Society for Economics Research in India
Working Paper Series

PREFERENCES OR EXPECTATIONS: UNDERSTANDING THE GENDER GAP IN MAJOR CHOICE

Aparajita Dasgupta
(Ashoka University, Sonepat)
Anisha Sharma
(Ashoka University, Sonepat)

Working Paper No. 2
http://seri-india.org/research

August 2020

Preferences or expectations: Understanding the gender gap in major choice

Aparajita Dasgupta[1] Anisha Sharma[2]

August 2020

Abstract

What explains the low enrolment of women in science and economics? We combine administrative and survey data on a sample of high-income workers in India who have completed the same elite graduate programme to estimate the return to studying different undergraduate degrees - across science, business and economics, and the humanities. We find evidence of a large earnings premium to studying science and economics, yet disproportionately low female enrolment in these subjects. Using data on the subjective expectations of undergraduate students who are in the process of selecting a major, we model major choice as a function of major-specific and job-specific attributes. We identify significant gender differences in the preferences for different attributes as well as in the expectations of future outcomes, especially of grades. Women are willing to pay twice as much as men for course enjoyment and higher grades, even as they expect lower grades in science and economics. This suggests that in addition to pervasive norms about which subjects are better suited for women, women also suffer from a relative confidence gap in their major-specific abilities.

JEL-Classification: D84; I23; I24; I26; J24
Keywords: higher education; returns to education; education and gender; major choice

[1]Ashoka University, Sonepat, Rai, Haryana 131029, India. aparajita.dasgupta@ashoka.edu.in
[2]Ashoka University, Sonepat, Rai, Haryana 131029, India. anisha.sharma@ashoka.edu.in
We would like to acknowledge helpful comments from Farzana Afridi, Amrita Dhillon, Seema Jayachandran, Shabana Mitra, Nishith Prakash, Sanchari Roy, and Soham Sahoo for their helpful comments and suggestions. This project received IRB approval from Ashoka University.

1 Introduction

In recent years, there have been substantial increases in tertiary school enrolment by women in both the developed and the developing world, so much so that women today outnumber men in universities in many countries. In the US, for example, there are 136 women enrolled in college for every 100 men, while in India, there are 109 women in college for every 100 men.[1] However, enrolment into different majors remains highly gendered: in the US, women are two-thirds as likely as men to study science, technology and business (Gemici and Wiswall, 2014), while in India, women are a third as likely as men to study computer science, and two-thirds as likely to study business (Ministry of Human Resource Development, 2019). These differences in enrolments by major appear to have significant financial consequences: evidence from the US suggests that there exist high returns to majors in science and technology-related fields (STEM) as well as business and finance (Altonji et al., 2016). While there has so far not been compelling evidence on major-specific wage differentials in developing countries, they are likely to exist there as well, implying that the closing of the gender gap in college enrolment alone may not have a large effect on the persistent and well-documented gender earnings gap around the world.(Blau and Kahn, 2000; Weichselbaumer and Winter-Ebmer, 2005).

A multidisciplinary literature spanning economics, psychology and sociology has attempted to explain the gendered enrolment of college students into majors, and particularly the under-representation of women in math-intensive fields like the STEM subjects (see Kahn and Ginther (2018) for a review of this literature). Recent research has quantified the importance of gendered preferences for majors and jobs in explaining major choice (Zafar, 2013; Wiswall and Zafar, 2015, 2017). However, almost all of the literature so far has been focused on the US context, despite the fact that recent research finds that cultural norms may have an important role to play in influencing educational choices (Friedman-Sokuler and Justman, 2020). Even on the question of whether there exists an earnings premium for STEM graduates, and whether these returns are as large for women as they are for men, there is limited evidence available from developing countries. It is in the context of a large developing country, India, that this paper asks the following questions: first, is there an earnings premium associated with studying subjects like science and business at university, compared to other majors, and does this earnings premium for science exist for women as well as for men? Second, given the answer to the first question, how do students decide what major to study, and, in particular, how important are gender differences in preferences over major- and job-related attributes compared to gender differences in expected outcomes? We

[1] Data available from the World Bank online database at `https://data.worldbank.org/`.

study these questions using novel data on young and highly productive professionals who have been educated in elite institutions in India and will go on to count among the highest earning individuals in the country. This is a demographic group of interest for a number of reasons: a narrowing of the gender gap at the top of the earnings distribution has been found to be associated with an increase in female promotions (Matsa and Miller, 2011) and even an improvement in firm performance (Smith et al., 2006). Moreover, gender inequality is likely to be lower at higher levels of income because of relatively gender progressive social norms (Bank, 2012), which implies that evidence of gender inequality in this sample reflects serious concerns about the pervasiveness of structural gender disadvantage.

To answer the first question, we study a group of comparable students who have all completed the same graduate programme after completing a range of different majors at the undergraduate level, across science and technology, economics and business, and the humanities and social sciences. We merge a rich administrative dataset including measures of cognitive ability, non-cognitive skills and socioeconomic and demographic background, with survey data we collect on educational and job histories as well as survey measures of personality traits. We find large returns of 17-32% to taking majors in science and economics, relative to social science and the humanities, even after controlling for a wide range of cognitive, non-cognitive and demographic characteristics. The returns to women in science are as large as they are for men. We also find that occupational choice is a key driver of the gender and the major earnings differences, with science and economics majors dominating the higher paying occupations, and men outnumbering women in these majors.

We next estimate a model of major choice by surveying a sample of undergraduate students who are in the process of making a decision about what major to study or have recently decided what major to enrol in. We collect information on their subjective expectations of major and job-specific outcomes conditional on major choice and using these data to estimate models of major choice by gender. We quantify the relative importance of *beliefs* about future outcomes and *preferences* for different attributes of majors and jobs in determining gender differences in major choice, finding that, unlike in the US, a much higher share of the gender differences in enrolment (50-75%) can be explained by differences in expectations, and the rest explained by differences in preferences. We find significant differences in preferences by gender, with women placing a much greater weight on enjoying coursework and grades and a lower weight on high earnings, compared to men. Women are willing to pay 8.1% of their income on graduation for a percentage point increase in the probability that they enjoy their coursework and 5.1% of their income for a percentage point increase in the probability that they get an above median grade. These estimates are twice as large as the corresponding estimates for men (4.3% and 2.3% respectively). These differences are most acute in the bottom half of the ability distribution: high-ability men and women are relatively similar to one another but low-ability women appear to have very different preferences from low-ability women. We also compare the expectations of outcomes with observed outcomes from the same university to find significant differences by gender in the size of the errors students make about future outcomes. In particular, men overestimate their grades in science and economics, compared to women.

Our paper contributes to and builds on a number of literatures. First, we provide robust estimates of the return to studying a given major at university by collecting a unique and detailed dataset of labour and educational histories of a comparable group of students all of whom completed the same graduate programme, and combining it with administrative data on the same students. In a developing country context, there are very few studies that estimate the returns to studying specific disciplines and many of them suffer from serious omitted variable bias concerns. A recent study using data from Chile finds high returns to studying science for men, but not for women (Aguirre et al., 2020); in contrast, we find high returns for women to studying science. Previous studies for India have all focused on the return to studying science in high school. Roychowdhury (2019) finds that studying science in high school increases earnings by between 10-26% for wage-earners and finds that the main channel for these increased earnings is through selection into high-paying professional occupations. Using the same dataset, Jain et al. (2019) find that men who studied science in high school earn 20% more than men who studied business. The returns to studying science are high for women as well: Sahoo and Klasen (2018) find that women who study science in high school are more likely to be employed, more likely to work in a male-dominated occupation, and earn 28% more than women who studied non-technical subjects, conditional on having completed secondary school. However, these estimates shed little light on the returns to studying science at the undergraduate level, and can offer limited insight into the trade-offs involved in major choice. Our institutional setting is unique and the data very rich, which allows to control for a wide range of potentially confounding characteristics. In addition, our sample of workers represents a segment of highly-qualified professionals, educated in elite institutions, and receiving high earnings relative to the national average. This allows us to examine gender gaps in earnings at the right tail of the ability and earnings distribution, where candidates are balanced on abilities by gender, and where social and financial constraints to women studying science in high school, and to working, are likely to be lower.

Second, we provide new subjective expectations data to explain differences in major choice across men and women in a developing country. To our knowledge, we are the first paper to explain major choice in the context of a developing country. Our context is

India, a country with strongly patriarchal social norms that may influence the educational and labour market opportunities available to women. In the context of the US, Gemici and Wiswall (2014) use large survey data to estimate dynamic model of major choice and find that the differences in subject choice across gender can be explained by differences in gender preferences. Similarly, Zafar (2013) and Wiswall and Zafar (2017) use subjective expectations data to estimate a model of college major choice to find that college major differences across gender can be explained by different gender preferences for job attributes of jobs associated with certain majors. Reuben et al. (2017) use experimental measures of personality traits such as competitiveness, confidence and risk aversion to explain varying major choices by gender. This paper adds to that literature by finding that even within a developing country, differences in gender preferences are important. However, differences in expectations are relatively more important than in the US, explaining 50-75% of the variation in major enrolment. We also find evidence of substantial gendered errors in the prediction of grades, suggesting a role for policy in correcting the beliefs of students about their future earnings and grades (Arcidiacono et al., 2012; Conlon, 2019). In terms of methods, we also build on the growing literature using stated preferences and subjective expectations survey data to estimate a model of decision choice in a developing countries, quantifying the relative importance of differences in preferences and expectations in explained gendered enrolment in majors (Manski, 2004; Delavande, 2014; Delavande and Zafar, 2019).

The paper is organised as follows. Section 2 describes the background and context. Section 3 describes the data and presents summary statistics. Section 4 describes the empirical methodology. Section 5 presents the results. Section 6 discusses gender differences in preferences for education as well as differences in subjective expectations of future outcomes.

2 Background and context

2.1 Determinants of major choice

Standard models of human capital investment emphasise the monetary return on the choice of major through the probability of getting a job in a given sector and the wages in that job (Altonji, 1993; Arcidiacono, 2004); if there exists an earnings premium to studying majors in STEM relative to others, then that would predict higher rates of enrolment in STEM majors. One reason that could explain the gap in gender enrolment into high-paying majors is if the expected earnings premium for STEM majors is larger for men than it is for women. While evidence from the US suggests that the opposite is true, and that gender earnings gaps are narrowest within STEM fields (Goldin, 2014), recent evidence from Chile finds that the high returns to studying STEM accrue only to men (Aguirre et al., 2020). On the other hand, empirical evidence suggests that while the earnings elasticity of major choice is positive and significant, it is not quantitatively large (Beffy et al., 2012; Wiswall and Zafar, 2015). Recent work emphasises that students, especially women, also value non-pecuniary attributes of jobs, such as getting a job in a desired sector, opportunities for work-life balance, the prospects of promotion and potential for earnings bonuses, and the gender diversity in the workplace (Zafar, 2013; Wiswall and Zafar, 2017).

Students also appear to value attributes of the major itself. Student abilities and tastes for a subject are important determinants of the choice they make at the undergraduate level (Altonji, 1993; Arcidiacono and Koedel, 2014; Weinberger, 2004; Stinebrickner and Stinebrickner, 2008). Ability is important because students care about their likelihood of graduation, their final grades, as well as the cost of the effort required for them to maintain those grades. Students also appear to value non-pecuniary attributes of subjects as well, electing to study subjects that they enjoy, or that they think their parents will approve of (Zafar, 2012).

Finally, individual characteristics that are fixed across majors and jobs also influence the choice of major, both through differences in preferences for attributes as well as expectations of outcomes conditional on major choice. While gender is the main focus of this paper, other characteristics may also influence major choice through preferences or expectations. First, cognitive and non-cognitive abilities are widely understood to affect major choice by changing the student's beliefs about their grades, the probability of graduating, and the effort required to to do well (Arcidiacono, 2004), and may vary by gender as well. More specifically, while verbal and language skills are important for any academic domain, they are relatively more important for the humanities and social sciences. Women outperform men on most tests of verbal ability (Halpern et al., 2007), so it might be the case that women pursue their comparative advantage and select into humanities and social science majors. Second, personality characteristics such as self-esteem, attitudes towards risk, and degree of competitiveness have been found to be correlated with gender as well as major choice (Reuben et al., 2017). Third, students of lower socioeconomic status may be less well-informed than students who are better off (Hastings et al., 2015). This informational advantage is amplified through selection into elite schools that provide better counselling to their students. Elsewhere, race and ethnic status have also been found to be associated with both college choice and college major choice (Arcidiacono and Koedel, 2014; Friedman-Sokuler and Justman, 2020).

In this paper, we collect major-specific subjective expectations data from students on several of these attributes so as to estimate a model of subject choice. We collect survey data on individual attributes to check that the results are robust to controlling for different characteristics. Since there is little evidence in India on relative earnings across majors, we also collect data from a sample of high-income earners to quantify the size of the earnings premium for studying different disciplines at university, as well as the associated gender gap. The data for both analyses is drawn from alumni and current students of an independent, private university in India. In the remainder of this section, we describe the context and institutional background for the data.

2.2 Context and institutional setting

In the Indian education system, children complete elementary education between the ages of 6 and 14 (grades 1-8), and secondary education between the ages of 14 and 18 (grades 9-12). In the first two years of secondary education, students study a wide range of subjects including language, mathematics, science and social science. After grade 10, students enter two years of higher secondary school during which they select a narrower range of subjects from among science, economics and business, or the humanities. At the end of grade 12, students can enter higher education to complete degrees or diplomas, which are typically 3-5 years long.[2]

According to Indian government statistics, as of 2018, 39 million students were enrolled in secondary school (grades 9-10) and another 25 million students in higher secondary school (grades 11-12). 29 million students were enrolled in an undergraduate or integrated undergraduate-graduate degree. At the undergraduate level, approximately 56% of students are enrolled in arts and business degrees (the two categories are not clearly separated in the government data), while 41% are enrolled in STEM degrees (Ministry of Human Resource Development, 2019). Female enrolment is lower than male enrolment at all levels of schooling at about 47% of all enrolled students in secondary school and at university.

