selection bias problem more formally and discusses other reasons for selection bias than parental gender preferences.

iment. Because parents do not know the gender of a subsequent child when they make the decision to progress to the next parity, we *can* causally estimate the effect of "future" children's gender on "current" children's outcomes. In absence of the ideal experiment, I leverage the random assignment of the second child's gender conditional on the gender of the first child and the parents wanting a second child. In this way, it is possible to estimate the causal effect of second-born children's gender on first-born children's outcomes. For this strategy to provide a valid estimate of the causal effect of sibling gender composition on STEM participation, the sex of the second child needs truly to be random.[6] Subsection 3.2 provides evidence, supporting the identifying assumption of random assignment.

Thus, I leverage two sets of comparisons: 1) I compare first-born women who have a second-born brother to first-born women who have a second-born sister and 2) vice versa for men. I always estimate the model separately for men and women, as they might come from different types of families and because the outcomes of men and women differ substantially. The empirical specification for the main analysis is:

$$Y_i^{First-Born} = \alpha_0 + \alpha_1 Opposite\ Sex_i^{Second-Born} + X_i'\delta + \nu_i, \tag{1}$$

where $Y_i^{First-Born}$ indicates whether individual i (who is first-born) participates in STEM education and the estimate of interest is α_1, i.e. the effect of having a second-born sibling of the opposite sex. X_i is a vector of fixed effects for birth municipality, year-by-month of birth, spacing in months to the second-born sibling, immigrant status, maternal age at birth, paternal age at birth, maternal level-by-field of education, and paternal level-by-field of education.[7] ν_i is the error term and is clustered at the year-by-month of birth level.[8]

As this strategy only relies on the random assignment of the second child's sex, parents can respond to the gender composition of their first two children in terms of subsequent fertility. Appendix Table A2 shows that, for the main sample of the analysis (described in Section 3), having a second-born sibling of the opposite sex reduces, on average, first-born women and men's family size by 0.07 and 0.08 siblings, respectively. Therefore, family size might mediate some of the effect of having a second-born opposite sex sibling if family size has an independent impact on STEM participation.

Existing studies find that family size does not affect educational attainment in

[6]Some evidence suggests that the Trivers-Willard hypothesis, which proposes that females in disadvantaged circumstances are less likely to bear male offspring, may hold in human populations through increased mortality among male relative to female fetuses, although the impacts of even extreme events are small (Almond and Edlund, 2007; Almond and Mazumder, 2011; Hamoudi and Nobles, 2014; Trivers and Willard, 1973). Given I condition on the first child's gender, however, this does not appear to be a valid threat for the identification.

[7]If the parent does not have a field-specific education, I use the field of occupation.

[8]However, the level of clustering does not make any difference for the results.

either Israel or Norway, using twins as an instrument for family size (Angrist et al., 2010; Black et al., 2005). In Appendix A.2, I replicate this finding in the Danish context and show that family size only has a borderline significant effect on women's probability of completing a STEM degree, while there is no effect for men or on STEM enrollment. The results in Appendix Table A1 suggest that having one additional sibling reduces women's probability of completing a STEM degree by 1.1 percentage points. Scaling this effect, for women, by the effect of having a second-born brother on family size suggests that family size mediates ($-0.07 \times -1.1 =$) 0.08 percentage points of any potential effect on STEM completion. As the main results suggest that women with a younger brother relative to women with a younger sister are less likely to participate in STEM, the effects of sibling gender composition for women might therefore be conservative. Moreover, I show that sibling gender composition does not affect educational attainment or achievement (Subsection 4.2). Finally, Subsection 6.1 further tests the robustness of the results to family size. Based on these different pieces of evidence, family size does not seem to be an important confounder of the effect of sibling gender, but might, if anything, bias the effect towards zero for women.

3 Data

3.1 Data and Sample Selection

I use Danish administrative data for the total population from 1980 through 2015. One central feature of this dataset, compared to most previous studies on sibling gender composition, is that I can link all children to their parents and siblings. Thus, I observe parents' complete fertility history and thereby, correctly measure the sibling gender composition. Furthermore, I have information on parents' date of birth; length, type, and field of education; labor market attachment; and occupation.[9] For the children, I observe every time a person enrolls in an education and have detailed information on the characteristics of the program, such as level, type, and field; data on enrollment is available since 1978. The educational registry further reports the highest completed degree at an annual basis. Throughout, I follow the International Standard Classification of Education (ISCED) for the definition of all educational measures. I include observations through age 27 for all enrollment measures and through age 30 for all completion measures to give people time to complete the education in which they enroll. Finally, I also observe the children's annual labor earnings and occupation.

I restrict the sample to cohorts born between 1962 and 1986 to allow for sufficient

[9]The registers started to report occupation in 1991. To characterize parental occupation, I use the mode occupation from 1991–2000.

time to enroll and complete an education. Moreover, I only include first-born children, who are the first child to both the mother and father; I exclude first generation immigrants to eliminate concerns about unobserved siblings in the data and because I might not observe all their educational history; I only consider individuals who have at least one full sibling (same mother and father) born less than four years apart and who survives the first year of life; I exclude families where either the first or second child is a twin; and finally, I exclude those few individual's who die before age 30 or do not live in Denmark at any time between age 26 and 30.[10] I refer to this sample of first-born children as the *main sample*.

Table 2 provides descriptive statistics on demographic characteristics of the main sample by sibling gender composition. Overall, first-born men and women come from very similar family backgrounds regardless of sibling gender. One-third of the sample is born in one of the two largest metropolitan areas in Denmark (Greater Copenhagen and Aarhus) and another third is born in Jutland outside the County of Aarhus. Average spacing to the younger sibling is 2.5 years and 1.2 percent are second generation immigrants. Mothers are, on average, 23.3 years at birth and have 11.2 years of education, while fathers are 26.0 years and have 12.0 years of education. A large share of parents are within very gender-typed fields. Thirty-two and 12 percent of mothers are respectively within low- or medium-level administration and health fields, while 43 percent of fathers are within STEM fields.[11]

3.2 Randomization Checks

Columns (1) and (2) in Panel A show the mean for first-born women with a second-born sister and brother, respectively. Column (3) provides the p-value from a two-sided t-test of significance between these two groups of first-born women. Columns (4) through (6) provide similar statistics for first-born men. Most differences are statistically insignif-

[10]The last restriction leads to the exclusion of 3,979 individuals due to death by age 30 (of whom 17 percent die before their first birthday) and 8,985 individuals due to emigration. Moreover, I also exclude those very few individuals (569) who do not live in Denmark for more than ten years between age 13 and 30. Sibling gender composition does not affect the attrition due to these restrictions.

[11]I have defined these very gender-typed fields based on having a great majority of graduates/workers of one gender within these fields. Maternal field in administration is defined as having a field-specific education within Education; Arts and Humanities; or Business, Administration, and Law (ISCED fields 2, 3, and 4) or having an occupation as Business and Administration Associate Professional (ISCO-08 sub-major group 33) or Clerical Support Workers (ISCO-08 major group 4) conditional on not having any field-specific education. Maternal field in health is defined as having a field-specific education within Health below Master level (ISCED field 9) or having an occupation as Health Professional or Health Associate Professional (ISCO-08 sub-major groups 22 and 32) conditional on not having any field-specific education. Paternal field in STEM is defined as having a field-specific education within STEM (ISCED fields 5–7) or having an occupation as Science and Engineering Professional; Science and Engineering Associate Professional; Craft and Related Trades Workers; or Laborer in Mining, Construction, Manufacturing, and Transport (ISCO-08 sub-major groups 21, 31, 71–75, and 93) conditional on not having any field-specific education.

Table 2
Descriptive Statistics and Balancing Test by Sibling Gender Composition

	First-Born Women			*First-Born Men*		
Panel A: Average by Gender of the Second-Born Sibling and t-test of Difference						
Second-Born	Sister (1)	Brother (2)	p-value (3)	Sister (4)	Brother (5)	p-value (6)
Region of Birth (pct.)						
Greater Copenhagen	23.86	23.70	0.46	23.97	23.81	0.44
Rest of Zealand	17.74	17.99	0.18	17.70	17.97	0.15
Funen	8.58	8.67	0.49	8.64	8.56	0.52
Aarhus	12.59	12.43	0.32	12.49	12.46	0.84
Rest of Jutland	36.93	36.90	0.90	36.97	36.96	0.98
Greenland	0.30	0.30	0.88	0.24	0.25	0.54
Spacing (months)	30.43	30.48	0.28	30.48	30.43	0.23
2^{nd} Gen. Immigrant (pct.)	1.17	1.15	0.76	1.15	1.08	0.13
Mother's age (years)	23.30	23.26	0.03	23.32	23.31	0.72
Father's age (years)	26.06	26.02	0.03	26.05	26.03	0.41
Mother's education (years)	11.21	11.20	0.86	11.21	11.23	0.16
Mother's edu unknown (pct.)	2.19	2.15	0.54	1.98	2.03	0.41
Father's education (years)	11.99	11.99	0.89	11.99	12.03	0.02
Father's edu unknown (pct.)	3.32	3.37	0.58	3.20	3.20	0.97
Lives with both biological parents at age 17 (pct.)	78.56	78.47	0.66	79.05	79.05	0.98
Parental field of education/occupation (pct.)						
Mother in Admin.	31.77	31.89	0.60	31.86	32.01	0.49
Mother in Health	11.97	12.21	0.13	12.00	12.32	0.04
Father in STEM	43.08	42.66	0.08	43.05	42.97	0.73
Observations	80,593	84,140		84,360	88,980	
Panel B: Balancing Test						
Joint F-statistic	0.90			0.95		
Prob > F	0.98			0.83		

Note: Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Panel A shows the average of family background characteristics for first-born women with a second-born female [Column (1)] and male sibling [Column (2)] and first-born men with a second-born female [Column (4)] and male sibling [Column (5)]. Columns (3) and (6) report the p-values from t-tests of significance between women and men with siblings of different gender, respectively. The balancing test tests whether the control variables included in X_i in Equation (1) can predict having a younger opposite sex sibling. F-test of joint significance of all control variables.

icant from each other at conventional levels and those differences that are statistically significant are small and, to some extent, expected due to the large sample size and the number of t-tests.[12] To account for these small baseline differences, as outlined in Section 2, I flexibly control for parental age and education among a wide range of other fixed effects in the analysis.

Panel B shows statistics from a balancing test, testing whether the demographic characteristics included in X_i in equation (1) can predict having a sibling of the opposite gender. More precisely, it reports the F-test of joint significance of all the covariates in a regression where the outcome is an indicator for having a second-born sibling of the opposite gender. The F-test strongly rejects joint significance for both samples. Thus, this balancing test supports the identifying assumption that the younger sibling's gender is random conditional on the first child's gender.

As I only have annual data since 1980, I do not observe time-varying parental characteristics before birth for most individuals in the main sample. However, for later cohorts, I can check whether parents with a second-born opposite sex child differ from parents with a second-born same sex child. The graphs in Appendix Figure A5 illustrate the estimates from an event study of the effect of having a second child of the opposite sex on a variety of parental SES characteristics. This is estimated separately by the gender of the first child, although shown in the same graph, from five years before the first child's birth through 14 years after for cohorts born between 1985 and 2002. The gender composition of children does not affect parental cohabitation, marital status, length of education, parental employment, or parental annual labor earnings before or around the birth of their first child.[13] This further supports the randomness of the second child's gender.

3.3 Education and Field of Study

In Denmark, children are required to attend primary school from age 7 through grade 9.[14] In the final year of 9^{th} grade, students decide whether they want to apply for secondary education or enter the labor market.[15] Secondary education (ISCED level

[12]Due to assortative mating, the difference for women is statistically significant for both parents' age.

[13]Note at the same time, that the only systematic difference in parental SES characteristics after the first child's birth by sibling gender composition is a positive effect of having mixed sex children on maternal labor earnings between six and nine years by three to five percent after the first child's birth (the measure of earnings does not include parental leave benefit, implying that the effect on total income is smaller than the estimated effect on labor earnings). Thus, the socio-economic conditions experienced during childhood do not, overall, seem to differ by sibling gender composition besides the increased probability of living in a larger family as shown in Appendix Table A2.

[14]For the cohorts of study, it was common to attend a so-called kindergarten class the year before starting first grade, although it was not mandatory.

[15]They can also choose to enroll in an optional 10^{th} grade, which is formally a continuation of primary school. In the analysis, I restrict the attention to enrollment in and completion of programs after primary

3) consists of two types: vocational training and academic high school. Within each of these types, students choose their broad field at the time of application. However, while vocational education is field-specific and prepares students for specific occupations, academic high school is generic and prepares students for tertiary education. Vocational education covers most fields and for this type of education, I group Information and Communication Technologies and Engineering (ISCED fields 61 and 71) as STEM. The academic high school has overall four tracks (language, math, technical, and commercial), of which I group the math and technical tracks as STEM-preparing. Note, however, that, consistent with the ISCED definition, I do not consider the academic high school as a field-specific education.

Tertiary education (ISCED levels 5–8) consists of three types: vocational, professional, and academic. I refer to the latter two jointly as *college*. Similarly, I group vocational secondary and vocational tertiary educations as *vocational education*.[16] An academic high school diploma gives access to all types of tertiary education, while a vocational secondary degree usually only gives direct access to vocational tertiary programs within the same specific field.[17] Though, many vocational secondary programs do not have a natural continuation at the tertiary level; 88 percent of men and women in the main sample with a secondary vocational degree do not complete another education at a higher level.[18]

An application to tertiary education is an application to a specific program. Most college STEM programs have certain STEM high school courses as prerequisites, such as advanced Math and intermediate Physics and Chemistry. Therefore, an academic high school STEM diploma gives much easier access to college STEM majors than other secondary school degrees. However, it is possible to take complementary courses after high school graduation to meet the admission criteria. Acceptance to college mainly depends on the grade point average (GPA) from high school. Most STEM programs admit all eligible applicants (or have very low GPA cutoffs), meaning that once fulfilling the high school STEM course requirements, good prior school performance is not necessary for enrolling within STEM fields in higher education. As women's underrepresentation in STEM is limited to math-intensive —and, generally, better paid—science fields (Kahn and Ginther, 2017), my preferred definition of STEM college majors exclude Biologi-

school, i.e. after grade 9 and 10.

[16]Distinguishing by level gives very similar results. Though, the effect for men is driven by STEM vocational education at the secondary level, while it for women is at the tertiary level.

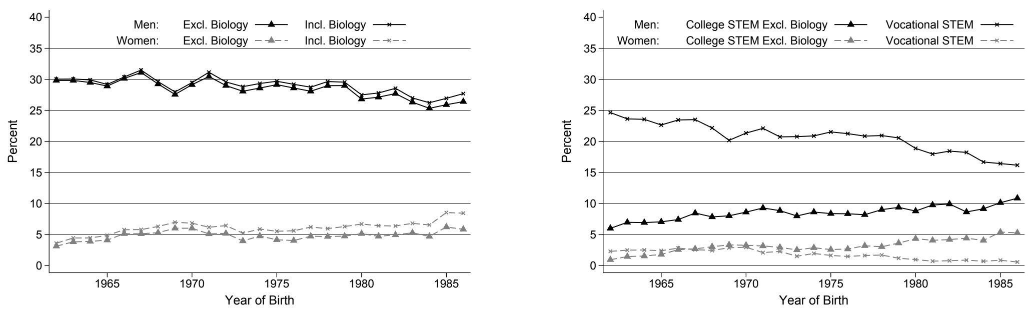
[17]Students with a vocational secondary degree will often be required to have taken one or two academic high school courses at a basic level, such as Math and English.

