

Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender

Ernesto Reuben

Columbia University and IZA, e-mail: ereuben@columbia.edu

Matthew Wiswall

Arizona State University, e-mail: matt.wiswall@gmail.com

Basit Zafar

Federal Reserve Bank of New York, e-mail: Basit.Zafar@ny.frb.org

Abstract

Using an experiment to measure overconfidence and preferences for competitiveness and risk, this paper investigates whether these measures explain gender differences in college major choices and expected future earnings. We find that individuals who are overconfident and overly competitive expect to earn significantly more. In addition, gender differences in overconfidence and competitiveness explain 18% of the gender gap in earnings expectations. These experimental measures explain as much of the gender gap in earnings expectations as a rich set of control variables. While expected earnings are related to college major choices, the experimental measures are not related with college major choice.

This version: October 2015

JEL Codes: J16, D81, D84, I21, C93.

Keywords: earnings; gender differences; subjective expectations; risk aversion; overconfidence; competitiveness; college majors.

Acknowledgements: We thank the NYU Center for Experimental Social Sciences (CESS) for providing assistance in conducting the survey and experiment. We thank Muriel Niederle, Uri Gneezy and participants at presentations at the 2012 North American ESA Meetings, 2014 ASSA Meetings, CUNY, MiSoC Workshop on Subjective Expectations and Probabilities, 2014 SITE Experimental Economics Workshop, 7th Maastricht Behavioral and Experimental Economics Symposium, University of Essex, and City University of New York. Grant Graziani and Victoria Gregory provided outstanding research assistance. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System as a whole.

Note: This is the authors' version of a work that was accepted for publication at *The Economic Journal*. Changes resulting from the publishing process may not be reflected in this document. A final version is published at <http://dx.doi.org/10.1111/eoj.12350>.

1 Introduction

While considerable progress has been made to explain gender differences in occupations and labour market trajectories, residual differences remain unaccounted for by standard variables, such as labour market experience, amount and quality of education, family background, as well as customary demographic characteristics (Blau and Kahn, 2000, 2012; Jarrell and Stanley, 2004; Black et al., 2008; Hegewisch et al., 2013). What accounts for the remaining differences? A recent and growing literature points at expectations as an important predictor of educational choices and attainment.¹ Not only are students more likely to self-select into fields in which they expect relatively higher earnings (Arcidiacono et al., 2013; Zafar, 2013; Wiswall and Zafar, 2015), but expectations can easily become self-fulfilling. For example, students with low expectations will have a smaller incentive to perform well academically (Jacob and Wilder, 2011; Beaman et al., 2012; Stinebrickner and Stinebrickner, 2012), or subsequently, they will be more willing to accept a low-paying job offer and less likely to negotiate for higher salary because it is in line with their beliefs. Consequently, studying gender differences in expected earnings can take us a long way in explaining the observed gender differences in career choices and success. In fact, given that realized earnings and other labour market outcomes can be affected by a number of unanticipated future events (and may suffer from the problem of reverse causality), we argue that investigating why young men and women form different expectations about future earnings is potentially *more* important than realized earnings for the purpose of understanding the role of gender in education and career choices.

In this paper, we evaluate whether well-documented gender differences in which men are more competitive (Niederle and Vesterlund, 2011), tend to be more overconfident (Bertrand, 2011), and are more willing to take risks (Eckel and Grossman, 2008; Croson and Gneezy, 2009) help explain gender differences in expectations about future earnings and educational choices. For a sample of undergraduate students from New York University (NYU), we combine a laboratory experiment to measure these preferences and biases with a survey of labour market expectations and education choices, including detailed demographic data, family background information, and test scores and grades. While our sample is from a high ability group, as evidenced by high average test scores and grades for our sample, we observe similarly large gender differences in college major choices in nationally representative US data. In addition, the availability of a large set of observable

¹Evidence from other domains—retirement savings, investment, health—also shows that expectations tend to be good predictors of choices, above and beyond standard determinants (Wolpin and van der Klaauw, 2008; de Paula et al., 2014; Armantier et al., 2015). The analysis of earnings expectations follows a larger literature that collects and uses subjective expectations data to understand decision-making under uncertainty (for a survey, see Manski, 2004).

individual characteristics in our survey allows us to control for multiple proxies of student ability.