The data for this study is drawn from an independent private university near Delhi, India, which offers two categories of programmes: a one-year graduate multidisciplinary programme and a 3-year undergraduate programme where students can select from one of 12 majors. The one-year graduate multidisciplinary programme is designed to identify and develop “leadership potential" in young people. Admission to the programme is highly

[2]Most BA and BSc degrees are three years long, some STEM degrees, such as engineering and architecture, 4 years long, and medical degrees 5 years long. In addition, many students elect for 5-6 year "integrated" degrees which include both an undergraduate degree and a graduate degree; these occur frequently in the natural sciences and in law.

selective and the curriculum [3] spans a wide range of subjects across science, business and the humanities. Students additionally receive mentoring from business and academic leaders to develop their professional skills and networks. The setting is of interest to us because incoming students – who all hold an undergraduate degree on entry to the programme – come from diverse academic backgrounds spanning the STEM disciplines, business, commerce, economics as well as the social sciences, arts and humanities, and law. However, all are comparable to one another because they have been admitted through the same, rigorous application process. They go on to work in largely professional roles in sectors such as consulting and finance, computer and data science, natural sciences, research and education, and earn a median income that is five times the national per capita income, categorising them as high income earners.[4]

Asking the same respondents about why they chose their own majors can lead to choice-supportive bias as respondents may seek to rationalise choices they made several years in the past by emphasising the positive aspects of their choice and playing down the negative aspects of their choice (Mather et al., 2000). To eliminate this bias, we turn to a group of undergraduate students from the same institutional setting who are either yet to declare an undergraduate degree or have recently done so. The same university has a standard 3-year undergraduate programme where students can choose, during the course of their second year, from among 12 majors, which we club together into four categories – science, economics, social sciences, and the humanities. These categories comprise the following majors:

- Science: Biology, Chemistry, Computer Science, Mathematics, Physics
- Economics and Business: Economics
- Social sciences: Political Science, Psychology, Sociology
- Humanities: English, History, Philosophy

To estimate the returns to different majors, we collect administrative and survey data on the alumni of the graduate programme (the "high-income earners sample"), and to estimate the model of major choice we collect detailed subjective expectations data from a survey of the undergraduate students ("the major choice sample").

[3]The curriculum and design of the programme has been developed in partnership with a number of international universities including the University of Pennsylvania, Carleton College, University of California Berkeley, University of Michigan, King's College London, Trinity College Dublin, Sciences Po Paris, Yale University and Wellesley College.

[4]The median income of this sample of earners is approximately Rs 677,000, compared to the national per capita income in 2019 at Rs 141,000. (Data retrieved from the Reserve Bank of India's data warehouse, accessible at `https://dbie.rbi.org.in/`.)

3 Data

3.1 High-income earners sample

Administrative data on students of the graduate programme between 2011-18 was collected as part of their online application. This includes detailed student-level information across many dimensions. First, it includes information on their academic background – school locations, subject choices and test scores in grades 10 and 12, and university locations, subject choices and Grade Point Averages in their undergraduate and postgraduate degrees. Second, applicants also complete an employment history with detailed information on the role, industry, location and income earned on their last two jobs. Third, applicants include information on their family backgrounds – parent's education, role, industry, and location of employment, and family income. This online data is matched with the detailed scores given to applicants during the admissions process. We also have data on the amount of need-based financial aid received by each student (which can cover 15%, 25%, 50%, 75% or 100% of total costs). The amount of financial aid is determined after reviewing detailed financial information on all students, including tax returns, salary slips and asset statements of all immediate family members, and is the strongest measure of the socioeconomic status of their household.

In addition to the administrative data, we contacted the approximately 1100 alumni of the programme by phone and by email and invited them to participate in an online survey; 326 completed the entire survey of which 176 are women. As compensation, the alumni were told that three of the people who successfully completed the survey would be randomly selected to each receive a gift voucher worth approximately $150 (Rs. 10,000). The survey was approximately an hour long, and included several modules. The first module collected information on a full educational history and socioeconomic and parental background characteristics. A second module on employment collected full employment histories for all respondents of all jobs they have ever held, income earned, hours spent at work on each job, and sector of occupation. A third module elicited information on some personality traits: self-esteem, attitudes towards competition, and attitude towards risk. For self-esteem, the survey measure is the widely used ten-question Rosenberg Self Esteem Scale (Gray-Little et al., 1997). For competitiveness, we use a five-question survey measure of comfort with competitiveness designed by Bönte et al. (2017). For risk-aversion, we use ten lottery questions based on those in Dohmen et al. (2010) as well as two vignette-style questions to measure the level of risk aversion.

3.2 Major choice sample

To explain major choice we survey a group of undergraduate students at the same private university. An electronic survey was distributed to all enrolled undergraduate students at the university and students were told that the study was about analysing major choice (survey instrument is available on request). The survey was approximately 15 minutes long and at the end of the survey respondents could choose whether they wanted to include their name in a random draw for a gift voucher worth approximately $150. Of the approximately 1000 students we wrote to, 351 students filled out the entire survey of which 193 are female.

We follow Zafar (2013) in designing our survey to elicit subjective expectations associated with graduating in any one of four categories of majors. Students were asked to rank their preferred major choices, after which they were asked a series of questions eliciting their subjective expectations about outcomes associated with graduating in each major. The surveyed outcomes included some discrete and some continuous variables on both major characteristics (probability of getting a high GPA, hours per week spent studying, probability of enjoying the degree, probability that parents will approve of the choice) and job characteristics (expected wage on graduation, expected wage ten years after graduation, probability of employment, probability of employment in preferred sector, probability of having a gender-balanced workplace, probability of enjoying work-life balance). For example, students were asked the probability that they would get a GPA of at least 3.3/4.0 (an above-average outcome at the university) if they graduated in a science major, in economics, in a social sciences major and in a major in the humanities. Similarly, students were asked what their expected income would be immediately after graduating from one of the four categories of degrees. There is a high degree of variation in expectations across both continuous and discrete outcomes suggesting that students understood the questions and were able to use the full distribution of outcomes from 0 to 100.

We also collected data on demographic background, such as parent's and siblings' educational background, subject area of study, and occupational status, and source of funding. Finally, we collected survey measures of some personality characteristics – self-esteem, attitude towards risk and competitiveness.

3.3 Summary statistics

3.3.1 High-income earners sample

For the sample of earners, we aggregate degrees at the undergraduate level into three categories: STEM (including science, engineering and architecture), business (including commerce, accounts and economics), and the humanities and social sciences (including the social sciences, art, design and law). Of the 326 surveyed alumni, 50% graduated with STEM degrees, 22% business and 26% humanities and social sciences. This distribution at the undergraduate level closely maps the total distribution of students in India across science and non-science majors (Ministry of Human Resource Development, 2019). However, since Indian government statistics do not report business degrees separately from humanities degrees, we cannot compare the distribution of students across these two disciplines to the all-India figures.

Table 1 presents summary statistics on the sample of high-earning individuals. The distribution of men and women across different majors (Table 1) is strikingly different. 39% of women graduate with STEM degrees, almost half the proportion of men at 68%. 27% of women and 16% of men study economics or business, while 35% of women and 16% of men study the humanities.

The average male salary is significantly higher than the average female salary by 20%. Men are also more likely to be currently employed (75%) compared to women (66%), and have more years of work experience (6.1 compared to 5.5 years for men and women respectively). These gaps in job outcomes cannot be explained by differences in our measure of ability, as women's grade 10 test scores are significantly higher than men. There is no gender difference in language fluency or in mathematical fluency, as measured by the proportion of students who had english and mathematics, respectively, as one of their top three subjects in high school.

There is significant variation by gender across some personality traits and socioeconomic characteristics: women in the sample have significantly lower levels of self-esteem and competitiveness but are no more risk averse than the men in the sample. Women are also more likely to have a mother who has completed an undergraduate degree. However, male and female students are equally likely to have received a need-based financial scholarship to fund their education, suggesting that income differences in the sample are not significantly different by gender.

The major-earnings gradient and gendered enrolment into different majors is evident in figure 1. There is a clear earnings gradient across different categories of majors with the highest earnings for science, followed by economics and then the humanities. The earnings gap between men and women is significant only for economics graduates and is insignificantly different from zero for science and humanities graduates. This is in line with evidence from the US as well, which shows that gender earnings gaps are narrowest within scientific occupations and widest for economics and business occupations (Goldin, 2014)

3.3.2 Major choice sample

Table 2 presents summary statistics on the sample of undergraduate students. The most popular major is economics, with 39% of all students ranking this major as their most preferred, followed by science (20%), social science (20%) and the humanities (12%). However, as with the older sample, there are substantial differences in the first choice of major by gender. More than half of the women prefer a major in social science or the humanities (37% and 17% respectively), compared to 26% of men (20% and 6% respectively). Correspondingly, three quarters of the men prefer a major in science or in economics. 15% of women report a desire to select a science major and 31% in economics, compared to 25% and 49% of men, respectively. Drop-out is rare – under 1% of undergraduates failed to complete their degrees in the preceding three years, and so, we do not consider it to be part of the choice set.

In terms of ability, female students have significantly higher grade 10 aggregate and english language test scores than men but there is no significant difference by gender in maths scores.

On personality measures, unlike in the sample of high-earning individuals, there are no significant differences between men and women on any of the three measures – risk aversion, competitiveness and self esteem. In terms of family background, 50% of women have a mother who has an undergraduate degree compared to 39% of men, and 60% have a working mother, compared to 48% for men. In this sample, men are also more likely to receive need-based financial scholarships than women.

We ask the undergraduate students about future expected outcomes – the probability of being employed, married or having children within ten years (when they will be between 28 and 30 years old). Perhaps surprisingly, there are few differences between men and women in these beliefs. In fact, women believe they are more likely to be employed compared to men, though the difference is marginal. This suggests that differences in future outcomes over a ten year horizon are not driving the differential selection by major.

Table 3 shows the differences in subjective outcomes across different major choices. Science majors are seen to be the most difficult, with only 55% of students predicting that they could easily score an above-average GPA, compared 64% in economics, 74% in the social sciences and 70% in the humanities. Students also anticipate needing to study for more hours on a science major (19 hours a week) compared to economics (17 hours), the social sciences (14 hours) or the humanities (15 hours each), and that they will enjoy the major less, with 51% expecting to enjoy coursework, compared to 58% in economics, 71% in the social sciences and 67% in the humanities. Students, however, report a much higher likelihood of their parents approving of their choices if they study either science or economics (87% each), compared to social sciences (72%) or the humanities (66%).

The increased effort required to study science and to get high grades in science subjects is traded off against the higher predicted wages and probability of getting a job after graduating in science. The highest average salaries are predicted for economics majors, which is 4% higher than the expected salary for science majors, 38% higher than the expected salary after a social sciences major and 55% higher than the expected salary after a humanities major. Economics majors are also believed to be more likely to get a job immediately after graduation (72%), compared to science (65%), social science (57%), and humanities majors (50%). At the same time, students do anticipate that jobs following science and economics majors are likely to be more intensive and less flexible, reflecting in the lower probability of being able to enjoy a work-life balance, and they are also less likely to be gender-balanced, compared to jobs after majoring in the social sciences or the humanities.

The male-female gender gap in expectations by major are also indicated in alternate columns of table 3, and some of these differences are significantly different from zero. In terms of major characteristics, there is a sharp gender divide across science and economics on the one hand, and social science and the humanities on the other. Men predict higher grades for themselves relative to women in science and economics while women predict higher grades for themselves than men in social sciences and the humanities. Men are much more likely to believe that their parents will disapprove of their choices if they study the social sciences or the humanities compared to women. Women are much more likely to believe they will enjoy studying the social sciences and the humanities compared to men, while men are more likely to believe they will enjoy studying science and economics.

For work characteristics, women predict much lower salaries for themselves than men across science and economics, but they do not predict significantly lower probabilities of getting a job immediately after graduation. Men are significantly more likely to expect that they will get a job in a preferred sector if they study science or economics while women are significantly more likely to be employed in their preferred sector if they study social sciences or the humanities.

4 Econometric models

4.1 Earnings premiums by major

The data on high-income earning individuals provides an unusually good setting to examine evidence on earnings differentials across different majors. All the wage-earners were admitted to the same highly selective graduate programme even while there is considerable variation across their choice of undergraduate major, and we have access to many determinants of earnings that are typically unobserved in a given population. We estimate variants of the following model:

$$y_{ikt} = \beta_1 M_{ik} + \beta_2 X_{it} + \delta_t + \epsilon_{ikt} \tag{1}$$

where y_{ikt} represents earnings y of individual i, who majored in k in year t, M_{ik} indicates whether the individual majored in STEM, economics and business or in arts and the humanities, X_{it} includes individual-level characteristics correlated with earnings and δ_t are year fixed effects. We consider three outcomes: earnings in a job, whether a person is employed at the time of the survey, and average hours of work per week spent on the current job.

There are several potentially confounding factors that we control for. First, cognitive and non-cognitive abilities are widely understood to affect educational and labor market outcomes (Heckman et al., 2006), and may vary by gender across our sample. We use standardised scores in grade 10 as a measure of cognitive ability. Admission to the programme is based on an admissions score, comprised of a measure of cognitive ability – demonstrated by academic performance – as well as psychosocial skills or "soft skills" – measured through an interview process. We control for this admissions score as well.

Second, as discussed earlier, women are often comparatively stronger in verbal and language skills, which are more closely associated with the humanities and social sciences (Halpern et al., 2007), and could explain why women are more likely to enrol in the humanities. In our administrative data, students report the three highest subject test scores in grade 10. We use an indicator of whether English in included in the top three as a measure of comparative ability in verbal skills, and an indicator of whether mathematics is included in the top three as a measure of comparative ability in mathematical skills.