[18]However, restricting the focus to vocational STEM education reveals some compositional differences between men and women. While 78 percent of men with a vocational STEM degree have their highest completed education at the secondary level, this number is only 24 percent for women.

cal and Related Sciences (ISCED field 51, henceforth *Biology*).[19] I follow Kahn and Ginther (2017) by defining hard sciences in college as Physical Sciences, Mathematics, Statistics, Economics, Information and Communication Technologies, and Engineering (ISCED fields 53, 54, 311, 61, 71) and refer to this definition as *STEM excluding Biology*. However, I also show the results when including Biology in the STEM definition and refer to this measure as *STEM including Biology*.

Figure 1

Share of Cohort with Field-Specific STEM Degree at Age 30 by Gender



(a) STEM excl. and incl. Biology **(b)** College vs. Vocational STEM Degree

Note: Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Graph (a) illustrates the share of a cohort by gender completing a field-specific STEM degree, excluding and including Biology in the definition of STEM. Graph (b) illustrates the share of a cohort completing a STEM vocational and STEM excluding Biology college degree by gender.

The main analysis of STEM education considers field-specific STEM educations in any type and at any level of education after primary school. This is to not potentially confound the results on STEM choice with educational attainment. Thus, the main outcomes of interest indicate whether the individual ever enrolls in and completes a field-specific STEM education preparing for the labor market, including secondary and tertiary vocational STEM programs and college STEM majors. However, because the results in Subsection 4.2 demonstrate that sibling gender composition does not affect educational attainment, I complement the main STEM measures with nine additional outcomes. I examine whether the first place of enrollment after primary school has a STEM focus, i.e. whether it is either secondary STEM vocational education or in the STEM-preparing track in the academic high school. In line with this, I consider the

[19] I group Environment (ISCED field 52) together with Biology, as very few study within that field. Appendix Figure A1 illustrates men and women's average earnings percentile by birth cohort at age 35 by type and field of highest completed education as well as the male share in each cell. Of those who have a college Biology major, only 33 percent are male and average male earnings are substantially lower than for other STEM degrees.

probability of ever enrolling in and completing the academic high school STEM track. Finally, I split field-specific STEM educations by type, thereby investigating effects on the probability of studying in and completing a vocational STEM program and a college STEM (both including and excluding Biology) major, separately.[20]

Figure 1 shows, for the main sample, the share of each cohort completing a field-specific STEM degree by gender over time. Graph (a) presents the development in any field-specific STEM degree when excluding and including Biology in the STEM definition, respectively. The share of women obtaining a STEM degree excluding (including) Biology has increased from 3.1 (3.6) to 5.8 (8.4) percent between the first and last cohorts in the sample. In contrast, the share of men with a STEM degree excluding (including) Biology has declined slightly across cohorts from around 29.8 (30.0) to 26.4 (27.7) percent. Graph (b) shows that this decrease is due to a smaller share of men attaining a STEM vocational degree (24.6 percent for the 1962-cohort vs 16.2 percent for the 1986-cohort), although the share of men obtaining a college STEM major has increased from 6.0 to 10.8 percent. Similarly, women have experienced a decrease from 2.3 to 0.6 percent in the share attaining a STEM vocational degree and a large relative increase from 0.9 to 5.3 percent in the share with a College STEM major.

4 Results

4.1 STEM Education

Figure 2 provides some first evidence on the effect of sibling gender composition on STEM choice. Each graph illustrates, by cohort, the raw difference between the share of first-born individuals who enroll in or complete any field-specific STEM program with a second-born opposite sex sibling and those with a same sex sibling. For women, the overall pattern is clear, illustrating that those with a brother are both less likely to ever enroll in and complete a field-specific STEM program compared to women with a sister. On average, these differences are 0.49 (0.53) and 0.54 (0.58) percentage points for respectively enrollment in and completion of STEM excluding (including) Biology and are statistically significant at the one percent level.[21] In contrast, the pattern is more noisy for men, though indicating that men with a sister compared to a brother are more likely to ever enroll in and might be more likely to complete a STEM program. The difference for any enrollment in (completion of) a STEM program excluding Biology between men with a sister and men with a brother is 0.81 (0.34) percentage points

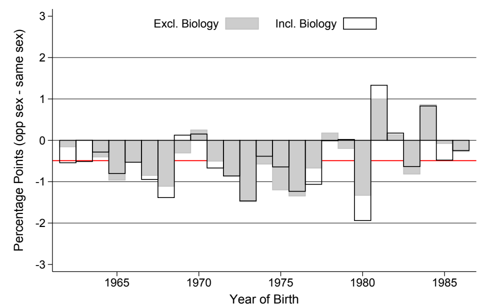
[20]Considering whether the highest completed education is within STEM reveals very similar results as for having any field-specific STEM degree (not reported).

[21]See the *No controls* estimates in Appendix Table A3. This table further illustrates that the estimates are almost identical with different control versions, supporting the assumption that sibling gender is random.

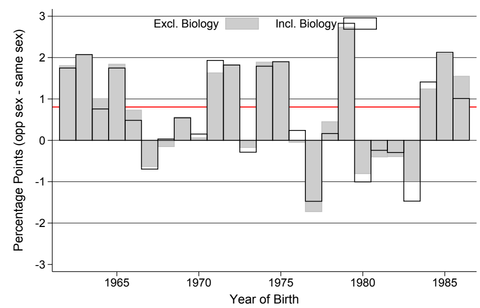
with a p-value of 0.001 (0.11).

Figure 2

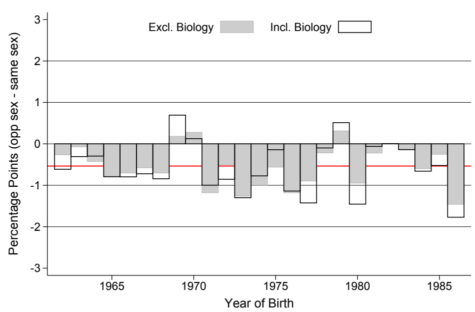
Field-Specific STEM Enrollment and Completion by Gender Across Cohorts: Opposite-Same Sex Sibling Differences



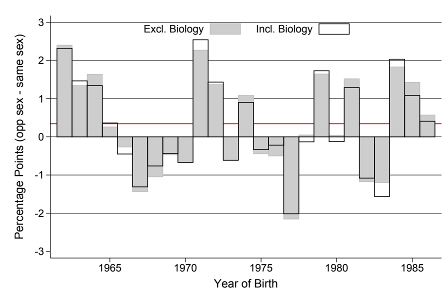
(a) First-born Women: Enrollment



(b) First-born Men: Enrollment



(c) First-born Women: Completion



(d) First-born Men: Completion

Note: Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Graphs (a) and (b) illustrate, by cohort, the raw difference between the share of individuals who enroll in a field-specific STEM program excluding (gray) and including (white) Biology with an opposite sex sibling and those with a same sex sibling for women and men, respectively. Graphs (c) and (d) illustrate, by cohort, the raw difference between the share of individuals who complete a field-specific STEM program excluding (gray) and including (white) Biology with an opposite sex sibling and those with a same sex sibling for women and men, respectively. The red, horizontal line in each graph represents the mean difference in STEM excluding Biology across cohorts.

Table 3 shows the main results on sibling gender composition and STEM education by gender, controlling for demographic and family background characteristics. First-born women with a second-born brother are 0.48 (0.53) percentage points less likely to ever enroll in any field-specific STEM program excluding (including) Biology relative to those with a sister. Given a baseline average of 8.7 (10.1) percent for women with a sister, the relative change corresponds to a decrease by 5.5 (5.1) percent. This effect persists into educational attainment, resulting in a decreased probability of ever completing a STEM degree by 10.5 (9.3) percent. The effects are very similar when considering STEM including Biology, though the percent effect is slightly smaller due to a larger baseline. These results consequently demonstrate that sibling gender has a powerful effect on women's likelihood of going into traditionally male-dominated STEM fields.

Table 3

Field-Specific STEM Enrollment and Completion

	STEM Enrollment		STEM Completion	
	Excl. Biology (1)	Incl. Biology (2)	Excl. Biology (3)	Incl. Biology (4)
Sample of First-Born Women				
Second-Born Brother	-0.48***	-0.51***	-0.53***	-0.58***
	(0.14)	(0.15)	(0.10)	(0.11)
Same Sex Baseline	8.7	10.1	5.0	6.2
Percent Effect	-5.5	-5.1	-10.5	-9.3
Observations	164,733			
Sample of First-Born Men				
Second-Born Sister	0.80***	0.77***	0.32	0.33
	(0.23)	(0.23)	(0.22)	(0.22)
Same Sex Baseline	40.9	41.5	28.5	29.1
Percent Effect	2.0	1.9	1.1	1.1
Observations	173,340			

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Each Panel-Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last educational observation. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. *Field-specific STEM* education refers to vocational and college programs.

In contrast, first-born men with a younger sister are more likely than men with a brother to enroll in an education within STEM excluding (including) Biology by 0.80

(0.77) percentage points, representing a relative effect of 2.0 (1.9) percent. The estimated impact is larger in magnitude than for women; however, due to a larger baseline (40.9 percent), the relative effect is only around one-third of the one for women. At the same time, sibling gender has no statistically significant effect on men's probability of completing a STEM education and the magnitude of the effect is small (0.32 percentage points; 1.1 percent).

To elaborate on the main results, Table 4 provides a more nuanced picture on the educational process related to STEM fields from first place of enrollment after grade 9 through age 30. Sibling gender already impacts women's first active educational choice: women with a younger brother are 3.5 percent less likely to enroll in a program with emphasis on STEM subjects as their first place of enrollment after compulsory schooling. As only very few women enroll in secondary vocational STEM programs, this effect is entirely driven by a decreased probability of enrolling in and completing the STEM tracks in the academic high school. After secondary schooling, women with a younger brother are again less likely to choose an education within STEM. Women with a brother compared to women with a sister are 9.0 percent less likely to complete a vocational STEM degree and 11.5 (9.5) percent less likely to complete a STEM college major excluding (including) Biology. Consequently, these results show that once women opt *out* of STEM fields—which already happens at the time of high school application—they do not opt *in* again and that women fall out of STEM for each educational transition. These findings stress that women's choice not to study within STEM fields originates to the time before exiting compulsory education but is not only limited to that period.

For men, the story is different. Men with a younger sister are only slightly more likely to enroll in a program with STEM focus as their first place of enrollment (0.8 percent). The percent effect increases, though, when restricting STEM enrollment to vocational programs (1.7 percent) and STEM college majors excluding Biology (2.6 percent). In line with the main results, sibling gender does not have a statistically significant impact on STEM completion for any of these separate types of education. Thus, although sibling gender affects men's likelihood of choosing an education within STEM fields, the effect does not persists into actual degree completion. As the next subsection shows that sibling gender composition does not affect school performance, the results support an interpretation of changed interests in STEM fields, but that sibling gender does not improve men's ability to actually succeed in STEM programs.

Table 4

STEM Education from end of Compulsory Schooling through Age 30

	STEM Focus in First	Academic High School STEM Track		Vocational STEM (Secondary and Tertiary levels)		College STEM Major			
	Enrollment	Enrollment	Completion	Enrollment	Completion	Excl.	Incl.	Excl.	Incl.
						Enrollment		Completion	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample of First-Born Women									
Second-Born Brother	-0.89***	-0.91***	-0.87***	-0.18*	-0.18***	-0.37***	-0.43***	-0.36***	-0.41***
	(0.21)	(0.22)	(0.19)	(0.11)	(0.07)	(0.10)	(0.11)	(0.08)	(0.10)
Same Sex Baseline	25.7	26.7	20.8	4.2	2.0	4.9	6.3	3.1	4.3
Percent Effect	-3.5	-3.4	-4.2	-4.3	-9.0	-7.6	-6.8	-11.5	-9.5
Observations					164,733				
Sample of First-Born Men									
Second-Born Sister	0.40*	0.22	0.12	0.53**	0.26	0.33**	0.30*	0.09	0.09
	(0.24)	(0.23)	(0.19)	(0.22)	(0.20)	(0.15)	(0.15)	(0.13)	(0.14)
Same Sex Baseline	51.5	36.0	25.2	30.4	21.0	12.5	13.2	8.3	8.8
Percent Effect	0.8	0.6	0.5	1.7	1.2	2.6	2.3	1.1	1.0
Observations					173,340				

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Each Panel-Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last educational observation. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. *STEM Focus in First Enrollment* indicates whether the first place of enrollment after compulsory education is within STEM (vocational secondary or academic high school). *Academic High School STEM Track* indicates enrollment in and completion of academic high school from the math or technical tracks. *Vocational STEM* indicates enrollment in or completion of a vocational STEM program either at the secondary or tertiary level. *College STEM Major* indicates enrollment in and completion of a college STEM program.

The results are, broadly, comparable to other studies on STEM choice both in terms of the magnitude of the effects and finding largest effects for women. For instance, having a one standard deviation larger proportion of female math and science teachers in high school (Bottia et al., 2015) and in introductory courses in the U.S. Air Force Academy (Carrell et al., 2010) increases women's probability of graduating with a STEM college major by almost 10 percent (no effect for men). However, the latter study only finds an effect for women with above median math ability. Similarly, increasing the proportion of female peers in lower secondary education by one standard deviation in Austria increases the probability that girls choose a typical male track by 13 percent (Schneeweis and Zweimüller, 2012). Fischer (2017) further find that women who are enrolled in a class with a one standard deviation larger share of high ability peers in college introductory Chemistry classes are seven percent less likely to graduate in STEM, while men's STEM persistence is unaffected. Other studies on gender peer composition, in contrast, find that a larger share of female peers increases both men and women's probability of choosing a more gender-stereotypical college major (Brenøe and Zölitz, 2017; Zölitz and Feld, 2017).

4.2 Educational Performance and Attainment

The findings on STEM education could be due to changes in ability and educational attainment. If sibling gender largely impacts ability, an effect on field choice could simply be a rational response even though the interest in STEM fields stays constant. Appendix Table A4 shows that sibling gender composition has no effect on either girls or boys' school performance, a proxy for ability.[22] Moreover, one might worry that sibling gender could affect the probability of any enrollment and thereby enrollment in any field. Table 5 shows that sibling gender composition does not impact educational enrollment in or completion of any type of post-compulsory, vocational, or college education.[23] Consequently, these results demonstrate that sibling gender composition does not affect educational achievement or attainment, supporting an interpretation that changes in interests are the channel for the effects of sibling gender.

[22] Appendix Table A4 also shows that sibling gender composition does not affect the probability of being observed with any of the GPA measures. Appendix Figure A3 illustrates the distributions of the three GPA measures by gender and sibling gender composition. The differences by sibling gender are extremely small and distributional effects do not seem to be important.

[23] Neither do I find any effect on the probability of ever completing grade 9 or completing grade 9 on time (not reported).

Table 5
Educational Enrollment and Attainment by Age 30

	Post-Compulsory		Vocational		College	
	Enroll-ment (1)	Com-pletion (2)	Enroll-ment (3)	Com-pletion (4)	Enroll-ment (5)	Com-pletion (6)
Sample of First-Born Women						
Second-Born Brother	0.00 (0.12)	-0.09 (0.17)	-0.06 (0.24)	0.04 (0.23)	-0.04 (0.21)	-0.20 (0.22)
Same Sex Baseline	95.2	85.7	54.3	40.5	45.7	38.5
Percent Effect	0.0	-0.1	-0.1	0.1	-0.1	-0.5
Observations	164,733					
Sample of First-Born Men						
Second-Born Sister	-0.06 (0.10)	-0.20 (0.19)	0.13 (0.21)	-0.06 (0.24)	-0.06 (0.20)	-0.28 (0.20)
Same Sex Baseline	94.7	82.4	66.2	50.9	34.0	26.8
Percent Effect	-0.1	-0.2	0.2	-0.1	-0.2	-1.0
Observations	173,340					

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Each Panel-Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last educational observation. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. *Post-Compulsory* indicates enrollment in and completion of any type of education after primary school. *Vocational* refers to enrollment in and completion of vocational secondary and tertiary programs. *College* refers to enrollment in and completion of college programs.