The key advantage of our research design is we combine experimental and survey data for the same sample of subjects, unlike research which uses experimental data or survey data in isolation. This allows us to quantify the importance of experimental measures of preferences and biases to the beliefs and outcomes we measure in the survey, and quantify the extent to which the experimental measures represent independent variation that is not well proxied by a relatively rich set of individual characteristics collected in the survey. In addition, because we measure multiple traits for each subject—competitiveness, overconfidence, and risk taking—we can assess the importance of each measure separately within the same sample of subjects.

Using our survey data, we first document a large gender gap in expected future earnings that increases with age: compared to men, on average, women expect to earn 31% less at age 30 and 39% less at age 45. We show that the observed gender wage gap is the result of gender differences in expected earnings within each major/occupation as well as gender differences in major/occupational choices. More specifically, in our sample, males are 82% more likely to major in business and economics and women are 62% more likely to major in the humanities, which mirrors observed gender differences in major choice in nationally representative data of the US.² To isolate the effect of major choice on earnings expectations, we collect data on students' expected earnings in *all* majors (as defined by aggregated major categories) and not simply their *chosen* major. On average across all majors, women expect to earn 19% less than men at age 30 and 23% less at age 45.³ Hence, even though college major choice explains an important part of the difference in the earnings expectations of men and women (as in Brown and Corcoran, 1997; Weinberger, 1998; Arcidiacono, 2004), an equally if not more important part is due to differences in expected earnings within majors.

Turning to our experimental data, we find substantial gender differences in each of

²For the cohort born in 1988 and surveyed in the 2012 American Community Survey (age 24 at time of survey), there is a 39 percent male advantage in business and a 56 percent female advantage in humanities. Note that the students in our survey were aged 18-22 in 2012 and born during 1990-94.

³While there is no direct counterpart to the expected earnings data in realized earnings—the survey data on expectations is about future, unrealized earnings—and our sample of expectations is from a high ability population at an elite private university, it is worth noting that expected earnings mirror gender gaps in realized earnings for all US college graduates, with women's average earnings being 17% lower than those of men at age 30 and 36% lower at age 45 (controlling for differences in major composition between genders). When we ask our sample a separate set of questions about their perceptions of average earnings in the US population, we find that the students' beliefs about average earnings are substantially lower than their beliefs about their own earnings, and their beliefs about the general population, on average, resemble the true average population earnings.

the experimentally-derived measures. We calculate a relative risk aversion coefficient for each student using a series of lotteries and find that the average coefficient for men is 56% lower than that for women, indicating that men are less risk averse. We also find that men are more than twice as likely as women to overestimate their true ability level, which we use to construct a measure of overconfidence.⁴ Finally, we find that men are twice as likely as women to pick a compensation scheme where rewards are allocated through competition with others (a tournament) rather than through non-competitive means. Moreover, the difference in competitiveness between men and women remains when we construct a competitiveness measure that controls for perceptions of relative abilities and risk preferences.

In analysing the combined experiment and survey data, we find that the competitiveness and overconfidence measures, but not the risk aversion measure, are significantly related to the student's expectations about future major-specific earnings, with earnings expectations increasing in the level of competitiveness and overconfidence. The experimentally-derived attributes alone explain 17% and 19% of the gender gap in earnings expectations for age 30 and age 45, respectively. Note that this analysis is conducted within major and hence the effect of these experimentally-derived attributes on earnings is not confounded by gender differences in major composition. Furthermore, these differences in earnings expectations are specific to the individual's beliefs about his or her own future earnings in a given major, as we find no statistically significant relationship between the experimental measures and the students' perceptions about the average earnings in the *population*. Two other findings underscore the importance of the relationship between the competitiveness and overconfidence measures and earnings expectations. First, the experimental measures explain as much of the gender gap in earnings expectations as a rich set of control variables, including the student's SAT (Scholastic Aptitude Test) scores, race, and family background. Second, the experimental measures are not well proxied by the control variables measuring ability and family background, as we find that they are not significantly related to the control variables. Thus, our findings highlight that a small number of individual attributes can explain a substantial portion of the gender gap in earnings expectations, a portion that would otherwise be unaccounted for by even a relatively rich set of control variables.