Third, as discussed earlier, personality characteristics such as self-esteem, attitudes towards risk, and degree of competitiveness have been found to be correlated with gender, major choice and earnings (Heckman et al., 2006; Buser et al., 2014; Niederle and Vesterlund, 2007; Dohmen et al., 2010). We use our survey measures to control for all three traits.

Fourth, socioeconomic background is likely to influence entry into different majors (Hastings et al., 2015). We control for whether the mother completed an undergraduate degree, and whether the student received a need-based financial scholarship.

4.2 Empirical model of subject choice

Following Zafar (2013), we use a standard random utility model of discrete choice, where students face a choice across M different majors. Student i derives utility U_{ikt} from their choice of major k at time t, which depends on major-related characteristics (M_{ikt}) – such as a student's relative ability to perform well on a particular major, measured by their expected grades and the hours of effort they will have to exert on the major, their taste for the subject, and whether their parents will approve of their choices; characteristics of the jobs they anticipate they will get on graduating from that major (J_{ikt}) – probability of getting a job, wages in that job, probability of getting a job in a preferred sector, and characteristics of that job; and individual characteristics that are fixed across jobs and majors such as gender, personality attributes such as self-esteem, attitudes towards risk and competition, socio-economic background (X_{it}).

At time t, major and job-related outcomes M_{ikt} and J_{ikt} are uncertain. Individual i will have subjective beliefs $\{P_{ikt}(M_{ikt}), P_{ikt}(J_{ikt})\}$ over all outcomes associated with k different majors and will select major m such that $m \equiv \arg\max_k \int U_{it}(\mathbf{M_{ikt}}, \mathbf{J_{ikt}}, X_{it}) dP_{ikt}(\mathbf{M_{ikt}}, \mathbf{J_{ikt}})$. Following Zafar (2013), we assume an additively separable utility function in binary and continuous outcomes $\mathbf{b_{ikt}}$ and $\mathbf{d_{ikt}}$ respectively:

$$m \equiv \arg\max_k (U_{it}(\mathbf{b_{ikt}}, \mathbf{d_{ikt}}, \mathbf{X_{it}})) \equiv \arg\max_k \sum_r u_r(b_r, \mathbf{X_{it}}) + \sum_q \gamma_q(\mathbf{X_{it}}) d_q + \epsilon_{ikt} \quad (2)$$

An individual i will choose major m with probability:

$$Pr(m \mid X_{it}, P_{ikt}(b_r), P_{ikt}(d_q)) =$$
$$Pr\left(\begin{array}{l} \sum_r \int u_r(b_r, X_{it}) dP_{imt}(b_r) + \sum_q \gamma_q(X_{it}) \int d_q dP_{imt}(d_q) + \epsilon_{imt} \\ \geq \sum_r \int u_r(b_r, X_{it}) dP_{ikt}(b_r) + \sum_q \gamma_q(X_{it}) \int d_q dP_{ikt}(d_q) + \epsilon_{ikt} \end{array}\right)$$
$$\forall m \in C_i, m \neq k \quad (3)$$

For binary outcomes, $\int u_r(b_r, X_{it}) dP_{imt}(b_r) = P_{imt}(b_r = 1)\Delta u_r$ where $\Delta u_r(\mathbf{X_{it}}) = u_r(b_r = 1, \mathbf{X_{it}}) - u_r(b_r = 0, \mathbf{X_{it}})$, the difference in utility between the binary outcome taking place and not. Probabilities $P_{imt}(b_r = 1)$ are directly elicited from the respondent. For continuous outcomes, $\int d_q dP_{imt}(d_q) = E_{ikt}(d_q)$, where $E_{ikt}(d_q)$ are expected outcomes directly elicited from the respondent. Then, the probability of choosing major m is given by:

$$Pr(m \mid X_{it}, P_{ikt}(b_r), P_{ikt}(d_q)) =$$
$$Pr\left(\begin{array}{l} \sum_r P_{imt}(b_r = 1)\Delta u_r + \sum_q \gamma_q(X_{it}) E_{imt}(d_q) + \epsilon_{imt} \\ \geq \sum_r P_{ikt}(b_r = 1)\Delta u_r + \sum_q \gamma_q(X_{it}) E_{ikt}(d_q) + \epsilon_{ikt}) \end{array}\right)$$
$$\forall m \in C_i, m \neq k \quad (4)$$

A standard assumption is that ϵ_{ikt}s are independent across individuals i and major choices k, with a Type 1 extreme value distribution, implying that the model can be estimated as a conditional logit model (Luce, McFadden 1974) which includes both alternative-specific and case-specific variables. The probability that the student will choose major m over k is

$$Pr(m \succ k) = \frac{exp(\beta x_{im})}{\sum_k exp(\beta x_{ik})} \quad (5)$$

In addition, survey respondents are also asked to rank all available majors in their choice set, and the probability of observing any given ranking of choices a, b, c, d can be estimated by an exploded logit model (Beggs et al., 1981):

$$Pr(a \succ b \succ c \succ d) = \frac{exp(\beta x_{ia})}{\sum_M exp(\beta x_{ij})} \cdot \frac{exp(\beta x_{ib})}{\sum_M exp(\beta x_{ij})} \cdot \frac{exp(\beta x_{ic})}{\sum_M exp(\beta x_{ij})} \tag{6}$$

Surveyed students were distributed across all three years of the programme with 51% from the first year, 21% in the second year and 26% in the third year. Students select their major by the end of their third semester, in the middle of their second year, which means that of the sample, only the third year students have already declared a major while the others are yet to do so. We estimate the model using data on all students combined as well as by using data on only first and second year students.

5 Results

5.1 Earnings premiums across majors

The results of the earnings premium model are presented in table 4. The results indicate that while there is no significant difference in earning levels across STEM and economics/business majors, there is a substantial wage penalty associated with graduating in the arts and the humanities of 20-30% relative to STEM graduates. Comparing columns 1 and 2, we see that this earnings premium for science accounts for one-fifth of the gender earnings gap. Moreover, these earnings differences persist even after including controls for cognitive ability, math and language skills, as well as our personality measures such as level of risk aversion, competitiveness and self-esteem. The wage penalty to graduates in the arts and to women in general only becomes negligibly different from 0 when we include controls for seven occupational categories, which suggests that the wage penalty to women operates through the occupational choice channel, which in turn determines the choice of major. Notably, the earnings penalty for women does decline by a third with the inclusion of different controls but a sizeable gap persists across all specifications.

One reason for the differential enrolment of women into the humanities and social science degrees could be that the returns to different majors vary by gender: if women face less sharp trade-offs between the different majors compared to men, they will be less likely to flock to the relatively higher paying majors. The differential returns to women from a degree are estimated and the results presented in table 5. The most striking finding is that the gender earnings penalty relative to men is smallest for science: the estimated coefficients on FemXScience are 9-13% in columns 1-4, and are insignificantly different from zero. Conversely, the gender earnings penalty relative to men is highest for majors of economics and business, with large coefficients of 32-50%. This is similar to evidence from the US which shows that the gender wage gap is smallest in technical fields and the largest in business (Goldin, 2014).

In the appendix, we implement quantile regressions at different points of the earnings distribution to examine whether the STEM earnings premium varies with income, and we find that the gender penalty is the smallest for the lowest earning quintile in this sample, and increases with higher quintiles.

One concern could be that the jobs that were taken up after completing the graduate programme could differentially affect returns to different majors. To rule out this possibility, in the appendix A, we estimate these returns separately for jobs started before graduate school and find that the direction of the effects remains the same.

We also consider other labour market outcomes such as the probability of employment and hours spent at work per week. Table 6 shows the association between the likelihood of employment at the time of the survey and a respondent's undergraduate major. Arts and social science majors are significantly less likely to find themselves in full-time employment relative to science and economics majors. In this case, non-employment includes searching for a job while unemployed, being unemployed without searching for a job and working part-time or doing voluntary work.

Table 7 presents estimates of the association between hours spent at work per week at the time of the survey and a respondent's undergraduate major. Jobs following degrees in science are associated with significantly longer hours compared to economics and the humanities and social sciences, between which the difference is insignificant. After controlling for major and a range of other characteristics, men are estimated to work approximately three hours per week more than women but the difference is not significantly different from zero. The lack of variation by gender is not surprising in the context of this sample. Over 90% of the sample is not married and none of the respondents have children. Bertrand et al. (2010) find that labour supply of women starts to decline relative to men only after they start families and have children.

5.2 Major choice

In this section, we discuss the results from the estimation of the major choice models, using the sample of students who are in the process of selecting, or have recently selected, a major. Table 8 shows the estimated coefficients of a conditional logit model, using only the information on the stated major preferred by each survey respondent. For binary events, the estimated coefficients are the Δus described above, while for continuous events, the estimated coefficients are the γs. The first three columns contain estimates from the model that include all surveyed students, including those who have already selected into their chosen major, while the next three columns are restricted to the students who are yet to declare a major. In both sets of estimations, the largest, positive coefficients are on enjoying coursework. Other major-related characteristics such as getting an above median GPA, and having parents approve of the student's major choice are also positive and large, but substantially smaller than enjoying coursework. The effort required to do well on the course does not appear to matter at all for major choice. Since drop-out rates are so low at this university, we are not concerned that students are selecting majors to avoid dropping out. Among job-related characteristics, getting a job is not significantly important but getting a job in a desired sector has a large, positive and significant coefficient. Work-life balance and wages are also positive and significant, but much smaller than the other characteristics. Getting a job in a gender-balanced firm actually enters the model negatively, with students more likely to select a major associated with less-gender balanced work-spaces, even when holding other factors like salary and work-life balance constant.

The preferences between men and women are notably different. While all students place the highest importance on enjoying coursework, men seem to care much more about parental approval, while women relatively care more about getting higher grades. Since all students believe that parents are more likely to approve of their studying science or economics, the greater weight placed by men on parental approval would predict the relatively greater selection of men into subjects like science and economics, which parents are more likely to approve of. Among job characteristics, men care more about getting a job in a desired sector, where the coefficient is almost twice as large as that for women, and they also care much more about the salary they expect to earn on graduating from a given major, the coefficient on which is 80% larger than the coefficient for women.

The coefficients on job characteristics that are believed to be important for women – a gender-balanced workplace and the probability enjoying work-life balance – do not appear to matter for major selection. However, given the importance of employment in a preferred sector, there could be other characteristics of the workplace which are both important in driving major choice and vary by gender, a question we turn to in section 6.

The ranking of different factors by their importance in major selection remains the same whether we consider all surveyed students, or just consider students who are yet to declare a major. The coefficient on expected grades is higher for the younger students while the coefficient on getting a job in a desired sector is higher for the full sample of students, suggesting some subtle ways in which students who have already selected a major reflect on the reasons behind their choices. Even with these differences, the gender differences are also largely the same across the full sample of students and those who are yet to declare a major.

Table 9 shows estimated coefficients of the rank order logit model using the information on the full ranking of major sets from 1 to 4. The ranking of the different factors remains the same as in the earlier model: the largest coefficient is on enjoying coursework, followed by expected grades, and then parental approval. Again, job characteristics are less important than major characteristics, and the coefficient on getting a job in a preferred sector is more than three times as large as the coefficient on earnings.

In terms of gender differences, in this model too, women care more about grades than men; in fact the coefficient on expected grades is as large now as on enjoying coursework. Similar to the earlier case, women care relatively more about getting a job in a desired sector. For earnings, the coefficient for men is positive, significant, and twice as large as the coefficient for women, which is not significantly different from zero. Since expected wages for both men and women are higher after graduating from science and economics majors, the higher weight on wages for men predicts that men are more likely to select into these relatively high-paying majors.

For the ranked preference data model as well, the ranking of different factors by their importance in major selection remains the same whether we consider all surveyed students, or just consider students who are yet to declare a major. One notable difference is the decline in the size of the coefficient on log earnings, reflecting the decreased importance of earnings in the full ranking of the choice set.

5.3 Willingness to pay estimates

To make sense of the estimates of the models of subject choice in Tables 8 and 9, we carry out willingness to pay calculations. Given the assumed linear and additively separable utility function, we calculate willingness to pay in the following way: consider event j of getting a job in a preferred sector and consider the possibility of increasing the probability of this event taking place from P_{ij} to P'_{ij}. In exchange for the increased probability of getting a job in a preferred sector, and in order for utility to stay the same, the individual should be willing to give up some part of their expected earnings Y. This willingness to pay, WTP can be calculated from the following condition:

$$P_{ij}\Delta u + \gamma lnY = P'_{ij}\Delta u + \gamma ln(Y + WTP) \quad (7)$$

Solving for WTP:

$$WTP = \left(exp\frac{(P_{ij} - P'_{ij})\Delta u}{\gamma} - 1 \right) \times Y \tag{8}$$

Using the estimates from columns 1-3 of Table 8, we calculate the average willingness to pay using the estimated coefficients in the first three columns of table 8. These calculations are presented in table 10 as a percentage of average earnings (columns 1 to 3) and in absolute rupee terms (columns 4-6), computed at the average earnings for the sample. The average willingness to pay for the increased probability of better non-pecuniary outcomes such as the increased probability of enjoying a course or getting high grades is high. Students are willing to forgo up to 6% of their average earnings for a percentage point increase in the probability that they enjoy their coursework. Similarly they are willing to pay up to 4% of their average earnings for a percentage point increase in the probability of getting above median grades and of their parents approving of their choice of major. Working in the sector of their choice is also important: students are willing to reduce their income at graduation by 4% to get a job in their preferred sector.