4.3 Labor Market Outcomes

If the effects of sibling gender are truly due to changed interests, we would expect to see the effects on STEM choice to persist into labor market outcomes —and in particular, into occupational choice. To study this, I follow people well into their mid-career by observing their occupational choice and annual labor earnings at age 30, 35, and 40. I restrict the main sample to cohorts born between 1962 and 1977; however, the results for the age 30 outcomes are similar when including all cohorts. I define an individual to have a high-skilled STEM occupation if the mode occupation in a five year period up to the indicated age is within STEM fields.[24] Moreover, I consider the effects on the annual earnings percentile by age and cohort. The advantage of this measure is that it provides a standardized measure of relative earnings that includes individuals with zero earnings and is comparable across cohorts and ages.

Table 6 provides the results on labor market outcomes. The first three columns show the effect of sibling gender on the probability of being employed in a STEM occupation. For women, the effects on STEM education clearly persist into occupational choice. Women with a younger brother are respectively 6.5 and 9.0 percent less likely to work in a STEM occupation at age 30 and 40. Thus, the changes in educational STEM participation carry over into the labor market for women. In contrast, sibling gender composition does not affect men's probability of working within STEM at age 30 or 35. At age 40, men with a second-born sister are, however, borderline significantly more likely to work within STEM by 2.9 percent. Hence, they might be somehow more interested in STEM, as the results on STEM enrollment also suggest. This is, nevertheless, not a very robust finding.

Both men and women experience a negative effect of having an opposite sex sibling on earnings by around one-third of a percentile [Columns (4) to (6)].[25] The effect is, though, largest in magnitude and is more robust for women than men. Finding a negative effect of having a younger brother on women's earnings is not surprising, given the previous results of lower participation in (higher paying) field-specific STEM educations and STEM occupations. Similarly, Cools and Patacchini (2017) show that in the U.S. women with any brother earn less around the age of 30. Their estimated effect is, however, much larger in magnitude which might be due to the U.S. being a less gender equal society compared to Denmark, whereby the effect of sibling gender composition in the U.S. might be stronger on the development of gender identity. For men, the negative impact of having a younger sister on earnings might be explained by

[24] I use the Danish version of the International Standard Classification of Occupations (DISCO) to determine high-skilled STEM occupations. For the 08 version, I define these as sub-fields 21, 25, 31, and 35.

[25] Sibling gender composition has no effect on cumulated work experience or unemployment at the different ages (not reported).

a larger degree of educational mismatch given the effect on STEM enrollment does not persist into STEM completion; a negative impact is also consistent with the findings by Peter et al. (2015).

Table 6
STEM Occupation and Annual Labor Earnings Percentile by Cohort

	STEM Occupation			Earnings Percentile		
Age	30	35	40	30	35	40
	(1)	(2)	(3)	(4)	(5)	(6)
Sample of First-Born Women						
Second-Born	-0.23**	-0.30***	-0.40***	-0.30**	-0.41***	-0.38***
Brother	(0.11)	(0.11)	(0.12)	(0.14)	(0.14)	(0.14)
Same Sex Baseline	3.5	4.3	4.4	45.7	46.0	47.7
Percent Effect	-6.5	-6.9	-9.0	-0.7	-0.9	-0.8
Observations	120,621	119,967	119,034	120,621	119,967	119,034
Sample of First-Born Men						
Second-Born	0.08	0.12	0.35*	-0.18	-0.32**	-0.29*
Sister	(0.17)	(0.19)	(0.18)	(0.17)	(0.15)	(0.16)
Same Sex Baseline	8.8	11.8	12.1	63.3	63.6	62.5
Percent Effect	0.9	1.0	2.9	-0.3	-0.5	-0.5
Observations	126,983	126,354	124,933	126,983	126,354	124,933

All estimates are multiplied by 100 to express effects in percentage/percentile points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children with a younger biological sibling born within four years apart) restricted to cohorts born between 1962 and 1977. Each Panel-Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last observation for the specific outcome. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. *STEM Occupation* indicates whether mode occupation in a five year period through the indicated age is within STEM. *Earnings Percentile* is the percentile by cohort and age at the population level in annual labor earnings.

5 Mechanisms

5.1 Possible Mechanisms

So far, I have documented that the sibling gender composition in a family does matter for the formation of interests in STEM fields, especially for women. But *why* does sibling gender change men and women's likelihood of participating in STEM fields? —fields that are gender-stereotypical for men and the opposite for women. To investigate this question further, this subsection draws on the literature to identify relevant mechanisms, while the subsequent subsections provide some empirical evidence. Overall, I consider changes in identity to be the channel of the altered interests. The overarching argument is that individuals with an opposite sex sibling are more exposed to gender-stereotypical behavior and are therefore more inclined to acquire traditional gender norms. In this context, gender-stereotypical behavior could become more salient either through changes in the nature of child-sibling or child-parent interactions, including parental investments.[26]

First, parents might interact differently with their children depending on the gender composition in terms of attitudes and the quantity, quality, and content of time spent together. Assuming that both parents spend at least some time with their children, a traditional household specialization model suggests that parents gender-specialize their investment in children when having mixed sex children if mothers are more productive in creating female human capital and fathers are more effective in creating male human capital (Becker, 1973). Parents might also derive more utility from spending time with a same compared to an opposite sex child due to the type of activities done with the child. In both cases, parents of mixed sex children would gender-specialize, to a greater extent, than parents of same sex children.

McHale et al. (2003) suggest that because parents of mixed gender children have the opportunity to gender-differentiate their parenting, children with opposite gender siblings might have the strongest explicit gender-stereotypes. Endendijk et al. (2013) find some evidence that fathers with mixed sex children exhibit stronger gender-stereotypical attitudes than fathers with same sex children. Moreover, Stoneman et al. (1986) find that mothers of mixed sex children treat their children more gender-stereotypically than mothers of same sex children. Previous research has further documented that, overall, mothers talk more in general and more about interests and attitudes with daughters than sons (Maccoby, 1990; Leaper et al., 1998; Noller and Callan, 1990). Fathers, in contrast, talk more and spend more time with sons than daughters and have a greater emotional attachment to sons (Bonke and Esping-Andersen, 2009;

[26] Appendix A.3 provides a short overview of alternative mechanisms discussed in previous papers on sibling gender composition. These mechanisms are, however, not compatible with the empirical findings.

Morgan et al., 1988; Noller and Callan, 1990). Fathers are, furthermore, more likely to impose gender-stereotypical expectations on their sons than daughters and fathers dislike more often cross-gender-typed behavior among boys than do mothers (Burge, 1981; Freeman, 2007; Raag and Rackliff, 1998). Thus, these different pieces of evidence suggest that parents of mixed sex children gender-specialize their parenting more and thereby expose children to more gender-stereotypical behavior.

Second, the child and its sibling might interact differently depending on their gender composition. In particular, having a sibling of the opposite gender might make children more aware of "appropriate" behavior for their own gender and induce them to develop more gender-stereotypical attitudes and interests. Several studies have, for instance, shown that the presence of opposite gender peers increases gender-typed behavior in preschoolers [for references, see Raag and Rackliff (1998)]. The overall mechanism is in line with the same sex education literature, arguing that children, especially girls, acquire less gender-stereotypical interests when being together with same gender children only (Booth et al., 2014; Schneeweis and Zweimüller, 2012).[27] Previous studies show that same sex education makes girls relatively less risk-averse (Booth et al., 2014), that women tend to be less competitive when facing male competitors (Niederle and Vesterlund, 2011), and that STEM fields are perceived as more competitive (Buser et al., 2014). Therefore, having a sibling of the opposite gender might induce individuals —particularly women —to develop more gender-stereotypical preferences for STEM fields due to a greater awareness of gender through sibling interactions. In particular, Cools and Patacchini (2017) show that women with at least one brother develop more traditional gender attitudes relative to those without any brother. This mechanism is compatible with the results on STEM participation.

In sum, a particularly important mechanism for the observed effect of sibling gender on interests in STEM fields —that is possible to test for empirically—is differences in child-parent interactions. In the remainder of this section, I explore this mechanism in five different ways. First, in the daily child-parent interactions, we might observe that same gender parents of mixed sex children invest more quality time in their same sex child. Second, this might cause differences in the quality of the relationship between the child and its mother and father, respectively. Third, we might observe that mixed sex children exhibit more or are to a larger extent exposed to gender-stereotypical behavior due to differential parental behavior. Fourth, in the extreme case of parental divorce, we might expect that mixed sex children would be more likely to live with their same

[27] A further argument is that girls will perform better, especially in male-dominated subjects when taught in same sex classrooms. Some studies find improved (math) achievement among girl in same sex education and show evidence that mechanisms are a reduction in stereotype threat (Booth et al., 2013), improved self-confidence, and a more accurate self-assessment of math skills (Eisenkopf et al., 2015). However, other studies do not find an effect of same sex education on educational achievement (Doris et al., 2013; Jackson, 2012; Halpern et al., 2011).

sex parent compared to same sex children due to a larger degree of gender-specialized parenting. Fifth, if parents gender-specialize their parenting more when having mixed sex children, we would expect the effects of sibling gender on STEM preferences to be stronger for individuals with a more gender-stereotypical same gender parent, as more gender-stereotypical parents transfer more traditional preferences to their same sex children than less gender-stereotypical parents (Humlum et al., 2017). Thus, common for these predictions is that a parent of mixed sex children influences his or her same gender child more than a parent of same sex children.

5.2 Parental Time Investment

To investigate whether sibling gender composition affects child-parent interactions —and in particular, whether it affects parents' quality time investment—I draw on the Danish Longitudinal Survey of Children (DALSC).[28] The sample consists of 6,011 randomly sampled children born between September 15 and October 31, 1995 to a mother with Danish citizenship. The survey consists of five waves (1996, 1999, 2003, 2007, and 2011) and is unique due to its very detailed information on family socio-economic characteristics, family structure, and parental time use. For this analysis, I select first-born children who have a second-born sibling born within five calendar years apart.[29] I construct a parental time investment index measuring the number of times a week each parent does a particular quality time activity with the child. At age 7 and 11, both parents report how often they do different types of activities together with the child. I define quality time as playing with the child, helping with homework, doing out-of-school activities, reading/singing, and going on an excursion.[30]

Columns (1) through (4) in Table 7 provide the results on parental time investment by each parent for the two ages, separately. Mothers of a first-born daughter and a second-born son invest more time in their first-born daughter at both ages compared to mothers with two daughters. On average, mothers spend 0.7 activities more each week, corresponding to an increase of five and ten percent at age 7 and 11, respectively. In contrast, fathers invest 9–14 percent less time in their first-born daughter when having mixed sex children. This reduction in total paternal time investment is driven by decreased time spent helping with homework and reading for the daughter [Appendix Table A6]. This finding could indicate that girls with a younger brother receive less

[28] The study was designed by researchers from SFI, the Danish National Centre for Social Research, in cooperation with other research institutions.

[29] I only observe the year of birth of siblings and do therefore not have more precise information on the spacing. Restricting spacing to four years as for the main analysis gives similar results, although the smaller sample size reduces the precision of the estimates.

[30] Parents report how often they do these activities with the child; I code "almost daily" as 6 times a week, "2–3 times a week" as 2.5, "sometimes" as 0.5, and "never" as 0.

Table 7
Parental Time Investment and Housework at Age 7 and 11

	Parental Time Investment				Housework w Parents	
	Age 7		Age 11		Age 7	Age 11
	Mom	Dad	Mom	Dad		
	(1)	(2)	(3)	(4)	(5)	(6)
Sample of First-Born Girls						
Second-Born	0.68*	-0.76*	0.64*	-0.79**	0.17	-0.19
Brother	(0.38)	(0.45)	(0.37)	(0.37)	(0.30)	(0.29)
Same Sex Baseline	12.7	8.8	6.6	5.7	3.8	3.8
Percent Effect	5.4	-8.6	9.7	-13.8	4.5	-5.0
Observations	665	495	606	415	488	398
Sample of First-Born Boys						
Second-Born	-0.76*	-0.38	-0.47	0.06	-0.58**	-0.12
Sister	(0.39)	(0.42)	(0.36)	(0.41)	(0.28)	(0.30)
Same Sex Baseline	12.6	9.5	7.4	6.0	3.8	3.1
Percent Effect	-6.0	-4.0	-6.3	1.0	-15.4	-3.8
Observations	709	543	602	426	534	396

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. DALSC sample. Each Panel-Column represents the results from separate regressions. All models control for mother's and father's age (squared) and fixed effects for spacing to the younger sibling in years, parental marital status in 1996, parents having been together for at least 5 years in 1996, region of birth, maternal level of education, paternal level of education, and family income level in 1995. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. *Parental time investment* is measured as the total number of activities (playing, doing homework, doing out-of-school activities, reading/singing, going on an excursion) done together with the child at a weekly basis. *Housework with parents* measures the total number of housework activities (cooking, domestic chores) done by both parents together with the child at a weekly basis.

qualified help with STEM-related homework, which might prevent them from growing interests in these fields. Overall, girls receive the same amount of time investment regardless of their younger sibling's gender. These results clearly show that first-born girls with a second-born brother experience more gendered parenting.

For boys, the overall picture is similar. Mothers of mixed sex children invest less time in their first-born son relative to mothers with two sons; however, this is only statistically significant at age 7. This reduction in mothers' time spent with sons is mainly driven by a decrease in time spent playing and, to some extent, doing out-of-school activities with the son [Appendix Table A5]. In contrast, sibling gender composition does not affect fathers' total time investment in boys. This, however, masks some important findings when considering the individual components of the index: fathers of mixed sex children play less with their first-born son but seem to help more with homework and to read to the son at age 7 relative to fathers with two sons [Appendix Table A6]. Consequently, first-born boys with a second-born sister receive, on average, less total parental time investment, driven by less time spent playing. This might help explain why the effect of sibling gender does not persist into actual STEM graduation if boys are more responsive in their STEM pursuit to any decline in parental inputs relative to girls, regardless of the composition of the decline. Despite an overall decrease in parental time investment in sons, the findings still demonstrate that first-born boys with a second-born sister receive proportionally more male inputs. In conclusion, this analysis supports the hypothesis that parents of mixed sex children gender-specialize their parenting more than parents of same sex children.

5.3 Child-Parent Relations

Given the findings on parents' differential investment in first-born children by the second-born's gender, sibling gender might also affect the relationship between the child and its parents and thereby the strength of the transmission of parental preferences. Although measured at different ages, DALSC asks the mother, father, and child how each person perceives their relationship to the child/each parent. From these questions, I construct indexes based on principal component analysis with higher values reflecting better relationships [see Appendix Table A7]. Each index is standardized to have mean zero and standard deviation of one. Table 8 shows that fathers perceive the relationship to their first-born daughter worse at age 7 when having a second-born son compared to fathers with two daughters. Similarly, first-born girls with a second-born brother report worse quality of the relationship to their father at age 15. Meanwhile, sibling gender does not affect the relationship between parents and sons, although indications (not statistically significant) suggest that first-born sons have a slightly worse relationship to their mother when having a younger sister relative to a brother.

