NBER WORKING PAPER SERIES

INTERGENERATIONAL AND SIBLING SPILLOVERS IN HIGH SCHOOL MAJORS

Gordon B. Dahl
Dan-Olof Rooth
Anders Stenberg

Working Paper 27618
http://www.nber.org/papers/w27618

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2020, Revised January 2023

Previously circulated as "Intergenerational and Sibling Peer Effects in High School Majors" We thank colleagues and seminar participants at several universities and conferences for valuable feedback and suggestions. We are grateful to Hans Eric Ohlson at Statistics Sweden for help in locating and accessing the data. The project received generous financial support from the Institute for Social Research at Stockholm University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Intergenerational and Sibling Spillovers in High School Majors
Gordon B. Dahl, Dan-Olof Rooth, and Anders Stenberg
NBER Working Paper No. 27618
July 2020, Revised January 2023
JEL No. I21,J24

ABSTRACT

This paper estimates family spillovers in high school major choice in Sweden, where admission to oversubscribed majors is determined based on GPA. Using a regression discontinuity design, we find large sibling and intergenerational spillovers that depend on the gender mix of a dyad. Same-gender siblings copy one another, while younger brothers recoil from older sister's choices. Fathers and mothers influence sons, but not their daughters, except when a mother majors in the male-dominated program of Engineering. Back of the envelope calculations reveal these within family spillovers have sizable implications for the gender composition of majors.

Gordon B. Dahl
Department of Economics
University of California, San Diego
9500 Gilman Drive #0508
La Jolla, CA 92093-0508
and NBER
gdahl@ucsd.edu

Dan-Olof Rooth
Swedish Institute for Social Research (SOFI)
Stockholm University
10691 Stockholm
Sweden
dan-olof.rooth@sofi.su.se

Anders Stenberg
Swedish Institute for Social Research (SOFI)
Stockholm University
10691 Stockholm
Sweden
anders.stenberg@sofi.su.se

1 Introduction

Parents and siblings are key players in an adolescent's network. From early childhood to the teenage years, family interactions occur daily and during a time when individuals are forming their own identities. Research in psychology has posited positive role model effects within families, where children look up to older siblings or parents and emulate their choices. But there are also theories which imply an opposite-signed recoil effect due to differentiation (a desire to be unique, driven by sibling rivalry or rebellion against parental expectations). Families could also be a valuable for information transmission or support related to different options. On top of this, whether a major is gender conforming could play a role. A recurring prediction in these models is that both sibling and parental effects should be stronger for same-gender pairs. However, there is little consensus on which effects dominate or whether parents or older siblings have a stronger impact.[1]

Family influences during the teenage years are particularly consequential, as decisions are being made which will affect adult outcomes. One of the most important choices is what type of career path to pursue. In many countries throughout the world, including much of Europe, Latin America, and Asia, students choose a field specialization in secondary school (i.e., a "high school major").[2] This choice is made at the relatively young age of 15 or 16, when field preferences are still in flux and knowledge of different careers is limited, but before the natural distancing of siblings or full independence from parents. These high school major choices are consequential, as they play a foundational role in future labor market success. As we show in Dahl et al. (forthcoming), the differential earnings returns across high school majors often rival the return to an additional two years of schooling. Traditionally male-dominated majors such as Engineering yield higher earnings, while female-dominated majors such as Humanities result in lower earnings. These effects are primarily driven by individuals with different high school majors ending up in different occupations, and to a lesser extent, different college majors.

In this paper, we examine the role of both older siblings and parents on high school major choices. Given the prior literature's hypotheses on gender-specific interactions, we allow for different spillovers based on the gender mix of a dyad. Despite the potential importance of

[1]Section 5 reviews the literature on these competing theories. For more general overviews, see Bush and Peterson (2013), McHale et al. (2013), and Raley and Bianchi (2006).

[2]Countries requiring students to choose fields in secondary school in Europe include the Czech Republic, Denmark, France, Italy, Norway, Poland, Spain, Sweden, and the United Kingdom; in Latin America include Argentina, Chile, Colombia, Cuba, Mexico, Paraguay, and Venezuela; and in Asia include Indonesia, Iran, Malaysia, Pakistan, the Philippines, and Saudi Arabia.

families on choices, causal evidence on these types of spillovers remains scarce for two reasons. First, these effects are notoriously difficult to estimate given the challenges identified by Manski (1993). Correlated unobservables are particularly likely to create a bias when studying family spillovers.[3] High school major choices could be driven by common factors such as a shared environment or family income, rather than a spillover effect. Second, it is difficult to access and link datasets which contain the major choices of either siblings or of parents and their children. We overcome these challenges in the context of Sweden's secondary school system, where students choose between five academic majors: Engineering, Natural Science, Business, Social Science, and Humanities. Our setting is unique in that students were not allowed to choose which school to attend, so the choice concerns only majors and not institutions.

Academic majors are often oversubscribed, so admission is rationed based on a student's cumulative ninth grade GPA. Students rank their preferred majors and admission is determined using an allocation mechanism that is both Pareto efficient and strategy proof. Importantly, individuals around a major's GPA cutoff in a school region should be roughly similar on all observable and unobservable dimensions. This allows us to compare older siblings or parents just above versus just below major-specific admission cutoffs, and estimate whether (i) their younger siblings or (ii) their children are affected in their high school major choice. We use high-quality register data of all applicants to high school during 1977-1991 (for parents and siblings) and 2011-2019 (for children) to implement a regression discontinuity (RD) design.

We find that both older siblings and parents exert a significant influence on high school major choice. Throughout the paper, we define a younger sibling's or a child's "major choice" to be the major they list as their most preferred choice on their ranking list. A younger sibling is 2.5 percentage points more likely to choose the same major as the one their older sibling *enrolled in*. Since younger siblings choose the same major 20% of the time, this translates to a 13% increase relative to the mean. Similar effect sizes are found for parental influences, with children being 11% more likely to choose the same major as the one their parent enrolled in. However, these estimates mask a rich heterogeneity in spillover effects, as they do not account for the gender mix of siblings or parent-child pairs. This turns out to be important empirically, as we find spillover effects which are positive, negative, or zero depending on the gender mix of a dyad.

[3]The other issues Manski identifies are endogenous group membership and reflection. The first is not an issue for family spillovers since individuals do not choose their siblings or children, and the second is not an issue if older siblings and parents influence younger siblings and children, but not the other way around.

Starting with same-gender sibling pairs, we find strong positive spillover effects. Younger brothers are 25% more likely to choose a particular major if their older brother enrolled in it. Likewise, younger sisters are 18% more likely to choose the major their older sister enrolled in. However, an examination by birth spacing reveals that the underlying forces driving these effects differ by gender. For brother pairs, there is a relatively small and insignificant spillover effect if both brothers will be in school at the same time (i.e., within 3 years in age of each other). In contrast, when brothers are more than 3 years apart, there is a substantial 42% increase in same major choice. This pattern is consistent with the older brother serving primarily as a role model or information provider when the age gap is large, but with differentiation due to sibling rivalry outweighing any such effects when the two brothers are close in age. For sister pairs, there is the opposite pattern. There is a 23% increase in same major choice if two sisters will be in school at the same time, whereas for sisters more than 3 years apart, the effect is small and insignificant. This is the opposite of what sibling rivalry would have predicted and more consistent with sisters supporting each other when close in age.

Turning to mixed-gender sibling pairs, we find either a strong negative spillover effect or no effect. Younger brothers are 26% less likely to choose the major their older sister enrolled in, while younger sisters are not influenced by older brothers on average. Examining birth spacing provides further insight on the recoil effect. The estimated effect increases to a 41% reduction when younger brothers are within 3 years of age of their older sister, while there is no effect for those more distant in age. In other words, young brothers avoid their older sister's major, but only if they will overlap for a portion of their high school studies. This is a somewhat surprising finding, as sibling rivalry is predicted to be less important for opposite gender sibling pairs.

Our findings for siblings provide some support for theories from complementary literatures which argue that: (i) role model effects are largest for same-gender siblings more distant in age, (ii) brother pairs close in age have the greatest amount of sibling rivalry, and (iii) sister pairs are the most supportive of each other. But we also observe patterns not predicted by the literature.

Shifting focus to parental influences, which occur roughly 30 years later when their children are applying for high school, we find the strongest effects on sons. When a father enrolls in a given high school major, his son is 4.3 percentage points more likely to also choose it. If the father not only enrolls in a high school major, but also completes it, this effect rises to

5.6 percentage points.[4] This is a sizable 22% increase relative to the mean. Similarly, when a mother completes a high school major, her son is 18% more likely to choose it. In contrast, the effects on daughters are small and not statistically significant for either mothers or fathers.

These intergenerational patterns are not consistent with the general prediction in the literature that same-gender dyads should have the strongest effects. Instead, we find that fathers and mothers both serve as role models or information providers for sons, but not for daughters. This could be explained by the fact that the education landscape and labor markets have changed more for girls than boys in the time since parents made their choices, or that daughters want to differentiate themseleves more from their parents than sons do.

Correlations overstate family spillover effects by a factor of 4.1 for siblings and 2.7 for parents and their children. Equally worrisome is that correlations miss out on much of the heterogeneity by gender mix, finding large positive effects for *all* sibling and parent-child gender dyads. These sharp contrasts highlight the need to use a convincing research design and to separately estimate effects by gender dyad.

As a final exercise, we further break down these family spillovers based on the gender makeup of majors and whether they are gender conforming. When a major is gender-neutral, we find strong positive spillovers for both brother-brother and sister-sister pairs. Brothers also copy older brothers into the gender-conforming, male-dominated major of Engineering (i.e., they choose Engineering as their first choice if the older brother enrolled in it). Sisters, in contrast, do not copy older sisters into the gender-conforming, female-dominated majors. This exercise also provides further insight into the recoil effect we found for older sister-younger brother pairs. The recoil effect is concentrated in female-dominated majors, suggesting that younger brothers especially want to differentiate themselves from an older sister when the major is non-gender conforming. The intergenerational results also exhibit some heterogeneity by the gender makeup of majors. Sons copy fathers and mothers into male-dominated and gender-neutral majors, but not into female-dominated majors. For daughters, the only effect is that mothers who complete the non-gender conforming choice of Engineering influence their daughters to also choose Engineering. These sibling and intergenerational patterns have important implications for gender imbalances across different majors, and hint at the importance of same-gender role models for women pursuing non-gender conforming choices.

The intergenerational and sibling results complement each other well. Because an individual

[4]For parents, we report effects related to both enrollment and completion. For siblings, we focus on enrollment, since younger siblings close in age make their choice before their older sibling completes a major.

affects both their children and their younger siblings, we are uniquely able to compare magnitudes. We find that an individual's major choice has similarly-sized influences on their children and their siblings.[5] When interpreting these results, it is useful to consider the time gaps in education decisions. The sibling spillovers occur relatively close in time, when curricula and information about a major are most similar, but usually before an older sibling has finished college or begun a job. In contrast, the intergenerational spillovers occur over 30 years later on average, but this gives the child an opportunity to observe a parent's occupation and earnings history. Both sets of results display a consistent pattern of the family as a mechanism for reinforcing gender stereotypical norms for males, but having either no impact or breaking gender stereotypical norms for females.

In the broader education space, our paper relates to three strands of research: family spillover effects in years of schooling, course content, and choice of college.[6] To study the causal link between parent's and children's years of schooling, researchers have used school reforms, compulsory schooling laws, twins, and adoptions, often reaching different conclusions on the size of any intergenerational link (for reviews, see Black and Devereux 2011, Björklund and Jäntti 2009, Björklund and Salvanes 2011, and Holmlund et al. 2011). In contrast, there are few papers studying spillovers in siblings' years of schooling (Quereshi 2018b, Adermon 2013).

To study spillover effects in course content, Joensen and Nielsen (2018) takes advantage of a pilot program in Denmark which lowered the cost of choosing advanced math and science classes. They find spillover effects which are sizable and statistically significant for brother pairs, but not for other sibling-gender combinations. While no prior studies have looked at intergenerational links in course content, Hanushek et al. (2021) looks at intergenerational links in cognitive skills for math and language.

To study spillover effects in college choices for siblings, a recent set of papers have used RD designs based on GPA cutoffs.[7] Altmejd et al. (2021) looks at sibling spillovers in the U.S., Chile, Sweden, and Croatia. For the U.S., they find large effects of an older sibling barely

[5]It would also be interesting to compare whether an individual is more influenced by a sibling or a parent. This is unfortunately not possible with our data, as a series of reforms substantially reduced the number of oversubscribed majors after 1992 (see Section 2.2). We can still use major choices of those born after 1992 as outcomes in an intergenerational setting, but we do not have enough quasi-random variation due to oversubscription to identify sibling spillovers in majors for these later years.

[6]There is also a literature which looks at sibling spillovers in school achievement. For two recent examples, see Nicoletti and Rabe (2019) and Quereshi (2018a).

[7]Correlational estimates for sibling college choices also exist; for example, see Goodman et al. (2015).

getting into a college, both for whether the younger sibling attends any college or the same college. For the three other countries, there is compelling evidence that younger siblings follow their older siblings to the same college institution or same institution-major combination. But as Altmejd et al. write, "in contrast to the strong college-choice spillover effects, we find almost no influence on major choices" (p. 1862, see also Table III).[8] Similarly, Aguirre and Matta (2021) and Dustan (2018) find evidence for sibling effects in institution choice in Chile and Mexico, but not for major. There are no similar studies for parent-child spillovers in college choice, likely due to the difficulty in obtaining the relevant intergenerational data.

More broadly, our paper is related to an emerging literature on other spillover effects within the family, both across siblings and across generations. A sampling of recent papers includes family spillovers in risky behavior (Altonji et al. 2017), military service (Bingley et al. 2021), entrepreneurship (Lindquist et al. 2015), cognitive and noncognitive skills (Lundborg et al. 2014), medical treatments and diagnoses (Daysal et. al. 2022, Fadlon and Nielsen 2019, Persson et al. 2021), labor supply (Nicoletti et al. 2018), disability and test scores (Black et al. 2021), and use of social insurance programs (Dahl et al. 2014).

Our paper makes three fundamental contributions. We are the first to causally estimate parent-to-child spillovers in the choice of major, finding large intergenerational effects which show up decades later. Second, we find strong evidence that the choice of high school major is influenced by siblings, a new margin which complements prior findings for sibling spillovers in college choice. Third, we find that family spillover effects are not monolithic, with the gender mix of dyads playing an important role for the size and direction of effects.

Back of the envelope calculations reveal that these family spillovers have nontrivial effects on the gender composition of majors, especially for Engineering. The influence of older brothers serves to reinforce gender norms by increasing the male share in Engineering by 4 percentage points, while older sisters help to break down the gendered nature of Engineering by raising the female share by 8 percentage points. Turning to parental influences, both mothers and fathers help make Engineering a less male-dominated field, in part because mothers pull daughters into Engineering and in part because when sons follow fathers into fields other than Engineering, they are drawn away from Engineering.

[8] Prospective students in Chile, Sweden, and Croatia do not apply to colleges individually as in the U.S., but instead send in a single ranking of their college *plus* major preferences to a centralized administration that allocates students for the entire country. Students often prioritize majors over college institution, for example, by ranking Engineering at College A as their first choice and Engineering at College B as their second choice, rather than a different major being involved in their second choice.

It follows that the family spillovers we identify matter for education policy, as they will add to any direct effects of reforms. One policy is to expand slots in the higher-paying major of Engineering. But a gender-blind expansion will risk reinforcing traditional gender norms for boys through family spillovers. More targeted policies could nudge girls into Engineering through outreach campaigns, by making it easier for them to gain admission, or by providing exposure to female role models (see Bettinger and Long 2005, Breda et al. 2020, Buckles 2019, Carrell et al. 2010, and Porter and Serra 2020). This would not only increase representation of women by affecting those directly targeted, but also help decrease gender segregation indirectly via family spillovers.

The remainder of the paper proceeds as follows. The next section describes Sweden's secondary education system and our data. Section 3 discusses our empirical design and Section 4 presents first stage results. Section 5 reviews theories in the literature on familial influences. Section 6 presents our aggregate estimates of sibling and intergenerational spillovers, followed by analyses across gender dyads in Sections 7 and 8. Section 9 provides robustness checks. Sections 10 and 11 document heterogeneous effects based on the gender conformity of a major and the resulting impact on gender imbalances across majors. The final section concludes.

2 Setting and Data

2.1 Admission to High School Majors in Sweden

After nine years of compulsory schooling, in the year individuals turn 16, students in Sweden can apply to a field of study in secondary school (i.e., a high school major). In this paper, we focus on the five academic majors which are preparatory for university studies or direct entry into the labor market: Engineering, Natural Science, Business, Social Science, and Humanities. These academic majors take three years to complete, with Engineering having an optional fourth year.

As shown in Appendix Table A1, the curricula for these majors differ substantially. Engineering and Natural Science require more math and science (with Engineering having more technology-related courses), Business requires specialized courses such as law and accounting, and Social Science and Humanities require more social science (with Humanities having more language courses). These differences highlight that the choice of a high school major is one of the most important decisions a teenager will make regarding their future career path.[9]

[9] College majors do not require specific high school majors for admission, but some college majors do require the completion of specific classes. For example, to major in engineering in college, one needs advanced math

Individuals can also choose from a variety of non-academic majors which take two years to complete. These programs focus on vocational skills or general education, but not at the level required for admission to college. Since these non-academic programs are usually not oversubscribed, we cannot use our RD design for them, apart from including them as possible next-best options for older siblings and parents, and as possible choices for younger siblings and children. During the periods we study, roughly half of students enroll in an academic versus non-academic major.

During our sample period, the majors did not experience large changes and students were only choosing majors and not which school to attend. In some regions, majors could still be tied to specific institutions – for example in regions where there are two schools and each school specializes in a different set of majors. However, for most school regions there is only one school (on average 88 out of 129, although this varies by year). In a robustness check, we limit our sample to older siblings whose top two major choices are both available at the school they are admitted to and find almost identical estimates (see Section 9).

If a major is oversubscribed, students compete for slots based on their application GPA. The application GPA is the average grade across 10-12 school subjects as of ninth grade. Grades in each subject range from 1 to 5 and have an approximate mean of 3 and standard deviation of 1 in the entire population of ninth grade students. Applicants have an extra 0.2 bonus points added to their GPA if they apply to a major which accepted 30% or less of their gender nationally in the prior year.

Students rank their preferences for up to 6 majors, and a central administration office then allocates students. Admission decisions are made sequentially, with the highest-GPA applicant being admitted to their first-choice major, the second-highest GPA applicant being admitted to their highest-ranked major among the set of majors which still have space in them, and so forth. This "serial dictatorship" mechanism of allocating slots is both Pareto efficient and strategy proof, as long as 6 choices is not a binding constraint (Svensson 1999). Empirically, six choices does not appear to be a binding constraint.[10]

The key factor which determines whether a major will be oversubscribed is the lumpiness of class sizes. Classes, and therefore majors, are often capped at 30 students. If there is

classes; these are included in the Engineering and Natural Science high school curricula. In Sweden, students can also take adult education courses after high school to satisfy the course requirements for college majors.

[10]Only 0.2% of all applicants are admitted to their 6th choice and only 1.6% even list a sixth choice. Between 1982-84, bonus GPA points were also given for first and second choices on a ranking list, so students may not have revealed their true preferences. As we show, excluding these years does not materially affect our estimates.

only one class for a given major and 33 students list the major as their first choice, it will be impacted. In contrast, if only 27 students list it as their first choice, everyone will be admitted. Depending on expected demand for a major, there could be two or even three classes for a given major. This lumpiness often leads to a major being oversubscribed in a school region in one year, but not the next. It also means that the most popular majors are not necessarily the ones which will be oversubscribed.

In our setting, it is important not to confuse "oversubscription" with "highly competitive." There is not a universal or persistent ordering in which majors have higher cutoffs or are more likely to be oversubscribed, either across or within school regions. Moreover, average cutoffs (conditional on having a cutoff) are broadly similar across majors. After we introduce our data, we will empirically document the variation in relative cutoffs within the same school region over time in Section 2.3.