There are differences in the willingness to pay by gender. Women are willing to forgo 8.1% of their income on graduation for a percentage point increase in the probability that they enjoy their coursework and 5.1% of their income for a percentage point increase in the probability that they get an above-median grade. These shares are twice as large as the corresponding shares for men (4.3% and 2.3% respectively). Men are willing to give up a somewhat higher share of their salary to increase the probability that their parents approve of their choice and that they get a job in their desired sector (4.2% and 3.7% respectively) than women (3.8% and 3.5%). In absolute terms, women are willing to give up 50% more than men for greater course enjoyment (Rs 44,000 to Rs 29,000), and almost twice as much as men for higher grades (Rs 28,000 to Rs 15,000).

The fact that women are willing to pay much more than men in both proportional and absolute terms for non-pecuniary characteristics, such as course enjoyment, predicts their selection out of high paying jobs and majors associated with them, relative to men. We have already estimated a large and positive earnings premium associated with jobs that open up to holders of degrees in science and economics. From our survey data on undergraduate students, we find that both male and female students expect lower grades in majors in science and economics. Moreover, female students also report beliefs that they will enjoy science and economics which are lower than men. The elicited beliefs and estimated preferences together support the relative selection of women out of science and economics compared to men.

5.4 Preferences and ability

A common reason that is often cited to explain differential gender enrolment into science and economics is ability, and particularly, ability in mathematical and technical subjects. In the earnings regressions described in section 4, we find that while ability is correlated with earnings, including measures of high-school test scores in the regressions does not affect the estimated earnings premium associated with science and the humanities or the male earnings gap. For the undergraduate sample, women have significantly *higher* high-school test scores than men. Women's high-school maths scores are also higher than men's, though this different is not statistically significant.

Could the gender differences in preferences be driven by high ability students? Given that gender-based stereotyping particularly around women's supposedly weak mathematical abilities starts young (Hyde, 2014), it might be the case that women who do well in mathematics despite these stereotypes are more similar to high-ability men, while gender differences between relatively low ability men and women are more pronounced. Although there are no systematic differences in grade 10 mathematics scores across men and women in our sample, the male-female difference in preferences in the top half of the grade distribution may be different from the same difference in the bottom half of the grade distribution. Table 11 shows estimated coefficients of a rank order logit model separately estimated for men and women with above and below median high school math test scores. There is some evidence to suggest the gender gap in preferences is more acute among low-ability students. In particular, low-ability women in terms of maths scores care much more about getting a job at all, rather than employment in a preferred sector, both compared to all men and compared to high-ability women. Low-ability women also care less about parental approval in major choice and more about expected grades compared to low-ability men. This suggests that the gender differences we document in tables 8 and 9 are relatively muted compared to the gender differences in the bottom half of the ability distribution.

6 Discussion

6.1 Decomposition and counterfactual analysis

The analysis of the previous section identifies differences in the preferences of men and women, while table 3 indicates significant differences in the beliefs of male and female students about major-specific outcomes. To identify the relative contributions of differences in preferences and differences in expected outcomes in explaining gendered enrolment into different majors, we implement the Fairlie (2005) decomposition method. This method decomposes differences in the probability of observing the selection of a given major across men and women into that part explained by differences in expectations and differences in preference parameters. Further details are provided in the appendix.

The results are in Table 12. Differences in expectations of outcomes conditional on majoring in a given subject can explain approximately half the gender gap in enrolment in science and the social sciences, and almost three-quarters of the gap in enrolment in economics and the humanities. Among these belief distributions, it is the beliefs about characteristics that relate to the major, as opposed to the jobs that follow after graduation, that are the most important for explaining the gender gap in enrolment. The most important variation in beliefs that explains differential gender enrolment is the probability that a student will enjoy a course, which explains between 30%-60% of the differential gender enrolment into a subject. Job characteristics are relatively unimportant in explaining differences in major enrolment. In particular, future incomes and the probability of finding a job are relatively less important in explaining gender differences, and only the probability of getting a job in a desired sector has any explanatory power at all, explaining 20-35% of the gender difference in enrolment in the sciences, economics, and the social sciences, but not in the humanities.

Similar results are seen in table 13 which presents counterfactual estimates of the change in the gender gap in enrolment predicted by the model under different scenarios. Column 1 presents the estimated gender gap in enrolment predicted by the model estimated in column 1 of Table 9. Columns 2 to 6 present the estimated gender gaps in enrolment if the female distribution of beliefs is substituted with the male distribution of beliefs about different outcomes. We order the male and female observations by the predicted probabilities of selecting a given major estimated from the base model, and then match the observations pairwise in order to replace the female beliefs with the male beliefs. Since there are more female observations than male observations, the results in columns 2-6 are estimated on a smaller sample of observations.

For enrolments in science, the gender gap declines by over 50% if the distributions for expected enjoyment of coursework are equalised, but the other variables are of limited importance. The probability of getting a job in a desired sector does reduce the gender enrolment gap as well but by less than 10%. Equalising the distribution of beliefs about grades, however, has a large impact on the gender enrolment gap in economics and the social sciences, reducing the gap by 22% and 31% respectively. Parental approval has a small impact on the gender gap in social sciences and the humanities, through the fact that men expect to have less parental support than women if they study these choices, so equalising these beliefs lead fewer women to select into these subjects. Among work characteristics, most strikingly, equalising expected income does not reduce the gender gap across any of the categories of majors. Equalising the beliefs about getting a desirable job does reduce the gender gap

The decomposition results contrast with similar studies in the US, where differences in preferences explain a larger share of gendered enrolment, compared to differences in beliefs. This distinction is important, because it suggests that differences in the preferences of men and women are comparatively less important than the beliefs students have about whether or not they think they will be able to enjoy coursework in a particular discipline. This has important policy implications, since it might be easier to shift beliefs about future outcomes than it is to change underlying preferences, particularly with respect to expected grades. However, the most important factor driving enrolment is whether students think they will enjoy studying a particular subject, which is likely to be deeply rooted in prevailing social norms. We discuss these differences in expectations in the next section.

6.2 Understanding differences in expectations

Since differences in expectations explain more than half of the gender difference in enrolments, it is important to understand where these differences emerge from and if they vary by individual characteristics. We consider expectations of four important outcomes: course enjoyment, grades, future earnings, and the probability of getting a job in their desired sector.

Across all models, the most important factor influencing major choice is how much students expect to enjoy the associated coursework. There are significant gender differences in the level of these expectations (table 3). The male-female gap in expectations about course enjoyment in science is significant and positive; that is, women believe they are less likely to enjoy studying science compared to men. With economics, the difference is also positive but insignificantly different from zero. For social science and humanities, the male-female gap in expected enjoyment is negative: women believe they are more likely to enjoy studying these subjects compared to men. There is some evidence from social psychology that suggests that there exist gender differences in interests in engineering, science and mathematics (Su et al., 2009). However, these differences are not hardwired in biology but are themselves socially constructed, and influenced by prevailing gender stereotypes, discrimination, and other cultural and social constraints to women entering fields such as science and economics (Hyde, 2014; Bertrand, 2020). Differences in enjoyment, could also be correlated with expected grades and preferred sector of employment: if women believe they are less likely to do well on a course or get a desired job, they may be less likely to believe they will enjoy studying that course. However, the fact that all three factors are significant in all of the estimated models suggests that the enjoyment question is picking up something different from grades and occupational sorting. The correlation coefficient between expected enjoyment and grades is 0.76 for women and 0.6 for men, and the coefficient between enjoyment and preferred sector is 0.53 for women and 0.42 for men, all of which are well below 1. (The full set of correlations is presented in table A3 in the appendix).

There are substantial differences in the way men and women predict their grades. Table 14 presents average expected grades for students by gender and by major and compares these to the observed grades of older cohorts of students who graduated between 2017 and 2020. Men significantly overestimate their grades across all majors, with the gap between average expected grade and average observed grade the largest for economics. Women, too, overestimate their grades in the social sciences and the humanities. However, for science, women underestimate their grades, and for economics, the difference is insignificantly different from zero.

With respect to earnings, women predict lower earnings for themselves compared to men across all majors. The size of the expected earnings gap is large in the case of science and economics: women expect 30% lower earnings in science and 23% lower earnings in economics, compared to men. The expected gender gap is positive in favour of men but not significant in social science and the humanities. However, as our earnings regressions suggest (column 1 of table 5), significant gender gaps in earnings can be found within both economics and the humanities, with gender discounts of 32% and 43% respectively, but *not* within science. This implies that women are relatively pessimistic with respect to their earnings within science and relatively optimistic with respect to their earnings within social science and the humanities, compared to men. To the extent that expected earnings are driving major choice, this would predict their differential selection out of science and into social science and the humanities.

Finally, with respect to getting a job in a desired sector, men predict an 8 percentage higher probability of getting a job in their preferred sector if they study economics, while women predict higher probabilities for social sciences and the humanities respectively. Table 15 shows the distribution of first-ranked occupational sectors by gender and major. Science students tend to select into jobs within computer sciences, natural sciences, and education and research. Economics students overwhelmingly opt for jobs in consulting and finance, while social science and humanities students disproportionately opt for jobs in the social sector, and in the media and arts. The earnings data on occupation indicates a hierarchy of earnings that places the occupations of women-dominated majors in the bottom half of the distribution, while consulting and finance, computer science and the natural sciences are in the top half of the distribution (Figure 2). Strikingly, even conditional on major choice, women are less likely to opt for the highest paying jobs in the consulting and finance sectors.

Why do women prefer these sectors? Our survey does not offer clear answers. The probability of better work-life balance and a gender-balanced work-space are not significant in the estimated models of work choice. Additionally, there are no gender differences in the probability that students see themselves as being married or having children ten years later (Table 2), suggesting that a desire to exit the labour force or work part-time is not driving this selection. Research by sociologists (Charles and Bradley, 2009) suggests that in post-industrial economies, students value self-expression more highly than economic benefits: when it comes to choosing a career, students are keen to find an occupation about which they are passionate and which they think allows them to best express themselves. In the context of persistent gendered stereotypes and aspirations, this could lead to further gender segregation across occupations as women opt out of technical fields to pursue jobs in fields like social service, development and government, even though these are associated with lower pay. Perhaps it is these stereotypes and other unobserved characteristics of the sectors that students see themselves working in that are appealing to the women in our study.

Could individual characteristics be driving these gender differences in expectations? Measures of ability are likely to be correlated with expected beliefs about grades, probability of employment and earnings. Differences in personality traits such as self-esteem, attitudes towards risk and attitudes towards competitiveness could explain why women are less likely to predict high grades or high salaries for themselves. Similarly, socioeconomic background could be correlated with the future expectations of grades or earnings of students. Table 16 presents regressions of the gender differences in expectations controlling for these characteristics. For consistency with the earnings regressions in table 5, we include the same controls as in those regressions. Expectations continue to be significantly correlated with gender. In particular, women predict lower enjoyment and lower grades in science and economics, and higher enjoyment and higher grades in social studies and humanities, compared to men. In terms of job characteristics, women predict lower wages across all majors and lower probabilities of working in their desired sector across all majors except science. However, here again, the estimated size of the expected gender gap in wages is much larger than in the observed data with the same set of controls (column 4 of table 5). In the expectations data, women expect earnings that are between 40-55% lower than men, while in the observed data, a significant earnings difference of 50% is seen only within economics graduates, and the estimated coefficients for science and the humanities is 13-22%, neither of which are significant. Finally, the expected probability of getting a job in a desired sector also varies by gender within economics, social sciences and the humanities, with women believing they are less likely to get a job in their preferred sector if they study economics, and more likely to get a job in their preferred sector if they study the social sciences or the humanities.

In sum, gendered expectations of future outcomes clearly divide across gendered majors: with women predicting relatively improved outcomes when they study social science and the humanities, and relatively worse outcomes when they study science and economics, compared to men.

7 Conclusion

In this paper we study the gender-wise earnings premium associated with studying science and business as compared to studying the humanities and the social sciences in the context of a developing country. Combining rich data on student's subjective expectations and administrative records, we decompose the gender differences in major choice into differences in preferences between men and women, and differences in expected outcomes, particularly pertaining to students' beliefs about their own ability. We believe we are the first such study from a developing country perspective, which provides an important theoretical and policy-relevant contrast to much of the literature that is based on US data.

Using unique survey data, we find significant differences in the return to studying science and economics compared to the humanities and social sciences. In contrast to recent research from Chile (Aguirre et al., 2020), earnings differentials are high for both men and women, and there are in fact, no gender differences in the return to science. We provide evidence that the source of these differences is the entry to relatively higher-paying occupations. Accounting for major selection reduces the male wage premium by as much as a fifth, but the male premium remains strikingly high at 16% even with added controls for a wide range of characteristics, including cognitive and non-cognitive ability, personality traits, and socioeconomic background.

Women, however, despite not facing within-major or within-occupation wage penalties relative to men, disproportionately enrol into the humanities and social sciences. Our estimation of a model of subject choice sheds light on the question of what explains this differential enrolment. We find evidence of both differences in preferences and in beliefs about future outcomes by gender, but unlike previous research from the US, the differences in beliefs are quantitatively much more important. On preferences, we find that women value enjoying a course of study and getting high grades on the course more than men, while men value higher earnings more than women. Working in a desired sector is also important, but job-related characteristics such as work-life balance and gender-balanced work-spaces are not significant determinants of gender disparities in enrolment in our results. Since women's beliefs about their employment in ten years time is no different than men, and given that most women do not foresee themselves as being married or having children in ten years suggests that the desire to exit the workforce or seek flexibility in working hours is not driving the results. However, that women are disproportionately selecting into low-paying occupations even conditional on major suggests that there are other unobserved characteristics of these jobs over which women have relatively strong preferences.

These results are particularly interesting in the context of a developing country where many research and policy efforts have been focused at arresting and reversing a recent decline in female labour force participation. Our results suggest that even in an environment where women are confident that they will be working at the age of 30, the gender gap in major enrolment will remain large and persistent. Rather than simply expanding opportunities for women to work, equalising female representation across different majors and occupations will require transforming workplaces within specific sectors to make them more attractive to women.