Table 8

Quality of Child-Parent Relations (Mean 0, SD 1)

	Mother's	Fathers'	Child's relationship to	
	Relationship to Child		Mother	Father
Child Age	11/15	7	15	15
	(1)	(2)	(3)	(4)
Sample of First-Born Girls				
Second-Born	-0.08	-0.23***	0.01	-0.17*
Brother	(0.09)	(0.09)	(0.09)	(0.10)
Same Sex Baseline	0.1	0.1	0.1	0.1
Observations	494	485	560	551
Sample of First-Born Boys				
Second-Born	0.08	0.05	-0.08	0.01
Sister	(0.09)	(0.10)	(0.08)	(0.07)
Same Sex Baseline	0.0	0.0	0.0	0.1
Observations	513	529	596	588

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. DALSC sample. Each Panel-Column represents the results from separate regressions. All models control for mother's and father's age (squared) and fixed effects for spacing to the younger sibling in years, parental marital status in 1996, parents having been together for at least 5 years in 1996, region of birth, maternal level of education, paternal level of education, and family income level in 1995. All child-parent relationship indexes represent the first component from principal component analyses shown in Appendix Table A7, are standardized such that a higher value reflects a better relationship, the mean is zero, and the standard deviation is one.

5.4 Exposure to Gender-Stereotypical Activities

As a further investigation of mechanisms, I consider whether first-born boys and girls are differentially affected by having a younger sibling of the opposite gender in terms of exposure to female-typed activities. With data from the DALSC, I construct an index measuring the total number of times a week the parents involve the child in housework activities (cooking and other domestic chores reported by each parent), which are traditionally perceived as female-typed activities. Sibling gender does not affect girls' involvement in housework with parents [Table 7, Columns (5) and (6)]. In contrast, at age 7, first-born boys with a second-born sister are 16 percent less involved in housework activities. This difference in housework involvement fades, however, out by age 11. These results suggest that boys with a younger sister at young ages are more exposed to gender-stereotypical behavior than those with a brother.[31]

5.5 Family Structure at Age 17

In the extreme case of parental divorce or separation (henceforth *divorce*), the living arrangement between parents and children might additionally help shed light on child-parent interactions in the main sample. If parents of mixed sex children gender-specialize more than parents of same sex children, first-born children with a second-born opposite sex sibling might be more likely to live with their same sex parent (SSP) in case of parental divorce. Moreover, a family living arrangement where the oldest child lives with the same sex parent and the younger child lives with the opposite sex parent (OSP) might be more prevalent. Yet, sibling gender composition might also affect the likelihood of living in a traditional family, defined as living with both biological parents. Table 9 studies how sibling gender composition affects family structure at age 17 for the main sample.[32] From this, it is clear that sibling gender composition does not alter the probability of living in a traditional family at age 17, neither for women nor for men [Columns (1) and (4), respectively].

Conditional on living in a non-traditional family, the results show a pattern consistent with the predictions. First-born girls with a second-born brother are more likely to live with their mother [Column (2)]. Furthermore, both first-born men and women with a second-born opposite sex sibling are more likely to live in a living arrangement in which they live with their same sex parent and their younger sibling lives with the opposite sex parent [Columns (3) and (6)]. For women (men) the estimated effect is 5.2 (3.6) percentage points, corresponding to an increase of 120 (26) percent relative to

[31] An alternative interpretation is that boys exhibit more gender-stereotypical behavior. However, I cannot test for this distinction.

[32] I observe the family structure on January 1^{st} each year and use the observation for the year the person turns 18 years or the last year in which the child lives with at least one biological parent.

Table 9
Family Structure at Age 17

	First-Born Women			*First-Born Men*		
Sample	All	Non-Traditional		All	Non-Traditional	
First-Born lives w	Both parents	SSP	SSP, sib w OSP	Both parents	SSP	SSP, sib w OSP
	(1)	(2)	(3)	(4)	(5)	(6)
Second-Born	-0.04	0.91**	5.23***	-0.07	0.47	3.55***
Opposite Sex	(0.18)	(0.39)	(0.27)	(0.20)	(0.47)	(0.38)
Same Sex Baseline	78.6	78.2	4.4	79.1	29.2	13.8
Percent Effect	-0.1	1.2	119.6	-0.1	1.6	25.7
Observations	162,564	34,922	34,745	171,416	35,913	35,736

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children with a second-born biological sibling born within four years apart), born 1962–1986. Each Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at observation of family structure. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. *All* includes everybody who lives with at least one biological parent, while *Non-Traditional* excludes those living with both biological parents. *SSP* indicates that the firstborn child lives with its biological same sex parent. *SSP, sib w OSP* indicates that the first-born child lives with its same sex parent and the second-born child lives with the opposite sex parent (opposite sex compared to the first child's gender).

the mean for women (men) with a same sex sibling. These results consequently show a strong effect on the living arrangement among non-traditional families and thereby support the previous findings (based on the much smaller DALSC sample) on more gender-specific parenting and time investment in families with mixed sex children.

5.6 Heterogeneity

Finally, this subsection studies heterogeneity in the effects of sibling gender composition on STEM preferences and ties the findings to the discussion on mechanisms. Table 10 explores heterogeneity by parents' field. For women, the effects are strongest among those with a mother who has a heavily female-dominated education or occupation, i.e. within either administration (e.g. secretary and office work) and health (e.g. nursing). In contrast, for men, the effects are concentrated among those with a father within STEM. Thus, those individuals who have a same sex parent with gender-specific human capital are the ones driving the effect of sibling gender. Meanwhile, having a gender-stereotypical opposite sex parent seems unimportant for heterogeneity in the effect.[33] This is consistent with the hypothesis that same sex parents of mixed sex children invest more time in their same sex child than parents of same sex children, as we would expect that parents with more gender-stereotypical human capital would reinforce gender-specialization to a larger extent than those parents with less gender-specific human capital. Additionally, Appendix Table A9 shows that the effect on field-specific STEM enrollment is particularly large for men who come from families where the parents have a traditional division of labor during childhood. Consequently, these heterogeneities indicate that differences in child-parent interactions are important for the effects of sibling gender composition on STEM interests.

Expanding the sample to include individuals spaced up to 15 years from their second-born sibling shows that sibling gender only affects STEM education for first-born women with spacing of less than five years and less than three years for men [Appendix Table A10; Appendix Figure A4]. Meanwhile, the estimated effects by spacing are not statistically significantly different from each other, probably due to the small fraction of children with very long spacing to their second-born sibling. This finding that individuals with long spacing to their younger sibling do not experience an effect of sibling gender might indicate the importance of sibling interactions. However, it could also be because parents with children spaced far apart treat the first-born child similarly regardless of the younger sibling's gender.[34]

[33] Appendix Table A8 shows heterogeneity by parental length of education. The effects are, generally, concentrated among individuals with a high educated (≥ 12 years of schooling) same sex parent.

[34] Despite large changes in society over these 25 birth cohorts, the effects do not differ systematically by decade of birth (Appendix Table A11). This is consistent with the finding by Haines et al. (2016) that gender-stereotypes have not changed over the last three decades in the U.S.

Table 10

Field-Specific STEM Education: Heterogeneity by Parental Field

Sample of	*First-Born Women*		*First-Born Men*	
	Enroll-ment (1)	Com-pletion (2)	Enroll-ment (3)	Com-pletion (4)
Second-Born	-0.08	-0.21	0.62	0.08
Opposite Sex (SBOS)	(0.20)	(0.15)	(0.38)	(0.34)
SBOS × Mom Admin	-1.41***	-0.71*	-1.00	-0.69
	(0.49)	(0.38)	(0.75)	(0.68)
SBOS × Mom Health	-0.66**	-0.63***	-0.04	-0.13
	(0.29)	(0.22)	(0.53)	(0.48)
SBOS × Dad STEM	-0.05	-0.08	0.76*	0.88**
	(0.28)	(0.22)	(0.45)	(0.42)
N	164,632		173,262	

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Each Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, age at last educational observation, and the parental fields, which are interacted with SBOS. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. *Field-specific STEM* education refers to vocational and college programs within STEM excluding Biology.

6 Robustness Checks

6.1 Family Size

As discussed in Section 2, sibling gender composition affects family size but family size does not seem to strongly affect STEM participation (Appendix A.2). However, to further test the robustness of the main results to family size, this subsection applies three different strategies: 1) to flexibly control for family size, 2) to divide the sample by family size, and 3) to study the effect of having a co-twin of the opposite sex. Although family size is endogenous to sibling gender composition, strategy (1) and (2) is useful to the degree that it informs about the sensitivity of the results while keeping the potential biases in mind. These robustness analyses, together with the evidence of no differential effect by sibling gender on educational attainment and the supplementary analysis of the effect of family size on STEM participation, provide convincing evidence that family size does not confound the effects of sibling gender composition.

The first strategy, controlling for family size, may bias the estimates of sibling gender because family size is an outcome of sibling gender composition. Therefore, accounting for family size might lead to a bad control problem. In other words, if the effect of having an opposite sex sibling goes through family size, the estimate of sibling gender would be attenuated when controlling for family size. Yet, one could also view family size as an omitted variable if family size has an independent effect on STEM participation. In such case, the estimated effect (when omitting family size) would be upward biased if the effect of family size on STEM preferences is negative and downward biased if it is positive. In Table 11, the first row in each panel repeats the main results, while the second row shows the estimates of sibling gender when flexibly accounting for family size.[35] Overall, the estimates are extremely similar.

[35] I flexibly account for family size by including dummies for the number of biological siblings and dummies for the number of children the mother and father potentially have, respectively, from later relationships.

Table 11

STEM Education: Controlling for and Splitting by Family Size

	Field-Specific STEM				Academic HS		Vocational STEM		College STEM	
	Enrollment		Completion		STEM Track		(Any Level)		Excl. Biology	
	Excl. Biology	Incl. Biology	Excl. Biology	Incl. Biology	Enroll-ment	Com-pletion	Enroll-ment	Com-pletion	Enroll-ment	Com-pletion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sample of First-Born Women										
Main Estimates	-0.48***	-0.51***	-0.53***	-0.58***	-0.91***	-0.87***	-0.18*	-0.18***	-0.37***	-0.36***
($N = 164,733$)	(0.14)	(0.15)	(0.10)	(0.11)	(0.22)	(0.19)	(0.11)	(0.07)	(0.10)	(0.08)
Family Size Controls	-0.50***	-0.53***	-0.55***	-0.59***	-0.87***	-0.82***	-0.21*	-0.19***	-0.37***	-0.36***
($N = 164,733$)	(0.14)	(0.15)	(0.10)	(0.11)	(0.22)	(0.19)	(0.11)	(0.07)	(0.10)	(0.08)
1 Sibling	-0.70***	-0.71***	-0.78***	-0.81***	-1.08***	-0.96***	-0.24*	-0.22**	-0.52***	-0.56***
($N = 93,285$)	(0.20)	(0.21)	(0.14)	(0.16)	(0.27)	(0.25)	(0.14)	(0.09)	(0.15)	(0.11)
2+ Siblings	-0.28	-0.33	-0.34**	-0.42**	-0.71*	-0.77**	-0.09	-0.13	-0.32*	-0.21
($N = 54,634$)	(0.23)	(0.24)	(0.17)	(0.18)	(0.37)	(0.34)	(0.17)	(0.11)	(0.18)	(0.13)
Sample of First-Born Men										
Main Estimates	0.80***	0.77***	0.32	0.33	0.22	0.12	0.53**	0.26	0.33**	0.09
($N = 173,340$)	(0.23)	(0.23)	(0.22)	(0.22)	(0.23)	(0.19)	(0.22)	(0.20)	(0.15)	(0.13)
Family Size Controls	0.66***	0.63***	0.18	0.18	0.19	0.11	0.39*	0.12	0.32**	0.08
($N = 173,340$)	(0.23)	(0.23)	(0.22)	(0.22)	(0.23)	(0.19)	(0.22)	(0.20)	(0.15)	(0.13)
1 Sibling	1.02***	1.01***	0.59*	0.59*	0.38	0.40	0.73**	0.54**	0.38*	0.09
($N = 96,248$)	(0.32)	(0.32)	(0.31)	(0.31)	(0.30)	(0.26)	(0.30)	(0.27)	(0.21)	(0.18)
2+ Siblings	0.15	0.07	-0.20	-0.19	0.02	-0.10	-0.16	-0.33	0.27	0.14
($N = 59,556$)	(0.40)	(0.40)	(0.37)	(0.37)	(0.35)	(0.32)	(0.38)	(0.34)	(0.25)	(0.22)

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Each cell presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last educational observation. *Family Size Controls*-models further include dummies for the number of biological siblings and dummies for the number of children the mother and father potentially have, respectively, from later relationships. *1 Sibling*-models restrict the sample to those who only have one full sibling and no half-siblings. *2+ Siblings*-models restrict the sample to those who have at least two full siblings and no half-siblings.

The second strategy is to split the sample by family size.[36] However, note that since family size is endogenous, this robustness check comes with a selection problem. Imagine that those parents with two same sex children who are very gender-stereotypical and have a gender preference for the opposite gender compared to the gender of their children always progress to the third parity. In that case, first-borns with a second-born same sex sibling who only have one sibling would come from less gender-stereotypical families compared to those who have at least two siblings. Therefore, we would expect the effect of having an opposite sex sibling to be larger in magnitude for the one-sibling sample than for the entire sample. Reversely, we would expect individuals with a younger same sex sibling who have at least two younger siblings to come from more gender-stereotypical families, implying that the effect of having an opposite sex sibling would be smaller in magnitude than for the total sample. This is exactly what the results in the third and fourth rows show in Table 11. In fact, the estimates for the sample with at least two siblings are much smaller in magnitude and insignificant in most cases.[37]

Finally, to circumvent potential confounding effects from family size, I examine the effect of having a co-twin of the opposite gender as an alternative empirical strategy.[38] The key empirical feature of the sample of twins is that twin gender composition only has a very limited impact on family size [Column (1)], especially for twins born at the first parity. Overall, the effects of having a co-twin of the opposite gender on STEM choice, both for the sample of all twins and twins born at the first parity, are very similar to the main results [Appendix Table A12]. The magnitude of the effects is, however, much larger. This might be due to the much greater intensity of the exposure to a co-twin compared to a younger sibling.

6.2 The Effect of an Older Sibling's Gender

Despite the potential problems with selection bias from estimating the effect of an older sibling's gender, as discussed in Section 2, we would still expect the direction of the effect on STEM participation to be the same. Such analysis can thus serve as a robustness check. Considering the potential bias, if parents who prefer a son want their son to be more gender-stereotypical than parents without son-preferences, the effect of having an older sister on the sample of second-born sons would be upward biased, according to the example on parental gender preferences in Section 2. Appendix Table

[36] I restrict the sample to individuals who only have biological siblings, i.e. none of their parents have children with another person than the parent; though the results are similar when including those with half-siblings.

[37] However, the insignificance might partly be due to smaller sample sizes.

[38] This approach is similar to the one in Peter et al. (2015) with the caveat that I do not have information on zygocity.

A13 shows the results from an analysis of the associations between having an older opposite sex sibling and STEM participation for a sample of second-born children. Overall, these results are similar to the main results on STEM education. However, for men, the effects on any field-specific STEM enrollment and completion are around three times larger than for the main results, which might both be due to selection bias and to a role model effect of the older sibling. These results are also closer to the ones in Anelli and Peri (2014) who do not find a significant effect for women although the magnitude of their estimate (-1.3 percentage point for enrollment in a high earnings college major) is larger than my corresponding estimate (-0.2 percentage points for enrollment in a STEM excluding Biology college major).