The relationship between overconfidence and competitiveness and earnings expectations provides important insight into the mechanisms underlying gender differences in the

⁴Following the practice in behavioural economics, we use the term overconfidence to refer to an individual's tendency to believe that their performance relative to others is better than it actually is (e.g., Malmendier and Tate, 2008; Niederle and Vesterlund, 2011). This tendency is also referred to as optimism, especially in the psychology literature (Moore and Healy, 2008).

labour market. Our results are consistent with either overconfident (underconfident) students sorting themselves into (out of) higher-paying occupations within a major category, and/or overconfident (underconfident) students expecting to be more (less) successful in higher-paying occupations. Disproportionately overconfident men may pursue different occupations on the extensive margin and more aggressively negotiate for salary on the intensive margin than women. Our results also suggest that the gender gap in earnings expectations are partly driven by overly competitive individuals, who are disproportionately men, who presumably seek occupations with tournament-based pay, whilst individuals who are averse to competition, who are disproportionately women, shy away from such higher-paying jobs.⁵ These findings, based on a sample of high ability students attending an elite university (i.e., precisely the kind of students who are more likely to make it to the higher echelons of their professions), provide a possible explanation for the glass ceiling phenomenon (Bertrand and Hallock, 2001; Albrecht et al., 2003; Bertrand et al., 2010), whereby higher earning and higher prestige positions require aggressive negotiation and compensation is based on relative performance.

In contrast to the results on future earnings expectations, we find that our experimental measures of competitiveness and overconfidence are not systematically related with major choice, as defined in our survey by four aggregated major categories. Consistent with risk preferences affecting schooling decisions (Nielsen and Vissing-Jorgensen, 2006; Belzil and Leonardi, 2007), we do find that risk averse students are less likely to select into majors with greater earnings uncertainty, but the result is not statistically significant at conventional levels. Using the students' perceptions of the characteristics of each major (e.g., prevalence of bonus pay, earnings uncertainty, and other job attributes), we find that the lack of a relationship between the experimental measures and major choice is not because students think that all majors are equally competitive or equally risky. We conclude from the fact that while competitiveness and overconfidence affect future earnings expectations, the channel through which they operate is *within* broad college major categories (humanities, business and economics, natural sciences, and engineering) and through the level of specific majors, occupations or jobs. This is consistent with the evidence that there are large gender differences at more narrowly defined college majors and occupations (for a recent review see Goldin, 2014), as for example with the very different gender compositions in medical fields between surgeons and general practitioners.

⁵There are various reasons why occupations with tournament-based pay might have higher expected earnings. For example, if performance pay is used in markets with adverse selection to differentiate employees according to their ability, it can lead to overly high wages for the most talented workers (Moen and Rosén, 2005; Bénabou and Tirole, 2015). Alternatively, if most people find competition to be inherently distasteful (e.g., in our sample only 14% of the students are classified as overly competitive), then there can be a compensating wage differential for competitive jobs.

Our results complement the findings of the concurrent study of Buser et al. (2014), which correlates the same type of competitiveness measure to high school track choice among Dutch students.⁶ They find that controlling for ability, confidence, and risk attitudes, laboratory measures of competitiveness explain about 15% of the gender gap in the “prestige” of high school track choice, with boys more likely to choose the prestigious science and math tracks over the less prestigious humanities tracks. While our sample shares the general gender gap in human capital investments, with women more likely to choose humanities fields over science and business fields, we do not find a similar relationship between competitiveness and major choice. However, the two studies are not strictly comparable given that our sample is different (high ability American college students versus Dutch high school students), and our measure of education is at the university level. Importantly, the relationship between prestige and fields of study may differ in the two settings. Buser et al. (2014) study a context where the prestige of educational profiles highly correlates with their math and science intensity. By contrast, our sample of participants might not rank majors in such a precise way. Our study also contributes to the literature on individual characteristics and labor market outcomes by showing that competitiveness and confidence measures strongly relate to earnings expectations, and that these measures can even explain gender differences *within* careers.

The paper is organized as follows. Section 2 describes the design and implementation of the experiment and survey, and it presents descriptive statistics of our sample. In section 3 we analyze the students’ decisions during the experiment and explain how we construct our measures of competitiveness, overconfidence, and risk aversion. We investigate the link between our experimental measures and earnings expectations in section 4, and with educational choices in section 5. We conclude in section 6.

2 Study design

Our study consists of two parts, an experiment and a survey.

2.1 The Experiment

The main goal of the experiment is to obtain individual-specific measures of competitiveness, overconfidence, and risk preferences. Our design is an adaptation of the setup implemented in Niederle and Vesterlund (2007) and the risk preferences elicitation task

⁶Another study that tests the relation between competitiveness and educational outcomes is Zhang (2013), which examines whether competitiveness predicts middle school students in rural China taking an important high school entrance exam. Zhang (2013) finds that competitive students are more likely to take the exam but finds no relationship between gender and either competitiveness or exam taking.

used by Dohmen et al. (2010).