On beliefs, we find that women predict a much larger gender earnings penalty than is observed from graduates of the same university, even within major. With respect to grades, while all students struggle to predict their grades accurately, men are far more optimistic about their grades than women, and this gap is particularly large when it comes to science and economics. Recent research by Shi (2018) finds that even in the US context, belief in abilities can explain a significant part of the enrolment of women into STEM fields. Can these beliefs be corrected? In recent research, Baker et al. (2018) and Conlon (2019) find that students are responsive to information they receive about the earnings distribution conditional on major choice both in terms of updating their own expectations as well as changing their choice of major. Whether the provision of grade distribution information can close the gender gap in major choice is an area of future research.

References

Aguirre, J., J. Matta, and A. M. a. Montoya (2020). Joining the Men's Club: The Returns to Pursuing High-earnings Male-dominated Fields for Women. *Working paper*.

Altonji, J., P. Arcidiacono, and A. Maurel (2016). The analysis of field choice in college and graduate school. In *Handbook of the Economics of Education*, pp. 305–396. Elsevier.

Altonji, J. G. (1993, January). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics 11*(1, Part 1), 48–83.

Arcidiacono, P. (2004, July). Ability sorting and the returns to college major. *Journal of Econometrics 121*(1), 343–375.

Arcidiacono, P., V. J. Hotz, and S. Kang (2012, January). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics 166*(1), 3–16.

Arcidiacono, P. and C. Koedel (2014, July). Race and college success: Evidence from missouri. *American Economic Journal: Applied Economics 6*(3), 20–57.

Baker, R., E. Bettinger, B. Jacob, and I. Marinescu (2018). The effect of labor market information on community college students' major choice. *Economics of Education Review 65*, 18–30.

Bank, W. (2012). World Development Report 2012: Gender Equality and Development. Technical report, World Bank.

Beffy, M., D. Fougere, and A. Maurel (2012). Choosing the field of study in postsecondary education: Do expected earnings matter? *Review of Economics and Statistics 94*(1), 334–347.

Beggs, S., S. Cardell, and J. Hausman (1981, September). Assessing the potential demand for electric cars. *Journal of Econometrics 17*(1), 1–19.

Bertrand, M. (2020, May). Gender in the Twenty-First Century. *AEA Papers and Proceedings 110*, 1–24.

Bertrand, M., C. Goldin, and L. F. Katz (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics 2*(3), 228–55.

Blau, F. D. and L. M. Kahn (2000, November). Gender differences in pay. *Journal of Economic Perspectives 14*(4), 75–100.

Bönte, W., S. Lombardo, and D. Urbig (2017). Economics meets psychology: Experimental and self-reported measures of individual competitiveness. *Personality and Individual Differences 116*, 179–185.

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

Charles, M. and K. Bradley (2009). Indulging Our Gendered Selves? Sex Segregation by Field of Study in 44 Countries. *American Journal of Sociology 114*(4), 924–976.

Conlon, J. J. (2019, September). Major Malfunction: A Field Experiment Correcting Undergraduates' Beliefs about Salaries. *Journal of Human Resources forthcoming*.

Delavande, A. (2014). Probabilistic Expectations in Developing Countries. *Annual Review of Economics 6*(1), 1–20.

Delavande, A. and B. Zafar (2019). University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints. *Journal of Political Economy 127*(5), 2343–2393.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review 100*(3), 1238–60.

Fairlie, R. W. (2005). An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement 30*(4), 305–316.

Friedman-Sokuler, N. and M. Justman (2020, February). Gender, culture and STEM: Counter-intuitive patterns in Arab society. *Economics of Education Review 74*, 101947.

Gemici, A. and M. Wiswall (2014). Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors. *International Economic Review 55*(1), 23–56.

Goldin, C. (2014, April). A Grand Gender Convergence: Its Last Chapter. *American Economic Review 104*(4), 1091–1119.

Gray-Little, B., V. S. L. Williams, and T. D. Hancock (1997). An Item Response Theory Analysis of the Rosenberg Self-Esteem Scale:. *Personality and Social Psychology Bulletin 23(5)*, 443–451.

Halpern, D. F., C. P. Benbow, D. C. Geary, R. C. Gur, J. S. Hyde, and M. A. Gernsbacher (2007, August). The Science of Sex Differences in Science and Mathematics. *Psychological science in the Public Interest 8*(1), 1–51.

Hastings, J., C. A. Neilson, and S. D. Zimmerman (2015, June). The effects of earnings disclosure on college enrollment decisions. Working Paper 21300, National Bureau of Economic Research.

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics 24*(3), 411–482.

Hyde, J. S. (2014). Gender Similarities and Differences. *Annual Review of Psychology 65*(1), 373–398.

Jain, T., A. Mukhopadhyay, N. Prakash, and R. Rakesh (2019). Science Education and Labour Market Outcomes in a High STEM Economy. IZA Discussion Paper No. 11908.

Kahn, S. and D. Ginther (2018). Women and Science, Technology, Engineering, and Mathematics (STEM): Are Differences in Education and Careers Due to Stereotypes, Interests, or Family? In *The Oxford Handbook of Women and the Economy*. Oxford University Press.

Manski, C. F. (2004). Measuring expectations. *Econometrica 72*(5), 1329–1376.

Mather, M., E. Shafir, and M. K. Johnson (2000). Misremembrance of options past: Source monitoring and choice. *Psychological Science 11*(2), 132–138.

Matsa, D. A. and A. R. Miller (2011). Chipping away at the glass ceiling: Gender spillovers in corporate leadership. *American Economic Review 101*(3), 635–39.

Ministry of Human Resource Development (2019). All India Survey on Higher Education. Technical report.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Reuben, E., M. Wiswall, and B. Zafar (2017, January). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal 127*(604), 2153–2186.

Roychowdhury, P. (2019, November). (Em)powered by Science? Estimating the Causal Effect of High School Major Choice on Labor Market Earnings in India. SSRN Scholarly Paper ID 3488320, Social Science Research Network, Rochester, NY.

Sahoo, S. and S. Klasen (2018, August). Gender Segregation in Education and its Implications for Labour Market Outcomes: Evidence from India. IZA Discussion Paper No. 11660.

Shi, Y. (2018, June). The puzzle of missing female engineers: Academic preparation, ability beliefs, and preferences. *Economics of Education Review 64*, 129–143.

Smith, N., V. Smith, and M. Verner (2006). Do women in top management affect firm performance? A panel study of 2,500 Danish firms. *International Journal of Productivity and Performance Management 55*(7), 569–593.

Stinebrickner, R. and T. R. Stinebrickner (2008). The causal effect of studying on academic performance. *The B.E. Journal of Economic Analysis & Policy 8*(1).

Su, R., J. Rounds, and P. I. Armstrong (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin 135*(6), 859–884.

Weichselbaumer, D. and R. Winter-Ebmer (2005). A meta-analysis of the international gender wage gap. *Journal of Economic Surveys 19*(3), 479–511.

Weinberger, C. J. (2004). Just ask! Why surveyed women did not pursue IT courses or careers. *IEEE Technology and Society Magazine 23*(2), 28–35.

Wiswall, M. and B. Zafar (2015, April). Determinants of College Major Choice: Identification using an Information Experiment. *The Review of Economic Studies 82*(2), 791–824.

Wiswall, M. and B. Zafar (2017, August). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics 133*(1), 457–507.

Zafar, B. (2012). Double Majors: One for Me, One for the Parents? *Economic Inquiry 50*(2), 287–308.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources 48*(3), 545–595.

A Appendix

A.1 Survey questionnaire

In the survey administered to the undergraduate students, the following questions were asked about each of four major choices: science, economics, social science and the humanities.

1. Suppose you majored in the following academic field. What probability do you place on getting a CGPA equal to or higher than 3.3/4.0. (Please indicate the probability as a percentage). For example, if you think there is a 50% chance you will get a CGPA higher than 3.3/4.0 in the indicated academic field, drag the slider to the 50% mark.

2. Suppose you majored in the following academic field. How many hours per week do you expect you would need to work to get a GPA of 3.3/4.0 in the given academic field? For example, if you believe you will need to spend 1 hour per week to get a GPA of 3.3 in a given academic field, enter 1 next to it.

3. Suppose you majored in the following academic field. What probability do you place on enjoying your academic work during your undergraduate degree? (Please indicate the probability as a percentage). For example, if you think there is a 50% probability that you will enjoy your academic work in the indicated field, drag the slider to the 50% mark.

4. Suppose you majored in the following academic field. What probability do you place on your parents approving of your choice? (Please indicate the probability as a percentage). For example, if you think there is a 50% probability that your parents will approve of your major choice if you were to complete a major in the indicated academic field, drag the slider to the 50% mark.

5. Suppose you majored in the following academic field. Now imagine that you were to look for a job immediately after college. What probability do you place on getting employed right after you complete your undergraduate degree? (Please indicate the probability as a percentage). For example, if you think there is a 50% chance that you will get a job right after you complete your undergraduate degree in the indicated academic field, drag the slider to the 50% mark.

Now, we will be asking you some questions on expected earnings in the future.

6. Please indicate below the units you would like to use to answer these questions. For example, assume you believe you will earn Rs. 1000 per month. You could either choose 'Monthly Salary in Rupees' and fill in "1000" on the next page or choose 'Annual Salary in Rupees' and fill in "12000" on the next page.

7. You have chosen to fill your answers according to Monthly Salary. Please answer the following question accordingly. Supposed you have majored in the following academic fields. What do you think will be your salary in Rupees 10 years after completing your undergraduate degree? For example, assume you believe you will earn Rs 1000 per month after studying this field, please enter "1000" against that particular academic field.

8. You have chosen to fill your answers according to Annual Salary. Please answer the following question accordingly. Suppose you majored in the following academic field. What do you think will be your first salary in Rupees if you work right after completing your undergraduate degree? For example, assume you believe you will earn Rs 12000 per year after studying in this field, and as you chose to report 'Annual Salary in Rupees', and please enter "12000" against that particular academic field.

9. Suppose you majored in the following academic field. What probability do you place on having your desired level of work-life balance? (Please indicate the probability as a percentage). For example, if you think there is a 50% probability that you will have your desired level of work-life balance if you were to complete a major in the indicated academic field, drag the slider to the 50% mark.

10. Suppose you majored in the following academic field. What probability do you place on working in a balanced gender space (equal number of men and women)? (Please indicate the probability as a percentage). For example, if you think there is a 50% probability that you will work in a balanced gender space if you were to complete a major in the indicated academic field, drag the slider to the 50% mark.

11. Which sector would you like to work in? Rank the following sectors in order of your preference where 1 is your most preferred sector and 6 is your least preferred sector.

12. Given your most preferred sector (whatever you ranked first on the previous page), what is the probability that you will be employed in that sector given that you had graduated from each of the following academic fields? For example, if you chose 'Computer Science/IT/Data Analytics' as your first-ranked preference and you think there is a 50% probability that you will be employed in this sector if you completed

a major from the indicated academic field, drag the slider to the 50% mark for that field.

Now, we would like to ask you some questions about what you expect your life to be like 10 years from now.

13. What is the probability (percent chance) that you will be married 10 years from now?
14. What is the probability (percent chance) that you will have children 10 years from now?
15. What is the probability that you will be working 10 years from now?
16. Suppose you majored in the following academic field. What probability do you place on being employed 10 years after you complete your undergraduate degree? (Please indicate the probability as a percentage). For example, if you think there is a 50% chance that you will have a job 10 years after you complete your undergraduate degree in the indicated academic field, drag the slider to the 50% mark.
17. Supposed you have majored in the following academic fields. What do you think will be your salary in Rupees 10 years after completing your undergraduate degree? For example, assume you believe you will earn Rs 12000 per month after studying a specific academic field, and as you chose to report 'Annual Salary in Rupees', and please enter "12000" against that particular academic field.

A.2 Earnings regressions for pre-graduate jobs

One concern with running estimations of (1) is that the completion of the graduate fellowship could differentially affect the returns to different majors. However, the direction in which this bias operates is not immediately clear. For example, the skills transferred through the course could complement the the skills taught in science degree to a greater or lesser extent than through a humanities degree. To rule out this possibility, we estimate (1) on a sample of all jobs taken up before starting on the graduate degree. 198 jobs in the data were held before these students joined the programme. To estimate the return to the major using these jobs alone will eliminate any impact the programme may have had on their earning ability. These results are presented in table A1. The coefficient on humanities is still negative but larger, suggesting that the impact of the programme did reduce wage gaps across majors. The coefficient on arts/humanities is only reversed when occupational fixed effects are included, emphasising again that the earnings benefit of majors emerges from the subsequent access to high-earning occupations.

A.3 Quantile earnings regression

The return to different majors could also vary at different points in the earnings distribution. Table A2 presents the results of quantile regressions at the 10th, 25th, 50th, 75th and 90th percentile of incomes. The earnings premium for science relative to humanities does not vary significantly by quantile, but the earnings premium for economics and business increases substantially relative to both the humanities as well as science with the earnings quantile. This suggests that among high-ability students, there is a positive return to studying economics, and among low-ability students, there is a positive return to studying science. The earnings penalty associated with the humanities and social sciences increases along the earnings distribution.