After admission decisions are sent out in July, there can be reallocations of students to different fields of study. This can happen for a variety of reasons. For example, a student admitted to Engineering may change their mind and transfer to another major, such as Social Science, that still has open slots. This move will also open up a slot in Engineering, which another student can take. While changes can happen at any time, it becomes more difficult to switch after the fall of the first year given curriculum differences.

These reallocations are not necessarily random, as they depend on individuals changing their minds and potentially discretion on the part of the local high school principal. Luckily, we observe the actual admission decision, which is a mechanical and binary function of the GPA cutoff. We can use the admission decision cutoff in a RD design as an instrument for either major enrollment or completion. We can also use the sharp cutoff in admission decisions to estimate the reduced form effects of admission itself.

2.2 Data

We have information on the major ranking lists submitted by all students going back to 1977. This allows us to observe which majors they prefer and which major they are admitted to.[11] Our data and empirical design have some similarities to Dahl et al. (forthcoming), which

[11]If an individual is admitted to either their first or second ranked choice, which happens 96% of the time, then we define these as the individual's preferred and next-best alternative majors, respectively. For individuals who are admitted to a third or lower ranked choice, their preferred choice is defined as the lowest-GPA choice above their accepted choice, and the next-best alternative as their accepted choice. For ease of exposition, in what follows we refer to the preferred major as the first-best choice, even if it wasn't the first choice on their list, and likewise the next-best alternative as the second-best choice.

studies how different high school majors affect future earnings.

Several factors influence which years we can use to study sibling and intergenerational spillovers. In 1992, Business, Social Science, and Humanities were merged into one major, only to re-emerge as separate majors again in 2011. The introduction of private schools and school choice, as well as other reforms, also substantially reduced the number of oversubscribed majors after 1992. These factors mean we can use the 1977-1991 period to study sibling spillovers, and 1977-1991 data for parents merged with 2011-2019 data for their children to study intergenerational spillovers. But we do not have enough quasi-random variation due to oversubscription to identify sibling spillovers during the 2011-2019 time period.[12]

We use data from several registers housed at Statistics Sweden. To identify siblings, we use the sibling register (Statistics Sweden 2021g); to identify parents, we use the multiple generation register (Statistics Sweden 2021e). Using personal identification numbers we merge in preferred major choices from the high school application database (Statistics Sweden 2021a). Additionally, we merge in grades, demographic, and socioeconomic variables using personal identification numbers (Statistics Sweden 2021b, 2021c, 2021d, 2021f). We limit our analysis to full siblings (same biological mother and father) and biological children.

There are a total of 1.3 million individuals applying to any type of high school program between 1977-1991. There are a median of 432 applicants within a school region in a year; depending on the year, there are between 114 and 137 high school regions. Our estimation samples for older siblings and parents are limited to those who have a first-best academic choice where demand exceeds supply. This includes 611,837 individuals applying to one of the five academic programs, of which 326,211 had a first-best choice which was oversubscribed. As Appendix Table A2 shows, the characteristics of students applying to oversubscribed programs is broadly similar to those applying to non-impacted programs. We further condition on the applicants who have a second choice (reducing the sample to 288,357), school regions and years where two or more majors were not combined (284,329), and GPA being non-missing, between 2.0 to 5.0, and within a sample window of -1.0 to +1.5 points around the normalized cutoff (278,968). We then drop individuals whose GPA lands exactly at a cutoff where only some applicants are admitted (263,878); for details see Appendix A.

From these 263,878 observations we create our sibling and intergenerational samples. For

[12]For 2011-2019, we have roughly 6,000 observations where the older sibling applied to an oversubscribed program, which is far too small to be useful. In comparison, our 1977-1991 sibling sample has over 88,000 observations.

the sibling sample, we first condition on the applicant having a younger sibling, which yields 154,283 observations.[13] In addition, we require the older sibling to be at least one year ahead in school; older siblings need to apply between 1977-1990 and younger siblings between 1978-1991. This results in our baseline estimation sample of 88,174 sibling pairs.[14] For the intergenerational sample, we condition on the applicant having a child who applies between 2011-2019. This results in our baseline estimation sample of 168,933 parent-child pairs.[15]

2.3 Major Choices and Cutoffs

Appendix Figure A1 shows the distribution of major choices for applicants to an academic track for both of our time periods (panel A). Between 1977-1991, Engineering and Business were the most popular choices, with over one-fourth of applicants choosing each of these majors. Humanities was the least popular with fewer than 10% of individuals listing it as their first choice. The Engineering and Business shares decline substantially by 2011-2019, with Social Science seeing the largest increase.

A key difference across majors is the fraction of male versus female applicants. Panel B reveals that fewer than 20% of applicants who listed Engineering as their first-best choice were female. On the other end of the spectrum, in 1977-1991, Social Science was 70% female and Humanities was 85% female. In between are Natural Science and Business, which attract a roughly equal sex mix. This variation will allow us to explore whether male-dominated, gender-balanced, and female-dominated majors induce different types of family spillover patterns.

We next turn to describing the cutoffs for the oversubscribed programs. Details on how we determine the GPA cutoffs, which are not recorded in our dataset, are found in Appendix A. The distribution of cutoff GPAs is plotted in panel A of Figure 1. The figure also overlays the distribution of individual GPAs for applicants to oversubscribed academic majors. The mean cutoff GPA for oversubscribed majors is 3.4, a value which corresponds to the 18th percentile of GPAs among applicants to oversubscribed academic majors. To put this in further perspective, the mean cutoff GPA also corresponds to roughly the 60th percentile of GPAs in the sample of all ninth graders, including those who apply to nonacademic majors or do not continue on

[13] We cap family size at 5 siblings for the sibling sample and 5 children for the intergenerational sample, which drops 1.6% and 0.5% of the data respectively.

[14] There are 2.7 siblings on average in a family during 1977-1991, with older siblings having 1.1 younger siblings on average in our estimation sample.

[15] For the intergenerational sample, parents have 2.4 children on average, of which 1.5 are observed during 2011-2019 in our estimation sample.

to high school. While the cutoffs vary from year to year, they are generally only binding for applicants with GPAs in the bottom half of our estimation sample.

As we noted earlier, it is important not to confuse "oversubscription" with "highly competitive." This is because the cutoff for a major is determined by local supply versus demand, which varies from year to year within a school region. There is not a consistent ordering over time for which majors are more likely to have higher admission cutoffs. For example, Engineering has a higher cutoff than Natural Science in 37% of years within the same school region on average, while the reverse is true in 25% of years. In 38% of years both programs either have open enrollment, or less commonly, identical cutoffs. Similar patterns are found for the other major combinations as reported in Appendix Table A3. Moreover, as panel B in Figure 1 shows, the distribution of cutoffs are fairly similar across the different academic majors, with mean cutoffs differing across majors by less than 0.2 GPA points, a small amount relative to the distribution of students' GPAs. These facts regarding the major cutoffs are useful to keep in mind when interpreting the estimates, which will capture local average treatment effects for applicants around the cutoffs.

3 Model

3.1 Using Preferred Choices in an RD Design

Our goal is to estimate family spillovers in high school majors. In this section, we talk about modeling the effect of older brothers on their younger brothers, but the same ideas apply to other sibling and intergenerational dyads. Our setting and empirical approach is closely related to ideas first found in Jackson (2010) and later used in Beuermann et al. (2022), which study effects of attending a better public school. Their allocation mechanism closely mirrors ours, with students being assigned to schools, in part, based on their rankings and scores on a standardized exam, which generates discontinuity variation around cutoffs.

The idea behind our estimation approach can best be understood by starting with a simple example. Consider brother pairs where the older brother chose Engineering as his first choice. The reduced form RD is a sharp design which compares whether a younger brother chooses Engineering as his first choice depending on whether his older brother barely was admitted versus not admitted to Engineering. It is a sharp design because all older brothers whose GPA exceeds the cutoff are admitted while none whose GPA is below the cutoff are admitted. The reduced form RD can be written as:

$$y_{st} = 1[x < c_{st}]g^l(c_{st} - x) + 1[x > c_{st}]g^r(x - c_{st}) + 1[x > c_{st}]\theta + w'\gamma + e_{st} \quad (1)$$

where we have omitted the subscript identifying sibling pairs for convenience. The cutoff (c_{st}) for the running variable of GPA (x) depends on the school region (s) and year of application (t) for each major, and is normalized to be zero to allow for pooling across school regions and years. The outcome variable y_{st} is a dummy for whether the younger brother chooses the same first-choice major as his older brother (i.e., Engineering). The functions g^l and g^r are allowed to differ to the left and right of the cutoffs, w is a set of pre-determined controls (region and year fixed effects), and e is an error term. In our setting, every applicant whose GPA exceeds the cutoff is accepted to their first choice major, while every applicant below is not. The main coefficient of interest is θ, which is the jump at the cutoff.

Continuing with the example, Figure 2 plots the probability a younger brother chooses Engineering against the running variable of the older brother's GPA. There are different slopes to the left and right of the cutoff, and more importantly, there is a sharp and statistically significant jump in the probability a younger brother chooses Engineering at the cutoff.[16] We estimate a precise spillover effect in this example because over 50% of older brothers choose Engineering as their first choice. To gain power so that we can estimate and compare effects for different gender dyads, we stack data across all majors.

Stacking across different first-choice majors, where we restrict slopes and jumps to be common across first-choice majors, yields the following reduced form RD:

$$y_{jst} = 1[x < c_{jst}]g^l(c_{jst} - x) + 1[x > c_{jst}]g^r(x - c_{jst}) + 1[x > c_{jst}]\theta + \delta_j + w'\gamma + e_{jst} \quad (2)$$

where the notation is the same as in equation 1, except now the subscript j is added to indicate the first-choice major of the older brother. The cutoffs are first-choice major specific, and as are the intercepts. As before, the outcome variable y_{jst} is a dummy for whether the younger brother chooses the same first-choice major as his older brother.

Our baseline model adds extra flexibility by allowing the slopes and intercepts to account for the major an older brother is admitted to. We allow for different slopes to the right of the cutoff for each of the first-choice majors (corresponding to the major an older brother is admitted to), and different slopes to the left of the cutoff for each second-choice major

[16]As Appendix Figure A2 shows, the increase in the probability a younger brother chooses Engineering is offset by a reduced probability of choosing Natural Science, and to a lesser extent, Business.

(corresponding to the major an older brother is admitted to).[17] We also allow separate first and second choice major intercepts.

To gain precision, we always combine the four sibling dyads into a single regression. We start by showing aggregate regressions, and then proceed to models which allow the jump at the cutoff and the dummies for preferred major to be dyad specific. We do the same for our intergenerational regression. We further explore heterogeneity based on whether a first-choice major is gender conforming. We probe robustness on a variety of dimensions, including using the less flexible common slope model of equation 2. We scale the reduced form effects using a fuzzy RD, with enrollment in a major (for older siblings and parents) or completion of a major (for parents) in first stage regressions.

3.2 Threats to Validity

To have a valid RD design, the running variable cannot be perfectly manipulated around the cutoff. There is little chance of such manipulation in our setting. One reason is that the required GPA to gain admission to a major is not known in advance, and varies from year to year as a function of the number of applicants. Thresholds differ 83% of the time for majors with a cutoff in successive years, as illustrated in Figure 3.

One way to test for manipulation is to check for balance in pre-determined characteristics around the admission cutoff. We do this in Appendix Figure A3 for all applicants to an oversubscribed major during 1977-1991. There are no noticeable jumps at the cutoff, a finding which is confirmed with formal tests.[18] Another common test for manipulation is based on jumps in the density of the running variable at the cutoff. However, a standard McCrary (2008) test or the newer test proposed by Cattaneo et al. (2018) are not applicable since the cutoff is based on an order statistic.[19]

To identify the causal spillover effects of enrollment or completion, we need exclusion, monotonicity, and irrelevance conditions to hold. These are extra conditions which are not required for the reduced form of admission. The exclusion restriction requires that crossing the admission threshold for a major only affects outcomes through enrollment (or completion). For enrollment, being admitted but not enrolling provides little information or experience with a major, and so it seems likely that the condition holds both for siblings and parents.

[17] There are 5 first choices but 7 second choices, because we allow vocational and non-academic general majors as second choices, but only academic majors as first choices.

[18] For the 8 estimated jumps, the largest t-statistic is 1.4 and six are below 1.

[19] In ongoing research, Cattaneo, Dahl, and Ma are working on a proof for there being a spurious density discontinuity and ways to modify a density test to account for this.

For completion, it is possible that parents take some specialized courses because they are admitted to a major, but then do not complete the major. However, given the large differences in curricula across majors, most switching occurs early on during the fall of the first year so there is limited scope for this channel. The monotonicity assumption requires that crossing an admission threshold does not make an individual less likely to enroll in (or complete) that major, which seems plausible.

Finally, we require the irrelevance condition discussed in Kirkeboen et al. (2016). In our context, this condition means that if crossing the GPA threshold for admission to a major (e.g., Engineering) does not cause a parent or older sibling to complete that major, then it does not cause them to complete a different major (e.g., Humanities) either. While this condition seems plausible in our setting, it is in theory possible that it does not hold for *completion* of a major. In contrast, we note that the irrelevance condition holds by construction for *admission* to a major. This is because we have a sharp discontinuity for admissions, where everybody above the GPA cutoff is admitted to the major. It is only when we use program completion to scale our reduced form estimates using a fuzzy RD that this issue arises. Likewise, it seems likely to hold for *enrollment* in a major, as almost everyone who is admitted to their first choice major initially enrolls in the major (see Figures 5 and A3) with switching mostly happening after enrollment.

To further probe the validity of our design, we conduct a placebo test. Specifically, we test whether a younger sibling's acceptance into a major affects their older sibling's ex-ante choice. Since the older sibling makes their major choice before their younger sibling knows if they have been accepted, there should not be an effect. In Table 1, we conduct this placebo test using the same RD regression specification as our main estimates (as explained in detail below). Both combined and for each gendered sibling pair, the placebo estimates are close to zero and not statistically significant. Note that a similar exercise cannot be performed for parents and children, since children made their major choices during a time period when few majors were oversubscribed.

4 First Stages

We begin our presentation of empirical results by documenting the first stages for both older siblings and parents. The first stages capture how admission to a major affects either enrollment or completion. Admission exhibits a sharp discontinuity as a function of an individual's GPA, jumping from 0 to 1 at the relevant cutoff. In other words, every applicant whose GPA exceeds the cutoff is accepted to their first choice major, while every applicant below is not. This is illustrated in Figure 4. Hence, we have a sharp RD design for the question of how admission to a major affects the choices of younger siblings or of children.

To scale this reduced form effect, we use a fuzzy RD (e.g., IV) design. For siblings, the relevant first stage is how an older sibling's admission affects their enrollment in a major. For parents, there are two possible first stages: one for enrollment in a major and one for completion of a major. We do not use major completion as a first stage for siblings because the older sibling often has not completed high school before the younger sibling submits their ranking list of preferred majors.

Figure 5 illustrates the first stages for siblings, separated by gender mix. Consider the top-left graph, which plots the probability an older brother enrolls in his first-best choice in normalized GPA bins. Everyone to the right of the vertical line is (initially) admitted to their first choice major, while everyone to the left is not (initially) admitted. Some people switch to other majors, usually before the school year begins. This reshuffling opens up slots for other students and explains why some individuals to the left of the admission cutoff are able to enroll in their first-best choice. Note that the density of normalized GPA is such that there are relatively few observations to the left of the cutoff, so a small drop in the enrollment shares in a GPA bin to the right of the cutoff can explain a large increase in the enrollment shares to the left of the cutoff.[20] Switching majors is not necessarily random which is why we need to instrument for enrollment using the admission decision. There is a roughly 60% jump in the probability of enrollment at the cutoff; similarly-sized jumps occur in the other panels for the other sibling gender combinations.

The first-stage graphs for parental enrollment look similar to those for siblings, and are therefore relegated to the appendix (Appendix Figure A4). This is not surprising, as both the older sibling and parent samples span the same time period of 1977-1991. Parents have an additional first stage for completion of a major. In Figure 6 we plot this alternative first stage. Consider the top-left graph, which plots the probability a father completes his first-best choice in normalized GPA bins. A nontrivial fraction of fathers admitted to their first-best major do not complete it. This is mostly due to individuals switching to other majors, usually in their first year of studies; few drop out of high school entirely (approximately 5%). There is

[20]For example, suppose there is an excess of applicants to the Engineering program, two classes of size 30, and 5 applicants just to the left of the cutoff. If 2 of the initially accepted students switch out of Engineering, this will open up 2 slots. If 2 of of the 5 individuals just to the left of the cutoff take these slots, then the enrollment rate for this group will be 40%.

a roughly 45% jump in the probability a father completes his first-best choice at the cutoff. Similar jumps are found in the other panels of Figure 6.

In Table 2 we report the corresponding first stage regression estimates. The first stage estimates do not vary much by the gender mix of a dyad. The first stage enrollment estimates for older siblings or parents range from .59 to .64, while the first stage completion estimates for parents range from .43 to .47. The estimates are all highly significant, indicating there will not be a problem due to weak instruments.

5 Theories of Family Spillovers

Before presenting our results, we review three leading theories posited in the literature for family spillovers: (i) role model effects, (ii) differentiation due to rivalry or rebellion, and (iii) information transmission/support. A recurring hypothesis within each of the these theories is that the gender mix of a dyad matters, with the strongest influence coming from same-gender matches (Russell and Saebel 1997). Another recurring theme is that fathers and mothers have different types of influences on their children, and that these effects can additionally depend on the gender of a child (e.g., Bush and Peterson 2013, Holmlund et al. 2011, Raley and Bianchi 2006). Most of this work is conceptual in nature, often motivated by experiential knowledge, surveys, or correlations. A more recent literature adds the extra layer of whether choices are gender conforming or not. Below we outline what has been hypothesized for each of these forces, recognizing that multiple theories could be in play simultaneously and could even work against each other.

Role model effects. The first use of the term "role model" was recorded in 1947. A role model is defined as "a person whose behavior, example, or success is or can be emulated by others, especially by younger people" (Random House, 2022). A closely related concept is "social learning theory", first developed by Bandura and Walters (1977), where individuals observe and imitate the behavior of family members, friends, or famous figures. The literature has pointed to older siblings and parents as potential role models. Early on, Brim (1958) argued that older siblings act as an example which younger siblings follow. McHale et al. (2012) reviews settings where younger siblings choices are impacted by older siblings, such as for prosocial and risky behavior. Beginning with Koch (1960), psychologists have argued that the impact of role models is larger for same-gender siblings and those further apart in age. Parents who are the same gender as their children are also thought to be more influential role models (e.g., Russel and Saebel 1997).

Differentiation. A spillover posited to have an opposite-signed effect is "differentiation", which is the process of building one's own identity separate from those around you (McHale et al. 2012). Siblings could choose different niches or activities to participate in as a way to reduce competition and develop their own sense of self. The literature hypothesizes that sibling rivalry should be especially strong for those close in age (Adams 1972, McHale et al. 2013) or of the same gender (Conley 2000, McHale et al. 2013). At the intersection of these two forces, two closely spaced brothers are predicted to have the most rivalry of any sibling pair (Dunn and Kendrick 1982). Differentiation by children from their parents could be driven by rebellion, where children fight against the authority and expectations of their parents. In particular, later-born children are thought to rebel more, while first born children identify with parents (Sulloway 1996).

Information and support. Family members could be important information providers. In our context, older siblings could provide gender-neutral information on teachers, course material, and the difficulty of a major. Parental information about majors should be less accurate, as teachers and course content will have changed in the intervening decades. But parents will be able to convey information on their labor market experiences, including how their high school major helped prepare them for their occupation. Siblings and parents could also play supportive roles, for example, by helping with subject-specific homework. In terms of closeness, Buist et al. (2002), Floyd (1995), and Pulakos (1989) all argue that sisters and mother-daughters have more intimate relationships, and hence, play more supportive roles compared to other dyads.