In the experiment, students are asked to perform a real task under different compensation schemes. The task consists of computing sums of four two-digit numbers for four minutes. The two-digit numbers are randomly drawn, with the same draw for all group members. After each answer, students are told whether their answer was correct and their total number of correct answers. We chose an addition task because it requires both effort and skill, and prior research suggests there are no gender differences in ability on easy math tasks (Hyde et al., 1990).

At the beginning of the experiment, students are informed that they will be randomly assigned to groups of four and that the experiment is divided into eight rounds, one of which will be randomly chosen for payment at the end of the study.⁷ In each round, students first read the instructions for that round. Subsequently, they make the required choices and, if necessary, perform the addition task. In this paper, we analyse the data of the first four rounds, which we describe below.⁸ Importantly, although students are informed of their own performance after each addition task, they do not receive any information about the performance or choices of others before the fifth round.

1. Tournament: In this round, students are compensated for performing the addition task in following way: the student with the highest number of correct answers in the group earns \$2.00 per correct answer while the remaining three students earn \$0.00 (ties are broken randomly).

2. Choice: In this round, prior to performing the addition task, students *choose* whether they are compensated according to a piece rate, whereby they earn \$0.50 per correct answer, or according to a tournament, whereby they earn \$2.00 per correct answer if they answer correctly more sums in this round than each of the other group members did in the previous round and \$0.00 otherwise (again, ties are broken randomly). Note that this design ensures that the students' earnings in this round do not depend on the (expected) choices of others.

3. Piece-rate: In this round, students are compensated for performing the addition task according to a piece rate of \$0.50 per correct answer.

⁷To familiarize students with the screen and the addition task, a two-minute practice round was also conducted. Performance in this round did not affect earnings.

⁸Of the additional four rounds, one is similar to the “submission task” in Niederle and Vesterlund (2007), where participants decide whether they want to be paid for their performance in the piece-rate round according to a tournament or a piece-rate compensation scheme. The other rounds consisted of providing students with information concerning their rank in the piece-rate round and eliciting their updated beliefs about their rank and re-eliciting their choice for the submission task.

4. *Beliefs about Tournament:* In this round, students do not perform the addition task. Instead, they are asked to estimate their performance in the first round relative to the performance of others in their group. Specifically, students are reminded of the number of sums they answered correctly in round 1 and are then asked “*For each of the ranks below, what is the percent chance (or chances out of 100) that you think you got that rank in Task 1?*” Responses across all ranks needed to add up to 100. A quadratic scoring rule is used to incentivize the true reporting of beliefs, with a maximum compensation of \$20.00 if the subjective rank distribution matches the students’ actual rank.

Our design differs from Niederle and Vesterlund (2007) in three ways. First, instead of asking participants for their expected rank, we elicit their subjective beliefs about their entire rank distribution. Hence, in our analysis, we do not need to assume that participants report the same statistic of their subjective distribution and that there are no gender differences in the statistic they choose to report (see Manski, 2004). This allows us to investigate overconfidence (i.e., biases in beliefs) at the individual level and incorporate potentially biased beliefs into the construction of the measure of competitiveness, as we show in the next section. Second, we use a slightly different order of compensation schemes—in their design, participants first perform under piece-rate, then tournament, and then choice. We moved the piece-rate compensation scheme to a later round because the remaining rounds of our experiment relate to the participants’ performance under that scheme. Third, while in Niederle and Vesterlund (2007) participants could see the other participants in their group, in our study participants could not identify their group members.

Lastly, since risk attitudes may be an important determinant of labour market outcomes and women are usually found to be more risk averse than men (Eckel and Grossman, 2008; Croson and Gneezy, 2009), we measure the students’ willingness to take risks. Specifically, at the end of the experiment, we give students an incentivized task similar to that in Dohmen et al. (2010). It entails ten choices, one of which is randomly chosen for payment. Each choice consists of selecting between a lottery and a certain payoff. The lottery is the same in all choices (winning either \$5 or \$1, each with a 0.50 probability), but the certain payoff increases from \$1.25 in the first choice to \$3.50 in the tenth choice in increments of \$0.25. If students maximize expected utility, they should prefer the lottery up to a specific certain payoff and then switch to the certain payoff in all subsequent choices. For example, a risk neutral individual chooses the lottery over the certain payoff when it is between \$1.25 and \$2.75, is indifferent when it equals \$3.00, and prefers the certain payoff when it equals \$3.25 or more.