A.4 Fairlie Decomposition Method

The Fairlie (2005) decomposition method is an extension of the standard Oaxaca-Blinder decomposition applied to non-linear equations. The Fairlie decomposition for a non-linear equation, $Y_j = F(X\hat{\beta})$, where $j = F, M$, is written as

$$\overline{Y_F} - \overline{Y_M} = \left[\sum_{i=1}^{N_F} \frac{F(X_{iF}\hat{\beta_F})}{N_F} + \sum_{i=1}^{N_M} \frac{F(X_{iM}\hat{\beta_F})}{N_M}\right] + \left[\sum_{i=1}^{N_M} \frac{F(X_{iM})\hat{\beta_F})}{N_M} + \sum_{i=1}^{N_M} \frac{F(X_{iF}\hat{\beta_M})}{N_M}\right] \quad (9)$$

where N_F and N_M are the total number of female and male observations respectively. The first term in the square brackets is the contribution to the gender gap in enrolment of the differences in the distribution of the beliefs, X, by gender. The second term in the square brackets is the contribution to the gender gap in enrolment of the differences in the preferences, β, by gender. To identify the contribution of a specific variable, X_i, to the gender gap in enrolment, first, assume for simplicity that $F(X)$ is a function of two variables, X_1 and X_2. Assume further that $N_F = N_M = N$ and that there exists a natural mapping of female to male observations. Then, the independent contribution of X_1 to the gender gap in enrolment is given by:

$$\frac{1}{N}\sum_{i=1}^{N} F(X_{1Fi}bet\hat{a}_1F + X_{2Fi}\hat{\beta_2}F) - F(X_{1Mi}bet\hat{a}_1F + X_{2Fi}\hat{\beta_2}F) \quad (10)$$

Similarly, the contribution of X_2 is given by:

$$\frac{1}{N}\sum_{i=1}^{N} F(X_{2Mi}\hat{\beta_1 F} + X_{2Fi}\hat{\beta_2 F}) - F(X_{1Mi}\hat{\beta_1 F} + X_{2Mi}\hat{\beta_2 F}) \tag{11}$$

Thus, the contribution of any variable, X_i, to the gender gap is given by the difference in the average predicted probability from replacing the female distribution of X_i with the male distribution, while holding the distribution of the other variable constant.

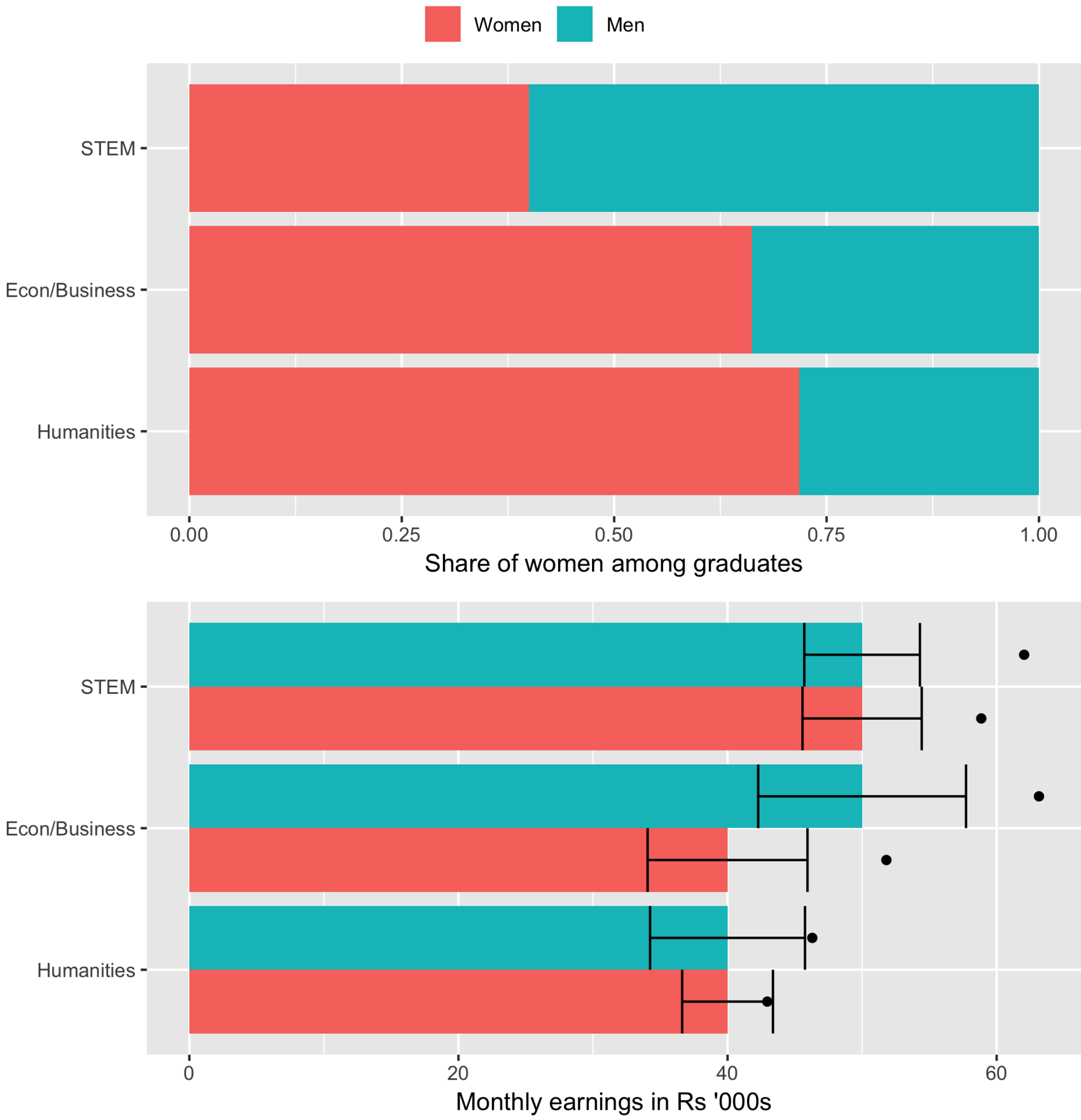
As Fairlie points out, this decomposition is sensitive to the ordering of the variables, so we randomise the order of the variables in each replication and, therefore, approximate average results over all possible orderings.

In practice, the sample size of male and female observations in our data is not equal. To address this, we randomly draw from the female sample, the same number of observations as in the male sample, and order the observations within each sample by predicted probability so as to match males and females with the same ranking of probability. We draw 100 such samples such that the mean value of estimates from the separate decompositions is used to approximate the results for the entire female sample. Standard errors are calculated as in Fairlie (2005), using the Delta method.

A.5 Correlations between expectations of outcomes

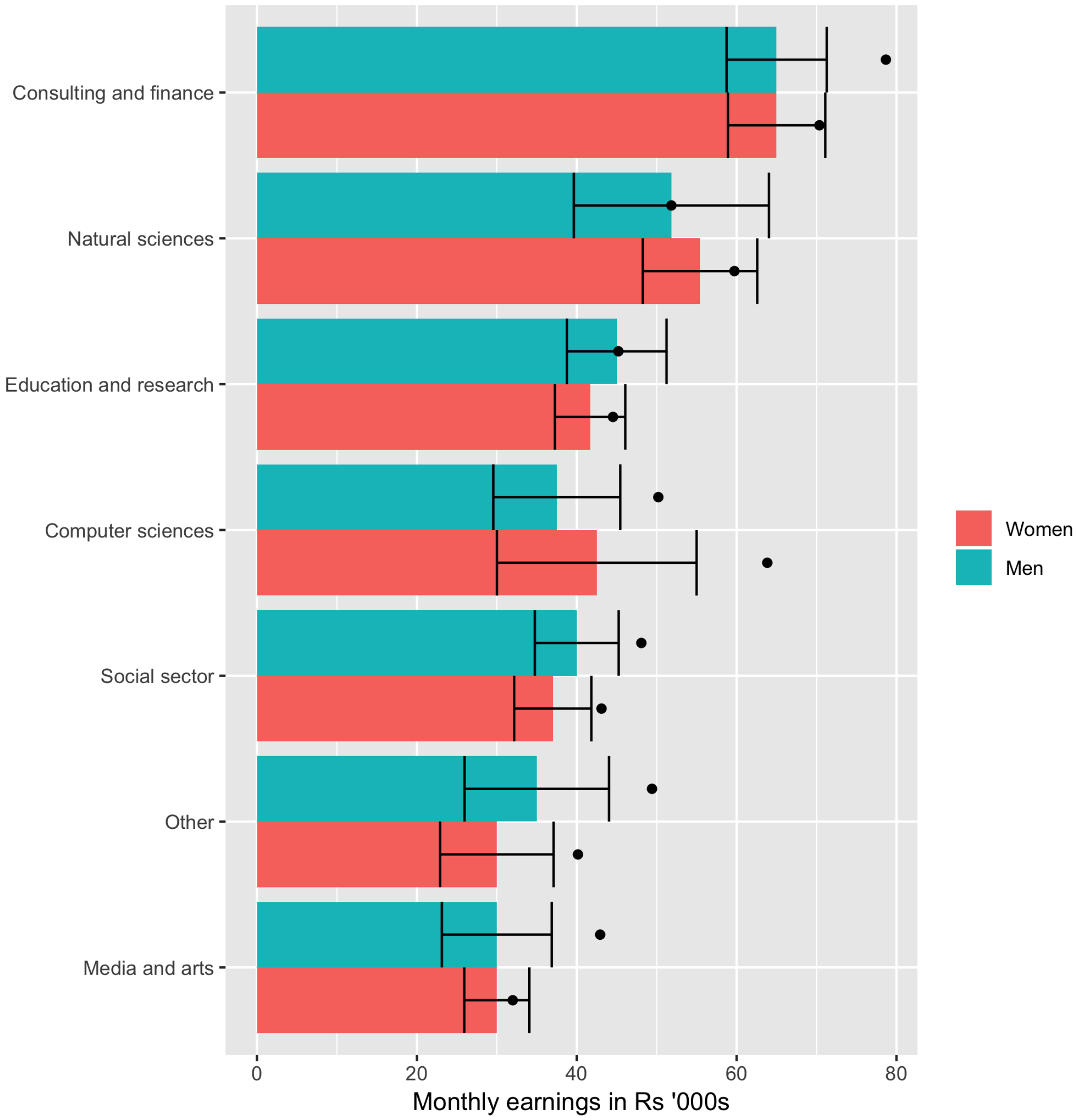
To establish whether the elicited subjective beliefs are highly correlated with one another, table A3 shows pairwise correlations (Pearson coefficient) between all attributes over which subjective expectations are elicited. The correlation coefficients are all well below 1. For enjoying a course and getting high grades, the coefficient is 0.6 for men and 0.76 for women.

Figure 1: Earnings gradient by gender and major



The dots represent mean earnings while the bars represent median earnings. 95% confidence intervals for the median earnings are also indicated.

Figure 2: Earnings gradient by gender and occupation



The dots represent mean earnings while the bars represent median earnings. 95% confidence intervals for the median earnings are also indicated.

Table 1: Sample statistics – High-income earners

	All	Men	Women	p-value
Job characteristics				
Salary (Annual, Rs.)	676,706	741,118	612,485	0.002
	(527,965)	(580,571)	(461,677)	
Work experience	5.82	6.14	5.52	0.000
	(2.37)	(2.41)	(2.30)	
Currently employed (%)	0.70	0.75	0.66	0.064
Hours per week	47.2	49.9	44.6	0.000
	(16.0)	(15.4)	(16.1)	
Ability measures				
Grade 10 Scores	9.00	8.91	9.08	0.010
	(0.58)	(0.68)	(0.46)	
Admissions score	5.55	5.57	5.54	0.799
	(1.25)	(1.37)	(1.15)	
English fluency (%)	0.82	0.79	0.84	0.268
Maths fluency (%)	0.72	0.74	0.71	0.609
Personality measures				
Risk aversion	5.21	5.10	5.30	0.523
	(2.83)	(2.77)	(2.87)	
Competition aversion	-2.55	-2.65	-2.46	0.020
	(0.73)	(0.74)	(0.72)	
Esteem	26.87	27.44	26.38	0.044
	(4.75)	(4.82)	(4.65)	
Mother completed UG (%)	0.84	0.77	0.91	0.000
Mother works (%)	0.53	0.49	0.57	0.142
Received scholarship (%)	0.63	0.66	0.61	0.388
Undergraduate major				
Science (%)	0.52	0.68	0.39	0.000
Economics (%)	0.22	0.16	0.27	0.020
Humanities (%)	0.26	0.16	0.35	0.000
Observations	326	150	176	

Standard deviations in parentheses for continuous variables. P-value reported for a pairwise t test (Chi square test) in equality of means or proportions across male and female outcomes.

Table 2: Sample statistics – Undergraduate students

	All	Men	Women	p-value
Ability				
Grade 10 score	91.49	90.74	92.09	0.087
	(7.27)	(9.54)	(4.64)	
Grade 10 english score	91.08	89.67	92.21	0.003
	(7.85)	(9.94)	(5.43)	
Grade 10 math score	90.72	90.30	91.07	0.453
	(9.22)	(11.52)	(6.75)	
Personality measures				
Risk aversion	1.73	1.80	1.66	0.130
	(0.84)	(0.86)	(0.83)	
Competition aversion	3.03	3.03	3.03	0.986
	(0.75)	(0.69)	(0.79)	
Self-esteem	15.11	15.09	15.13	0.886
	(2.37)	(2.43)	(2.31)	
Expected outcomes in 10 years (%)				
Employed	0.92	0.90	0.93	0.046
	(0.14)	(0.15)	(0.13)	
Married	0.57	0.56	0.57	0.908
	(0.29)	(0.29)	(0.29)	
Has children	0.33	0.35	0.31	0.266
	(0.27)	(0.27)	(0.27)	
Family background				
Father has UG	0.43	0.40	0.46	0.283
Father studied science	0.43	0.42	0.43	0.978
Mother has UG	0.45	0.39	0.50	0.049
Mother studied science	0.35	0.32	0.37	0.365
Mother works	0.55	0.48	0.60	0.025
Sibling studied science	0.11	0.11	0.10	0.640
Received scholarship	0.34	0.42	0.28	0.009
Undergraduate major (preferred)				
Science (%)	0.20	0.25	0.15	0.016
Economics (%)	0.39	0.49	0.31	0.000
Social science	0.30	0.20	0.37	0.000
Humanities (%)	0.12	0.06	0.17	0.001
Observations	351	158	193	

Standard deviations in parentheses for continuous variables. P-value reported for a pairwise t test (Chi square test) in equality of means or proportions across male and female outcomes.