6.3 Alternative Measures of Field of Study

As a final test of the robustness of the main findings, I consider alternative measures of field of study. First, I use OECD (2016)'s definition of STEM to include Natural Sciences, Mathematics, Statistics, Information and Communication Technologies, Engineering, Manufacturing, and Construction (ISCED fields 51–73). Compared to my preferred definition of STEM, this alternative definition includes Biology, Manufacturing, and Construction and excludes Economics. Appendix Table A14 shows that the results for this alternative definition are very similar to the main findings [Columns (1) to (4)]. Second, instead of considering traditionally male-dominated fields, I consider care fields (Education, Health, and Welfare; ISCED fields 11, 91, and 92), which are traditionally female-dominated. The results on choosing an education within care fields stress the main finding that having an opposite sex sibling makes both men and women's interests more gender-stereotypical [Columns (5) to (8)]. These results demonstrate larger percent effects, due to a lower baseline, for men than women and compared to the main results on STEM choice and display a more consistent finding that men with a younger sister are more likely to opt out of female-typed fields. Third, I consider the narrow field of the highest completed education by age 30 (28 mutually exclusive groups). For women [Appendix Figures A6], the negative effect of having a younger brother on STEM completion is driven by Economics and Engineering.[39]

7 Conclusion

This study documents that the family environment has a powerful long-run impact on especially women's participation in traditionally male-dominated STEM fields. The results suggest that having an opposite sex sibling increases the probability of choosing

[39] Appendix Figure A7 shows the corresponding results for men without any consistent pattern.

a gender-stereotypical field of education. Women opt out of STEM already at the time of their first active educational choice at the end of 9^{th} grade. Men, on the contrary, show an increased interest in STEM fields, but are not more likely to complete a STEM education. The altered participation in STEM fields persist into occupational choice for women and has negative consequences for their earnings. An important mechanism for these findings is the effect on child-parent interactions. Parents with mixed sex children gender-specialize their parenting more and invest more quality time in their same sex child than parents with same sex children.

My findings emphasize that if policy makers want to increase the number of people —and particularly women—within STEM fields, they need to focus on early educational choices made already at the end of compulsory schooling. However, attention to decisions at this educational stage is not sufficient. As my analysis of mechanisms stresses, the family —representing a central aspect of the social environment —influences the formation of STEM preferences throughout childhood. Moreover, no evidence shows that men possess an inherent advantage over women in math ability: boys and girls start school with similar math performance; yet, around the time of puberty, the gender difference in average math performance (favoring boys) stabilizes (Kahn and Ginther, 2017). This suggests that social environmental factors influence the way in which boys and girls develop interests and abilities within STEM fields already during early school grades. Consequently, if society wants to give boys and girls the same opportunities in terms of labor market performance in adulthood, policy makers would need to focus on how to counteract the transmission of gender norms across generations and thereby the development of gender-stereotypical behaviors, attitudes, and preferences.

References

Almond, D. and L. Edlund (2007). Trivers–willard at birth and one year: evidence from us natality data 1983–2001. *Proceedings of the Royal Society of London B: Biological Sciences 274*(1624), 2491–2496.

Almond, D. and B. Mazumder (2011). Health capital and the prenatal environment: the effect of ramadan observance during pregnancy. *American Economic Journal-Applied Economics 3*(4), 56.

Altonji, J. G., E. Blom, and C. Meghir (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annu. Rev. Econ. 4*(1), 185–223.

Amin, V. (2009). Sibling sex composition and educational outcomes: A review of theory and evidence for the UK. *Labour 23*(1), 67–96.

Anelli, M. and G. Peri (2014). Gender of siblings and choice of college major. *CESifo Economic Studies 61*(1), 53–71.

Anelli, M. and G. Peri (2016). The effects of high school peers' gender on college major, college performance and income.

Angrist, J., V. Lavy, and A. Schlosser (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics 28*(4), 773–824.

Angrist, J. D. and W. N. Evans (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review 88*(3), 450.

Atkinson, R. D. and M. J. Mayo (2010). Refueling the us innovation economy: Fresh approaches to science, technology, engineering and mathematics (stem) education.

Bauer, T. and I. Gang (2001). Sibling rivalry in educational attainment: The german case. *Labour 15*(2), 237–255.

Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political economy 81*(4), 813–846.

Bedard, K. and O. Deschenes (2005). Sex preferences, marital dissolution, and the economic status of women. *Journal of human Resources 40*(2), 411–434.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). The more the merrier? the effect of family size and birth order on children's education. *The Quarterly Journal of Economics 120*(2), 669–700.

Blau, F. D. and L. M. Kahn (2016). The gender wage gap: Extent, trends, and explanations.

Bonke, J. and G. Esping-Andersen (2009). Parental investments in children: how educational homogamy and bargaining affect time allocation. *European Sociological Review 10*(20), 1–13.

Booth, A., L. Cardona-Sosa, and P. Nolen (2014). Gender differences in risk aversion: Do single-sex environments affect their development? *Journal of economic behavior & organization 99*, 126–154.

Booth, A. L., L. Cardona, and P. J. Nolen (2013). Do single-sex classes affect exam scores? an experiment in a coeducational university.

Bottia, M. C., E. Stearns, R. A. Mickelson, S. Moller, and L. Valentino (2015). Growing the roots of stem majors: Female math and science high school faculty and the participation of students in stem. *Economics of Education Review 45*, 14–27.

Brenøe, A. A. and S. Lundberg (2017). Gender gaps in the effects of childhood family environment: Do they persist into adulthood? *European Economic Review*.

Brenøe, A. A. and U. Zölitz (2017). High school gender composition and STEM study choice.

Brim, O. G. (1958). Family structure and sex role learning by children: A further analysis of helen koch's data. *Sociometry 21*(1), 1–16.

Burge, P. L. (1981). Parental child-rearing sex-role attitudes related to social issue sex-role attitudes and selected demographic variables. *Family and Consumer Sciences Research Journal 9*(3), 193–199.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness and career choices. *The Quarterly Journal of Economics 129*(3), 1409–1447.

Butcher, K. F. and A. Case (1994). The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics 109*(3), 531–563.

Carrell, S. E., M. E. Page, and J. E. West (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics 125*(3), 1101–1144.

Cheng, A., K. Kopotic, and G. Zamarro (2017). Can parents growth mindset and role modelling address stem gender gaps? *EDRE Working Paper*.

Conley, D. (2000). Sibship sex composition: Effects on educational attainment. *Social Science Research 29*(3), 441–457.

Cools, A. and E. Patacchini (2017). Sibling Gender Composition and Women's Wages. *IZA Working Paper* (11001).

Cyron, L., G. Schwerdt, and M. Viarengo (2017). The effect of opposite sex siblings on cognitive and noncognitive skills in early childhood. *Applied Economics Letters*, 1–5.

Dahl, G. B. and E. Moretti (2008). The demand for sons. *The Review of Economic Studies 75*(4), 1085–1120.

Doris, A., D. ONeill, and O. Sweetman (2013). Gender, single-sex schooling and maths achievement. *Economics of Education Review 35*, 104–119.

Eisenkopf, G., Z. Hessami, U. Fischbacher, and H. W. Ursprung (2015). Academic performance and single-sex schooling: Evidence from a natural experiment in switzerland. *Journal of Economic Behavior & Organization 115*, 123–143.

Endendijk, J. J., M. G. Groeneveld, S. R. van Berkel, E. T. Hallers-Haalboom, J. Mesman, and M. J. Bakermans-Kranenburg (2013). Gender stereotypes in the family context: Mothers, fathers, and siblings. *Sex Roles 68*(9-10), 577–590.

Favara, M. (2012). The cost of acting'girly': Gender stereotypes and educational choices. *Working Paper*.

Fischer, S. (2017). The downside of good peers: How classroom composition differentially affects men's and women's stem persistence. *Labour Economics*.

Freeman, N. K. (2007). Preschoolers perceptions of gender appropriate toys and their parents beliefs about genderized behaviors: Miscommunication, mixed messages, or hidden truths? *Early Childhood Education Journal 34*(5), 357–366.

Gallen, Y., R. Lesner, and R. Vejlin (2017). The labor market gender gap in denmark: Sorting out the past 30 years.

Gielen, A. C., J. Holmes, and C. Myers (2016). Prenatal testosterone and the earnings of men and women. *Journal of Human Resources 51*(1), 30–61.

Haines, E. L., K. Deaux, and N. Lofaro (2016). The times they are a-changing or are they not? a comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly*, 0361684316634081.

Halpern, D. F., L. Eliot, R. S. Bigler, R. A. Fabes, L. D. Hanish, J. Hyde, L. S. Liben, and C. L. Martin (2011). The pseudoscience of single-sex schooling. *Science 333*(6050), 1706–1707.

Hamoudi, A. and J. Nobles (2014). Do daughters really cause divorce? stress, pregnancy, and family composition. *Demography 51*(4), 1423–1449.

Hauser, R. M. and H.-H. D. Kuo (1998). Does the gender composition of sibships affect women's educational attainment? *Journal of Human Resources*, 644–657.

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

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. *National Bureau of Economic Research*.

Humlum, M. K., A. Nandrup, and N. Smith (2017). Closing or reproducing the gender gap? parental transmission, social norms and education choice.

Jackson, C. K. (2012). Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in trinidad and tobago. *Journal of Public Economics 96*(1), 173–187.

Kabtek, J. and D. Ribar (2017). Teenage daughters as a cause of divorce.

Kaestner, R. (1997). Are brothers really better? sibling sex composition and educational achievement revisited. *Journal of Human Resources 32*(2).

Kahn, S. and D. Ginther (2017). Women and stem. Technical report, National Bureau of Economic Research.

Kirkebøen, L., E. Leuven, and M. Mogstad (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, qjw019.

Koch, H. L. (1955). Some personality correlates of sex, sibling position, and sex of sibling among five-and six-year-old children. *Genetic Psychology Monographs*.

Lavy, V. and E. Sand (2015). On the origins of gender human capital gaps: Short and long term consequences of teachers stereotypical biases.

Leaper, C., K. J. Anderson, and P. Sanders (1998). Moderators of gender effects on parents' talk to their children: a meta-analysis. *American psychologist 34*(1), 3–27.

Lee, S., L. J. Turner, S. Woo, and K. Kim (2016). The impact of school and classroom gender composition on educational achievement.

Maccoby, E. E. (1990). Gender and relationships: A developmental account. *American psychologist 45*(4), 513–520.

McHale, S. M., A. C. Crouter, and S. D. Whiteman (2003). The family contexts of gender development in childhood and adolescence. *Social development 12*(1), 125–148.

Morgan, S. P., D. N. Lye, and G. A. Condran (1988). Sons, daughters, and the risk of marital disruption. *American journal of sociology 94*(1), 110–129.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annu. Rev. Econ. 3*(1), 601–630.

Noller, P. and V. J. Callan (1990). Adolescents' perceptions of the nature of their communication with parents. *Journal of Youth and Adolescence 19*(4), 349–362.

OECD (2016). *Education at a Glance 2016: OECD Indicators.*

Oguzoglu, U. and S. Ozbeklik (2016). Like father, like daughter (unless there is a son): Sibling sex composition and womens stem major choice in college.

Peri, G., K. Shih, and C. Sparber (2015). Stem workers, h-1b visas, and productivity in us cities. *Journal of Labor Economics 33*(S1), S225–S255.

Peter, N., P. Lundborg, and D. H. Webbink (2015). The effect of sibling's gender on earnings, education and family formation.

Powell, B. and L. C. Steelman (1989). The liability of having brothers: Paying for college and the sex composition of the family. *Sociology of Education*, 134–147.

Raag, T. and C. L. Rackliff (1998, May). Preschoolers' awareness of social expectations of gender: Relationships to toy choices. *Sex Roles 38*(9), 685–700.

Schneeweis, N. and M. Zweimüller (2012). Girls, girls, girls: Gender composition and female school choice. *Economics of Education review 31*(4), 482–500.

Stoneman, Z., G. H. Brody, and C. E. MacKinnon (1986). Same-sex and cross-sex siblings: Activity choices, roles, behavior, and gender stereotypes. *Sex Roles 15*(9), 495–511.

Trivers, R. L. and D. E. Willard (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science 179*(4068), 90–92.

Zölitz, U. and J. Feld (2017). The effect of peer gender on major choice and occupational segregation. *Mimeo*.

A Appendix

A.1 The Selection Bias Problem

To show the selection bias problem more formally, I here follow Peter et al. (2015). Assume a latent outcome $Y_i^* = \alpha + \beta G_i^{old} + X_i'\gamma + \epsilon_i$, where G_i^{old} is the gender of the older sibling and X_i is a vector of observable exogenous characteristics. ϵ_i contains other relevant unobservable variables, such as parental gender preferences denoted by P_i, and $E[\epsilon_i] = 0$. The bias arises because of the latent nature of Y_i^*, as we only observe the outcome if child i is born. In other words, $Y_i = Y_i^*$ if the child is born ($S_i = 1$) and Y_i is missing if the child is not born ($S_i = 0$). The selection depends both on parental preferences and the older child's gender, $S_i = f(P_i, G_i^{old})$. We can only estimate the effect for the sample of children who are born which gives the expected value of Y_i:

$$E[Y_i|S_i = 1, G_i^{old}, X_i] = \alpha + \beta G_i^{old} + \gamma X_i + E[\epsilon_i|S_i = 1, G_i^{old}, X_i] \tag{2}$$

$$= \alpha + \beta G_i^{old} + \gamma X_i + E[\epsilon_i|f(P_i, G_i^{old}) = 1, G_i^{old}, X_i].$$

As long as selection depends on the first child's gender and parental preferences affect the way in which parents raise their children $E[\epsilon_i|f(P_i, G_i^{old}) = 1, G_i^{old} = 1, X_i] \neq E[\epsilon_i|f(P_i, G_i^{old}) = 1, G_i^{old} = 0, X_i]$. This implies that the estimate of the older sibling's gender is biased.

A selection problem could also arise in the absence of parental gender preferences. Assume that first-born children have n normally-distributed traits, such as how easy the child is to take care of and how well it behaves. Suppose parents only want a second child if their first child has a value of each trait above a certain threshold. The threshold for or the distribution of each trait could be gender-specific. In both cases, parents who progress to the next parity would, on average, have different types of first-born children depending on the child's gender. For instance, if boys and girls have the same distribution of how well they behave but parents require girls to behave better than boys to have a second child, second-born children would, on average, have a better behaving older sibling if they have a sister compared to a brother. In this example, the estimated effect of the older sibling's gender on the younger child's outcomes might thus be due to the older sibling's behavior rather than due to his or her gender.

A.2 Does Family Size affect STEM Education?

Black et al. (2005) use twins as an instrument for family size to show that family size does not affect educational attainment, using Norwegian registry data; Angrist et al. (2010) find the same for Israel. However, they only consider length of schooling and

Table A1
Family Size and STEM Education using Twins as Instrument

	First Stage	Second Stage				
	# of	Years of	Field-spec STEM		College STEM	
	Siblings	educa-tion	Enroll-ment	Com-pletion	Enroll-ment	Com-pletion
	(1)	(2)	(3)	(4)	(5)	(6)
Sample of First-Born Women ($N = 166,213$)						
Twins at 2^{nd} parity	0.74***					
	(0.018)					
# of Siblings		0.03	-0.87	-1.13*	-0.15	-0.93*
		(0.07)	(0.92)	(0.63)	(0.76)	(0.52)
F-statistic of IV	1735.29					
Prob>F	< 0.001					
Average	1.6	13.5	8.5	4.8	4.7	2.9
Effect×0.07		0.00	-0.06	-0.08	-0.01	-0.07
Percent effect×0.07		0.02	-0.72	-1.66	-0.23	-2.22
Sample of First-Born Men ($N = 175,032$)						
Twins at 2^{nd} parity	0.72***					
	(0.016)					
# of Siblings		-0.11	-0.67	0.81	0.25	-0.01
		(0.08)	(1.58)	(1.56)	(1.01)	(0.88)
F-statistic of IV	2042.57					
Prob>F	< 0.001					
Average	1.7	13.3	41.2	28.7	12.7	8.3
Effect×0.08		-0.01	-0.05	0.06	0.02	0.00
Percent effect×0.08		-0.07	-0.13	0.22	0.16	-0.01

All estimates for binary outcomes (enrollment and completion) are multiplied by 100. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample including twin siblings born at second parity (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Each Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last educational observation. *# of Siblings* measures the total number of siblings the individual has, including full and half siblings. The effects are multiplied by 0.07 (*Effect×0.07*) for women and by 0.08 for men, because these are the effects of having an opposite sex sibling on the number of siblings.

not the probability of enrolling in or completing a field-specific STEM degree. In this supplementary analysis, I show, consistent with their findings, employing a similar strategy in the Danish context, that family size does not affect educational attainment or field-specific STEM enrollment.