Gender conforming choices. A more recent literature has added the extra layer of whether a behavior is gender conforming.[21] First, role models are thought to be especially important for women and non-gender conforming choices, but less so for men.[22] For example, parents in science occupations influence whether their daughters choose a STEM major (Anaya, Stafford and Zamarro, 2022). Second, to our knowledge, there are no theories for how rivalry and

[21] A related literature studies how the interaction of family influences and gender norms are affected by changes in the gender mix of children (Brenøe 2022, Humlum et al 2019, Ogzoglu and Ozbeklik 2016, Phillipp 2022). These theories are beyond the scope of this paper, as we lack the precision to test them.

[22] For example, female college students who have a female professor in math and science classes are more likely to major and graduate with a STEM degree (see, for example, Bettinger and Long 2005, Carrell et al. 2010). See also Dryler (1998).

rebellion interact with whether a choice is gender conforming. Third, those who are a minority in their major may be uniquely positioned to provide information on what it is like to pursue a non-gender conforming choice (e.g., females in Engineering or males in Humanities). Older siblings could provide information on what it is like to be a gender minority in the class, while parents could provide information on the types of barriers one is likely to face in the labor market.

6 Aggregate Estimates

We start our presentation of results by lumping all sibling pairs together and all child-parent pairs together regardless of gender. The dependent variable is a dummy for whether a younger sibling's preferred major (or child's preferred major), defined as the first choice on their ranking list of majors, matches their older sibling's (or parent's) first choice. The RD regressions use the baseline model described in Section 3.1, a window of -1.0 to 1.5, and triangular weights. The regression also includes the demographic variables listed in Appendix Table A2. Standard errors are clustered at the family level.

Panel A of Table 3 reports both reduced form and fuzzy RD estimates for all sibling dyads combined. When we write that a child "copies" their parent, we mean that the child's major choice matches either the major a parent was admitted to, enrolled in, or completed (and similarly for siblings). The reduced form estimate of 1.5 percentage points captures the increased probability a younger sibling chooses the same first-choice major as the one their older sibling was admitted to. On average, younger siblings choose the same major as the one their older sibling was admitted to 19.6% of the time,[23] so the reduced form estimate represents an 8% increase. The IV estimate (i.e., the fuzzy RD estimate) scales the reduced form by the first stage probability an older sibling enrolls in a program, and is 2.5 percentage points, for a 13% increase relative to the mean.

In panel B, we report the reduced form and fuzzy RD estimates for all parent-child pairs combined. The magnitudes are remarkably similar to the sibling estimates in panel A. The reduced form indicates that children are 1.4 percentage points more likely to choose their parent's first-choice major if their parent's GPA is just above the admission threshold. As a reminder, there are two possible first stages for parents: enrollment in a major and completion of a major. Starting with the enrollment IV estimate, there is a 2.3 percentage point increase

[23] We calculate this average using the sample where an older sibling's GPA is within plus or minus 0.2 GPA points of the threshold for admission to a major.

in children choosing the same major as the one their parent enrolled in. Since children copy their parent's major choice 21.4% of the time on average, this translates to an 11% increase.[24] The other possible IV estimate captures how parental completion of a major affects a child's choices. The effect size rises to a 3.1 percentage point effect, or a 14% increase.

Comparison to Correlational Estimates. How do these spillover effects compare to correlational estimates? To see how we construct the correlational estimate, consider siblings and the Engineering major. We regress whether the younger sibling chooses Engineering on whether their older sibling enrolled in Engineering. This is a correlational estimate because whether the older sibling enrolled in Engineering is not randomized, and hence could suffer from selection bias. Since the right hand side variable is dichotomous, the regression coefficient in this simple regression is equivalent to the difference in means for younger siblings whose older brother did versus did not enroll in Engineering.

Building on this example, for siblings, we calculate the fraction of younger siblings who list a major as their first choice if it is the one their older sibling enrolled in minus the fraction who choose it when their older sibling did not enroll in it. Doing this for each of the 5 majors and then taking the average across majors (weighted by the number of older siblings enrolled in each of the majors), yields our combined sibling correlational estimate. This correlational estimate can be compared to the RD enrollment estimate appearing in Table 3, as it uses the same outcome and independent variable. As reported in Table 4, for siblings, we find a 10.2 percentage point increase in younger siblings copying their older sibling's enrollment choice. This is 4.1 times larger than the IV estimate of 2.5 percentage points. Doing a similar exercise for parents and their children, we find a correlational estimate of 6.3 percentage points, which is 2.7 times larger than the IV estimate. The differences between the correlational and IV estimates are statistically different from each other at the 1% significance level.

The upshot of these comparisons is that the correlational estimates vastly overstate the magnitude of within-family spillovers.[25] This could be because unobserved factors which are common in the family drive similar major choices, rather than similar major choices reflecting a spillover effect. However, there are still sizable roles for sibling and parental influences

[24]We calculate the copying average using the sample where a parent's GPA is within plus or minus 0.2 GPA points of the threshold for admission to a major.

[25]The differences between the correlational and IV estimates are not primarily due to IV capturing a local average treatment effect. If we restrict the correlational samples to an older sibling's or parent's GPA being within plus or minus 0.2 GPA points of the threshold for admission to a major, the correlational estimates are 3.0 times larger for siblings and 2.1 times larger for parents and their children.

– just not as large as raw correlations would suggest. This parallels what has been found for correlational versus causal estimates of the intergenerational link in years of schooling. Summarizing papers in the Scandinavian context, Holmlund et al. (2011) report that the observed correlation across generations in years of schooling is between 0.2 and 0.3, while similar causal effects are markedly smaller at approximately 0.1.

While these aggregate effects are interesting, the theories of family spillovers outlined in Section 5 hypothesize that the gender mix of dyads should play an important role. We explore these heterogeneous effects in the next two sections.

7 Sibling Spillovers by Gender Mix

This section reports reduced form and IV estimates of sibling effects as a function of both gender mix and birth spacing. The first two columns of Table 5 report sibling effects by gender mix. The magnitudes and even the direction of spillovers depend heavily on the gender composition of a dyad. Start with brother pairs. The reduced form shows that if an older brother is admitted to his first-best choice, there is a 4.3 percentage point higher probability his younger brother will choose the same major. The corresponding IV estimate implies a younger brother is 6.3 percentage points more likely to choose the same first-choice major if his older brother enrolled in it. Relative to the mean, this translates to a 25% increase. Turning to sister pairs, there is also a strong same-sex sibling effect. An older sister's enrollment increases the chances her younger sister will choose the same major by 3.9 percentage points, or 18% relative to the mean. These are sizable same-sex sibling spillovers, consistent with what the literature hypothesizes.

The results for opposite-gender sibling pairs stand in sharp contrast to the large, positive spillovers for same-sex pairs. Consider older sisters and their younger brothers. In these families, the brothers recoil from the majors of their older sisters. A younger brother is 3.0 percentage points *less* likely to choose a major his older sister enrolls in. This amounts to a 26% reduction compared to the mean. When interpreting this effect, it is important to keep in mind that younger brothers are less likely to choose similar majors as older sisters (11.7%) versus older brothers (25.3%) due to the gender makeup of majors. Hence, an equally sized percentage point reduction for older sisters will translate to a larger percent reduction relative to the mean. Turning to older brother - younger sister sibling pairs, there is a small and statistically insignificant effect.

An important question in the literature is whether the four gender-specific dyads could be reduced to just two groups – same-sex and opposite-sex – regardless of gender. We conduct a joint test for whether the IV coefficient on brother-brother equals sister-sister and the coefficient on sister-brother equals brother-sister. We reject the null hypothesis that gender doesn't matter for these groupings (p-value = .04). We conclude that there is not a simple same-sex versus opposite-sex division in spillover effects, but rather more detailed gender-specific patterns.

Graphs of reduced form effects by sibling gender mix can be found in Figure 7, using raw data which do not include any controls. To understand these RD graphs, focus on the older brother - younger brother group. As an older brother's GPA increases, there is an increasing chance a younger brother will choose the same major. This indicates that younger siblings are more likely to copy older siblings who are better students. But more importantly, there is a discrete jump up in this probability of choosing the same major at the cutoff. There are likewise noticeable jumps which accord with the regression results for older sister - younger brother and older sister - younger sister pairs.

Turning to correlational estimates by gender mix (Appendix Table A4 panel A), these paint a biased and misleading picture, just as we found when doing the corresponding comparison at the aggregate level. The same-gender (brother-brother and sister-sister) correlational estimates are larger than the causal RD estimates by a factor of 3. Moreover, the correlations indicate positive effects of older sisters on younger brothers, while the IV estimates reveal a negative recoil effect, and the strong positive correlational estimates of older brothers on younger sisters is not present in the IV results. All of these differences are statistically significant at the 1% level.

Estimates by Birth Spacing. To gain insight into which of the theories described in Section 5 are driving the gender-mix patterns we observe, we estimate effects as a function of birth spacing. Differentiation due to sibling rivalry is hypothesized to be stronger for siblings close in age, while role model effects are thought to be stronger for those further apart in age. Moreover, sibling rivalry is predicted to be strongest for brother pairs, while sister pairs are thought to be more supportive of each other. We estimate separate regressions for siblings born within 3 years of each other versus further apart. This split is in part motivated by the fact that high school typically takes 3 years to complete for an academic major.[26] In our setting,

[26] All academic major take three years to complete, except for Engineering which can involve an optional fourth year. In Sweden, the age cutoff for starting school as a child is January 1. Hence if siblings are born within three calendar years of each other, they will attend high school at the same time (unless the older sibling has skipped a grade or the younger sibling has repeated a grade).

attending school concurrently should further increase sibling rivalry, while older siblings who have finished their degree should provide better information about the major.

Start with same-gender sibling pairs. For brother pairs, columns 3-6 in Table 5 show that the positive effects observed in columns 1-2 are largely driven by those who will not be attending school at the same time. There is a 10.7 percentage point spillover effect, or a 42% relative increase, for brothers spaced more than 3 years apart. In contrast, younger brothers with concurrent school attendance experience an effect one-fourth as large, indicating that rivalry outweighs any role model or information effects. This difference in magnitudes is statistically significant (p-value for difference < 0.01). As predicted by the literature, the results are consistent with sibling rivalry being most pronounced for brothers close in age, and with brothers more distant in age either serving as stronger role models or providing better information. Sister pairs also exhibit heterogeneity by birth spacing, but with the reverse pattern. The positive spillover effect is roughly twice as large for sisters attending school concurrently, although the estimates are noisy enough that the difference is not statistically significant. This pattern aligns with the prediction that sister pairs are more attached to each other and supportive of each other's goals. This is not to say that rivalry does not exist among sisters, but that other mechanisms dominate.

Now turn to mixed gender sibling pairs. The recoil effect documented in columns 1-2, where a younger brother turns away from his older sister's first-choice major, is largest when the two siblings are close in age. The negative spillover effect amounts to a 41% reduction for those who will be attending school at the same time, but is close to zero otherwise (p-value for difference $= 0.13$). This result was not predicted by the literature, which hypothesized that mixed gender siblings should have less differentiation compared to same-gender siblings. For older brother - younger sister sibling pairs, there is no spillover effect for concurrent school attendance and a hint of a positive effect for siblings spaced further apart. This second result is consistent with the literature's claim that opposite gender siblings should have smaller spillover effects.

While the heterogeneity by gender mix is interesting in itself, it also highlights that looking at spillover effects without separating by gender mix can lead to misleading conclusions. If we ignore gender mix and instead impose a common effect for sibling spillovers, we estimate a more modest aggregate 13% spillover effect (Table 3). But this masks the fact that some effects are positive, some are negative, and some are zero.

Estimates by Parent's Educational Background. In Appendix Table A5, we present estimates by parent's educational background. The rationale for this split is that if neither parent attended an academic program in high school, the older sibling may have a larger effect on their younger sibling's major choice. However, we find no statistically significant evidence for differential effects by parental education. In unreported results, we also find little evidence of heterogeneity by either the ability (as measured by GPA) of older siblings or parents, or by the ability of younger siblings or children.

8 Intergenerational Spillovers by Gender Mix

We now shift focus to intergenerational effects by gender mix. As outlined in Section 5, same-gender parents are hypothesized to be the strongest role models. Differentiation due to rebellion could also cause intergenerational spillovers. Moreover, parents could provide information about their experiences in the labor market, which could be specific to their gender. In terms of closeness, mothers and daughters are thought to have the strongest bond.

We start by estimating effects for same-gender dyads. There are large spillovers from fathers to sons. Table 6 reports a 2.9 percentage point increase in a son's probability of choosing a major if his father was admitted to it, a 4.3 percentage point increase if his father enrolled in it, and a 5.6 percentage point increase if his father completed it. The IV estimate based on completion translates to a 22% increase relative to the mean. In sharp contrast, the estimates for mother-daughter pairs are close to zero and statistically insignificant.

Turning to opposite gender dyads, we find that mothers have a strong influence on their sons. Table 6 reports a 1.5 percentage point increase in a son's probability of choosing a major if his mother was admitted to it, a 2.4 percentage point increase if his mother enrolled in it, and a 3.1 percentage point increase if his mother completed it. The IV estimate for completion translates to an 18% increase relative to the mean. When comparing effect sizes across father and mothers, it is important to keep in mind that sons are more likely to choose similar majors as their fathers than their mothers due to the gender makeup of majors (25.4% versus 17.2%). In contrast, we find no evidence that fathers influence their daughters' choices.

These estimates reinforce the lesson learned from the sibling analysis: estimating spillovers which do not distinguish by gender can lead to incomplete and misleading conclusions. The patterns are not consistent with the literature's prediction that a parent's influence should be strongest when the parent and child are the same gender. Instead, sons copy their parents' first-choice major, while daughters do not. An F-test for whether sons and daughters are equally affected by their parents is rejected with a p-value of 0.02 using the IV-completion estimates. This is a test for whether the father-son coefficient equals the father-daughter coefficient and the mother-son coefficient equals the mother-daughter coefficient.

The positive spillovers for sons could be due to role model effects or information, but in ways which are largely similar for fathers and mothers. The absence of spillovers for daughters is surprising. One possible explanation is that the education landscape and labor markets have changed more for girls than boys in the time since parents made their choice. This would make a parent's knowledge about majors less informative for daughters than sons. Consistent with this explanation, the Duncan indices (Duncan and Duncan, 1955) for Sweden have fallen from .55 in 1980 to .37 in 2010 for high school majors and similarly from .66 to .52 for occupations (Halldén and Stenberg, forthcoming).

RD graphs illustrating the reduced form effects are found in Figure 8. There are noticeable jumps at the GPA admission thresholds for sons choosing the same first-choice major as their fathers and as their mothers. In contrast, there are no noticeable jumps for daughters for either parent gender. While not the focus of our analysis, the slopes as a function of the running variable provide additional evidence that parents have less impact on daughters versus sons in their major choices. Starting with the first figure, the higher a father's GPA, the more likely a son is to copy his father's first-best major choice. A similar pattern emerges for mothers and sons. But the slopes are relatively flat for daughters copying either their mothers or their fathers major choices. In other words, the pattern of the slopes suggest that sons are heavily influenced by their parent's GPA when deciding whether to choose a similar major, but daughters are not.

Panel B of Appendix Table A4 reports corresponding correlational estimates. As we found when doing a similar comparison at the aggregate level (see Table 4), these correlations overstate effects – by a factor of more than 2 for sons and 4 for daughters. The correlations suggest that daughters are as influenced by their parents' choices as are sons, but the IV estimates reveal this is not the case. All of the correlational versus RD differences are statistically significant at the 5% level. The contrast between the correlational and causal estimates again highlights the need to use a convincing research design to estimate family spillovers.

Estimates by Birth Order. In Appendix Table A6, we report estimates based on whether a child is the firstborn in a family. The idea behind this split is that firstborn children do not have an older sibling making a high school major choice before they do, and so perhaps a parent's influence will be stronger. We find some evidence for this expected pattern. The IV estimates for each of the parent-child gender pairs are larger for firstborn versus non-firstborn children. However, these differences are not statistically significant. The largest gap is for sons of fathers, with firstborns experiencing a 7.8 percentage point increase in choosing the same major as the one their father completed compared to a 3.3 percentage point effect for later born children (p-value for difference = 0.14).

9 Specification Checks

Before continuing, we present a variety of robustness checks in Table 7, both for the sibling and intergenerational analyses. For simplicity, we focus on the reduced form results; IV robustness checks yield similar conclusions.

We start by probing the parameterization of the regression model. Column 1 repeats our baseline estimates from earlier tables for ease of comparison. In column 2, we use second-order polynomials instead of linear trends. The reduced form estimates are similar in magnitude, but the standard errors increase by roughly 30%.[27] Column 3 reveals the results are robust to cutting the window size in half. In column 4 we exclude observations with low cutoffs (i.e., cutoffs having fewer than 50 observations below) or high cutoffs (i.e., cutoffs having fewer than 50 observations above).[28] This reduces the sibling sample by 7% and the intergenerational sample by 3%, but does not change the results. The estimates are also robust to omitting 1982-84 (see footnote 10) or excluding the demographic control variables (columns 5 and 6). Column 7 limits the sibling sample to dyads from two-sibling families (i.e., families with exactly two children), and finds similar results.

We next show estimates for both a less and more flexible parameterization of the slopes which are a function of the running variable. Using either common slopes regardless of major choices or slopes which vary as a function of each first-second choice combination yields similar results (columns 8 and 9). In Appendix Figure A5, we illustrate that the common slope model is too restrictive for some major choices, which motivates why we use our more flexible baseline model. The graphs in the first column plot averages of the binned outcome variable for younger siblings (panel A) and children (panel B) against the running variable, allowing for separate

[27]Although not shown, the IV estimates become slightly larger since the first stage estimates shrink with a second-order polynomial. In this sense, the linear estimates we report as our baseline IV can be viewed as conservative.

[28]For example, for the brother-brother sibling sample, this excludes observations where the GPA cutoff is below 3.0 or above 4.4.

slopes for each of the five first choice majors to the right of the cutoff (i.e., those corresponding to the major an older sibling or parent is admitted to in case their GPA is above the cutoff), and a common slope to the left of the cutoff. The second column shows similar plots, but allowing separate slopes for each of the seven second-choice majors to the left of the cutoff (i.e., those corresponding to the major an older sibling or parent is admitted to in case their GPA is below the cutoff), and a common slope to the right of the cutoff. We emphasize that Appendix Figure A5 is for illustrative purposes only.

In Appendix Table A7, we perform a different set of robustness exercises, where we explore alternative measures for whether a younger sibling or child copies their older sibling or parent. Our baseline definition uses whether the first-choice major on a younger sibling's or child's ranking list matches their older sibling's or parent's preferred major. If we instead use whether the younger sibling or child included their older sibling's or parent's preferred major in *any order* on their ranking list, the results hardly change. For the sibling analysis, the results are also similar if we use whether the younger sibling was accepted to or enrolled in the same major as their older sibling's preferred major, and become slightly smaller if we use major completion. For the intergenerational analysis, the magnitudes drop by roughly 30-40% if we use whether a child's acceptance, enrollment, or completion of the same major is the same as their parent's preferred choice. These patterns hold for both the reduced form and IV specifications.

Our baseline sample includes younger siblings and children who apply to high school (including both academic and vocational majors). A relatively small fraction of younger siblings and children do not apply to high school at all, but drop out after ninth grade (11 and 4 percent, respectively). If we include those who drop out after ninth grade, and code them as not copying their older sibling or parent, the pattern of results is unchanged (see Appendix Table A8).[29]

In our setting, students list their preferred majors but not their preferred schools. However, it is possible that some majors are only found in some schools, which potentially confounds major and school choice. As a final robustness exercise, we limit our sample to older siblings whose first and second best major choices are both available at the school they are admitted to. For most school regions (on average 88 out of 129) there is only one school. In some of the remaining regions, however, different majors are found at different schools − for example a

[29]A related issue for the intergenerational analysis is that parents with an academic major could be more or less likely to have a child appearing in our analysis sample. However, we find no evidence that a parent getting into their first choice major had any effect on the probability of having a child who appears in our sample, with an estimated effect of -.005 (s.e. = .004).

region with two schools may have one school for Engineering and Natural Science and another for Business, Social Science and Humanities. In this example, our sample restriction would include older siblings whose first and second best choices were Business and Social Science, but not include those whose first and second best choices were Engineering and Business (since then it is unclear whether a younger sibling is copying a major or an institution). By excluding such cases, but retaining the others, we ensure that the only relevant margin is major choice for the younger sibling or child.