2.2 The Survey

In the survey, we collect basic demographic data from the students, including their choice of college major (or intended major) and a number of beliefs about various majors, including their beliefs about the future earnings that they would receive if they were to complete different majors. In order to keep the survey manageable, we aggregated the various college majors into five categories: 1) Business and economics, 2) Engineering and computer science, 3) Humanities and other social sciences, 4) Natural sciences and math, and 5) Never graduate/drop out.⁹ Conditional on graduating in each of these major categories, students are asked for their own expected earnings at different points in time (at ages 30 and 45), and the probability that they will earn more than \$35k and \$85k at age 30. For *each* of the potential majors, we also ask a series of questions about the perceived difficulty of each major and the student's relative ability to complete the major. In addition to collecting data about beliefs about their own future earnings, students also specify their beliefs about the mean earnings of current 30 year old workers in the population, conditional on college major. Finally, we collect data about various job characteristics associated with each major category. The specific wording of these questions is provided when we analyse the results.

2.3 Procedures

The study was administered to New York University (NYU) undergraduate students. Our experimental subjects were recruited in the same fashion as the other subjects who participated in experiments at the NYU Center for Experimental Social Science (CESS) lab. Students who wished to participate in experiments could sign up for experiments on-line. Given the survey that accompanied our lab experiment, we are able to provide detailed demographic and academic background information on the sample of students who participated in our experiment (discussed below).

Student participants were informed that the study consisted of a simple economic experiment and a survey about educational and career choices. We used standard experimental procedures, including anonymity and neutrally worded instructions. The experiment took 45 minutes and was followed by the survey, which took 30 minutes to complete.

In addition to earnings from the experiment, students were given a \$10 show-up fee and received \$20 for successfully completing the survey. Total compensation varied between \$31 and \$82, with an average of \$43. Fifteen sessions were held in total. Each session had between 8 and 24 students. Detailed procedures and the instructions of the experiment

⁹We provided students with a link where they could see how each college major maps into our aggregate categories.

are available in the supplementary materials.¹⁰

2.4 Sample characteristics

A total of 257 students participated in the study. Table 1 presents the descriptive statistics of key demographic variables. The first column reports the data for the whole sample and the next two columns report the statistics by gender (35% of our sample is male and 65% is female).¹¹ The last column reports p -values from tests of equality of distributions between males and females, based on a Wilcoxon rank-sum tests for the ordinal variables and χ^2 tests for the categorical variables (all tests are two-sided).

Judging by their SAT scores and parental characteristics, our sample represents a high ability group of college students from a high socioeconomic group. There are no statistically significant demographics differences between male and female students except for their SAT math score, where males score significantly higher than females ($p = 0.004$). For a comparison to the broader NYU population of students, for all incoming freshman in 2010, the 25th and 75th quartiles of the SAT math are 630 and 740 and for the SAT verbal are 610 and 710 (Integrated Post-Secondary Education Data System, IPEDS). The equivalent quartiles in our sample are 650 and 770 for math and 620 and 730 for verbal. Using SAT scores, the distribution of math and verbal abilities in our sample is quite similar to the broader NYU student population.

3 Experimental measures

In this section we provide a brief overview of the experimental data and then describe how we use them to obtain individual-specific measures of risk aversion, overconfidence, and competitiveness. Panel A in Table 2 provides descriptive statistics of the variables from the experiment. The first column reports statistics for all students and the next two columns report the statistics by gender, and the last column reports p -values from Wilcoxon rank-sum tests comparing the distributions of males and females.

We see that the mean number of sums answered correctly is higher for males than for

¹⁰Since the survey followed the experiment, one concern is that feedback from the experiment could influence survey responses. We designed the experiment so that respondents received no feedback about earnings or tournament rank until *after* they completed the survey. Respondents did receive feedback about their own performance in the task and performance relative to one group member in one of the rounds, but we control for this feedback in the analysis discussed below. We find that the inclusion of feedback does not change the results. See Section 4.2.

¹¹Compared to the NYU population, our sample is not particularly female: 63% of students in NYU are female compared to 66% in our sample (data from the Integrated Post-Secondary Education Data System, IPEDS).