I use a sample with similar sample restrictions as for the main sample (see Subsection 3.1) with the exception that I include firstborn singleton children who have younger twin siblings born at the second parity.[40] The instrument for family size is having twins at the second parity. Column (1) in Appendix Table A1 shows that the instrument is strong and relevant; see Angrist et al. (2010) and Black et al. (2005) for a discussion of the validity of the instrument.

Appendix Table A1 shows that family size does not significantly affect the length of highest completed education by age 30 or the probability of any field-specific or college STEM enrollment. Moreover, family size does not affect men's probability of STEM completion. Meanwhile, the effect of family size is borderline statistically significantly negative for women, suggesting that having more siblings reduces the probability of completing any STEM degree. This might suggest that for women the estimates of having a younger brother on STEM completion in the main analysis are conservative. This potential downward bias would, however, only be small, as first-born women with a second-born sister, on average, have 0.07 more siblings than first-born women with a second-born brother. This is what the statistic *Effect*$\times$*0.07* illustrates in the table.

A.3 Alternative Mechanisms

This appendix describes alternative mechanisms to the ones discussed in Subsection 5.1. These mechanisms are, however, not compatible with the empirical findings.

The effect of sibling interactions might also go in the opposite direction for two reasons. First, the spillover model in developmental psychology hypothesizes that siblings imitate and influence each other with their gender-specific traits. For instance, Brim (1958) and Koch (1955) show that mixed sex siblings exhibit more traits of the opposite gender and fewer of their own gender compared to same sex sibling pairs. Second, the reference group theory in sociology suggests that as soon as a same sex sibling is present in the family, the same sex sibling will be the child and parents' reference group (Butcher and Case, 1994). Therefore, having a same sex sibling might induce the child to behave more gender-stereotypically.

Studies examining the relationship between sibling gender composition and educational attainment have argued that budget constraints may play an important role (Amin, 2009; Butcher and Case, 1994). If parents face no borrowing constraints, they

[40] I include all multiple birth; twins, however, represent the vast majority of all multiple births.

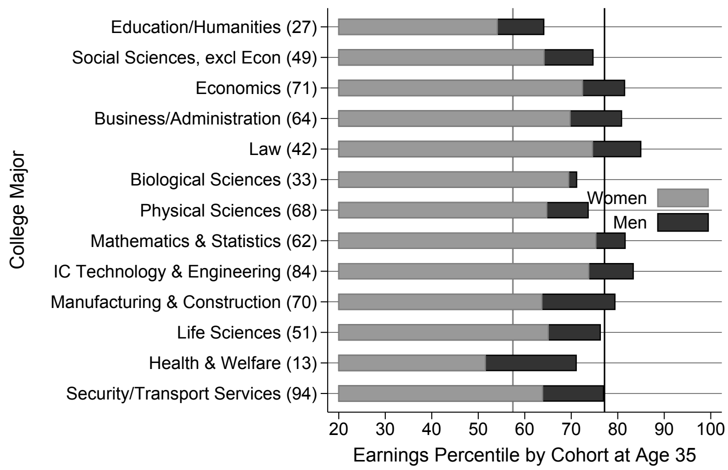
should, according to standard economic theory, invest in each child until marginal costs equal marginal benefits. However, if parents face borrowing constraints, they might decide to allocate their financial resources depending on the gender composition of their children. If parents want income equality between their children and the returns to education are smaller for women than men, then having a brother instead of a sister would be beneficial. However, parental aversion to income inequality cannot be the dominating channel, as we would otherwise have observed that having a sibling of the opposite gender should make the educational choice less gender-stereotypical.

In contrast, parents might want to maximize the total income of their children, thereby investing more in the child with the greatest returns to education. If returns to education are larger for men than women, having a brother would have adverse effects on educational attainment. In support of this argument, Powell and Steelman (1989) find for students enrolled in one college in the U.S. that the number of brothers puts more pressure on parents' financial support than do the number of sisters. Nevertheless, this is not a likely mechanism in the Danish context because there is no tuition fee at any educational level. Moreover, students in vocational training typically receive apprenticeship wages and students in tertiary education receive governmental student grants and loans to cover living expenses. For all cohorts in the analysis, students in tertiary education have at least had access to a combination of grants and loans of 1,000 USD a month in 2017-prices. It is also less clear how borrowing constraints should affect field choice, given sibling gender composition has no effect on the probability of enrolling in any type of program after compulsory education. Moreover, a more recent study shows that, for later generations in the U.S., parents to at least one son compared to parents with no sons do not differentially invest in their daughters (Cools and Patacchini, 2017).

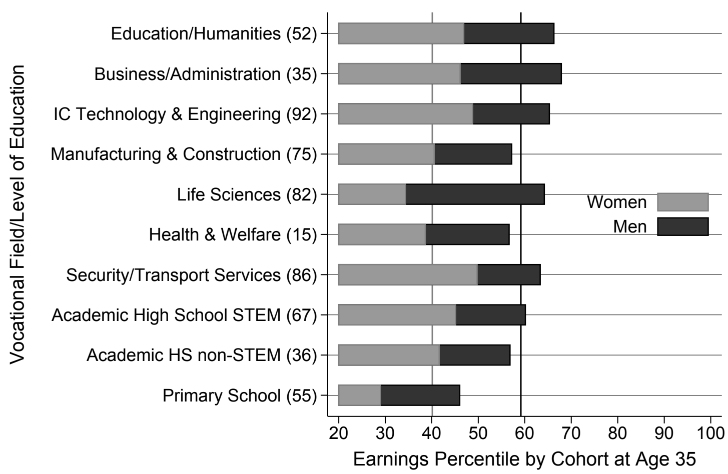
A.4 Appendix Figures and Tables

Figure A1

Average Earnings Percentile at age 35 by Field and Type of Education



(a) College Major

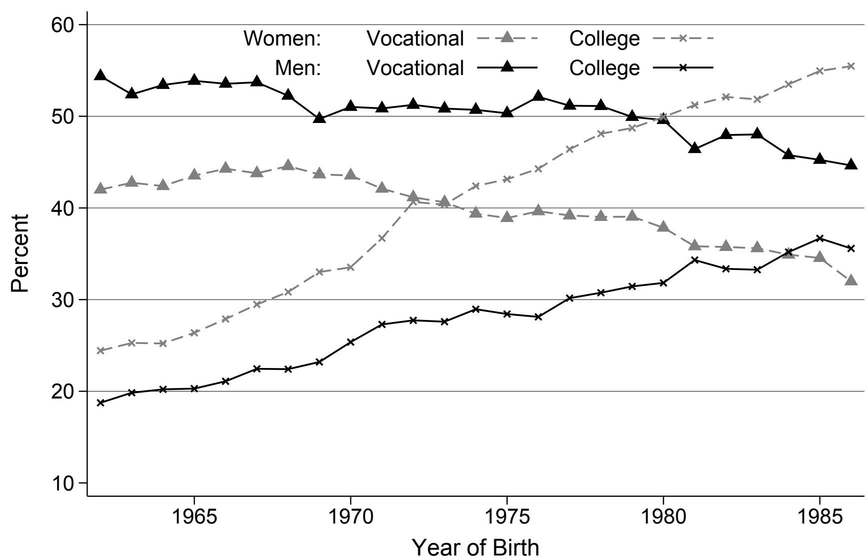


(b) Vocational Field/Level of Education

Note: Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Both graphs show separately by gender the average earnings percentile by birth cohort at age 35 by field and level of highest completed education by age 30. The number shown in parenthesis for each field label indicates the proportion of men in the specific group. The vertical lines indicate the mean earnings percentile for women (gray) and men (black) in each graph.

Figure A2

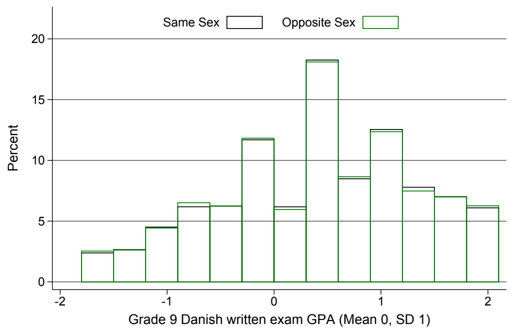
Educational Attainment at Age 30 by Gender Across Cohorts



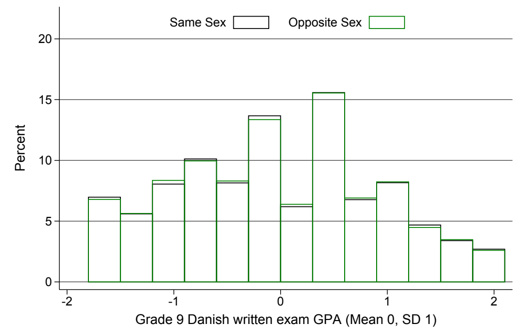
Note: Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). The graph illustrates the share of a cohort by gender completing at least vocational (secondary/tertiary) education and at least college education, respectively.

Figure A3

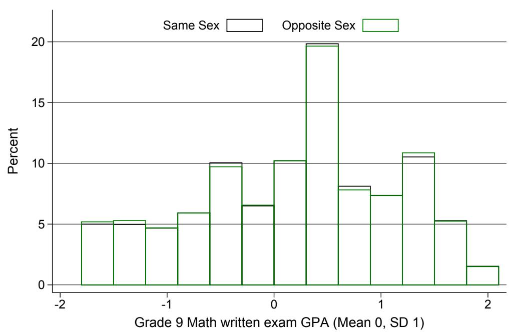
Distribution of Ability by Sibling Gender Composition



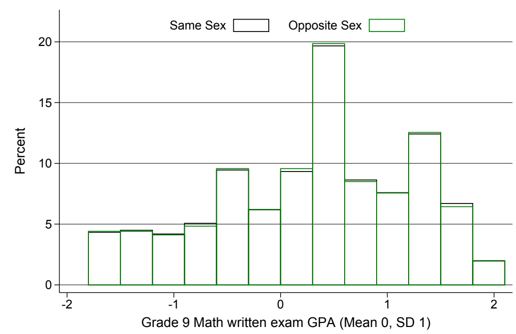
(a) Girls: Grade 9 Danish written exam



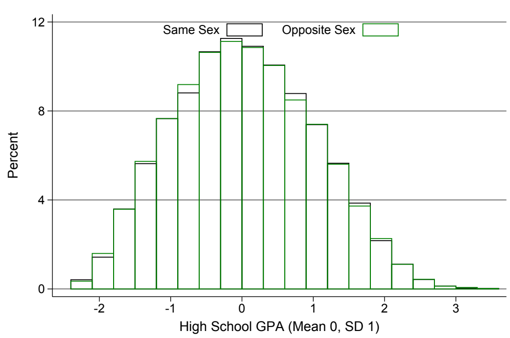
(b) Boys: Grade 9 Danish written exam



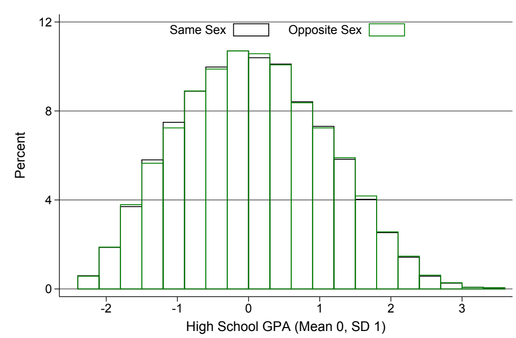
(c) Girls: Grade 9 Math written exam



(d) Boys: Grade 9 Math written exam



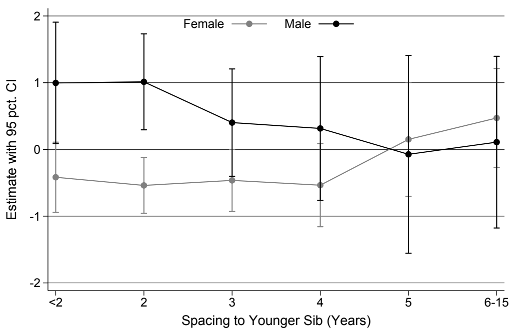
(e) Women: Academic HS GPA



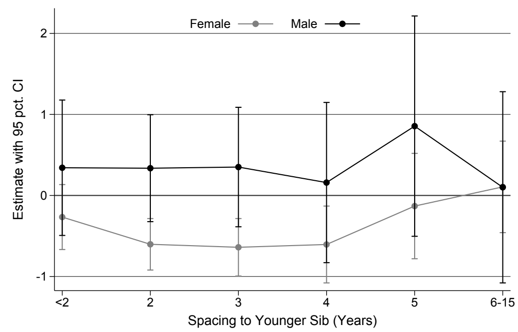
(f) Men: Academic HS GPA

Note: Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart) for academic high school GPA; children born between 1986 and 1999 with the same selection criteria as for the main sample for the grade 9 outcomes. The Grade 9 GPA measures come from the written exam at the end of grade 9 in respectively Danish and Math. *Academic HS GPA* is observed for students completing the academic high school; before 1999, this is only observed for those in the language and math tracks. The standardized GPA measures are standardized by year of graduation (for the high school GPA track-by-year of graduation) for the total population with mean zero and standard deviation of one. All graphs plot the distribution of the three measures of school performance by individuals with a same sex (black) and with an opposite sex sibling (green), respectively. The tails are truncated in order to have at least five observations within each cell.

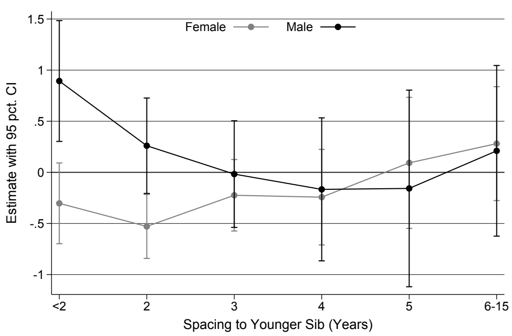
Figure A4
STEM Education by Spacing



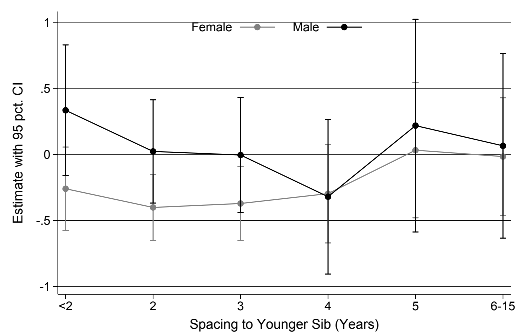
(a) Field-Specific STEM Enrollment



(b) Field-Specific STEM Completion



(c) STEM (excl. Bio.) College Enrollment

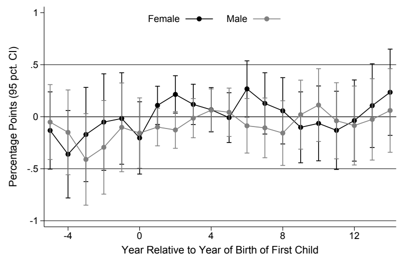


(d) STEM (excl. Bio.) College Completion

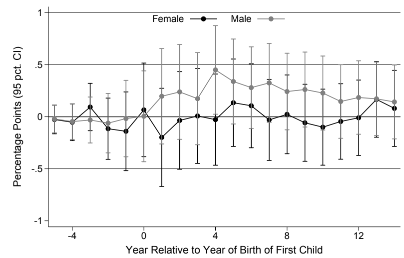
Note: Main sample (first-born children born 1962–1986) including individuals with a second-born biological sibling born up to 15 years apart. All graphs illustrate the estimated effect of having an opposite sex sibling by birth spacing; the estimates come from separate regressions by gender and are also displayed Appendix Table A10. The whiskers represent the 95 percent confidence interval.