Appendix Tables A9 and A10 repeat the baseline regressions found in Tables 5 and 6, but using the restricted sample which holds constant the school the older sibling was admitted to. The restricted samples have 27% and 28% fewer observations for the sibling and intergenerational analyses, respectively. While the standard errors are somewhat larger, the estimates are almost identical for both tables. Twelve out of 13 estimates remain statistically significant at conventional levels for the sibling table, and 5 out of 6 remain significant for the intergenerational table. We conclude that our findings are driven by major choice, and not confounded by school choice.

10 Estimates by Gender Makeup of Majors

As a final exercise, we look at major-specific spillovers within our gender mix dyads. We estimate separate sibling and intergenerational effects by the gender share in a major. This is motivated by hypotheses that gender-specific role-modeling matters, and in particular that mothers in STEM fields can have a strong influence on their daughters. We continue to estimate the effect on a younger sibling (child) choosing the same major as their older sibling (parent). But instead of imposing a common treatment effect, we estimate separate effects by major type. We define male-dominated, female-dominated, and gender-neutral majors according to the national criteria the government uses to assign gender bonuses for underrepresented genders in majors.[30] Engineering is male-dominated, Social Science and Humanities are grouped as female-dominated, and Natural Science and Business are grouped as gender-neutral.[31]

In Appendix Table A12 we characterize the student peer group an older sibling or parent is exposed to when admitted to each of these major groupings. We first explore how the gendered environment changes when an individual is admitted to a male-dominated, female-dominated, or gender-neutral major. We define the female share as the older sibling's (or parent's) fraction

[30]As mentioned in Section 2, applicants have an extra 0.2 bonus points added to their GPA if they apply to a major which accepted 30% or less of their gender nationally in the prior year.

[31]Appendix Table A11 shows separate estimates by each of the 5 academic majors.

of women in the same year and region for the major they are admitted to. If an older sibling's or parent's GPA exceeds the cutoff, they are admitted to their first choice major, whereas if their GPA is below the cutoff, they are admitted to their second choice major. As the table shows, if a father is admitted to the male-dominated major of Engineering, his female share falls by roughly 9% (whether or not he has a son or a daughter). In sharp contrast, if a mother is admitted to Engineering, her female share falls by a much larger amount – roughly 30%. In other words, due to different next best choices for fathers and mothers, a mother's gendered environment changes much more than a father's when she is admitted to Engineering. Turning to the female-dominated majors, we find the opposite pattern. When a father is barely admitted to Social Science or Humanities, his female share *rises* by 18%, whereas when a mother is barely admitted, her female share rises by a smaller 13%. The patterns are similar for older brothers and fathers as well as for older sisters and mothers.

Another way to characterize an older sibling's or parent's peer group is by GPA rank. GPA rank is defined as the older sibling's (or parent's) GPA position relative to their student peers in the same year and region for the major they are admitted to. Relative rank falls by roughly 30 rank points if an individual is barely admitted to their first choice versus their second choice. This is perhaps not surprising, as applicants who barely exceed a major's cutoff are the last students admitted to a major. What is more surprising is that the drop in relative rank is roughly the same regardless of whether a major is male-dominated, female-dominated, or gender-neutral. Likewise, the relative rank depends little on whether an older sibling or parent is male or female. So while the gendered environment in terms of fraction female varies widely across the types of majors, relative rank does not.

We now turn to estimates by the gender makeup of majors. Results for siblings by gender dyads are found in Table 8. Beginning with brother pairs, brothers copy brothers for the heavily male-dominated major of Engineering (i.e., they choose Engineering as their first choice if the older brother enrolled in it) and also into the gender-neutral majors of Natural Science and Business. There is a 8.6 percentage point effect for Engineering (28% increase) and a 6.0 percentage point effect for gender-neutral majors (25% increase). But there is no statistical evidence that younger brothers copy older brothers into the female-dominated majors of Social Science or Humanities. Sister pairs, in contrast, do not exhibit such gender-conforming patterns. Sisters copy sisters into gender-neutral majors, but not the female-dominated majors. There is a large, but statistically insignificant, effect for Engineering.

The patterns are equally interesting for opposite-gender siblings. Younger brothers are less willing to choose a female-dominated major if their older sister enrolled in it. There is a 4.2 percentage point drop in the likelihood a younger brother chooses Social Science or Humanities if this would copy his sister. This translates into a 56% drop relative the mean; the percent effect is large because a relatively small fraction of brothers choose a female-dominated major to begin with. In contrast, older brothers have no significant effect on their younger sisters, regardless of the share of females in a major. These patterns are consistent with younger brothers being influenced by whether a choice is gender conforming, while younger sisters are more immune to such forces. While many of the differences across major groupings within the dyads are large, we can only reject they are statistically different for 2 out of 4 gender dyads (using a joint F-test and the 10% significance level).

In Table 9 we perform a similar exercise for parents and their children. For sons, we find strong spillovers from both fathers and mothers, but only if the major is not female-dominated. Sons are 5.9 and 6.6 percentage points, respectively, more likely to choose the major their father completed if it was male-dominated or gender-neutral. These effects translate to 22% and 26% effects relative to their respective means. Likewise, if the mother's completed major was male-dominated or gender-neutral, the estimates for sons yield an increase of 18.0 and 4.1 percentage points, respectively. For the male-dominated major of Engineering, the effect is particularly large: it translates to a 69% increase. In contrast, for female-dominated majors there is no statistical evidence that sons copy parents.

Turning to daughters, there is not much evidence for spillovers from either fathers or mothers. The lone exception is that mothers may serve as a role model for Engineering, as predicted by the literature. Daughters are 11.0 percentage points more likely to choose Engineering. Since so few daughters choose Engineering to begin with (6.9%), this is a substantial 159% increase relative to the mean. This quasi-experimental evidence lines up with the correlational evidence found by Jacobs et al. (2017), which documents that mothers with engineering careers strongly influence the chances their children, and especially their daughters, plan to pursue Engineering. While many of the differences across major groupings within dyad are large, none are statistically significant.

Multiple Inference Adjustments. The findings of Tables 8 and 9 are further robust to multiple inference adjustments using the False Discovery Rate (FDR) control (see Anderson 2008). Following the logic of Anderson (2008), it is important to define domains. In our setting each gender-specific dyad is the domain, which implies adjusting tests for three different hypotheses (one for each major grouping) within each dyad. For the reduced form results in Table 8, all four estimates remain statistically significant at the 5 percent level, while for the reduced form results in Table 9, three out of five estimates remain significant at the 10 percent level (see Appendix Table A13). Of the two estimates that are no longer statistically significant, the estimate for mother-son is on the margin of significance with a q-value of .101.

We also conducted a similar multiple inference adjustment for Table 5 for birth spacing. In this case, there are two different hypotheses (close versus far birth spacing) within each dyad, and all three statistically significant estimates remain significant. A similar adjustment for Appendix Table A6 for birth order (two different hypotheses: first versus later born within each dyad) finds that the one statistically significant estimate remains significant.

11 Impact on Gender Imbalances across Majors

How much do these family spillovers in the previous section contribute to either reinforcing or ameliorating gender imbalances across majors? To answer this question for siblings and parents, we compare the observed gender distributions across majors (which includes any sibling or parental spillover effects) to the implied counterfactual distributions (which removes any sibling or parental spillover effects) using the estimates appearing in Tables 8 and 9.

To understand how we construct the counterfactual, take the example of older brothers. If we remove the influence of these older brothers, the IV estimate for Engineering in Table 8 implies an 8.6 percentage point (pp) drop in younger brothers who no longer choose Engineering. These younger brothers who no longer choose Engineering will instead end up in other majors. We assume these younger brothers are redistributed to the other majors using the means in the final columns of Tables 8 and 9. Of the 8.6 pp who no longer choose Engineering, 2.3 pp are reallocated to the gender-neutral majors, 1.4 pp to the female-dominated majors, and the remaining 4.9 pp to non-academic majors. Similar redistributions happen when considering the IV estimates for older brothers for the gender-neutral majors (Natural Science or Business) and the female-dominated majors (Social Science or Humanities).[32] Combining these effects, removing the influence of older brothers reduces the fraction of younger brothers who choose Engineering from 30% to 24%. Doing a similar exercise where we remove the influence of older brothers on younger sisters increases the fraction of younger sisters who choose Engineering from 11% to 13%. Combining all of this information, the fraction of men in Engineering drops

[32] We assume no spillover effects for non-academic majors in these calculations, as we do not have estimates for non-academic majors.

from the observed 73% to the counterfactual 69% reported in Table 10. The same logic applies when constructing counterfactual male shares when removing spillovers from older sisters, fathers, or mothers. If older brothers exerted no influence on their younger siblings of either gender, then the male share in the female-dominated fields of Social Science and Humanities would have been 35% instead of 33%. In other words, these patterns imply older brothers reinforce gender norms.

If older sisters exerted no influence, then the male share in Engineering would have been 81% (instead of 73%) and the male share in the female-dominated fields would have been 36% (instead of 33%). Thus, older sisters break down gender norms for Engineering, but not for female-dominated majors.

Turning to parental impacts, we calculate that if fathers had no effect on their children, then the male share would have been 81% (instead of 79%) in Engineering and the male share 40% (instead of 39%) in the female-dominated fields. Likewise, in the absence of spillover effects from m others, the male share would have been 83% (instead of 79%) in Engineering and the male share 41% (instead of 39%) in the female-dominated fields. In other words, both fathers and mothers serve to break down male dominance in Engineering, but slightly increase female dominance in Social Science and Humanities. Thinking of a father's influence, the male share in Engineering falls despite the estimated tendency for sons to copy their fathers into Engineering; this is mostly because sons also copy their fathers into gender-neutral and female-dominated majors which in turn draws them away from Engineering. For mothers, the main factor is that their daughters follow them into Engineering, with only small effects on a daughter's other major choices. The combined effect if neither fathers nor mothers had an influence would have been almost a halving of the fraction of daughters in Engineering (a counterfactual 12% female share versus the observed 21%).

These back of the envelope calculations have important policy implications. For example, if the goal is to increase female representation in Engineering, policymakers can harness family spillovers by targeting girls specifically to enroll in Engineering through outreach campaigns, easier admissions, or exposure to female-role models. This will not only affect an older sister's probability of choosing Engineering, but will also reduce male dominance in the Engineering major through her effect on younger siblings. In contrast, a gender-blind expansion of the higher paying major of Engineering will have a more muted effect, as the influence of older brothers will counteract that of older sisters in terms of achieving a more equal gender balance.

12 Conclusion

This paper provides the first causal evidence on parental spillovers in major choice. Likewise, this paper is the first to document causally that siblings influence the margin of high school major choice, complementing prior findings for college. There is not a simple same-sex versus opposite-sex division in spillover effects, nor is there a common pattern of sibling and intergenerational influences.

Our findings confirm many, but not all, of the predictions hypothesized in the literature regarding the importance of role models, differentiation due to sibling rivalry or rebellion, and information transmission/support. One of the main predictions is that these forces should be strongest for same-gender dyads. Consistent with the literature, we find sizable same-sex spillovers for both brothers and sisters due to role models or information transmission/support. However, we do not find support for this prediction in intergenerational spillovers. While sons are influenced by both fathers and mothers, daughters are not influenced by either parent. The lone exception is when the mother completes Engineering, in which case her daughter copies her. This is consistent with a literature that predicts that same-sex role models are especially important for non-gender conforming choices. We also find evidence of sibling rivalry, with an unexpected effect for younger brothers recoiling from their older sister's choices.

Gendered pathways drive many of our findings, both in terms of the sex makeup of sibling and parent-child pairs, and in terms of whether a major is gender conforming. Interestingly, we find that an individual's major choice has at least as large an influence on their sibling as their child, despite the literature's larger focus on parental influences (Björklund and Jäntti 2012, McHale et al. 2012). The patterns we observe also highlight that families play a key role in the development of gender imbalances across majors, consistent with the importance of families for gender norms as emphasized by Bertrand (2011) and Brenøe (2018).

References

Adams, B. N. (1972): "Birth order: A critical review," *Sociometry*, 411–439.

Adermon, A. (2013): "Essays on the Transmission of Human Capital and the Impact of Technological Change," Ph.D. thesis, Uppsala universitet.

Aguirre, J. and J. Matta (2021): "Walking in your footsteps: Sibling spillovers in higher education choices," *Economics of Education Review*, 80.

Altmejd, A., A. Barrios-Fernández, M. Drlje, J. Goodman, M. Hurwitz, D. Kovac, C. Mulhern, C. Neilson, and J. Smith (2021): "O brother, where start thou?

Sibling spillovers on college and major choice in four countries," *Quarterly Journal of Economics*, 136, 1831–1886.

Altonji, J. G., S. Cattan, and I. Ware (2017): "Identifying sibling influence on teenage substance use," *Journal of Human Resources*, 52, 1–47.

Anaya, L., F. Stafford, and G. Zamarro (2022): "Gender gaps in math performance, perceived mathematical ability and college STEM education: The role of parental occupation," *Education Economics*, 30, 113–128.

Anderson, M. L. (2008): "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training projects," *Journal of the American Statistical Association*, 103, 1481–1495.

Bandura, A. and R. H. Walters (1977): *Social Learning Theory*, vol. 1, Englewood Cliffs Prentice Hall.

Bertrand, M. (2011): "New perspectives on gender," in *Handbook of Labor Economics*, Elsevier, vol. 4, 1543–1590.

Bettinger, E. P. and B. T. Long (2005): "Do faculty serve as role models? The impact of instructor gender on female students," *American Economic Review*, 95, 152–157.

Beuermann, D. W., C. K. Jackson, L. Navarro-Sola, and F. Pardo (2022): "What is a good school, and can parents tell? Evidence on the multidimensionality of school output," *The Review of Economic Studies*, rdac025.

Bingley, P., P. Lundborg, and S. V. Lyk-Jensen (2021): "Brothers in arms: Spillovers from a draft lottery," *Journal of Human Resources*, 56, 225–268.

Björklund, A. and M. Jäntti (2009): "Intergenerational income mobility and the role of family background," in *Oxford Handbook of Economic Inequality*, Oxford University Press Oxford, 491–521.

——— (2012): "How important is family background for labor-economic outcomes?" *Labour Economics*, 19, 465–474.

Björklund, A. and K. G. Salvanes (2011): "Education and family background: Mechanisms and policies," in *Handbook of the Economics of Education*, Elsevier, vol. 3, 201–247.

Black, S. E., S. Breining, D. N. Figlio, J. Guryan, K. Karbownik, H. S. Nielsen, J. Roth, and M. Simonsen (2021): "Sibling spillovers," *The Economic Journal*, 131, 101–128.

Black, S. E. and P. J. Devereux (2011): "Recent developments in intergenerational mobility," in *Handbook of Labor Economics*, North Holland, vol. 4, 1487–1541.

Breda, T., J. Grenet, M. Monnet, and C. Van Effenterre (2020): "Do female role models reduce the gender gap in science? Evidence from French high schools," IZA Discussion Paper No. 13163.

Brenøe, A. (2018): "Origins of gender norms: Sibling gender composition and women's choice of occupation and partner," IZA Discussion Paper No. 11692.

Brenøe, A. A. (2022): "Brothers increase women's gender conformity," *Journal of Population Economics*, 35, 1859–1896.

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

Buckles, K. (2019): "Fixing the leaky pipeline: Strategies for making economics work for women at every stage," *Journal of Economic Perspectives*, 33, 43–60.

Buist, K. L., M. Deković, W. Meeus, and M. A. Van Aken (2002): "Developmental patterns in adolescent attachment to mother, father and sibling," *Journal of Youth and Adolescence*, 31, 167–176.

Bush, K. R. and G. W. Peterson (2013): "Parent–child relationships in diverse contexts," in *Handbook of Marriage and the Family*, Springer, 275–302.

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

Cattaneo, M. D., M. Jansson, and X. Ma (2018): "Manipulation testing based on density discontinuity," *The Stata Journal*, 18, 234–261.

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

Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014): "Family welfare cultures," *Quarterly Journal of Economics*, 129, 1711–1752.

Dahl, G. B., D.-O. Rooth, and A. Stenberg (forthcoming): "High School Majors and Future Earnings," *American Economic Journal: Applied Economics*.

Daysal, N. M., M. Simonsen, M. Trandafir, and S. Breining (2022): "Spillover effects of early-life medical interventions," *The Review of Economics and Statistics*, 104, 1–16.

Dryler, H. (1998): "Parental role models, gender and educational choice," *British Journal of Sociology*, 375–398.

Duncan, O. D. and B. Duncan (1955): "A methodological analysis of segregation indexes," *American sociological review*, 20, 210–217.

Dunn, J. and C. Kendrick (1982): *Siblings*, vol. 1, Cambridge MA; Harvard University Press.

Dustan, A. (2018): "Family networks and school choice," *Journal of Development Economics*, 134, 372–391.

Fadlon, I. and T. H. Nielsen (2019): "Family health behaviors," *American Economic Review*, 109, 3162–91.

Floyd, K. (1995): "Gender and closeness among friends and siblings," *The Journal of Psychology*, 129, 193–202.

Goodman, J., M. Hurwitz, J. Smith, and J. Fox (2015): "The relationship between siblings' college choices: Evidence from one million SAT-taking families," *Economics of Education Review*, 48, 75–85.

Halldén, K. and A. Stenberg (forthcoming): *Könssegregering i utbildningsval och yrke i Sverige under ett halvt sekel*, Government Offices of Sweden, Stockholm.

Hanushek, E. A., B. Jacobs, G. Schwerdt, R. Van der Velden, S. Vermeulen, and S. Wiederhold (2021): "The intergenerational transmission of cognitive skills: An investigation of the causal impact of families on student outcomes," NBER Working Paper 29450.

Holmlund, H., M. Lindahl, and E. Plug (2011): "The causal effect of parents' schooling on children's schooling: A comparison of estimation methods," *Journal of Economic Literature*, 49, 615–651.

Humlum, M. K., A. B. Nandrup, and N. Smith (2019): "Closing or reproducing the gender gap? Parental transmission, social norms and education choice," *Journal of Population Economics*, 32, 455–500.

Jacobs, J. A., S. Ahmad, and L. J. Sax (2017): "Planning a career in engineering:

Parental effects on sons and daughters," *Social Sciences*, 6, 2.

Joensen, J. S. and H. S. Nielsen (2018): "Spillovers in education choice," *Journal of Public Economics*, 157, 158–183.

Kirabo Jackson, C. (2010): "Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago," *The Economic Journal*, 120, 1399–1429.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016): "Field of study, earnings, and self-selection," *Quarterly Journal of Economics*, 131, 1057–1111.

Koch, H. L. (1960): "The relation of certain formal attributes of siblings to attitudes held toward each other and toward their parents," *Monographs of the Society for Research in Child Development*, 1–124.

Lindquist, M. J., J. Sol, and M. Van Praag (2015): "Why do entrepreneurial parents have entrepreneurial children?" *Journal of Labor Economics*, 33, 269–296.

Lundborg, P., A. Nilsson, and D.-O. Rooth (2014): "Parental education and offspring outcomes: evidence from the Swedish compulsory School Reform," *American Economic Journal: Applied Economics*, 6, 253–78.

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

McCrary, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142, 698–714.

McHale, S. M., K. A. Updegraff, and S. D. Whiteman (2012): "Sibling relationships and influences in childhood and adolescence," *Journal of Marriage and Family*, 74, 913–930.

——— (2013): "Sibling relationships," in *Handbook of Marriage and the Family*, Springer, 329–351.

Nicoletti, C. and B. Rabe (2019): "Sibling spillover effects in school achievement," *Journal of Applied Econometrics*, 34, 482–501.

Nicoletti, C., K. G. Salvanes, and E. Tominey (2018): "The family peer effect on mothers' labor supply," *American Economic Journal: Applied Economics*, 10, 206–34.