Table 1: Sample characteristics

Note: For the continuous outcomes, means are reported in the first cell and standard deviations are reported in parentheses. The rightmost column reports p -values from tests of equality of distributions between males and females, based on a Wilcoxon rank-sum tests for ordinal variables and χ^2 tests for categorical variables.

		All ($n = 257$)	Males ($n = 89$)	Females ($n = 168$)	p-value
Age		21.47 (1.45)	21.68 (1.81)	21.36 (1.21)	$p = 0.305$
Race:	White	28.79%	32.58%	26.79%	$p = 0.318$
	Asian	50.58%	51.69%	50.00%	
	Other	20.62%	15.73%	23.21%	
Parents income (\$1000s)		136.94 (121.78)	140.53 (125.05)	135.03 (120.35)	$p = 0.760$
Mother with B.A. or more		67.32%	73.03%	64.29%	$p = 0.155$
Father with B.A. or more		69.26%	70.79%	68.45%	$p = 0.700$
SAT math score		696.42 (80.76)	717.75 (71.05)	684.97 (83.52)	$p = 0.004$
SAT verbal score		676.45 (75.83)	680.38 (70.15)	674.36 (78.82)	$p = 0.818$
GPA		3.46 (0.32)	3.46 (0.33)	3.46 (0.32)	$p = 0.921$
School year:	Freshman	11.28%	10.11%	11.90%	$p = 0.689$
	Sophomore	10.51%	11.24%	10.12%	
	Junior	36.58%	32.58%	38.69%	
	Senior or more	41.63%	46.07%	39.29%	

females (the difference is statistically significant in the Tournament and Piece-rate rounds but not in the Choice round). Hence, unlike in Niederle and Vesterlund (2007), in our sample the average man performs slightly better than the average woman, which is in line with men also having higher average SAT math scores. However, judging by the size of the standard deviations, performance varies considerably within each gender. Consistent with the difference in performance, we see that the elicited belief of being ranked first in their group is significantly higher for men than for women (45% vs. 28%, $p < 0.001$). Moreover, we find a clear gender difference in the tendency to enter competitive environments: 54% of the male students choose to be compensated according the tournament versus only 27% of females students ($p < 0.001$).¹² Finally, we also find that men choose the lottery over the certain payoff significantly more often than women in the risk elicitation task.

Next, we use the data from the experiment to construct individual-specific measures of risk preferences, overconfidence, and competitiveness.

¹²This gender difference has been reported in many experiments with a similar design (e.g., Niederle and Vesterlund, 2007; Cason et al., 2010; Healy and Pate, 2011; Balafoutas and Sutter, 2012; Niederle et al., 2012) as well as in experiments that vary the design in important ways (e.g., Gneezy et al., 2009; Dohmen and Falk, 2011; Booth and Nolen, 2012; Andersen et al., 2013; Gupta et al., 2013).

Table 2: Descriptive statistics of the experiment

Note: Means are reported in the first cell and standard deviations are reported in parentheses. The rightmost column reports p -values from Wilcoxon rank-sum tests comparing the distributions of males and females.

	All	Males	Females	p -value	
PANEL A: EXPERIMENTAL OUTCOMES					
Earnings	39.68 (10.63)	40.69 (12.14)	39.13 (9.74)	$p = 0.621$	
Correct answers:	Tournament	11.91 (3.76)	12.82 (4.60)	11.42 (3.13)	$p = 0.028$
	Choice	12.66 (4.05)	13.51 (4.95)	12.21 (3.41)	$p = 0.132$
	Piece-rate	12.97 (4.30)	14.11 (4.91)	12.37 (3.83)	$p = 0.009$
Subjective probability of ranking 1 st	0.34 (0.26)	0.43 (0.31)	0.28 (0.21)	$p < 0.001$	
Proportion choosing Tournament	36.58%	53.93%	27.38%	$p < 0.001$	
Number of lottery choices	6.76 (2.04)	7.19 (1.75)	6.54 (2.14)	$p = 0.008$	
PANEL B: INDIVIDUAL-SPECIFIC MEASURES					
CRRA coefficient	0.61 (0.97)	0.42 (0.66)	0.72 (1.09)	$p = 0.015$	
Overconfidence	0.08 (0.27)	0.12 (0.28)	0.06 (0.26)	$p = 0.070$	
Competitiveness	-0.13 (0.64)	0.00 (0.63)	-0.20 (0.63)	$p = 0.018$	