Figure A5

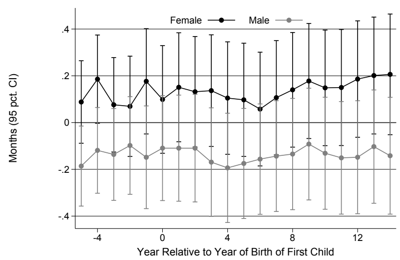
Family Structure and Parental Education, Employment, and Earnings



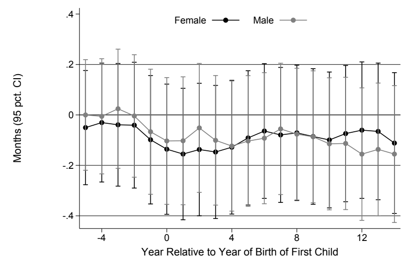
(a) Parents Cohabit/are Married



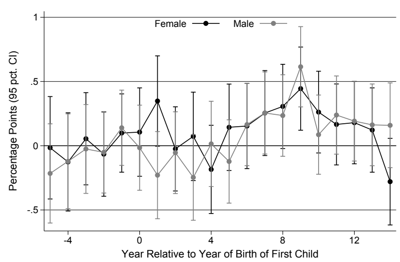
(b) Parents are Married



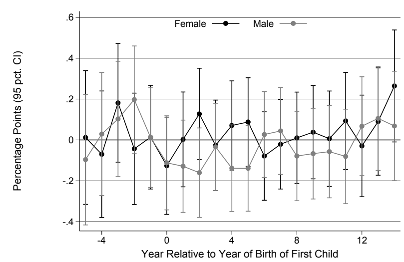
(c) Mother's Edu (months)



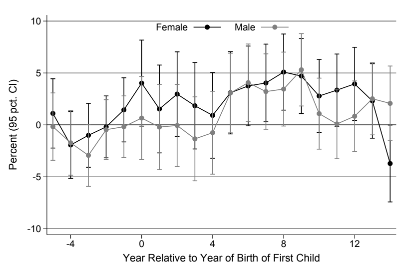
(d) Father's Edu (months)



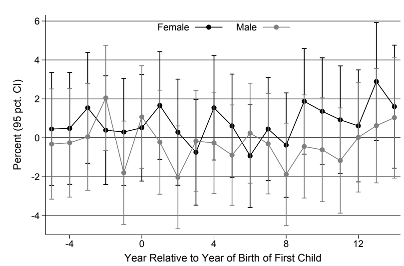
(e) Mother Employed



(f) Father Employed



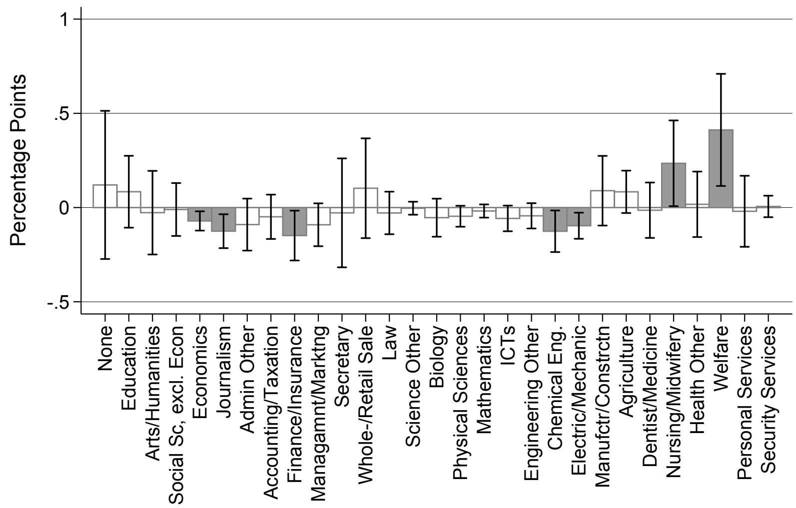
(g) Mother's Log(earnings)



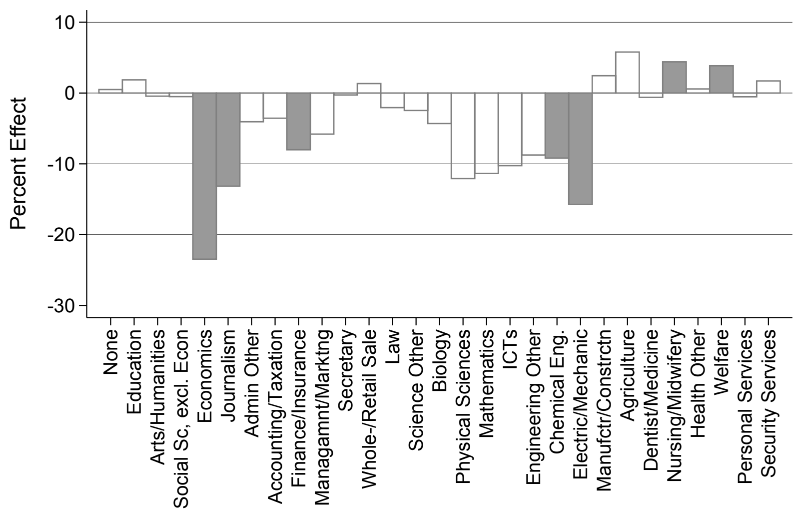
(h) Father's Log(earnings)

Note: Sample of first-born children born between 1985 and 2002 with a second-born biological sibling born within four years apart. The whiskers represent the 95 percent confidence interval. All graphs illustrate the estimates from an event study of the effect of having a second-born child of the opposite sex by gender of the first-born child. All models absorb time-specific fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, and paternal level-field of education.

Figure A6
Women: Narrow Field of Highest Completed Education



(a) Estimate w 95 pct. CI

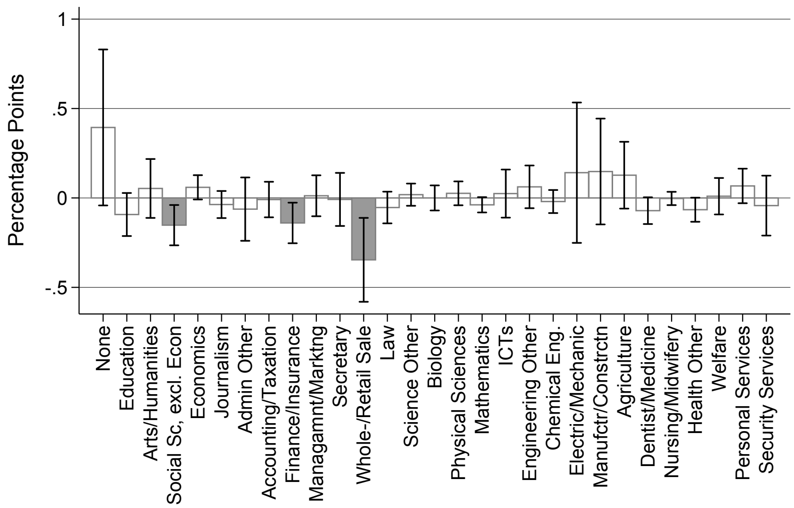


(b) Percent Effect

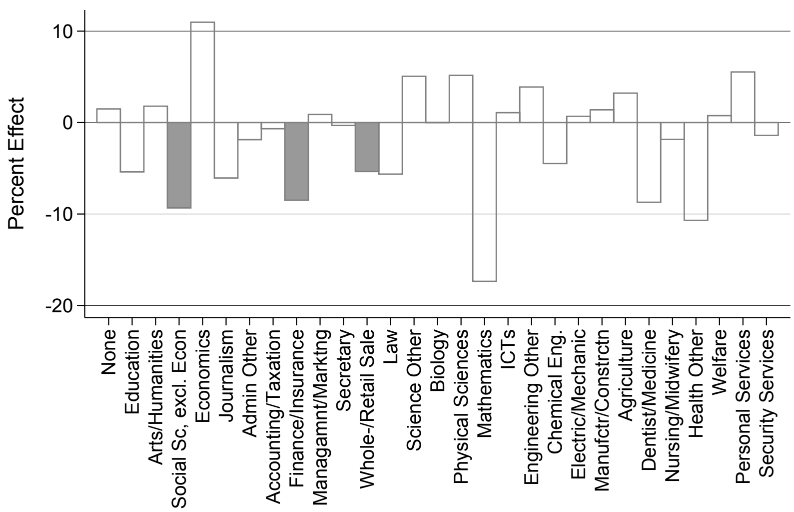
Note: Women in main sample (first-born daughters born 1962–1986 with a second-born biological sibling born within four years apart). Each bar represents the estimate from a separate regression. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, and paternal level-field of education. Graph (a) shows the estimates measured in percentage points together with the 95 percent confidence interval. Graph (b) shows the percent effect evaluated relative to the mean for individuals with a same sex sibling. Each outcome indicates whether the highest completed education by age 30 is within the indicated field.

Figure A7

Men: Narrow Field of Highest Completed Education



(a) Estimate w 95 pct. CI



(b) Percent Effect

Note: Men in main sample (first-born daughters born 1962–1986 with a second-born biological sibling born within four years apart). Each bar represents the estimate from a separate regression. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, and paternal level-field of education. Graph (a) shows the estimates measured in percentage points together with the 95 percent confidence interval. Graph (b) shows the percent effect evaluated relative to the mean for individuals with a same sex sibling. Each outcome indicates whether the highest completed education by age 30 is within the indicated field.

Table A2

Sibling Gender Composition and Number of Siblings

Sample of	*First-Born Women*			*First-Born Men*		
	# of Siblings (1)	≥ 2 Siblings (2)	≥ 3 Siblings (3)	# of Siblings (4)	≥ 2 Siblings (5)	≥ 3 Siblings (6)
Second-Born Opposite Sex	-0.07*** (0.00)	-4.96*** (0.22)	-1.43*** (0.13)	-0.08*** (0.00)	-6.89*** (0.23)	-1.33*** (0.13)
Same Sex Baseline	1.7	38.1	8.5	1.7	40.1	8.4
Percent Effect	-4.2	-13.0	-16.9	-4.7	-17.2	-15.8
Observations		164,733			173,340	

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Each Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, and paternal level-field of education. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. *# of Siblings* measures the total number of siblings the individual has, including full and half siblings. ≥ 2(3) *Siblings* takes the value one if the person has at least two (three) full siblings and zero otherwise.

Table A3

Different Control Versions: Field-Specific STEM Enrollment and Completion

	Excluding Biology			*Including Biology*		
	No controls	Controls excl. parental education	All controls	No controls	Controls excl. parental education	All controls
	(1)	(2)	(3)	(4)	(5)	(6)
Sample of First-Born Women: Enrollment						
Second-Born Brother	-0.49***	-0.48***	-0.48***	-0.53***	-0.51***	-0.51***
	(0.14)	(0.14)	(0.14)	(0.15)	(0.15)	(0.15)
Sample of First-Born Women: Completion						
Second-Born Brother	-0.54***	-0.53***	-0.53***	-0.58***	-0.57***	-0.58***
	(0.10)	(0.10)	(0.10)	(0.11)	(0.12)	(0.11)
Sample of First-Born Men: Enrollment						
Second-Born Sister	0.81***	0.80***	0.80***	0.77***	0.77***	0.77***
	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)
Sample of First-Born Men: Completion						
Second-Born Sister	0.34	0.32	0.32	0.34	0.32	0.33
	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Each cell presents estimates from separate regressions. All models include a constant. Models with the control version *Controls excl. parental education* absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, and age at last educational observation. Models with *All controls* further include fixed effects for maternal level-field of education and paternal level-field of education (i.e. a replication of the results in Table 3.)

Table A4

Educational Performance

Sample of	*First-Born Girls*			*First-Born Boys*		
	Grade 9 written exam		Academic	Grade 9 written exam		Academic
	Danish	Math	HS	Danish	Math	HS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Standardized GPA (Population Mean 0, SD 1)						
Second-Born	-0.009	-0.009	-0.009	0.002	0.004	0.009
Opposite Sex	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)
Average	0.411	0.191	0.042	-0.031	0.288	0.064
Observations	87,070	86,383	85,524	88,631	88,465	58,608
Panel B: Probability of having GPA observation (multiplied by 100)						
Second-Born	-0.077	-0.177	-0.237	-0.278	-0.215	-0.159
Opposite Sex	(0.175)	(0.181)	(0.220)	(0.196)	(0.197)	(0.200)
Average	91.4	90.7	51.9	87.6	87.4	33.8
Observations	95,226	95,226	164,733	101,223	101,223	173,340

Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart) for academic high school GPA; children born between 1986 and 1999 with the same selection criteria as for the main sample for the grade 9 outcomes. Each Panel-Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, and paternal level-field of education. The Grade 9 GPA measures come from the written exam at the end of grade 9 in respectively Danish and Math. *Academic HS GPA* is observed for students completing the academic high school; before 1999, this is only observed for those in the language and math tracks. The standardized GPA measures are standardized by year of graduation (for the high school GPA track-by-year of graduation) for the total population with mean zero and standard deviation of one.

Table A5

Components of Maternal Time Investment at Age 7 and 11

	Play	Home-work	Out-of-School Activity	Read/Sing	Excur-sion
	(1)	(2)	(3)	(4)	(5)
Sample of First-Born Girls					
			Age 7 ($N = 665$)		
Second-Born Brother	0.22	0.18	-0.01	0.14	0.14**
	(0.16)	(0.18)	(0.10)	(0.17)	(0.07)
Same Sex Baseline	2.4	4.1	1.1	4.3	0.9
Percent Effect	9.0	4.4	-0.9	3.3	16.4
			Age 11 ($N = 606$)		
Second-Born Brother	0.13	0.28	-0.03	0.08	0.18
	(0.11)	(0.19)	(0.12)	(0.07)	(0.14)
Same Sex Baseline	1.2	2.9	1.1	0.6	0.8
Percent Effect	11.2	9.5	-2.7	12.5	23.5
Sample of First-Born Boys					
			Age 7 ($N = 709$)		
Second-Born Sister	-0.32**	-0.17	-0.15*	-0.13	0.02
	(0.16)	(0.19)	(0.09)	(0.16)	(0.07)
Same Sex Baseline	2.9	3.5	1.1	4.3	0.9
Percent Effect	-11.2	-4.9	-13.4	-3.0	2.3
			Age 11 ($N = 602$)		
Second-Born Sister	-0.28**	0.11	-0.10	-0.09	-0.12
	(0.11)	(0.19)	(0.10)	(0.06)	(0.15)
Same Sex Baseline	1.3	3.5	1.0	0.6	1.1
Percent Effect	-22.4	3.2	-10.2	-13.9	-11.3

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. DALSC sample. Each Panel-Age-Column represents the results from separate regressions. All models control for mother's and father's age (squared) and fixed effects for spacing to the younger sibling in years, parental marital status in 1996, parents having been together for at least 5 years in 1996, region of birth, maternal level of education, paternal level of education, and family income level in 1995. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. Each of the individual components of maternal time investment is measured as the total number of activities done together with the child at a weekly basis.