Oguzoglu, U. and S. Ozbeklik (2016): "Like father, like daughter (unless there is a son): Sibling sex composition and women's stem major choice in college," *IZA Discussion Paper No. 10052*.

Persson, P., X. Qiu, and M. Rossin-Slater (2021): "Family spillover effects of marginal diagnoses: The case of ADHD," NBER Working Paper 28334.

Philipp, J. (2022): "Gendered university major choice: The role of intergenerational transmission," *Journal of Population Economics*, 1–49.

Porter, C. and D. Serra (2020): "Gender differences in the choice of major: The importance of female role models," *American Economic Journal: Applied Economics*, 12, 226–254.

Pulakos, J. (1989): "Young adult relationships: Siblings and friends," *The Journal of Psychology*, 123, 237–244.

Qureshi, J. A. (2018a): "Siblings, teachers, and spillovers on academic achievement," *Journal of Human Resources*, 53, 272–297.

——— (2018b): "Additional returns to investing in girls' education: Impact on younger sibling human capital," *The Economic Journal*, 128, 3285–3319.

Raley, S. and S. Bianchi (2006): "Sons, daughters, and family processes: Does gender of children matter?" *Annual Review of Sociology*, 32, 401–421.

Random House (2022): *"Role model" definition from dictionary.com [retrieved 26 October 2022].*

Russell, A. and J. Saebel (1997): "Mother–son, mother–daughter, father–son, and father–daughter: Are they distinct relationships?" *Developmental review*, 17, 111–147.

Statistics Sweden (2021a): *Sökande/antagna till gymnasiet 1977-1991 [database].*

——— (2021b): *Ak9-elevregistret 1988-1991 [database].*

——— (2021c): *Income and taxation registers 1978-1989 [database].*

——— (2021d): *Longitudinal integrated database for health insurance and labour market studies (LISA) 1990-2014 [database].*

——— (2021e): *Multiple generation registers (Flergenerationsregistret) [database].*

——— (2021f): *Population and housing census 1970, 1980 and 1985 [database].*

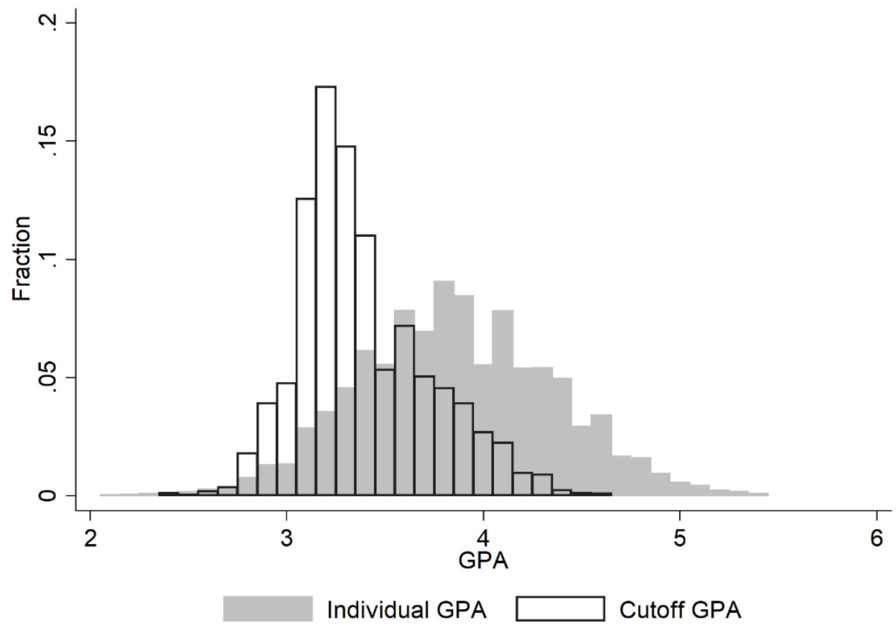
——— (2021g): *Sibling register [database].*

Sulloway, F. J. (1996): *Born to rebel: Birth order, family dynamics, and creative lives*, Pantheon Books.

Svensson, L.-G. (1999): "Strategy-proof allocation of indivisible goods," *Social Choice and Welfare*, 16, 557–567.

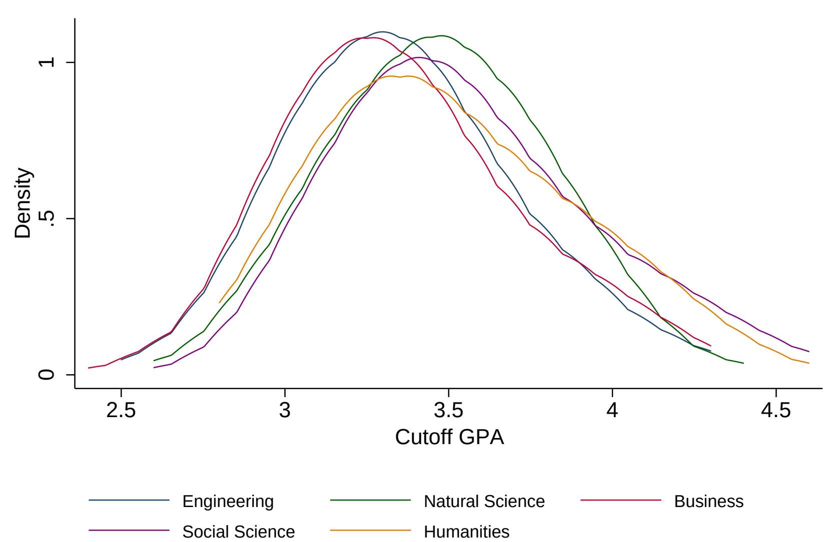
Figure 1. Cutoff GPA distributions.

Panel A: Cutoff distribution versus individual GPA distribution



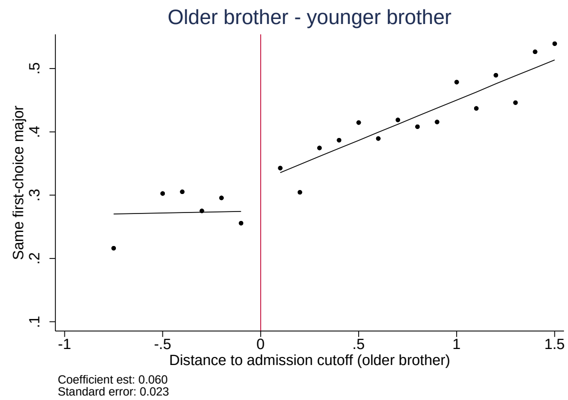
Notes: The white bars plot the distribution of cutoff GPAs for competitive programs, which vary by major, year, and school region. There are 3,487 competitive programs in our estimation sample. The grey bars plot the distribution of GPA for individuals in oversubscribed programs (N=263,878).

Panel B: Distribution of GPA cutoffs by high school major



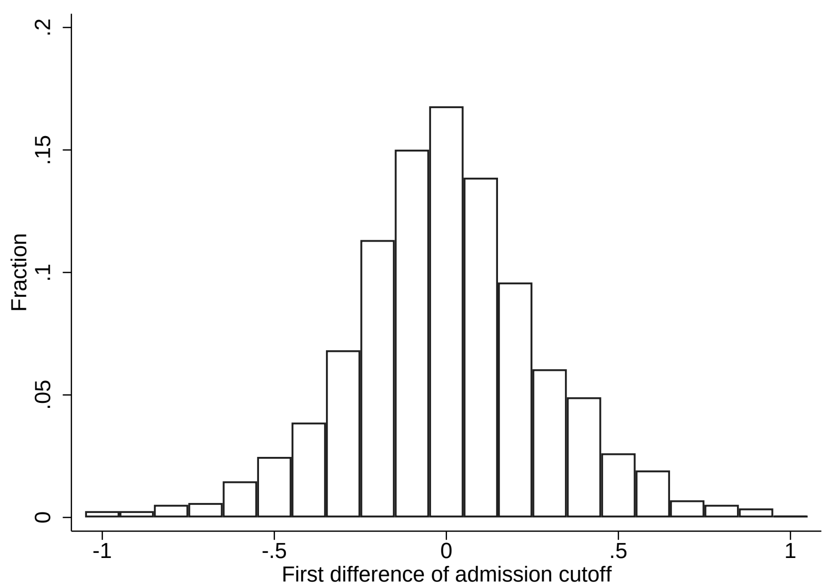
Notes: Kernel density estimates of GPA cutoffs by major, using an Epanechnikov kernel and a bandwidth of 0.2.

Figure 2. Probability a younger brother chooses Engineering if their older brother chose Engineering.



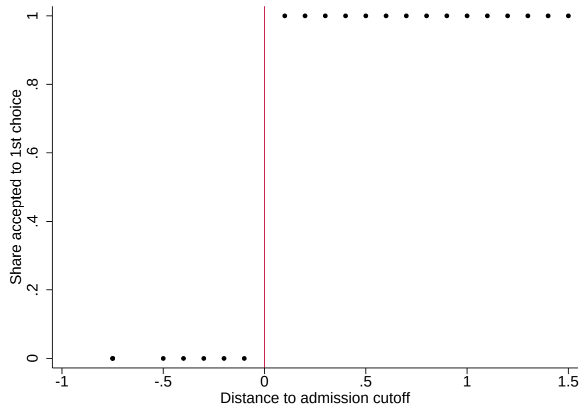
Notes: Sample limited to brother-brother pairs where the older brother chose Engineering as their first choice. Each observation is the average share of younger brothers who choose Engineering as their first choice as a function of their older brother's GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff for older brothers (normalized to 0). The estimated slopes are based on the common slope model, linear functions of GPA, a window of -1.0 to 1.5, and triangular weights. The number of observations is 11,706.

Figure 3. Distribution of current minus lagged admission cutoff GPA, 1977-1991.



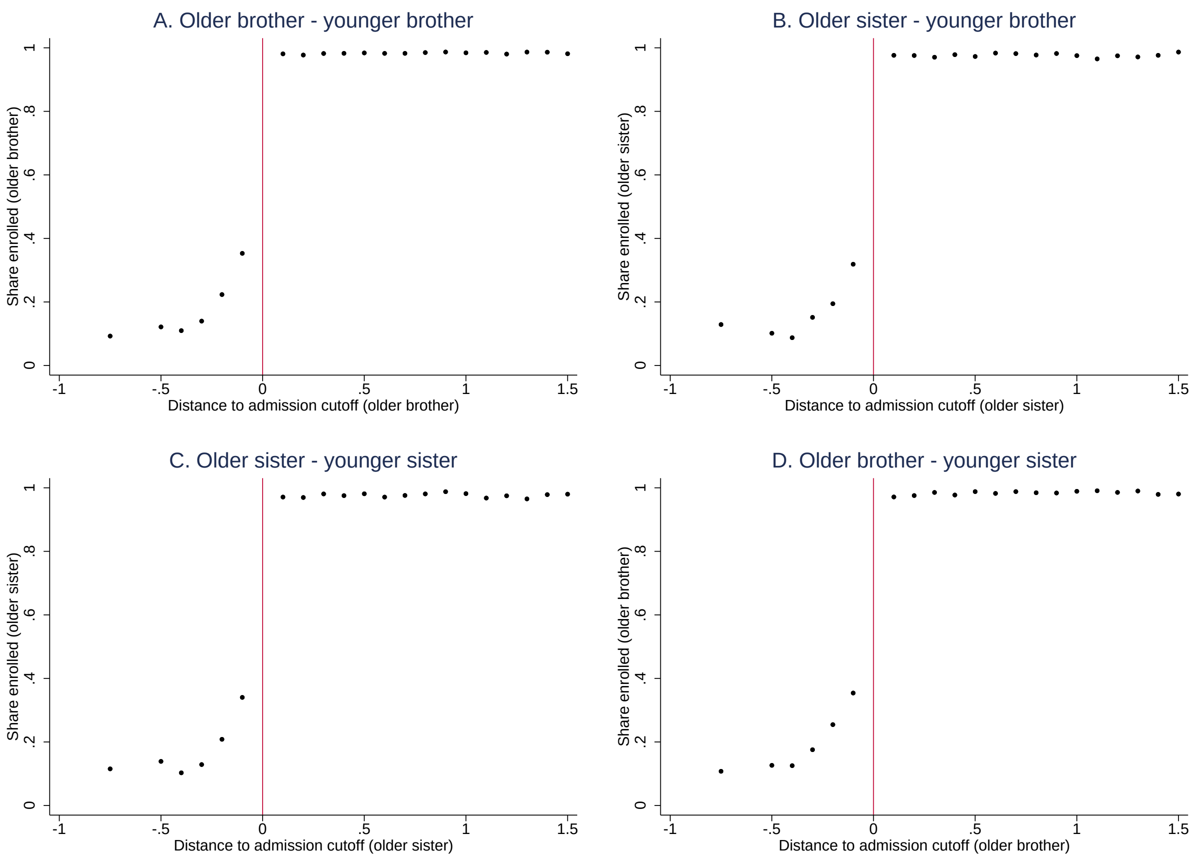
Note: Sample limited to academic majors which are oversubscribed two years in a row in a school region.

Figure 4. Share of individuals admitted to their first-choice major, 1977-1991.



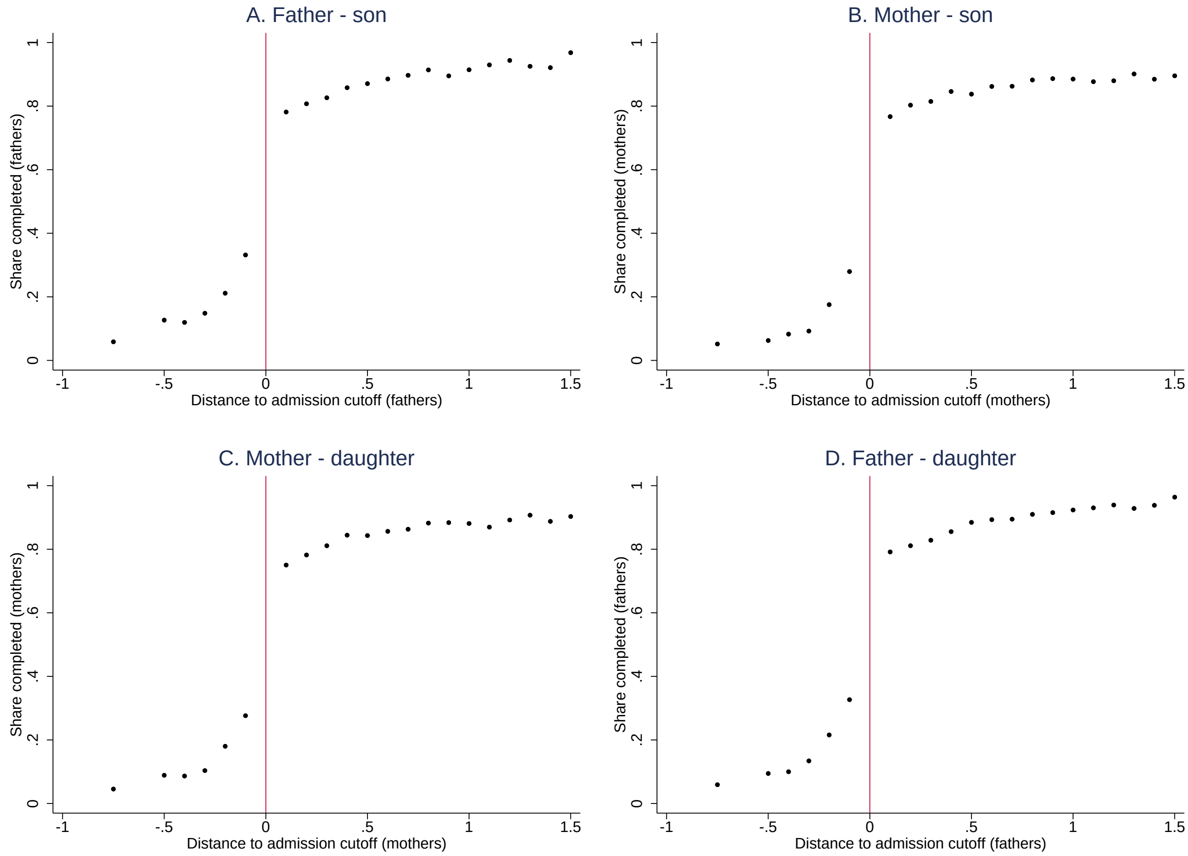
Notes: Each observation is the average share of individuals accepted to their first-best major choice as a function of their GPA in a 0.1 GPA bin, except for the leftmost observation which is a 0.5 bin due to sparsity. The vertical line denotes the admissions GPA cutoff (normalized to 0).

Figure 5. First stage: Share of older siblings enrolling in their first-choice major, by gender mix.



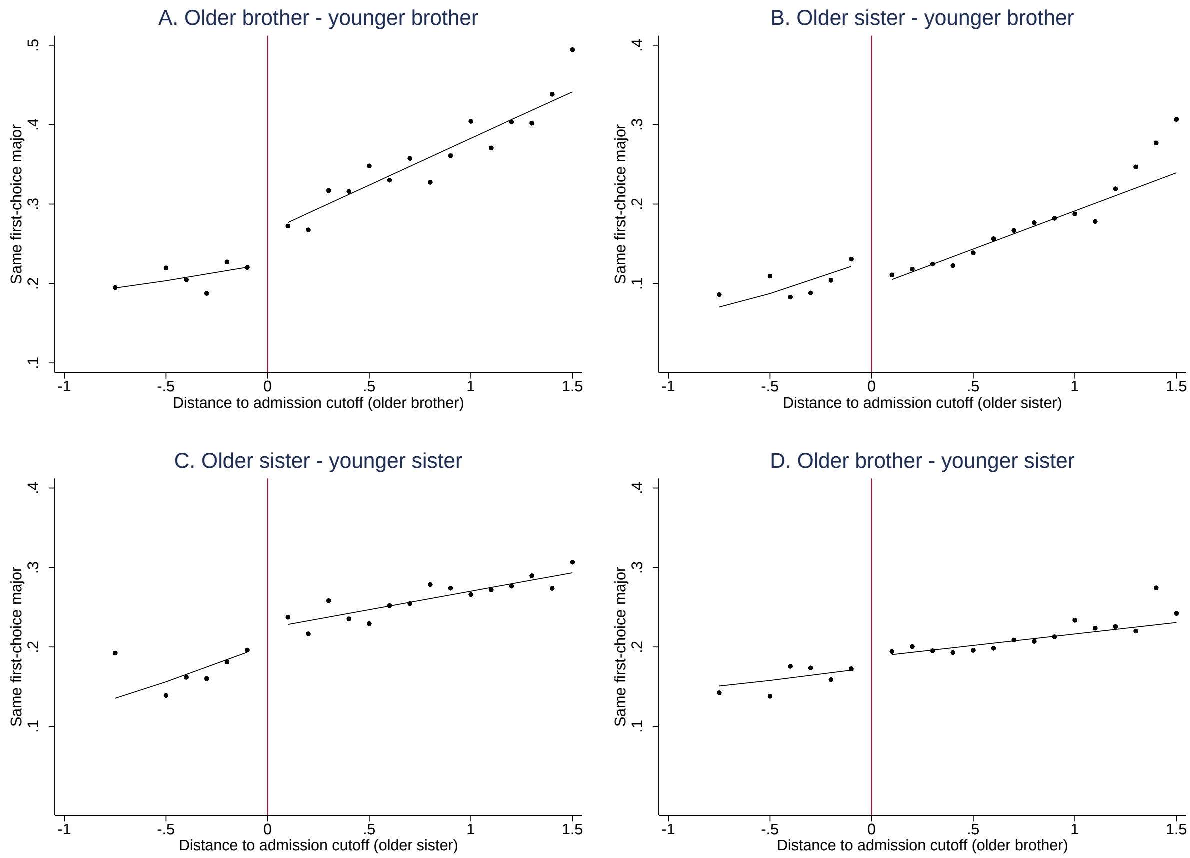
Notes: Each observation is the average share of older siblings who enroll in their first-best major choice as a function of their GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff (normalized to 0). The number of observations in panel A is 22,841 (older brother - younger brother sample), in panel B 22,926 (older sister - younger brother sample), in panel C 21,020 (older sister - younger sister sample), and in panel D 21,377 (older brother - younger sister sample).