Table 3: Major-specific expectations

	Science		Econ		Soc. Sci.		Hum	
	All	Men-Women	All	Men-Women	All	Men-Women	All	Men-Women
Major characteristics								
GPA ≥ B	0.55	0.045	0.64	0.092	0.74	-0.067	0.70	-0.056
	(0.26)	(0.102)	(0.22)	(0.000)	(0.20)	(0.001)	(0.22)	(0.016)
Hours studying	19.48	-2.908	16.75	-2.245	14.25	-0.503	14.67	0.008
	(18.06)	(0.134)	(16.44)	(0.204)	(12.42)	(0.706)	(13.06)	(0.996)
Parents approve	0.87	0.014	0.87	-0.006	0.72	-0.070	0.66	-0.058
	(0.20)	(0.507)	(0.20)	(0.773)	(0.27)	(0.016)	(0.30)	(0.070)
Enjoys coursework	0.51	0.095	0.58	0.141	0.71	-0.111	0.67	-0.088
	(0.29)	(0.002)	(0.27)	(0.000)	(0.22)	(0.000)	(0.24)	(0.000)
Job Characteristics								
Salary at 20	697.03	211.344	725.60	168.617	525.46	93.557	468.17	55.402
	(711.21)	(0.005)	(726.56)	(0.030)	(556.73)	(0.117)	(642.60)	(0.422)
Salary at 30	2,729.62	254.811	3,145.96	853.889	2,011.15	-215.028	1,843.66	238.863
	(3,578.30)	(0.508)	(5,854.33)	(0.175)	(2,746.05)	(0.467)	(2,886.35)	(0.442)
Employed at 20	0.65	0.015	0.72	0.014	0.57	-0.034	0.50	-0.035
	(0.25)	(0.583)	(0.21)	(0.541)	(0.23)	(0.163)	(0.22)	(0.144)
Employed at 30	0.79	-0.025	0.83	-0.001	0.77	-0.088	0.71	-0.084
	(0.25)	(0.353)	(0.22)	(0.963)	(0.22)	(0.000)	(0.23)	(0.001)
Job in desired sector	0.61	0.047	0.67	0.079	0.67	-0.116	0.59	-0.059
	(0.27)	(0.105)	(0.24)	(0.002)	(0.24)	(0.000)	(0.25)	(0.024)
Enjoy work-life balance	0.51	0.005	0.54	0.038	0.69	-0.014	0.68	-0.008
	(0.24)	(0.858)	(0.22)	(0.105)	(0.19)	(0.478)	(0.19)	(0.708)
Gender-balanced workspace	0.43	0.069	0.50	0.088	0.66	0.077	0.66	0.082
	(0.23)	(0.004)	(0.22)	(0.000)	(0.20)	(0.000)	(0.20)	(0.000)

Columns 1, 3, 5, 7 have means and standard deviations in parentheses. Columns 2, 4, 6, 8 present the male-female gap in expectations and p-value in parentheses for a pairwise t test (Chi square test) in equality of means or proportions across male and female outcomes.

Table 4: Returns to major

	Dependent variable: log income in '000s of Rs						
	1	2	3	4	5	6	7
Econ & Business		–0.111	–0.061	0.104	0.077	0.055	0.143
		(0.098)	(0.093)	(0.099)	(0.107)	(0.110)	(0.105)
Arts & soc sci		–0.328***	–0.258***	–0.207**	–0.164*	–0.169*	0.050
		(0.092)	(0.089)	(0.083)	(0.090)	(0.091)	(0.091)
Male	0.292***		0.229***	0.158**	0.203***	0.211***	0.159**
	(0.083)		(0.076)	(0.075)	(0.076)	(0.076)	(0.068)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Work Experience	No	No	No	Yes	Yes	Yes	Yes
Individual controls	No	No	No	No	Yes	Yes	Yes
Parental controls	No	No	No	No	No	Yes	Yes
Occupation FE	No	No	No	No	No	No	Yes
Observations	675	675	675	675	611	611	611
R^2	0.16	0.22	0.23	0.29	0.33	0.34	0.43

Year fixed effects indicate the year in which the job was started. Work experience includes a quadratic term in years of work experience. Individual controls include normalised grade 10 test scores, a binary indicator of relative fluency in English based on whether English was one of the student's top 3 subjects, a binary indicator of relative fluency in mathematics based on whether mathematics was one of the student's top 3 subjects, the admissions score for the graduate programme, and three survey measures of attitude towards risk, competitiveness and self-esteem. Parental controls include a binary indicator of whether the individual's mother is employed and a categorical variable indicating the level of need-based scholarship the student received (with categories of no scholarship, 25-75% of the fee, 75-100% of the fee). Occupational fixed effects includes fixed effects for one of 7 occupational categories. *, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively. Standard errors clustered by individual.

Table 5: Differential returns to major by gender

	Dependent variable: log income in '000s of Rs				
	1	2	3	4	5
femalescience	–0.089	–0.079	–0.120	–0.132	–0.049
	(0.102)	(0.104)	(0.098)	(0.096)	(0.079)
femaleecon	–0.324**	–0.363**	–0.502***	–0.502***	–0.477***
	(0.158)	(0.159)	(0.167)	(0.168)	(0.145)
femalehum	–0.431***	–0.164	–0.170	–0.170	–0.160
	(0.153)	(0.137)	(0.140)	(0.144)	(0.155)
Year FE	Yes	Yes	Yes	Yes	Yes
Work Experience	No	Yes	Yes	Yes	Yes
Individual controls	No	No	Yes	Yes	Yes
Parental controls	No	No	No	Yes	Yes
Occupation FE	No	No	No	No	Yes
Observations	675	675	611	611	611
R^2	0.18	0.30	0.34	0.35	0.44

All regressions include fixed effects for economics and arts and social sciences. Year fixed effects indicate the year in which the job was started. Work experience includes a quadratic term in years of work experience. Individual controls include normalised grade 10 test scores, a binary indicator of relative fluency in English based on whether English was one of the student's top 3 subjects, a binary indicator of relative fluency in mathematics based on whether mathematics was one of the student's top 3 subjects, the admissions score for the graduate programme, and three survey measures of attitude towards risk, competitiveness and self-esteem. Parental controls include a binary indicator of whether the individual's mother is employed and a categorical variable indicating the level of need-based scholarship the student received (with categories of no scholarship, 25-75% of the fee, 75-100% of the fee). Occupational fixed effects includes fixed effects for one of 7 occupational categories. *, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively. Standard errors clustered by individual.

Table 6: Probability of employment by major

	Dependent variable: Has a full-time job					
	1	2	3	4	5	6
Econ & Business		0.035	0.036	0.038	0.053	0.052
		(0.043)	(0.044)	(0.045)	(0.047)	(0.048)
Arts & soc sci		–0.093**	–0.091**	–0.075*	–0.083*	–0.081
		(0.041)	(0.042)	(0.044)	(0.049)	(0.050)
Male	0.014		0.003	0.014	0.040	0.042
	(0.034)		(0.035)	(0.036)	(0.039)	(0.039)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Work Experience	No	No	No	Yes	Yes	Yes
Individual controls	No	No	No	No	Yes	Yes
Parental controls	No	No	No	No	No	Yes
Occupation FE	No	No	No	No	No	No
Observations	248	248	248	248	230	230
R^2	0.03	0.06	0.06	0.07	0.11	0.12

Year fixed effects indicate the year in which the job was started. Work experience includes a quadratic term in years of work experience. Individual controls include normalised grade 10 test scores, a binary indicator of relative fluency in English based on whether English was one of the student's top 3 subjects, a binary indicator of relative fluency in mathematics based on whether mathematics was one of the student's top 3 subjects, the admissions score for the graduate programme, and three survey measures of attitude towards risk, competitiveness and self-esteem. Parental controls include a binary indicator of whether the individual's mother is employed and a categorical variable indicating the level of need-based scholarship the student received (with categories of no scholarship, 25-75% of the fee, 75-100% of the fee). *, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively. Standard errors clustered by individual.

Table 7: Hours per work by major

	Dependent variable: Weekly work hours						
	1	2	3	4	5	6	7
Econ & Business		–6.026**	–5.127**	–4.576*	–5.224*	–4.874*	–6.096**
		(2.516)	(2.580)	(2.630)	(2.792)	(2.802)	(2.830)
Arts & soc sci		–6.288***	–5.304**	–5.100**	–5.741*	–5.390*	–5.440*
		(2.409)	(2.490)	(2.558)	(2.921)	(2.904)	(3.016)
Male	4.664**		3.120	3.144	3.040	3.400	2.734
	(1.995)		(2.073)	(2.107)	(2.273)	(2.250)	(2.247)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Work Experience	No	No	No	Yes	Yes	Yes	Yes
Individual controls	No	No	No	No	Yes	Yes	Yes
Parental controls	No	No	No	No	No	Yes	Yes
Occupation FE	No	No	No	No	No	No	Yes
Observations	248	248	248	248	230	230	228
R^2	0.07	0.09	0.10	0.12	0.16	0.19	0.23

Year fixed effects indicate the year in which the job was started. Work experience includes a quadratic term in years of work experience. Individual controls include normalised grade 10 test scores, a binary indicator of relative fluency in English based on whether English was one of the student's top 3 subjects, a binary indicator of relative fluency in mathematics based on whether mathematics was one of the student's top 3 subjects, the admissions score for the graduate programme, and three survey measures of attitude towards risk, competitiveness and self-esteem. Parental controls include a binary indicator of whether the individual's mother is employed and a categorical variable indicating the level of need-based scholarship the student received (with categories of no scholarship, 25-75% of the fee, 75-100% of the fee). Occupational fixed effects includes fixed effects for one of 7 occupational categories. *, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively. Standard errors clustered by individual.

Table 8: Major choice using choice data

	Dependent variable: Stated choice					
	All students			1st and 2nd years		
	All	Men	Women	All	Men	Women
GPA ≥ B	4.648***	4.053***	5.199***	5.776***	7.079**	4.251**
	(1.019)	(1.482)	(1.591)	(1.458)	(2.916)	(1.942)
Hours studying	0.031	0.053	0.037	0.033	0.045	0.041
	(0.022)	(0.069)	(0.025)	(0.027)	(0.087)	(0.029)
Parents approve	4.675***	7.567***	3.883***	3.464***	8.628***	2.646**
	(0.854)	(2.003)	(0.969)	(0.949)	(3.073)	(1.062)
Enjoys coursework	7.904***	7.733***	8.413***	8.047***	9.714***	8.967***
	(0.927)	(1.507)	(1.391)	(1.127)	(2.644)	(1.644)
Log salary	1.224***	1.754**	0.990*	1.716***	2.758**	1.506**
	(0.438)	(0.771)	(0.537)	(0.573)	(1.181)	(0.648)
Employed at 20	–0.235	–1.430	0.388	–0.491	–4.655*	0.940
	(0.896)	(1.551)	(1.142)	(1.138)	(2.601)	(1.330)
Job in desired sector	4.798***	6.576***	3.562***	3.913***	6.243***	2.281*
	(0.819)	(1.410)	(1.040)	(0.951)	(1.709)	(1.206)
Enjoy work-life balance	1.791**	4.252**	0.604	1.112	3.588	0.033
	(0.884)	(1.696)	(1.145)	(1.102)	(2.656)	(1.270)
Gender-balanced workspace	–1.744**	–1.720	–1.437	–1.503	–2.772	–0.425
	(0.791)	(1.337)	(1.004)	(0.978)	(1.834)	(1.229)
Observations	1,396	624	772	1,028	452	576
Individuals	349	156	193	257	113	144
Choices	349	156	193	257	113	144
Log pseudolikelihood	-165.88	-59.77	-100.48	-115.84	-33.33	-76.29

All estimates are from maximum likelihood models of choice data using stated choice data where students select one out of four classes of majors as their most preferred choice. *, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively.

Table 9: Major choice using ranked preference data

	Dependent variable: Ranked choice					
	All students			1st and 2nd years		
	All	Men	Women	All	Men	Women
GPA ≥ B	4.713***	4.198***	5.199***	6.179***	7.751***	4.962***
	(0.541)	(0.718)	(0.829)	(0.716)	(1.175)	(0.944)
Hours studying	0.008	–0.017	0.022	0.028**	0.012	0.036**
	(0.013)	(0.025)	(0.014)	(0.014)	(0.032)	(0.017)
Parents approve	1.952***	2.055***	1.811***	1.818***	2.712***	1.417***
	(0.376)	(0.613)	(0.493)	(0.439)	(0.790)	(0.547)
Enjoys coursework	5.221***	4.929***	5.370***	5.161***	4.732***	5.498***
	(0.415)	(0.570)	(0.647)	(0.517)	(0.757)	(0.748)
Log salary	0.428***	0.563***	0.250	0.581***	0.770**	0.278
	(0.150)	(0.215)	(0.230)	(0.190)	(0.300)	(0.296)
Employed at 20	0.628	1.118*	0.345	0.142	0.366	0.091
	(0.415)	(0.630)	(0.569)	(0.511)	(0.823)	(0.664)
Job in desired sector	1.847***	1.663***	2.025***	2.208***	1.556**	2.742***
	(0.366)	(0.546)	(0.521)	(0.475)	(0.716)	(0.681)
Enjoy work-life balance	0.482	0.518	0.375	0.454	0.757	0.347
	(0.420)	(0.608)	(0.617)	(0.511)	(0.810)	(0.704)
Gender-balanced workspace	–0.514	–0.820	–0.144	–0.740	–1.643**	0.032
	(0.409)	(0.610)	(0.569)	(0.507)	(0.779)	(0.686)
Observations	1,393	621	772	1,025	449	576
Individuals	349	156	193	257	113	144
Choices	1,047	468	579	771	339	432
Log pseudolikelihood	-565.10	-263.87	-295.88	-387.98	-162.97	-217.96

All estimates are from maximum likelihood models of choice data using preference ranking data where students rank the four classes of majors from 1 to 4. *, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively.