Table A6

Components of Paternal Time Investment at Age 7 and 11

	Play	Home-work	Out-of-School Activity	Read/Sing	Excur-sion
	(1)	(2)	(3)	(4)	(5)
Sample of First-Born Girls					
			Age 7 (N = 495)		
Second-Born Brother	-0.11	-0.23	-0.02	-0.41**	0.02
	(0.20)	(0.16)	(0.09)	(0.19)	(0.08)
Same Sex Baseline	3.0	1.8	0.7	2.5	0.9
Percent Effect	-3.6	-12.8	-2.9	-16.6	2.3
			Age 11 (N = 415)		
Second-Born Brother	-0.21	-0.37**	-0.06	-0.03	-0.13
	(0.15)	(0.16)	(0.12)	(0.05)	(0.11)
Same Sex Baseline	1.6	1.9	1.0	0.6	0.6
Percent Effect	-13.2	-19.5	-5.9	-5.0	-20.3
Sample of First-Born Boys					
			Age 7 (N = 543)		
Second-Born Sister	-0.48***	0.12	-0.08	0.17	-0.11
	(0.18)	(0.16)	(0.11)	(0.18)	(0.08)
Same Sex Baseline	3.5	1.5	1.2	2.4	0.9
Percent Effect	-13.6	7.9	-6.9	7.2	-12.3
			Age 11 (N = 426)		
Second-Born Sister	0.03	0.08	-0.19	-0.04	0.17
	(0.15)	(0.19)	(0.13)	(0.06)	(0.13)
Same Sex Baseline	1.6	1.7	1.5	0.6	0.6
Percent Effect	1.9	4.6	-13.1	-6.8	27.2

Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. DALSC sample. Each Panel-Age-Column represents the results from separate regressions. All models control for mother's and father's age (squared) and fixed effects for spacing to the younger sibling in years, parental marital status in 1996, parents having been together for at least 5 years in 1996, region of birth, maternal level of education, paternal level of education, and family income level in 1995. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. Each of the individual components of paternal time investment is measured as the total number of activities done together with the child at a weekly basis.

Table A7
Principal Component Analysis: Child-Parent Relations

	Mother's	Fathers'	Child's relationship to	
	Relationship to Child		Mother	Father
Child Age	11/15	7	15	15
First Principal Component				
Age 11: How close is the relationship between you and your son/daughter (1–4)?	0.707			
Age 15: How close is the relationship between you and your son/daughter (1–3)?	0.707			
Age 7: How close is the relationship between you and your son/daughter (1–4)?		0.707		
Age 7: Are you satisfied with the relationship between you and your son/daughter (1(yes)–2(no))?		0.707		
Age 15: Your mother/father plays a very big role in your life (1–5)			0.314	0.358
Age 15: Your relationship with your mother/father is important to you (1–5)			0.363	0.379
Age 15: Your mother/father loves you (1–5)			0.357	0.351
Age 15: You trust your mother/father (1–5)			0.396	0.398
Age 15: You can expect your mother/father to listen to you (1–5)			0.407	0.393
Age 15: You can go to your mother/father for advice (1–5)			0.406	0.375
Age 15: You can count on help from your mother/father if you have a problem (1–5)			0.395	0.388
Eigenvalue				
First Component	1.335	1.348	3.568	4.329
Second Component	0.665	0.652	1.004	0.754

DALSC sample. All questions are answered on a likert scale with lower values being better. Therefore, the standardized measures used in Table 8 are all reversed, such that a higher value reflects a better relationship.

Table A8

Field-Specific STEM Enrollment and Completion: Heterogeneity by Length of Parental Education

	STEM Enrollment		STEM Completion	
	Excl. Biology (1)	Incl. Biology (2)	Excl. Biology (3)	Incl. Biology (4)
Sample of First-Born Women				
Opp×$M_{<HS}, F_{<HS}$	0.23	0.14	-0.13	0.03
	(0.25)	(0.17)	(0.15)	(0.10)
Opp×$M_{\geq HS}, F_{<HS}$	-0.73	-0.81**	-0.52*	-0.58**
	(0.45)	(0.35)	(0.31)	(0.25)
Opp×$M_{<HS}, F_{\geq HS}$	-0.46*	-0.59***	-0.05	-0.23*
	(0.27)	(0.19)	(0.18)	(0.13)
Opp×$M_{\geq HS}, F_{\geq HS}$	-0.74***	-0.74***	-0.64***	-0.57***
	(0.23)	(0.19)	(0.20)	(0.17)
Observations		156,953		
Sample of First-Born Men				
Opp×$M_{<HS}, F_{<HS}$	0.43	0.49	0.17	0.22
	(0.51)	(0.43)	(0.21)	(0.16)
Opp×$M_{\geq HS}, F_{<HS}$	0.13	-0.23	1.23***	0.13
	(0.76)	(0.70)	(0.46)	(0.38)
Opp×$M_{<HS}, F_{\geq HS}$	1.06**	0.38	0.11	-0.20
	(0.51)	(0.46)	(0.27)	(0.22)
Opp×$M_{\geq HS}, F_{\geq HS}$	0.94***	0.48	0.49*	0.32
	(0.35)	(0.33)	(0.28)	(0.25)
Observations		165,547		

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart) for those with information on both parents' length of education. Each Panel-Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, indicators for the parents' combination of length of education, and age at last educational observation.

Table A9

Field-Specific STEM Enrollment and Completion: Heterogeneity by Parental Division of Labor

	STEM Enrollment		STEM Completion	
	Excl. Biology (1)	Incl. Biology (2)	Excl. Biology (3)	Incl. Biology (4)
Sample of First-Born Women				
Second-Born Brother (SBB)	-0.43**	-0.55***	-0.40***	-0.40***
	(0.17)	(0.12)	(0.13)	(0.10)
SBB×Traditional Division	-0.25	0.07	0.08	0.19
	(0.33)	(0.24)	(0.24)	(0.18)
Observations	162,575			
Sample of First-Born Men				
Second-Born Sister (SBS)	0.57**	0.21	0.26	0.10
	(0.26)	(0.25)	(0.17)	(0.15)
SBS×Traditional Division	1.26**	0.76	0.31	0.02
	(0.56)	(0.53)	(0.36)	(0.30)
Observations	171,082			

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart) for those with information of both parents' labor supply. Each Panel-Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last educational observation. *Traditional Division* takes the value one if paternal labor supply represents at least 75 percent of total parental labor supply during childhood and zero otherwise.

Table A10

Field-Specific STEM Enrollment and Completion: Heterogeneity by Spacing

Sample of	*First-Born Women*				*First-Born Men*			
	STEM Enrollment		STEM Completion		STEM Enrollment		STEM Completion	
	Excl. Biology (1)	Incl. Biology (2)	Excl. Biology (3)	Incl. Biology (4)	Excl. Biology (5)	Incl. Biology (6)	Excl. Biology (7)	Incl. Biology (8)
Opp × <2 years	-0.42	-0.27	-0.30	-0.26	1.00**	0.34	0.89***	0.33
	(0.27)	(0.20)	(0.20)	(0.16)	(0.46)	(0.43)	(0.30)	(0.25)
Opp ×2 years	-0.54**	-0.60***	-0.53***	-0.40***	1.01***	0.34	0.26	0.02
	(0.21)	(0.16)	(0.16)	(0.13)	(0.37)	(0.34)	(0.24)	(0.20)
Opp ×3 years	-0.46*	-0.64***	-0.22	-0.37***	0.40	0.35	-0.02	-0.00
	(0.24)	(0.18)	(0.18)	(0.14)	(0.41)	(0.38)	(0.27)	(0.22)
Opp ×4 years	-0.54*	-0.61**	-0.24	-0.30	0.31	0.16	-0.17	-0.32
	(0.32)	(0.24)	(0.24)	(0.19)	(0.55)	(0.50)	(0.36)	(0.30)
Opp ×5 years	0.15	-0.13	0.09	0.03	-0.07	0.86	-0.16	0.22
	(0.44)	(0.33)	(0.33)	(0.26)	(0.76)	(0.69)	(0.49)	(0.41)
Opp ×6–15 years	0.47	0.11	0.28	-0.02	0.11	0.10	0.21	0.07
	(0.38)	(0.29)	(0.28)	(0.23)	(0.66)	(0.60)	(0.43)	(0.36)
Observations	232,372				243,169			

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986) including individuals with a second-born biological sibling born up to 15 years apart. Each Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last educational observation. *Opp* indicates having a second-born sibling of the opposite gender.

Table A11
Field-Specific STEM Enrollment and Completion: Heterogeneity by Decade of Birth

	STEM Enrollment		STEM Completion	
	Excl. Biology (1)	Incl. Biology (2)	Excl. Biology (3)	Incl. Biology (4)
Sample of First-Born Women				
Second-Born	-0.51**	-0.41***	-0.40***	-0.26**
Brother	(0.20)	(0.15)	(0.15)	(0.11)
SBB×1970 − 79	-0.20	-0.27	-0.03	-0.14
	(0.30)	(0.22)	(0.22)	(0.17)
SBB×1980 − 86	0.53	-0.07	0.17	-0.21
	(0.41)	(0.29)	(0.30)	(0.26)
Prob>F	0.216	0.442	0.809	0.594
Observations		164,733		
Sample of First-Born Men				
Second-Born	0.91**	0.34	0.15	-0.12
Sister	(0.36)	(0.33)	(0.24)	(0.20)
SBS×1970 − 79	-0.03	-0.12	0.30	0.29
	(0.50)	(0.46)	(0.32)	(0.28)
SBS×1980 − 86	-0.47	0.14	0.27	0.50
	(0.65)	(0.64)	(0.44)	(0.40)
Prob>F	0.748	0.915	0.627	0.391
Observations		173,340		

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Each Panel-Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last educational observation. *Prob>F* reports the p-value from an F-test of joint significance of opposite sex sibling interactions with decade of birth.

Table A12

Effect of Co-Twin's Gender

	Next Birth	Field-Specific STEM Enrollment	Field-Specific STEM Completion	STEM Focus in First Enrollment	STEM in Highest Completion
	(1)	(2)	(3)	(4)	(5)
Panel A: Female Twins					
		Any Parity ($N = 12,755$)			
Co-Twin Brother	-1.32**	-1.56***	-1.40***	-2.81***	-1.26***
	(0.60)	(0.51)	(0.38)	(0.82)	(0.36)
Same Sex Baseline	24.1	8.0	4.7	23.3	4.3
Percent Effect	-5.5	-19.6	-29.9	-12.1	-29.6
		First Parity ($N = 4,730$)			
Co-Twin Brother	-0.48*	-1.95**	-2.05***	-0.79	-1.79***
	(0.29)	(0.90)	(0.68)	(1.50)	(0.64)
Same Sex Baseline	42.0	8.5	5.4	26.1	4.7
Percent Effect	-1.1	-23.0	-37.9	-3.0	-37.7
Panel B: Male Twins					
		Any Parity ($N = 13,067$)			
Co-Twin Sister	-1.83***	2.89***	1.28	2.10**	1.71*
	(0.61)	(0.97)	(0.89)	(0.99)	(0.88)
Same Sex Baseline	23.6	37.2	26.0	46.1	24.9
Percent Effect	-7.8	7.8	4.9	4.6	6.9
		First Parity ($N = 4,832$)			
Co-Twin Sister	-0.58*	3.14*	1.46	2.56	1.32
	(0.31)	(1.68)	(1.54)	(1.72)	(1.52)
Same Sex Baseline	40.2	37.8	26.4	48.1	25.5
Percent Effect	-1.4	8.3	5.5	5.3	5.2

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the mother level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Each Panel-Column-Sample presents estimates from separate regressions. The sample consists of twins born at respectively *any* and *first* parity. All models absorb fixed effects for birth county, year of birth, second generation immigrant status, mother's level and field of education, father's level and field of education, and age at last educational observation. The models further control for (cubed) mother's age at birth and (cubed) father's age at birth. The models for the *Any Parity*-sample further control for parity. *Next Birth* indicates if the parents get a subsequent child. *Same Sex Baseline* reports the mean outcome for individuals with a same sex co-twin. *Percent Effect* reports the estimated effect of co-twin gender relative to the baseline. *Field-Specific STEM* excludes Biology. *STEM Focus in First Enrollment* indicates whether the first place of enrollment after compulsory education is within STEM (vocational secondary or academic high school). *STEM in Highest Completion* indicates whether the highest completed education is a field-specific STEM education excluding biology.

Table A13

The Effect of an Older Sibling's Gender

	STEM Excluding Biology				Educational Attainment			
	Field-Specific		College		Vocational		College	
	Enroll-ment	Comple-tion	Enroll-ment	Comple-tion	Enroll-ment	Comple-tion	Enroll-ment	Comple-tion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample of Second-Born Women								
First-Born Brother	-0.29**	-0.17*	-0.21**	-0.20***	0.68***	0.75***	-0.77***	-0.76***
	(0.12)	(0.09)	(0.09)	(0.07)	(0.23)	(0.23)	(0.22)	(0.22)
Same Sex Baseline	7.8	4.1	3.9	2.5	58.2	43.0	41.5	35.2
Percent Effect	-3.7	-4.1	-5.4	-8.1	1.2	1.7	-1.9	-2.2
Observations	170,803							
Sample of Second-Born Men								
First-Born Sister	2.37***	1.67***	0.31**	0.16	0.55***	0.14	0.12	-0.01
	(0.24)	(0.23)	(0.14)	(0.12)	(0.21)	(0.23)	(0.20)	(0.19)
Same Sex Baseline	39.0	27.3	9.9	6.6	69.4	53.7	29.8	23.8
Percent Effect	6.1	6.1	3.1	2.4	0.8	0.3	0.4	0.0
Observations	178,306							

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Sample of second-born children born between 1962 and 1986 with an older biological sibling born within four years apart. Each Panel-Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to older sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last educational observation. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline.

Table A14

Alternative Measures of Field of Study

	OECD's STEM Definition				Care Fields			
	Field-Specific		College		Field-Specific		College	
	Enroll-ment (1)	Comple-tion (2)	Enroll-ment (3)	Comple-tion (4)	Enroll-ment (5)	Comple-tion (6)	Enroll-ment (7)	Comple-tion (8)
Sample of First-Born Women								
Second-Born Brother	-0.70***	-0.75***	-0.42***	-0.36***	0.74***	0.72***	0.57***	0.53***
	(0.25)	(0.21)	(0.11)	(0.10)	(0.23)	(0.22)	(0.21)	(0.19)
Same Sex Baseline	34.8	25.0	6.4	4.4	32.1	26.2	24.6	20.6
Percent Effect	-2.0	-3.0	-6.6	-8.2	2.3	2.7	2.3	2.6
Observations	164,733							
Sample of First-Born Men								
Second-Born Sister	0.50**	0.34	0.28*	0.07	-0.25**	-0.23**	-0.27**	-0.23**
	(0.23)	(0.24)	(0.16)	(0.14)	(0.12)	(0.10)	(0.11)	(0.09)
Same Sex Baseline	67.8	53.5	13.6	9.5	6.7	4.7	6.0	4.3
Percent Effect	0.7	0.6	2.1	0.7	-3.7	-4.9	-4.5	-5.4
Observations	173,340							

All estimates are multiplied by 100 to express effects in percentage points. Standard errors in parentheses, clustered at the year-month of birth level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Main sample (first-born children born 1962–1986 with a second-born biological sibling born within four years apart). Each Panel-Column presents estimates from separate regressions. All models absorb fixed effects for birth municipality, year-month of birth, spacing in months to younger sibling, second generation immigrant status, maternal age at birth, paternal age at birth, maternal level-field of education, paternal level-field of education, and age at last educational observation. *Same Sex Baseline* reports the mean outcome for individuals with a same sex sibling. *Percent Effect* reports the estimated effect of sibling gender relative to the baseline. *OECD's STEM Definition* includes Natural Sciences, Mathematics, Statistics, Information and Communication Technologies, Engineering, Manufacturing, and Construction (thereby including Biology, Manufacturing, and Construction and excluding Economics compared to the main definition). *Care Fields* include Education, Health, and Welfare.