Figure 6. First stage: Share of parents completing their first-choice major, by gender mix.



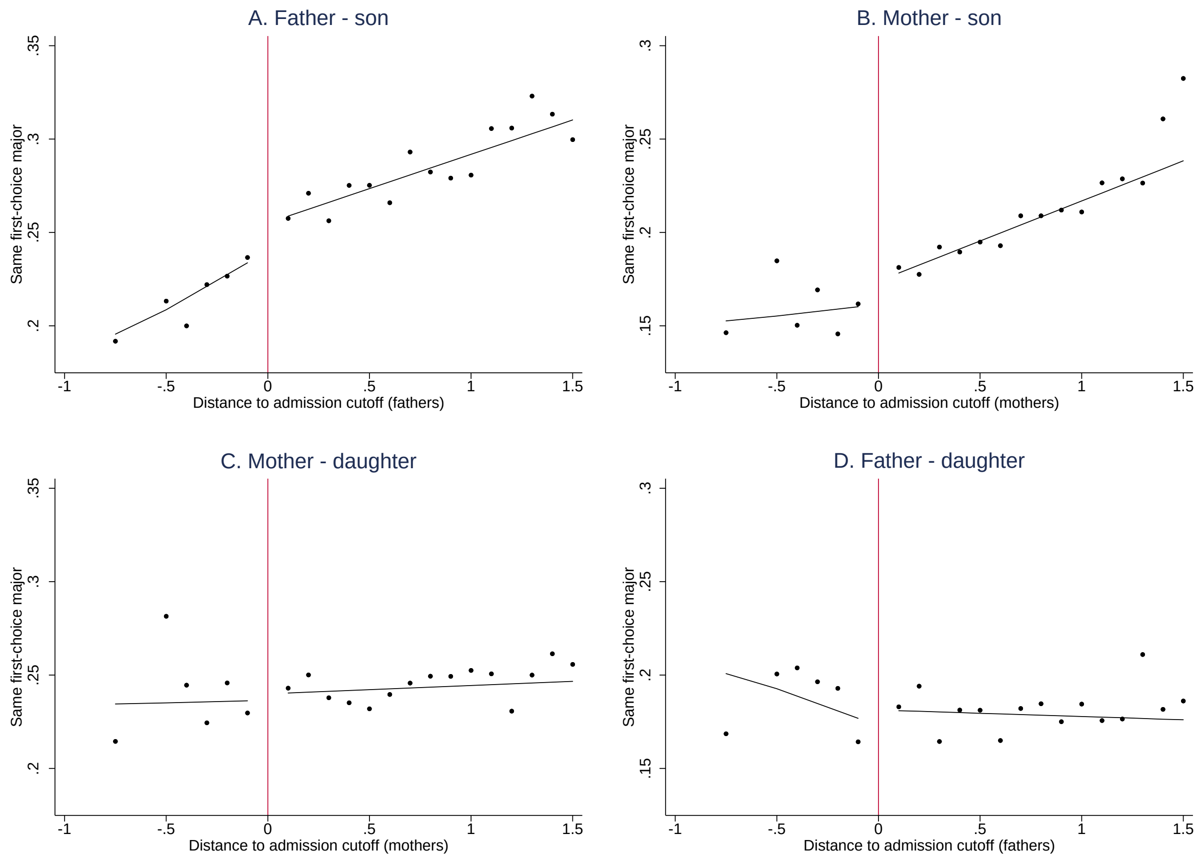
Notes: Each observation is the average share of parents who enroll in their first-best major choice as a function of their GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff (normalized to 0). The number of observations in panel A is 37,390 (father-son sample), in panel B 49,057 (mother-son sample), in panel C 46,791 (mother-daughter sample), and in panel D 35,695 (father-daughter sample).

Figure 7. Reduced form: Probability a younger sibling chooses the same first-choice major as their older sibling, by gender mix.



Notes: Each observation is the average share of younger siblings whose first choice on their preference list matches the first-best major choice of their older sibling, as a function of their older sibling's GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff for older siblings (normalized to 0). The estimated slopes are based on the common slope model, linear functions of GPA, a window of -1.0 to 1.5, and triangular weights. The number of observations in panel A is 22,841 (older brother - younger brother sample), in panel B 22,926 (older sister - younger brother sample), in panel C 21,020 (older sister - younger sister sample), and in panel D 21,377 (older brother - younger sister sample).

Figure 8. Reduced form: Probability a child chooses the same first-choice major as their parent, by gender mix.



Notes: Each observation is the average share of children whose first choice on their preference list matches the first-best major choice of their parent, as a function of their parent's GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical line denotes the admissions GPA cutoff for parents (normalized to 0). The slopes are based on the common slope model, linear functions of GPA, a window of -1.0 to 1.5, and triangular weights. The number of observations in panel A is 37,390 (father-son sample), in panel B 49,057 (mother-son sample), in panel C 46,791 (mother-daughter sample), and in panel D 35,695 (father-daughter sample).

Table 1. Placebo test: Does a younger sibling's first-choice major affect their older sibling's ex-ante choice?

	Reduced Form	IV-enrolled	Mean
(1) Impact on older sibling			
Younger sibling – older sibling	.000	.000	[.224]
	(.007)	(.012)	
(2) Impact on older brother			
Younger brother – older brother	.002	.003	[.324]
	(.010)	(.016)	
Younger sister – older brother	-.012	-.016	[.171]
	(.009)	(.014)	
(3) Impact on older sister			
Younger sister – older sister	.007	.009	[.232]
	(.011)	(.016)	
Younger brother – older sister	.004	.006	[.171]
	(.009)	(.015)	
N	91,688	91,688	

Notes: See notes to Table 5. The placebo regression estimates whether a younger sibling affects their older sibling, while Table 5 estimates whether an older sibling affects their younger sibling. Since the older sibling makes their major choice before their younger sibling, there should not be an effect in the placebo regression.
** p<0.10, ** p<0.05, *** p<0.01*

Table 2. First stage estimates for the probability of enrolling or completing a first-choice major.

Panel A: Siblings		**Panel B: Intergenerational**		
	Enrolled		Enrolled	Completed
(1) All		**(1) All**		
Older siblings	.618***	Parents	.604***	.449***
	(.008)		(.007)	(.007)
(2) By gender mix		**(2) By gender mix**		
Older brother – younger brother	.619***	Father – son	.602***	.432***
	(.011)		(.009)	(.009)
Older sister – younger brother	.636***	Mother – son	.609***	.465***
	(.012)		(.009)	(.009)
Older sister – younger sister	.624***	Mother – daughter	.612***	.458***
	(.013)		(.009)	(.009)
Older brother – younger sister	.602***	Father – daughter	.592***	.442***
	(.011)		(.009)	(.009)
N	88,174		168,933	168,933

Notes: The outcome variable is a dummy for whether the older sibling or parent enrolled (or completed) their first-best major, as a function of whether their GPA exceeded the admissions cutoff. Regressions use the baseline model, linear functions of GPA, a window of -1.0 to 1.5, triangular weights, fixed effects for year and school region, dummies for preferred major interacted with sibling/parent-child gender mix, dummies for next-best alternative major, and the demographic controls listed in Appendix Table A2. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01*

Table 3. Reduced form and IV estimates of the probability of choosing the same first-choice major as an older sibling or parent.

	Reduced form	IV-enrolled	IV-completed	Mean
Panel A: Siblings				
Older sibling – younger sibling	.015**	.025**	--	[.196]
	(.007)	(.011)		
N	88,174	88,174	--	
Panel B: Intergenerational				
Parent – child	.014**	.023**	.031**	[.214]
	(.006)	(.009)	(.012)	
N	168,933	168,933	168,933	

Notes: The outcome variable is whether a younger sibling's or child's first choice on their preference list matches the first-best major choice of their older sibling or parent. IV-enrolled uses as a first stage whether the older sibling or parent enrolled in their first-best major, as a function of whether their GPA exceeded the admissions cutoff. IV-completed uses whether the parent completed their first-best major. Regressions use the baseline model, linear functions of GPA, a window of -1.0 to 1.5, triangular weights, fixed effects for year and school region, dummies for preferred major interacted with sibling/parent-child gender mix, dummies for next-best alternative major, and the demographic controls listed in Appendix Table A2. Dependent mean is calculated for the sample where an older sibling's or parent's GPA is within plus or minus 0.2 GPA points of the admission cutoff. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01*

Table 4. Correlational estimates.

	Correlational estimate	IV-enrolled	Difference
Panel A: Siblings			
Older sibling – younger sibling	.102***	.025**	.078***
	(.002)	(.011)	(.011)
N	82,714	88,174	
Panel B: Intergenerational			
Parent – child	.063***	.023**	.040***
	(.001)	(.009)	(.009)
N	157,760	168,933	

Notes: See text for details on how the correlational estimates are constructed. Bootstrap standard errors based on 1,000 replications. The IV-enrolled estimates are taken from Table 3.
** p<0.10, ** p<0.05, *** p<0.01*

Table 5. Sibling estimates by gender mix and birth spacing.

	Any age gap		Age gap ≤ 3 years (concurrent school enrollment)		Age gap > 3 years		
	Reduced form	IV-enrolled	Reduced form	IV-enrolled	Reduced form	IV-enrolled	Mean
(1) Impact on younger brother							
Older brother – younger brother	.043***	.063***	.019	.027	.074***	.107***	[.253]
	(.010)	(.015)	(.013)	(.021)	(.015)	(.022)	
Older sister – younger brother	-.026***	-.030**	-.037***	-.048**	-.010	-.006	[.117]
	(.009)	(.014)	(.013)	(.019)	(.014)	(.020)	
(2) Impact on younger sister							
Older sister – younger sister	.025**	.039**	.036**	.049**	.013	.025	[.215]
	(.011)	(.016)	(.015)	(.021)	(.017)	(.025)	
Older brother – younger sister	.009	.017	-.002	-.001	.023	.039*	[.187]
	(.009)	(.014)	(.012)	(.019)	(.014)	(.021)	
N	88,174	88,174	52,274	52,274	35,900	35,900	

Notes: The outcome variable is whether a younger sibling's first choice on their preference list matches the first-best major choice of their older sibling. IV-enrolled uses as a first stage whether the older sibling enrolled in their first-best major, as a function of whether their GPA exceeded the admissions cutoff. Regressions use the baseline model, linear functions of GPA, a window of -1.0 to 1.5, triangular weights, fixed effects for year and school region, dummies for preferred major interacted with sibling gender mix, dummies for next-best alternative major, and the demographic controls listed in Appendix Table A2. Dependent mean is calculated for the sample where an older sibling's GPA is within plus or minus 0.2 GPA points of the admission cutoff. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01*

Table 6. Intergenerational estimates by gender mix.

	Reduced form	IV-enrolled	IV-completed	Mean
(1) Impact on sons				
Father – son	.029***	.043***	.056***	[.254]
	(.008)	(.012)	(.016)	
Mother – son	.015**	.024**	.031**	[.172]
	(.007)	(.011)	(.014)	
(2) Impact on daughters				
Mother – daughter	.004	.010	.015	[.244]
	(.008)	(.012)	(.015)	
Father – daughter	.007	.014	.021	[.185]
	(.008)	(.012)	(.015)	
N	168,933	168,933	168,933	

Notes: The outcome variable is whether a child's first choice on their preference list matches the first-best major choice of their parent. IV-enrolled uses as a first stage whether the parent enrolled in their first-best major, as a function of whether their GPA exceeded the admissions cutoff. IV-completed instead uses whether the parent completed their first-best major. Regressions use the baseline model, linear functions of GPA, a window of -1.0 to 1.5, triangular weights, fixed effects for year and school region, dummies for preferred major interacted with parent-child gender mix, dummies for next-best alternative major, and the demographic controls listed in Appendix Table A2. Dependent mean is calculated for the sample where a parent's GPA is within plus or minus 0.2 GPA points of the admission cutoff. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01*

Table 7. Sibling and intergenerational reduced form estimates using alternative specifications.

	Baseline	Quadratic	Smaller bandwidth	Omitting low & high cutoffs	Excluding 1982-84	No demo. controls	2-sibling families	Common slopes	First-second choice slopes
(1) Impact on younger brother									
Older brother – younger brother	.043***	.045***	.041***	.045***	.046***	.040***	.054***	.046***	.043***
	(.010)	(.013)	(.012)	(.010)	(.012)	(.010)	(.015)	(.010)	(.010)
Older sister – younger brother	-.026***	-.024*	-.020*	-.024**	-.021*	-.028***	-.025***	-.030***	-.025**
	(.009)	(.012)	(.011)	(.010)	(.012)	(.009)	(.014)	(.009)	(.010)
(2) Impact on younger sister									
Older sister – younger sister	.025**	.027*	.028**	.025**	.037***	.024**	.033*	.020*	.026**
	(.011)	(.014)	(.013)	(.012)	(.014)	(.011)	(.017)	(.011)	(.012)
Older brother – younger sister	.009	.011	.009	.008	.015	.006	-.002	.012	.010
	(.009)	(.012)	(.011)	(.010)	(.011)	(.009)	(.014)	(.009)	(.010)
N	88,174	88,174	62,214	81,714	62,384	88,174	42,186	88,174	88,174
(3) Impact on sons									
Father – son	.029***	.036***	.030***	.029***	.041***	.028***		.030***	.028***
	(.008)	(.010)	(.009)	(.008)	(.009)	(.008)		(.008)	(.008)
Mother – son	.015**	.022**	.021**	.016**	.021**	.014*		.016**	.016**
	(.007)	(.010)	(.008)	(.008)	(.009)	(.007)		(.007)	(.008)
(4) Impact on daughters									
Mother – daughter	.004	.011	.009	.003	.006	.003		.005	.005
	(.008)	(.010)	(.009)	(.008)	(.009)	(.008)		(.008)	(.008)
Father – daughter	.007	.014	.011	.007	.012	.006		.008	.006
	(.008)	(.010)	(.009)	(.008)	(.009)	(.008)		(.008)	(.008)
N	168,933	168,933	122,260	163,380	122,466	168,933		168,933	168,933

Notes: See notes to Tables 5 and 6. Column 2 includes quadratic functions of the running variable on both sides of the cutoff and column 3 reduces the bandwidth in half. Column 4 excludes observations with low cutoffs (fewer than 50 observations below the cutoff) or high cutoffs (fewer than 50 observations above). Column 5 excludes the years 1982-84, when bonus GPA points were added for the first and second choices on an individual's ranking list, and column 6 excludes the demographic controls listed in Table A2. Column 7 only includes families with two siblings. Columns 8 and 9 use the common slope and first-second choice slope models. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01.*

Table 8. Sibling estimates as a function of the gender makeup of majors.

	Reduced form	IV-enrolled	Mean
(1) Impact on younger brother			
Older brother – younger brother:			
Male-dominated major (E)	.059***	.086***	[.306]
	(.016)	(.023)	
Gender-neutral major (N+B)	.042***	.060***	[.242]
	(.014)	(.020)	
Female-dominated major (S+H)	.010	.016	[.139]
	(.017)	(.023)	
Older sister – younger brother:			
Male-dominated major (E)	.020	.032	[.296]
	(.067)	(.094)	
Gender-neutral major (N+B)	-.021	-.024	[.146]
	(.014)	(.020)	
Female-dominated major (S+H)	-.035***	-.042***	[.075]
	(.011)	(.015)	
(2) Impact on younger sister			
Older sister – younger sister:			
Male-dominated major (E)	.057	.100	[.138]
	(.047)	(.078)	
Gender-neutral major (N+B)	.041**	.059**	[.242]
	(.017)	(.023)	
Female-dominated major (S+H)	.009	.016	[.195]
	(.015)	(.021)	
Older brother – younger sister:			
Male-dominated major (E)	-.010	-.011	[.090]
	(.011)	(.017)	
Gender-neutral major (N+B)	.024	.036	[.269]
	(.016)	(.022)	
Female-dominated major (S+H)	.024	.036	[.252]
	(.021)	(.029)	
N	88,174	88,174	

Notes: See notes to Table 5. The regressions differ by allowing for heterogeneous effects based on the gender makeup of majors. Majors are denoted in parentheses by their first letters: E, N, B, S, H stand for Engineering, Natural Science, Business, Social Science, and Humanities, respectively. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01*

Table 9. Intergenerational estimates as a function of the gender makeup of majors.

	Reduced form	IV-enrolled	IV-completed	Mean
(1) Impact on sons				
Father – son:				
Male-dominated major (E)	.023*	.039**	.059**	[.268]
	(.012)	(.019)	(.026)	
Gender-neutral major (N+B)	.039***	.056***	.066***	[.250]
	(.011)	(.016)	(.019)	
Female-dominated major (S+H)	.015	.024	.035	[.229]
	(.016)	(.021)	(.029)	
Mother – son:				
Male-dominated major (E)	.079*	.166*	.180**	[.262]
	(.043)	(.094)	(.089)	
Gender-neutral major (N+B)	.023**	.034**	.041**	[.194]
	(.010)	(.014)	(.017)	
Female-dominated major (S+H)	.004	.009	.014	[.144]
	(.010)	(.014)	(.019)	
(2) Impact on daughters				
Mother – daughter:				
Male-dominated major (E)	.043*	.094*	.110**	[.069]
	(.025)	(.054)	(.053)	
Gender-neutral major (N+B)	.001	.005	.009	[.220]
	(.011)	(.015)	(.018)	
Female-dominated major (S+H)	.005	.011	.017	[.278]
	(.011)	(.016)	(.022)	
Father – daughter:				
Male-dominated major (E)	.005	.012	.027	[.065]
	(.008)	(.013)	(.018)	
Gender-neutral major (N+B)	.002	.007	.011	[.239]
	(.012)	(.017)	(.019)	
Female-dominated major (S+H)	.023	.035	.047	[.326]
	(.018)	(.024)	(.031)	
N	168,933	168,933	168,933	

Notes: See notes to Table 6. The regressions differ by allowing for heterogeneous effects based on the gender makeup of majors. Majors are denoted in parentheses by their first letters: E, N, B, S, H stand for Engineering, Natural Science, Business, Social Science, and Humanities, respectively. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01*

Table 10. Counterfactual gender shares for majors in the absence of family peer effects.

	Observed male share	Counterfactual male share		
		No spillover effect of:		
Panel A: Siblings		older brother	older sister	older brother or sister
Male-dominated major (E)	73	69	81	76
Gender-neutral major (N+B)	42	42	45	44
Female-dominated major (S+H)	33	35	36	38
Non-academic major	49	51	46	48
		No spillover effect of:		
Panel B: Intergenerational		father	mother	father or mother
Male-dominated major (E)	79	81	83	88
Gender-neutral major (N+B)	49	46	49	45
Female-dominated major (S+H)	39	40	41	42
Non-academic major	45	46	47	48

Notes: See text for details for how to construct the counterfactual. Note that observed male share does not line up with Appendix Figure A1 because the samples are different.

Appendix
for Online Publication

"Intergenerational and Sibling Spillovers in High School Majors"

Gordon B. Dahl, Dan-Olof Rooth, and Anders Stenberg

Appendix A: Inferring GPA Cutoffs

This appendix describes how we determine GPA cutoffs. While we observe the choice rankings for each individual and the admission decision, the GPA cutoff is not recorded in our dataset. Instead, we must infer the GPA cutoff from the data. Fortunately, the rules appear to have been strictly followed, so this is relatively straightforward.