3.1 Risk preferences

For our measure of risk preferences, we assume that the students' utility functions take the standard CRRA form and we use each student's choices in the risk elicitation task to calculate their coefficient of relative risk aversion. In other words, each choice by student i in the risk elicitation task consists of choosing between a certain payoff π_c , which gives utility $U_i(\pi_c) = (\pi_c^{1-\rho_i})/(1-\rho_i)$, and a lottery L that pays $\pi_h = \$5$ with 0.50 probability and $\pi_l = \$1$ otherwise, yielding expected utility $EU_i(L) = \frac{1}{2}(\pi_h^{1-\rho_i})/(1-\rho_i) + \frac{1}{2}(\pi_l^{1-\rho_i})/(1-\rho_i)$, where ρ_i is i 's coefficient of relative risk aversion. By looking at the value of π_c in the risk elicitation task at which student i switches from choosing the lottery to choosing the certain payoff, we obtain a range for the value of that student's relative risk aversion coefficient ρ_i . For simplicity, we take the midpoint of this interval as the value of ρ_i .¹³ Note that, 17 (6.61%) of our students had choice patterns that are inconsistent with expected utility maximization.¹⁴ Given that our analysis calls for an accurate measure of risk preferences, we decided it is more appropriate to drop these students from all subsequent data analysis. However, in the supplementary materials we show that the main results of the paper are robust to including these students.

¹³We set $\rho_i = -1$ for students who always chose the lottery and $\rho_i = 5$ for students who always chose the certain payoff. Our analysis is not sensitive to these parameterizations.

¹⁴It is a commonly found in the literature that a small fraction of participants, around 10%, either switch multiple times or switch once from the certain payoff to the lottery (see Holt and Laury, 2002).

Panel B in Table 2 provides the mean and standard deviation for the values of ρ_i . We can see that the mean coefficient of relative risk aversion is positive, indicating that students are risk averse on average. Moreover, consistent with the literature on risk preferences using monetary incentives (Eckel and Grossman, 2008; Croson and Gneezy, 2009), females exhibit significantly higher values of ρ_i indicating that they are more risk averse than men (0.72 vs. 0.42, Wilcoxon rank-sum test $p = 0.015$). Taking a closer look at the distribution of risk preferences reveals that 40% of females and 55% of males exhibit choices that are consistent with risk neutral preferences, 48% of females and 31% of males exhibit choices that are consistent with risk averse preferences, and 12% of females and 14% of males exhibit choices that are consistent with risk loving preferences.

3.2 Overconfidence

As has been done by others, we define overconfidence as overestimating one’s own abilities relative to others (e.g., Malmendier and Tate, 2008; Niederle and Vesterlund, 2011). This tendency is also referred to as optimism or overplacement (Moore and Healy, 2008). To measure it, we compare each student’s subjective probability of being ranked first in the Tournament round with their *true* probability of ranking first. To compute each student’s true probability of ranking first, we use the distribution of performance by all students in the Tournament round to draw 100,000 comparison groups for each student (draws within a comparison group are done without replacement). We then simply calculate the fraction of times each student is ranked first. Obtained this way, this fraction approximates the true probability of ranking first.¹⁵ Mirroring the gender difference in number of correct sums, men have a significantly higher true probability of being ranked first than women (32% vs. 21%, Wilcoxon rank-sum test, $p = 0.028$).

As our measure of overconfidence, we take the students’ subjective probability of being ranked first and subtract their true probability of attaining that rank. Positive (negative) values of this variable therefore indicate overconfidence (underconfidence). Panel B in Table 2 provides the mean and standard deviation of this variable. On average, both males and females overestimate their relative performance.¹⁶ However, consistent with the literature on gender differences in overconfidence (e.g., Beyer, 1990; Lundeberg et al., 1994; Bengtsson et al., 2005; Niederle and Vesterlund, 2007; Reuben et al., 2012, 2014),

¹⁵We use the probability of ranking first in the Tournament round because it is the most relevant for their choice between the tournament and piece-rate compensation schemes, which we use to construct our measure of competitiveness. Alternatively, one could compare their subjective expected rank to their true expected rank and/or their beliefs in the Piece-Rate round. Our results are qualitatively the same with these alternative measures of overconfidence.