Table 10: Willingness to pay estimates

	WTP (% of earnings)			WTP (Rs.)		
	All	Men	Women	All	Men	Women
GPA $\geq$ B	-3.7	-2.3	-5.1	-22,506.0	-15,463.2	-27,864.0
Hours studying	-0.0	-0.0	-0.0	-152.3	-204.9	-204.7
Parents approve	-3.7	-4.2	-3.8	-22,632.9	-28,582.0	-20,947.8
Enjoys coursework	-6.3	-4.3	-8.1	-37,768.5	-29,194.7	-44,372.8
Employed	0.2	0.8	-0.4	1,158.6	5,542.9	-2,132.2
Job in desired sector	-3.8	-3.7	-3.5	-23,218.5	-24,909.1	-19,247.6
Enjoy work-life balance	-1.5	-2.4	-0.6	-8,772.5	-16,210.7	-3,313.4
Gender-balanced workspace	1.4	1.0	1.5	8,665.7	6,669.4	7,963.9

Willingness to pay is reported as a percentage of average earnings (columns 1-3) and in absolute terms in Rs. (columns 4-6) computed at the average annual earnings for each group of students. The WTP is the income forgone to increase the probability of each discrete variable by 1 percentage point, and to increase the continuous variables by 1 unit (hours per week).

Table 11: Major choice by student ability

	Dependent variable: Ranked choice					
	Below median grades			Above median grades		
	All	Men	Women	All	Men	Women
GPA ≥ B	4.010***	2.584**	6.272***	5.300***	5.249***	5.182***
	(0.830)	(1.060)	(1.503)	(0.734)	(1.062)	(1.063)
Hours studying	–0.016	–0.059	0.019	0.017	0.005	0.025
	(0.027)	(0.042)	(0.028)	(0.015)	(0.030)	(0.018)
Parents approve	2.108***	2.422***	1.466	1.861***	1.665*	2.233***
	(0.594)	(0.897)	(0.905)	(0.500)	(0.887)	(0.632)
Enjoys coursework	4.677***	4.753***	4.125***	5.464***	5.181***	5.976***
	(0.595)	(0.787)	(1.078)	(0.583)	(0.878)	(0.834)
Log salary	0.386	0.478	0.348	0.421**	0.570**	0.152
	(0.333)	(0.492)	(0.504)	(0.173)	(0.245)	(0.272)
Employed at 20	1.575**	1.207	1.908**	–0.207	1.083	–1.133
	(0.623)	(0.918)	(0.882)	(0.565)	(0.863)	(0.789)
Job in desired sector	1.285**	2.466***	0.447	2.590***	1.491**	3.665***
	(0.563)	(0.915)	(0.765)	(0.518)	(0.745)	(0.772)
Enjoy work-life balance	0.701	0.887	0.375	0.364	0.273	0.164
	(0.671)	(0.981)	(0.986)	(0.550)	(0.795)	(0.819)
Gender-balanced workspace	0.227	–0.140	1.229	–1.203**	–1.473*	–0.850
	(0.615)	(0.963)	(0.904)	(0.571)	(0.867)	(0.791)
Observations	536	240	296	857	381	476
Individuals	134	60	74	215	96	119
Choices	402	180	222	645	288	357
Log pseudolikelihood	-222.46	-100.68	-116.50	-335.84	-160.25	-169.97

All estimates are from maximum likelihood models of choice data using preference ranking data where students rank the four classes of majors from 1 to 4. *, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively.

Table 12: Decomposition of gender differences

	Decomposition by gender			
	Science	Economics	Soc Sci	Humanities
GPA ≥ B	0.008	0.002	–0.014	0.006
	(0.009)	(0.012)	(0.012)	(0.016)
Hours studying	–0.001	–0.000	0.000	–0.000
	(0.003)	(0.002)	(0.004)	(0.005)
Parents approve	–0.002	0.006	0.008	–0.011
	(0.005)	(0.004)	(0.012)	(0.009)
Enjoys coursework	0.041***	0.092***	–0.045*	–0.077***
	(0.012)	(0.019)	(0.024)	(0.028)
Ln earnings	0.012	–0.008	0.023	0.001
	(0.013)	(0.017)	(0.016)	(0.010)
Employed at 20	0.004	–0.014**	0.001	0.001
	(0.006)	(0.007)	(0.007)	(0.007)
Job in desired sector	0.026***	0.054***	–0.037**	–0.009
	(0.009)	(0.014)	(0.015)	(0.010)
Enjoy work-life balance	0.002	–0.003	0.001	0.001
	(0.003)	(0.006)	(0.003)	(0.004)
Gender-balanced workspace	–0.015*	–0.025**	–0.021	–0.006
	(0.008)	(0.010)	(0.016)	(0.007)
Major characteristics	0.050***	0.098***	-0.050***	-0.083***
	(0.012)	(0.017)	(0.023)	(0.020)
Job characteristics	0.025	0.005	-0.032	-0.011
	(0.017)	(0.019)	(0.027)	(0.016)
Observations	351	351	351	351
Pr(1| Male)	0.29	0.42	0.20	0.08
Pr(1| Female)	0.16	0.27	0.36	0.21
Difference	0.14	0.14	–0.17	–0.12
Explained by characteristics	0.08	0.10	–0.08	–0.09

*, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively.

Table 13: Simulations of the change in gender gap in enrolment

	Base	GPA	Parents approve	Enjoy course	Income	Desired job
Science	11.8	11.2	12.0	6.6	12.5	10.9
Economics	16.7	13.0	16.3	11.2	19.0	17.2
Social Science	-18.2	-12.5	-16.8	-8.4	-22.0	-11.2
Humanities	-10.4	-11.7	-11.6	-9.5	-9.5	-16.9

Column 1 indicates the gender gap (male-female) in enrolment in the given category of majors predicted by the base model. Columns 2-6 indicate the predicted gender gap in enrolment if the female distribution is replaced with the male distribution of beliefs about the given outcome.

Table 14: Comparing expected grades to observed grades

	Expected grades		Observed grades		Difference		p-val	
	Men	Women	Men	Women	Men	Women	Men	Women
Science	0.580	0.530	0.440	0.570	0.140	-0.040	0.000	0.034
Economics	0.690	0.600	0.320	0.570	0.370	0.030	0.000	0.116
Soc Sci	0.700	0.770	0.470	0.540	0.230	0.230	0.000	0.000
Humanities	0.670	0.720	0.320	0.660	0.350	0.060	0.000	0.000

Columns 1-4 indicate the probability that a student will get a GPA above a B. The p-values in columns 7-8 are reported for a pairwise t test (Chi square test) in equality of means across expected and observed outcomes for men and women, respectively.

Table 15: Preferred sector by gender

	Men				Women			
	Sci	Econ	SocSci	Hum	Sci	Econ	SocSci	Hum
Consulting and finance	7.3	65.0	12.5	11.1	3.4	54.2	8.2	3.0
Social sector	2.4	21.2	53.1	44.4	6.9	37.3	41.1	21.2
Comp.Sci.	43.9	2.5	3.1	0.0	55.2	0.0	0.0	3.0
Nat.Sci.	17.1	1.2	6.2	0.0	17.2	0.0	27.4	0.0
Education and research	29.3	7.5	18.8	22.2	6.9	6.8	17.8	39.4
Media and Arts	0.0	2.5	6.2	22.2	10.3	1.7	5.5	33.3
N	162				194			

% of students within a major who rank a given sector as most preferred.

Table 16: Gender differences in expectations

	Dependent variable: Elicited expectation			
	Enjoy	GPA	(Log) Earnings	Pr(Emp) pref sector
FemXScience	–0.095***	–0.049*	–0.507***	–0.043
	(0.032)	(0.029)	(0.134)	(0.030)
FemXEconomics	–0.154***	–0.104***	–0.530***	–0.078***
	(0.029)	(0.024)	(0.123)	(0.026)
FemXSoc Sci	0.105***	0.060***	–0.372***	0.120***
	(0.025)	(0.023)	(0.129)	(0.026)
FemXHumanities	0.078***	0.048**	–0.419***	0.063**
	(0.025)	(0.024)	(0.123)	(0.027)
Demographic controls	Yes	Yes	Yes	Yes
Ability controls	Yes	Yes	Yes	Yes
Personality controls	Yes	Yes	Yes	Yes
Observations	1,360	1,360	1,360	1,360
R-squared	0.14	0.13	0.08	0.06

The dependent variable in columns 1, 2 and 4 is the expected probability that a student will enjoy the coursework, get a grade above a B, and find a job in a preferred sector after graduation, respectively. In column 3, the dependent variable is the expected earnings after graduation. *, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively.

Table A1: Returns to major: pre-graduate school jobs

	Dependent variable: log income in '000s of Rs						
	1	2	3	4	5	6	7
Econ & Business		–0.215	–0.215	0.102	0.147	0.199	0.445**
		(0.148)	(0.144)	(0.139)	(0.169)	(0.169)	(0.205)
Arts & soc sci		–0.613***	–0.506***	–0.299	–0.107	–0.131	0.479**
		(0.208)	(0.189)	(0.196)	(0.194)	(0.194)	(0.237)
Male	0.446***		0.325**	0.259*	0.287*	0.341**	0.296**
	(0.160)		(0.136)	(0.135)	(0.154)	(0.153)	(0.139)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Work Experience	No	No	No	Yes	Yes	Yes	Yes
Individual controls	No	No	No	No	Yes	Yes	Yes
Parental controls	No	No	No	No	No	Yes	Yes
Occupation FE	No	No	No	No	No	No	Yes
Observations	198	198	198	198	180	180	180
R^2	0.20	0.22	0.24	0.36	0.39	0.42	0.53

Year fixed effects indicate the year in which the job was started. All regressions include a binary indicator for whether the job was started after the graduate degree. Work experience includes a quadratic in years of work experience. Individual controls include normalised grade 10 test scores, a binary indicator of relative fluency in English based on whether English was one of the student's top 3 subjects, the admissions score for the YIF programme, and three survey measures of attitude towards risk, competitiveness and self-esteem. Parental controls include a binary indicator of whether the individual's mother is employed and a categorical variable indicating the level of need-based scholarship the student received (with categories of no scholarship, 25-75% of the fee, 75-100% of the fee). Occupational fixed effects includes fixed effects for one of 7 occupational categories. *, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively. Standard errors clustered by individual.

Table A2: Quantile regression of returns to major

	Dependent variable: log income in '000s of Rs				
	0.1	0.25	0.5	0.75	0.9
Econ & Business	0.105	0.004	0.006	0.166**	0.151
	(0.113)	(0.092)	(0.092)	(0.069)	(0.094)
Arts & soc sci	0.026	–0.091	–0.120*	–0.168**	–0.230***
	(0.126)	(0.085)	(0.069)	(0.078)	(0.079)
Male	0.310***	0.099	0.113**	0.130**	0.122*
	(0.071)	(0.074)	(0.056)	(0.055)	(0.063)
Year FE	Yes	Yes	Yes	Yes	Yes
Work Experience	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes
Parental controls	Yes	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes	Yes
Observations	611	611	611	611	611
R^2	0.30	0.23	0.17	0.17	0.22

Year fixed effects indicate the year in which the job was started. Work experience includes a quadratic in years of work experience. Individual controls include normalised grade 10 test scores, a binary indicator of relative fluency in English based on whether English was one of the student's top 3 subjects, the admissions score for the YIF programme, and three survey measures of attitude towards risk, competitiveness and self-esteem. Parental controls include a binary indicator of whether the individual's mother is employed and a categorical variable indicating the level of need-based scholarship the student received (with categories of no scholarship, 25-75% of the fee, 75-100% of the fee). Occupational fixed effects includes fixed effects for one of 7 occupational categories. *, **, *** denote significance at the 10%, 5%, and 1% confidence levels, respectively. Standard errors clustered by individual.

Table A3: Correlation coefficients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Men									
1. GPA	1								
2. Hours	-0.128**	1							
3. Parents approve	0.194***	0.126**	1						
4. Enjoy	0.599***	-0.0371	0.240***	1					
5. Log earnings	0.128**	0.0485	0.208***	0.164***	1				
6. Employed	0.250***	0.0254	0.426***	0.292***	0.274***	1			
7. Preferred sector	0.428***	0.0239	0.362***	0.426***	0.250***	0.389***	1		
8. Work-life balance	0.406***	-0.0872*	-0.0434	0.384***	0.0504	0.0988*	0.257***	1	
9. Gender balance	0.189***	-0.0257	-0.146***	0.180***	-0.0733	-0.0887*	0.0708	0.391***	1
Women									
1. GPA	1								
2. Hours	-0.230***	1							
3. Parents approve	0.141***	0.00150	1						
4. Enjoy	0.763***	-0.190***	0.109**	1					
5. Log earnings	0.141***	0.0445	0.186***	0.0582	1				
6. Employed	0.171***	-0.0296	0.405***	0.129***	0.465***	1			
7. Preferred sector	0.474***	-0.138***	0.265***	0.533***	0.112**	0.262***	1		
8. Work-life balance	0.426***	-0.200***	0.0329	0.517***	0.110**	0.124***	0.375***	1	
9. Gender balance	0.334***	-0.238***	-0.0254	0.341***	-0.0746*	-0.0538	0.239***	0.432***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table presents pairwise correlation coefficients between each pair of variables.