Each combination of year, school region, and major has the potential to be a competition for slots. We refer to these as "cells." Our RD design only applies to oversubscribed majors (i.e., competitive cells). If there are more applicants than slots, the admission GPA cutoff is inferred from the data. We limit our sample to cells where there is evidence for a sharp discontinuity, that is, where everybody above the GPA cutoff is admitted to the major and everybody below the cutoff is not.[33]

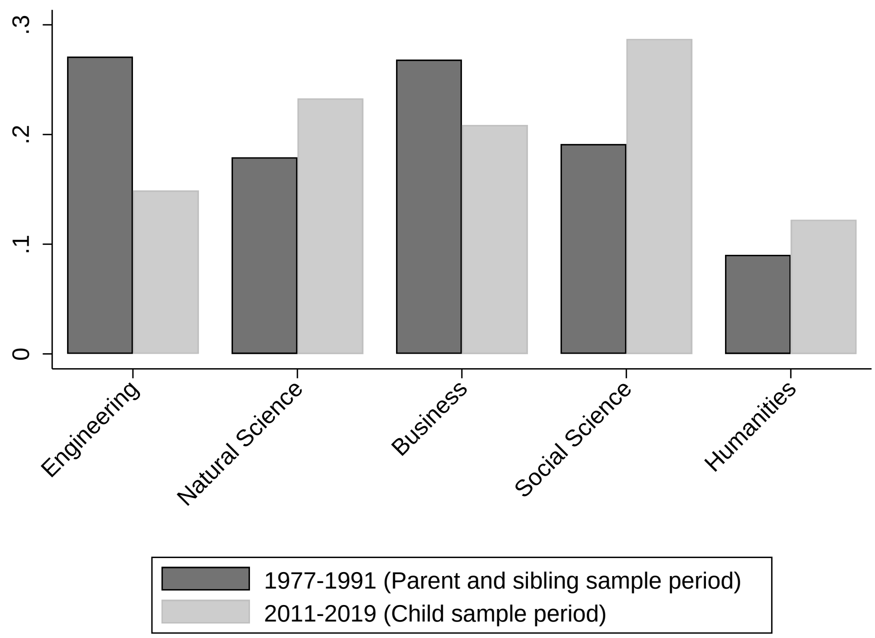
One complexity is that there can be a mix of accepted and non-accepted individuals at a cutoff GPA. For example, if the cutoff is 3.2 in a cell, there may only be slots for 3 out of the 5 applicants with a GPA of 3.2. Ties can happen since GPA is only recorded to the first decimal. In this case, it is important to know how people at the cutoff with the same GPA were admitted. We found some documentation which indicated admission was random, but also documentation which said that sometimes secondary criteria such as math grades were used to break ties. Since we do not know the criteria used to break ties, we discard the observations at the cutoff GPA. This should not create a problem, as we are still able to identify a sharp discontinuity above and below this mixed-cutoff GPA. Continuing with the example of a mixed cutoff at 3.2, we would drop all individuals with a GPA exactly equal to 3.2 in the cell, but define the cutoff as 3.2 for the remaining observations in the cell.

When there is not a mix of accepted and non-accepted individuals at a cutoff, we simply define the cutoff GPA as the average between the two adjacent GPAs. So for example, if everyone with a GPA of 3.3 is not admitted and everyone with a GPA of 3.4 is admitted, we define the GPA cutoff for the cell as 3.35. To enable pooling of data across regions and years, we normalize the cutoff GPA to 0 in our RD regressions.

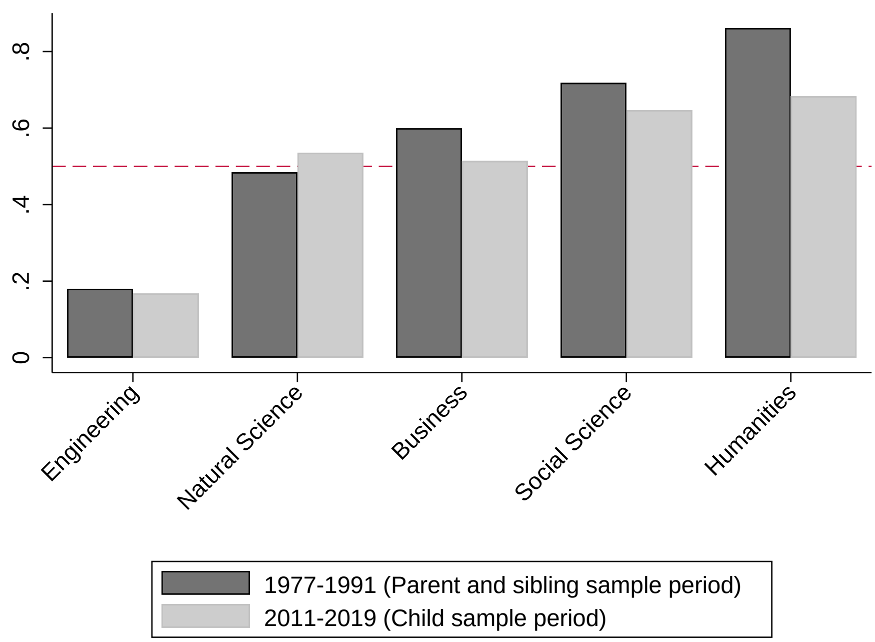
[33] We allow for a small amount of noise in the data due to measurement error, which is likely during this time period since most variables were transcribed and entered by hand. For example, if one observation with a GPA of 3.8 is recorded as not admitted while all of the remaining observations higher than 3.3 are recorded as admitted, it is likely that either GPA or major was erroneously recorded. Our rule is to retain the cell if the "miscoded" observations represent less than ten percent of the observations at the given side of the cutoff. If the condition is met, we retain the cell, but drop the "miscoded" observations. This procedure drops just 0.3 percent of the data. We also require at least 25 applicants in a cell and at least 3 observations to the left of the cutoff.

Figure A1. Applications to academic high school majors, 1977-1991 and 2011-2019.

Panel A: Share of applicants by major

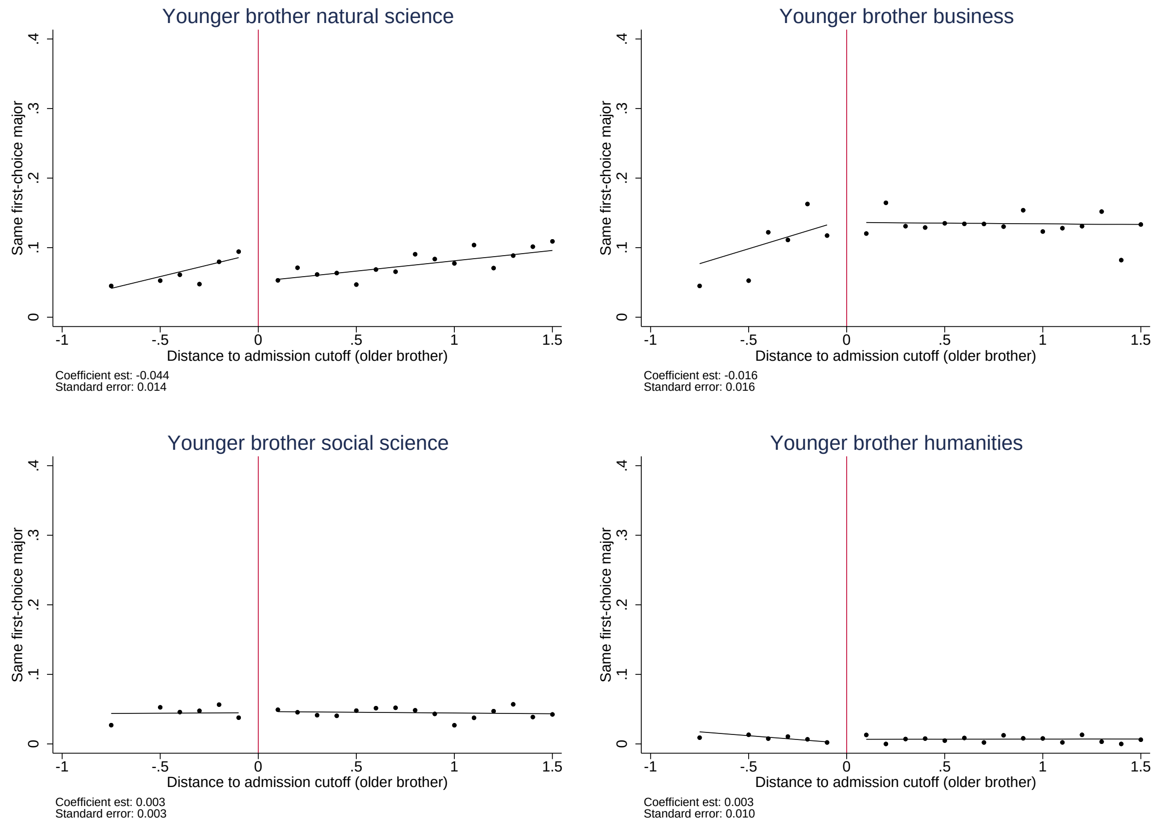


Panel B: Female share of applicants



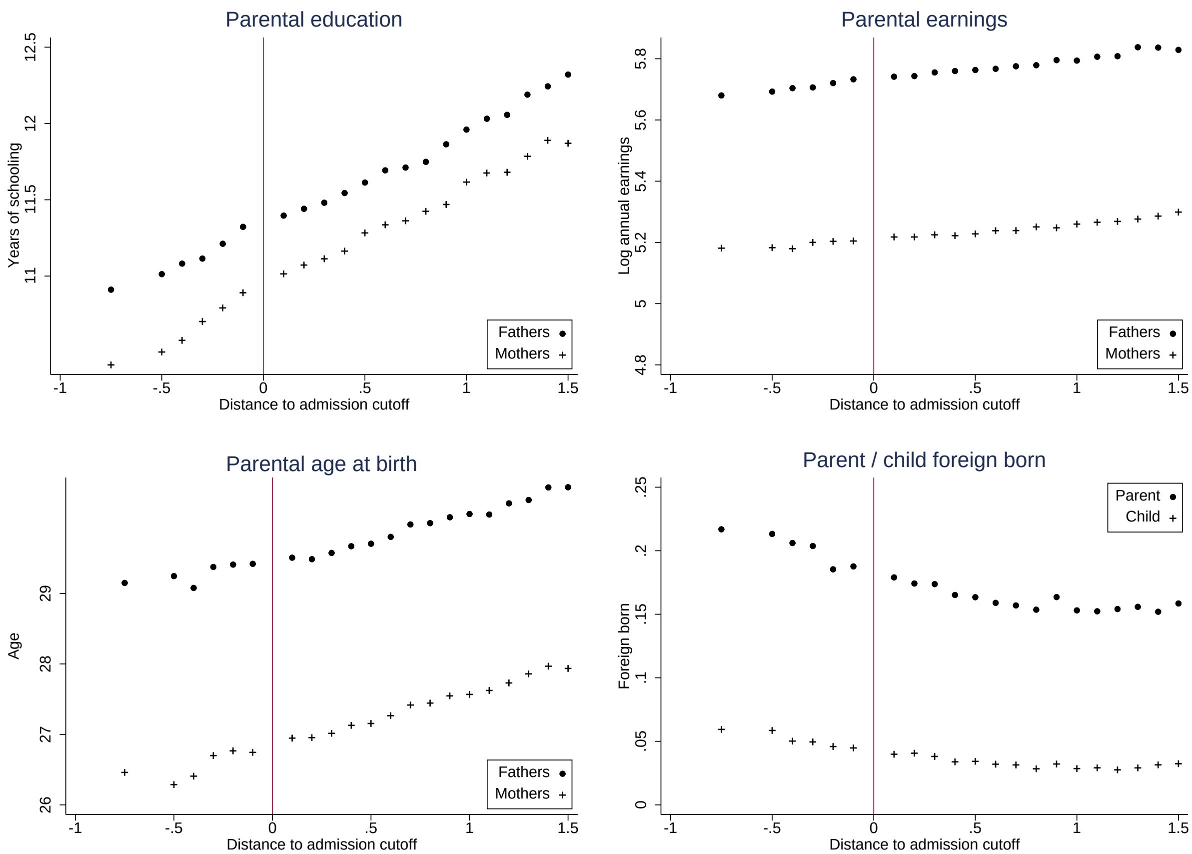
Notes: All applications to academic majors. For the years 1977-1991, N=607,767. For 2011-2019, N=558,442. The share in Humanities 2011-2019 also includes those in Arts. The dashed line marks a balanced gender composition.

Figure A2. Probability a younger brother chooses a non-Engineering major if their older brother chose Engineering.



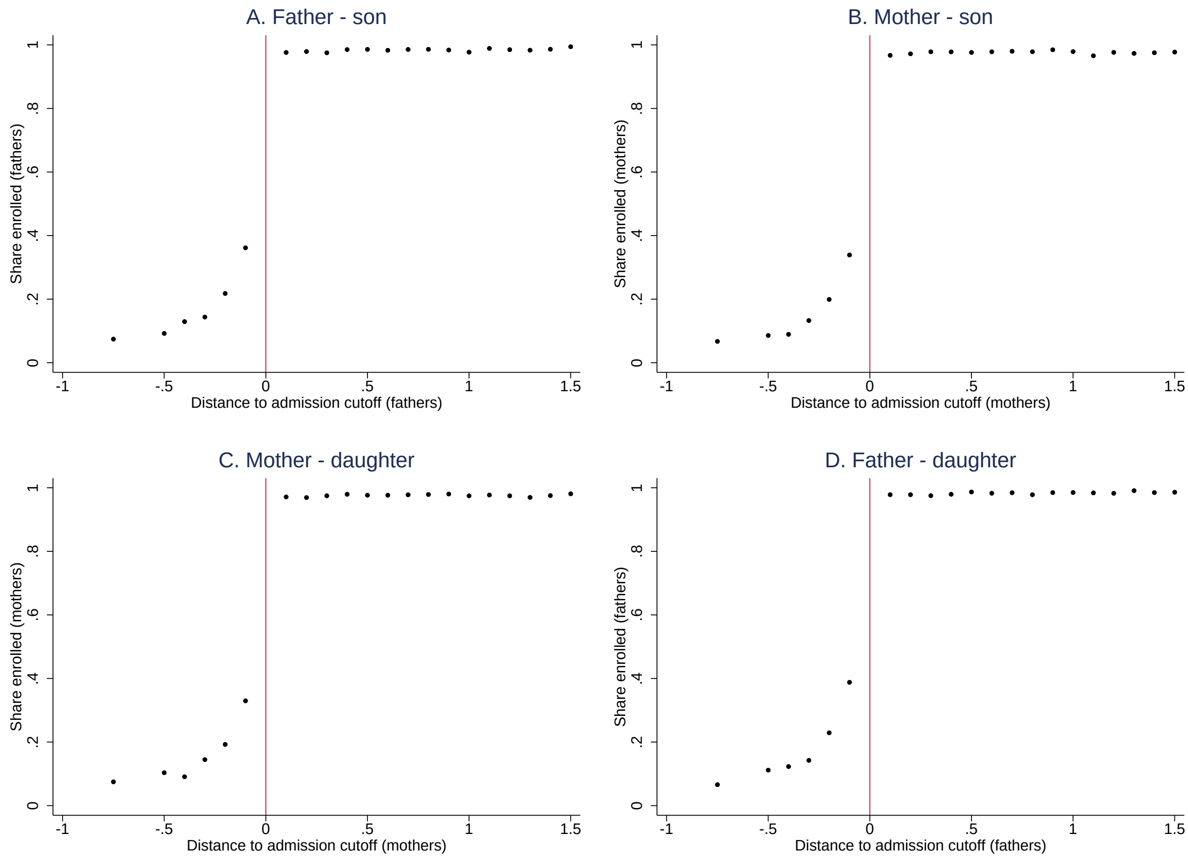
Notes: Sample limited to brother-brother pairs where the older brother chose Engineering as their first choice. Each observation is the average share of younger brothers who choose a non-Engineering major as their first choice as a function of their older brother's GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff for older brothers (normalized to 0). The estimated slopes are based on the common slope model, linear functions of GPA, a window of -1.0 to 1.5, and triangular weights. The number of observations is 11,706.

Figure A3. Smoothness of predetermined demographic variables at the cutoff.



Notes: Each marker is the average for the relevant outcome in a 0.1 GPA bin, except for the leftmost marker which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff for individuals in oversubscribed programs between 1977-1991, (normalized to 0). Parent foreign born is a dummy for whether at least one parent is foreign born. Parents here refer to the parents of applicants during 1977-1991 (i.e., these are the grandparents of the children in our intergenerational sample).

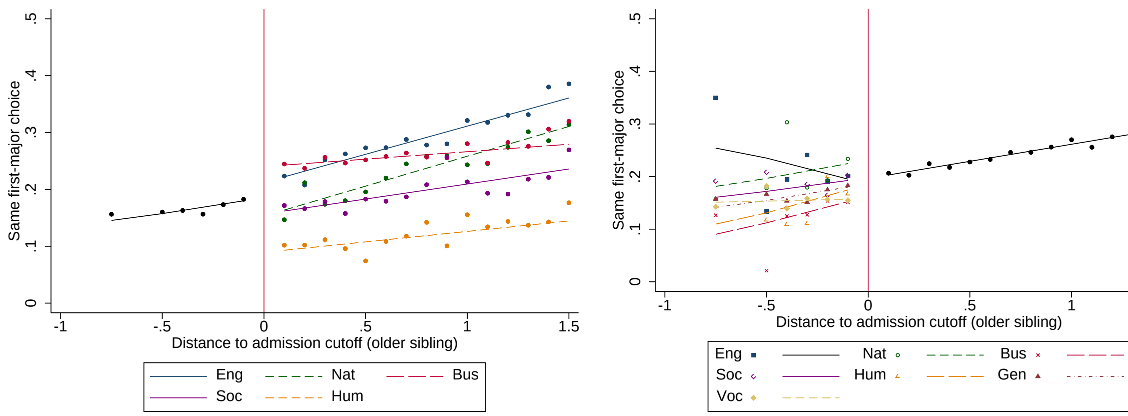
Figure A4. First stage: Share of parents enrolling in their first-choice major, by gender mix.



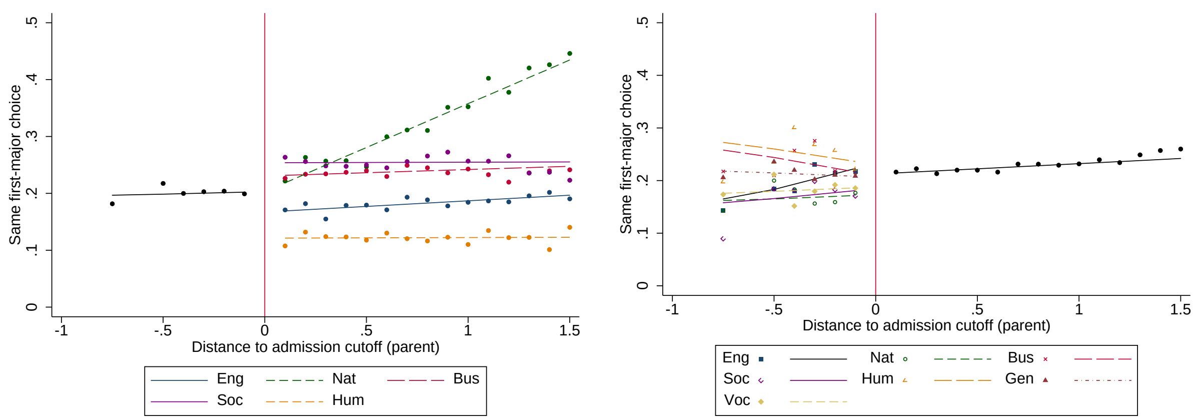
Notes: Each observation is the average share of parents who enroll in their first-best major choice as a function of their GPA. Each dot is a 0.1 GPA bin, except for the leftmost dot which is a 0.5 bin due to sparsity. The vertical lines denote the admissions GPA cutoff (normalized to 0). The number of observations in panel A is 37,390 (father-son sample), in panel B 49,057 (mother-son sample), in panel C 46,791 (mother-daughter sample), and in panel D 35,695 (father-daughter sample).

Figure A5. Comparison of the common slope versus baseline models.

Panel A: Siblings



Panel B: Intergenerational



Notes: The first column plots averages of the binned outcome variable for younger siblings and children against the running variable, allowing for separate slopes for each of the five first-best choices to the right of the cutoff and a common slope to the left of the cutoff. The second column shows similar plots, but allowing separate slopes for each of the seven second-best choices to the left of the cutoff and a common slope to the right of the cutoff. The trend lines are RD estimates using the underlying data, no covariates, and triangular weights. Note that these graphs are for illustrative purposes; we never mix the common slope and multiple slope models in estimation.

Table A1. Course requirements for each of the five academic programs.

	Weekly hours of course instruction				
Classes	Engineering	Natural Science	Business	Social Science	Humanities
Math	15^{adv}	15^{adv}	11	11	5
Natural science	17	22.5	3	9	7
Social science	11	16	16.5	25.5	25.5
Swedish	8	9	9	10	10
English	6	7	7	8	9
Additional languages	6	11	14	17	24
Art and music	-	4	-	4	4
Physical education	7	8	7	8	8
Technology related	22.5	-	-	-	-
Business related	-	-	25	-	-
Other	3.5	3.5	3.5	3.5	3.5
Total hours	96	96	96	96	96

Notes: The total amount of 96 hours consists of 34, 32, and 30 hours per week during the first, second, and third years, respectively. Engineering has an optional fourth year of 35 hours per week of mostly technology related courses. The superscript "adv" indicates that advanced math is required for Engineering and Natural Science. Business allows the possibility to exchange 3 hours of math with business-related courses. Natural science classes include physics, chemistry, and biology, while Social science classes include history, religion, philosophy, psychology, and social studies. These curricula are mandated by law and laid out in Lgy70 (Läroplan för gymnasieskolan); they remained unchanged during our sample period (1977-1991) but were modified in 1994.

Table A2. Summary statistics for all applicants with a first-choice academic program 1977-1991.

	Oversubscribed programs	Non-impacted programs
Father age	45.77	46.02
Mother age	43.21	43.34
Father schooling	11.63	11.32
Mother schooling	11.25	10.83
Father earnings	5.77	5.75
Mother earnings	5.23	5.20
Foreign born parent	0.17	0.17
Foreign born	0.04	0.04
Female	0.52	0.51
Age in year of applying	16.00	15.99
GPA	3.86	3.94
Observations	263,878	221,397

Notes: Parent and child characteristics are measured in the year of application (the child's 16th year since birth). Years of schooling inferred from highest education level. Parents here refer to the parents of applicants during 1977-1991 (i.e., these are the grandparents of the children in our intergenerational sample). Parent earnings are measured between the ages of 37-39 and are converted to year 2016 US dollars using an exchange rate of 8.5 SEK to 1 USD.

Table A3. Comparison of major cutoffs across years within the same school region.

	Fraction of years with a higher cutoff		
Major combinations	1st major	2nd major	No difference
Engineering vs. Natural Science	.37	.25	.38
Engineering vs. Business	.28	.42	.30
Engineering vs. Social Science	.21	.53	.27
Engineering vs. Humanities	.31	.38	.31
Natural Science vs. Business	.24	.46	.30
Natural Science vs. Social Science	.18	.51	.31
Natural Science vs. Humanities	.24	.38	.39
Business vs. Social Science	.24	.48	.28
Business vs. Humanities	.37	.32	.31
Social Science vs. Humanities	.47	.21	.32

Notes: The table reports the average fraction of years with a higher cutoff for one major compared to another within the same school region. If both majors have a cutoff in a given year in the same school region, we compare the two to determine which is higher. If one major has a cutoff, but the other does not, we record the major with the cutoff as having a higher cutoff. "No difference" can either reflect that both majors have cutoffs which are equal or that neither major was oversubscribed.

Table A4. Correlational estimates by gender mix.

	Correlational estimates	IV-enrolled	Difference
Panel A: Siblings			
Older brother – younger brother	.182***	.063***	.119***
	(.004)	(.015)	(.014)
Older sister – younger brother	.056***	-.029**	.085***
	(.003)	(.014)	(.013)
Older sister – younger sister	.108***	.039**	.069***
	(.004)	(.016)	(.016)
Older brother – younger sister	.098***	.017	.081***
	(.003)	(.014)	(.014)
N	82,714	88,174	
Panel B: Intergenerational			
Father – son	.111***	.043***	.068***
	(.003)	(.012)	(.021)
Mother – son	.045***	.024**	.021**
	(.002)	(.011)	(.011)
Mother – daughter	.054***	.010	.044***
	(.003)	(.012)	(.012)
Father – daughter	.061***	.014	.047***
	(.003)	(.012)	(.011)
N	157,760	168,933	

Notes: Correlational estimates are based on the fraction of younger siblings/children who list a major as their first choice if it is the one their older sibling/parent enrolled in minus the fraction who choose it when their older sibling/parent did not enroll in it. This is done for each of the 5 majors and averaged across majors (with weights equal to the number of older siblings/parents choosing each of the majors). Bootstrap standard errors based on 1,000 replications. For IV-enrolled estimates, see notes to Table 5 and 6.

** p<0.10, ** p<0.05, *** p<0.01*

Table A5. Sibling estimates by parent's educational background.

	At least one parent had an academic HS major		**Neither parent had an academic HS major**	
	Reduced form	IV-enrolled	Reduced form	IV-enrolled
(1) Impact on younger brother				
Older brother – younger brother	.032**	.048**	.053***	.075***
	(.014)	(.022)	(.014)	(.020)
Older sister – younger brother	-.028**	-.034	-.024*	-.027
	(.014)	(.021)	(.012)	(.018)
(2) Impact on younger sister				
Older sister – younger sister	.031*	.045*	.020	.031
	(.017)	(.024)	(.015)	(.023)
Older brother – younger sister	.013	.023	.003	.009
	(.013)	(.021)	(.013)	(.020)
N	51,957	51,957	36,106	36,106

Notes: See notes to Table 5. The sample is reduced slightly as 111 parents have missing values for education. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01*

Table A6. Intergenerational estimates by birth order.

	Firstborn child			Not firstborn child		
	Reduced form	IV-enrolled	IV-completed	Reduced form	IV-enrolled	IV-completed
(1) Impact on sons						
Father – son	.040***	.062***	.078***	.017	.025	.033
	(.011)	(.017)	(.021)	(.011)	(.017)	(.022)
Mother – son	.015	.027*	.035*	.013	.019	.025
	(.010)	(.016)	(.019)	(.010)	(.015)	(.019)
(2) Impact on daughters						
Mother – daughter	.010	.020	.027	-.002	-.001	.002
	(.011)	(.017)	(.021)	(.011)	(.017)	(.022)
Father – daughter	.010	.021	.030	.004	.007	.011
	(.010)	(.016)	(.020)	(.010)	(.015)	(.020)
N	87,272	87,272	87,272	81,661	81,661	81,661

Notes: See notes to Table 6. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01*

Table A7. Alternative measures for whether a younger sibling or child copies their older sibling or parent.

Panel A: Siblings

	Baseline	Same major any rank	Same major accepted	Same major enrolled	Same major completed
(1) Reduced form					
Older brother – younger brother	.043***	.041***	.041***	.042***	.037***
	(.010)	(.010)	(.009)	(.009)	(.009)
Older sister – younger brother	-.026***	-.029***	-.024***	-.023***	-.018**
	(.009)	(.010)	(.008)	(.008)	(.008)
Older sister – younger sister	.025**	.026**	.030***	.029***	.027***
	(.011)	(.012)	(.010)	(.010)	(.010)
Older brother – younger sister	.009	.005	.007	.005	.004
	(.009)	(.010)	(.009)	(.009)	(.008)
(2) IV-enrolled					
Older brother – younger brother	.063***	.059***	.060***	.061***	.053***
	(.015)	(.015)	(.014)	(.014)	(.013)
Older sister – younger brother	-.030**	-.034**	-.028**	-.026**	-.020*
	(.014)	(.014)	(.012)	(.012)	(.011)
Older sister – younger sister	.039**	.038**	.045***	.044***	.040***
	(.016)	(.017)	(.015)	(.015)	(.014)
Older brother – younger sister	.017	.011	.014	.012	.009
	(.014)	(.015)	(.014)	(.014)	(.013)
N	88,174	88,174	88,174	88,174	88,174

Panel B: Intergenerational

	Baseline	Same major any rank	Same major accepted	Same major enrolled	Same major completed
(1) Reduced form					
Father – son	.029***	.029***	.021***	.019***	.023**
	(.008)	(.008)	(.007)	(.007)	(.010)
Mother – son	.015**	.015*	.010	.008	.004
	(.007)	(.008)	(.007)	(.006)	(.009)
Mother – daughter	.004	.001	.006	.005	.004
	(.008)	(.008)	(.008)	(.007)	(.010)
Father – daughter	.007	.009	.002	.002	-.007
	(.008)	(.008)	(.007)	(.007)	(.009)
(2) IV-enrolled					
Father – son	.043***	.045***	.031***	.028**	.033**
	(.012)	(.013)	(.011)	(.011)	(.015)
Mother – son	.024**	.024**	.017*	.014	.007
	(.011)	(.012)	(.010)	(.010)	(.013)
Mother – daughter	.010	.006	.012	.009	.007
	(.012)	(.012)	(.011)	(.011)	(.014)
Father – daughter	.014	.017	.006	.005	-.007
	(.012)	(.012)	(.011)	(.011)	(.014)
(3) IV-completed					
Father – son	.056***	.057***	.040***	.036**	.041**
	(.016)	(.017)	(.015)	(.014)	(.019)
Mother – son	.031**	.031**	.022*	.018	.010
	(.014)	(.015)	(.013)	(.012)	(.017)
Mother – daughter	.015	.011	.016	.013	.010
	(.015)	(.016)	(.014)	(.014)	(.018)
Father – daughter	.021	.024	.011	.008	-.006
	(.015)	(.015)	(.014)	(.014)	(.018)
N	168,933	168,933	168,933	168,933	93,412

Notes: See notes to Tables 5 and 6. The baseline outcome variable is whether a younger sibling's or child's first choice on their preference list matches the first-best major choice of their older sibling or parent. Columns 2-5 replace this with whether the younger sibling or child (i) includes on their choice list, (ii) is accepted to, (iii) enrolls in, or (iv) completes the same major as the first-best choice of their older sibling or parent. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01.*

Table A8. Estimates corresponding to Tables 5 and 6, also including individuals who drop out after ninth grade.

	Reduced form	IV-enrolled	IV-completed
Panel A: Siblings			
Older brother – younger brother	.037***	.053***	
	(.009)	(.013)	
Older sister – younger brother	-.026***	-.031**	
	(.008)	(.012)	
Older sister – younger sister	.025**	.037***	
	(.010)	(.014)	
Older brother – younger sister	.004	.008	
	(.008)	(.013)	
N	99,384	99,384	
Panel B: Intergenerational			
Father – son	.031***	.046***	.058***
	(.008)	(.012)	(.015)
Mother – son	.017**	.026**	.033***
	(.007)	(.010)	(.013)
Mother – daughter	.006	.012	.017
	(.008)	(.011)	(.014)
Father – daughter	.009	.017	.023*
	(.007)	(.011)	(.013)
N	176,038	176,038	176,038

Notes: See notes to Table 5 and 6. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01*

Table A9. Estimates corresponding to Table 5, limiting the sample to older siblings whose first and second best major choices are both available at the school they are admitted to.

	Any age gap		**Age gap ≤ 3 years (concurrent school enrollment)**		**Age gap > 3 years**		
	Reduced form	IV-enrolled	Reduced form	IV-enrolled	Reduced form	IV-enrolled	Mean
(1) Impact on younger brother							
Older brother – younger brother	.032**	.045**	.008	.010	.064***	.088***	[.254]
	(.013)	(.018)	(.017)	(.025)	(.019)	(.026)	
Older sister – younger brother	-.027**	-.031**	-.040***	-.050**	-.009	-.007	[.111]
	(.011)	(.016)	(.015)	(.021)	(.016)	(.023)	
(2) Impact on younger sister							
Older sister – younger sister	.025*	.035*	.043**	.055**	.004	.010	[.219]
	(.014)	(.019)	(.018)	(.025)	(.021)	(.029)	
Older brother – younger sister	.007	.012	.000	.000	.016	.025	[.194]
	(.012)	(.017)	(.016)	(.023)	(.018)	(.026)	
N	63,533	63,533	37,353	37,353	26,180	26,180	

Notes: See notes to Table 5. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01*

Table A10. **Estimates corresponding to Table 6, limiting the sample to parents whose first and second best major choices are both available at the school they are admitted to.**

	Reduced form	IV-enrolled	IV-completed	Mean
(1) Impact on sons				
Father – son	.023**	.034**	.044**	[.257]
	(.010)	(.014)	(.018)	
Mother – son	.013	.021*	.028*	[.172]
	(.009)	(.013)	(.016)	
(2) Impact on daughters				
Mother – daughter	.006	.012	.017	[.246]
	(.009)	(.013)	(.017)	
Father – daughter	.004	.009	.014	[.196]
	(.009)	(.014)	(.017)	
N	123,103	123,103	123,103	

Notes: See notes to Table 6. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01*

Table A11. Sibling and intergenerational estimates for each of the five majors.

	Siblings		Intergenerational			
	Reduced form	IV enrolled		Reduced form	IV enrolled	IV completed
Brother-brother			**Father-son**			
Engineering	.058***	.086***	Engineering	.023*	.039**	.059**
	(.016)	(.023)		(.012)	(.019)	(.026)
Natural science	.042	.077	Natural science	.022	.044	.055
	(.031)	(.054)		(.030)	(.054)	(.060)
Business	.042***	.057***	Business	.042***	.057***	.068***
	(.016)	(.021)		(.012)	(.016)	(.019)
Social science	.010	.017	Social science	.015	.024	.033
	(.018)	(.025)		(.017)	(.024)	(.031)
Humanities	.007	.013	Humanities	.013	.020	.041
	(.037)	(.045)		(.029)	(.035)	(.067)
Sister-brother			**Mother-son**			
Engineering	.020	.032	Engineering	.043*	.094*	.110**
	(.067)	(.094)		(.025)	(.054)	(.053)
Natural science	-.126*	-.203	Natural science	.000	.006	.018
	(.069)	(.127)		(.044)	(.068)	(.084)
Business	-.015	-.016	Business	.001	.005	.009
	(.014)	(.019)		(.011)	(.015)	(.018)
Social science	-.038***	-.044**	Social science	.003	.007	.012
	(.013)	(.018)		(.014)	(.019)	(.024)
Humanities	-.028*	-.036	Humanities	.013	.025	.035
	(.015)	(.022)		(.018)	(.027)	(.035)
Sister-sister			**Mother-daughter**			
Engineering	.057	.099	Engineering	.005	.012	.027
	(.047)	(.078)		(.008)	(.013)	(.018)
Natural science	.024	.039	Natural science	.041	.075	.082
	(.044)	(.063)		(.030)	(.052)	(.056)
Business	.042**	.061**	Business	-.003	.000	.004
	(.018)	(.024)		(.012)	(.017)	(.019)
Social science	.017	.026	Social science	.028	.042	.052
	(.017)	(.024)		(.020)	(.027)	(.034)
Humanities	-.013	-.015	Humanities	-.003	.000	.006
	(.025)	(.036)		(.035)	(.045)	(.074)
Brother-sister			**Father-daughter**			
Engineering	-.010	-.011	Engineering	.079*	.166*	.180**
	(.011)	(.017)		(.043)	(.094)	(.089)
Natural science	.012	.027	Natural science	.002	.008	.019
	(.031)	(.056)		(.037)	(.057)	(.065)
Business	.025	.036	Business	.024**	.036**	.043**
	(.017)	(.023)		(.011)	(.015)	(.017)
Social science	.017	.027	Social science	.000	.005	.008
	(.023)	(.032)		(.012)	(.016)	(.021)
Humanities	.060	.080	Humanities	.013	.023	.033
	(.047)	(.059)		(.013)	(.020)	(.027)
N	88,174	88,174		168,933	168,933	168,933

Notes: See notes to Table 5. The regressions differ by allowing for heterogeneous effects based on the majors. Standard errors in parentheses, clustered at the family level.
** p<0.10, ** p<0.05, *** p<0.01*

Table A12. Changes in the characteristics of the student peer group an older sibling or parent is admitted to.

	Female share of peer group for older sibling	GPA rank relative to peer group for older sibling		Female share of peer group for parent	GPA rank relative to peer group for parent
Older brother – younger brother:			Father – son:		
Male-dominated major (E)	-.095***	-.363***	Male-dominated major (E)	-.086***	-.381***
	(.007)	(.009)		(.007)	(.008)
Gender-neutral major (N+B)	.046***	-.297***	Gender-neutral major (N+B)	.057***	-.314***
	(.006)	(.008)		(.005)	(.007)
Female-dominated major (S+H)	.173***	-.287***	Female-dominated major (S+H)	.175***	-.306***
	(.008)	(.012)		(.007)	(.010)
Older sister – younger brother:			Mother – son:		
Male-dominated major (E)	-.296***	-.318***	Male-dominated major (E)	-.273***	-.294***
	(.035)	(.036)		(.027)	(.030)
Gender-neutral major (N+B)	-.053***	-.315***	Gender-neutral major (N+B)	-.043***	-.322***
	(.005)	(.009)		(.003)	(.007)
Female-dominated major (S+H)	.133***	-.245***	Female-dominated major (S+H)	.126***	-.256***
	(.005)	(.008)		(.004)	(.007)
Older sister – younger sister:			Mother – daughter:		
Male-dominated major (E)	-.301***	-.317***	Male-dominated major (E)	-.345***	-.310***
	(.038)	(.043)		(.027)	(.034)
Gender-neutral major (N+B)	-.050***	-.310***	Gender-neutral major (N+B)	-.044***	-.315***
	(.005)	(.009)		(.003)	(.007)
Female-dominated major (S+H)	.134***	-.241***	Female-dominated major (S+H)	.129***	-.252***
	(.005)	(.009)		(.004)	(.007)
Older brother – younger sister:			Father – daughter:		
Male-dominated major (E)	-.091***	-.361***	Male-dominated major (E)	-.093***	-.380***
	(.008)	(.009)		(.007)	(.008)
Gender-neutral major (N+B)	.043***	-.301***	Gender-neutral major (N+B)	.049***	-.308***
	(.006)	(.009)		(.005)	(.007)
Female-dominated major (S+H)	.174***	-.287***	Female-dominated major (S+H)	.176***	-.297***
	(.008)	(.012)		(.006)	(.009)
N	88,174	82,611		168,933	154,881

Notes: See notes to Table 5. Female share is the older sibling's (or parent's) fraction of women in the same year and region for the major they are admitted to. GPA rank is the older sibling's (or parent's) GPA rank relative to their peers in the same year and region for the major they are admitted to. The regressions differ by allowing for heterogeneous effects based on the majors. Standard errors in parentheses, clustered at the family level.

** p<0.10, ** p<0.05, *** p<0.01*

Table A13. Multiple inference adjustments for the reduced form estimates in Tables 8 and 9.

Panel A: q-values after FDR control for Table 8		**Panel B:** q-values after FDR control for Table 9	
(1) Impact on younger brother		**(1) Impact on sons**	
Older brother – younger brother:		Father – son:	
Male-dominated major (E)	<.001	Male-dominated major (E)	.094
Gender-neutral major (N+B)	.005	Gender-neutral major (N+B)	.002
Female-dominated major (S+H)	.563	Female-dominated major (S+H)	.349
Older sister – younger brother:		Mother – son:	
Male-dominated major (E)	.764	Male-dominated major (E)	.101
Gender-neutral major (N+B)	.206	Gender-neutral major (N+B)	.078
Female-dominated major (S+H)	.004	Female-dominated major (S+H)	.716
(2) Impact on younger sister		**(2) Impact on daughters**	
Older sister – younger sister:		Mother – daughter:	
Male-dominated major (E)	.341	Male-dominated major (E)	.265
Gender-neutral major (N+B)	.045	Gender-neutral major (N+B)	.962
Female-dominated major (S+H)	.530	Female-dominated major (S+H)	.961
Older brother – younger sister:		Father – daughter:	
Male-dominated major (E)	.329	Male-dominated major (E)	.817
Gender-neutral major (N+B)	.329	Gender-neutral major (N+B)	.870
Female-dominated major (S+H)	.329	Female-dominated major (S+H)	.583

Notes: The table reports multiple inference corrected q-values (False Discovery Rate control) using the qqvalue package in Stata (method: simes).