¹⁶Both males and females significantly overestimate their probability of ranking first according to Wilcoxon signed-rank tests ($p < 0.002$)

Table 3: Choosing tournament compensation

Note: Marginal effects from probit regressions with robust standard errors. All regressions have 240 observations. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

Independent variables	I	II
Male	0.29*** (0.06)	0.18** (0.07)
True probability of ranking 1 st		-0.03 (0.13)
Subjective probability of ranking 1 st		0.85*** (0.16)
CRRA coefficient		-0.07* (0.04)
Wald χ^2	19.57	56.81

the mean level of overconfidence is larger for men than for women (12 percentage points vs. 6 percentage points, Wilcoxon rank-sum test, $p = 0.070$).

3.3 Competitiveness

Following Buser et al. (2014), we measure the effect of competitiveness on other variables by including in regressions the students' decision to enter the tournament in the second round of the experiment along with measures of the students' ability (their true probability of ranking first in the Tournament round), beliefs about relative performance (their subjective probability of being ranked first in the Tournament round), and risk preferences (their CRRA coefficient).

Are men in our sample more competitive than women after controlling for ability, beliefs, and risk preferences through a regression? The answer to this question is provided in Table 3, which presents the marginal effects of probit regressions with the choice of tournament compensation as the dependent variable. In regression I, the only independent variable is the students' gender. In regression II, we include the students' true probability of being ranked first, their subjective probability of being ranked first, and their CRRA coefficient. Both regressions are run with robust standard errors. Also, to facilitate the interpretation of the coefficients, we standardize the continuous independent variables to have a mean of zero and a standard deviation of one.

The coefficient for males is positive and statistically significant in both regressions. Importantly, the fact that the male coefficient is significant in regression II shows that

gender differences in beliefs and risk preferences are not enough to fully explain why men are more likely to choose tournament compensation. As Niederle and Vesterlund (2007), we interpret this remaining gender gap as being driven by competitiveness. In regression II, we can also see that the students' true probability of ranking first is not a significant predictor of their compensation choice while their belief of being ranked first and their CRRA coefficient are.

The drawback of this approach is that it implicitly assumes that the effect of each of these variables (ability, beliefs, and risk preferences) on tournament entry is linear and separable. For this reason, we also develop a measure of competitiveness that incorporates beliefs and risk preferences in a non-separable way through maximization of expected utility. To do so, we assume students maximize a CRRA utility function, which allows us to take into account heterogeneity in risk preferences.¹⁷ Let q_i be the number of sums i answered correctly in the Tournament round. Recall that the piece-rate compensation scheme pays \$0.50 per sum with certainty while the tournament compensation scheme pays \$2.00 per sum if the student is ranked first in her group and nothing otherwise. Then, the utility of the piece-rate (P) compensation scheme is $U_i^P(q_i) = (0.50 \times q_i)^{1-\rho_i} / (1-\rho_i)$, and the expected utility of the tournament (T) compensation scheme is $EU_i^T(q_i, p_i^{1st}) = p_i^{1st} (2.00 \times q_i)^{1-\rho_i} / (1-\rho_i)$, where ρ_i is i 's CRRA coefficient obtained from the risk elicitation task and p_i^{1st} is i 's subjective belief of being ranked first in her group in the Tournament round.¹⁸ In the absence of other considerations, utility-maximizing students would choose the tournament compensation scheme if $EU_i^T \geq U_i^P$, and the piece-rate compensation scheme otherwise. Now, let τ_i be a dummy that equals 1 if i chooses the tournament compensation scheme in the Choice task and 0 otherwise. Our second measure of competitiveness is then:

$$\text{Competitiveness}_i = \begin{cases} 1 & \text{if } \tau_i = 1 \text{ and } EU_i^T < U_i^P, \\ 0 & \text{if } \tau_i = 1 \text{ and } EU_i^T \geq U_i^P, \\ 0 & \text{if } \tau_i = 0 \text{ and } EU_i^T \leq U_i^P, \\ -1 & \text{if } \tau_i = 0 \text{ and } EU_i^T > U_i^P. \end{cases}$$

In other words, a student is overly competitive if she enters the tournament when she

¹⁷We also tried a measure of competitiveness assuming a linear utility function, which imposes risk neutrality for all students. Suggesting an important role of heterogeneous risk preferences in the measure of competitiveness, our results are weaker with linear utility.

¹⁸Technically, the belief that matters when deciding whether to pick the tournament compensation scheme or not is the probability that one's expected performance in the Choice round (conditional on choosing tournament) ranks first when compared with the performance of other group members in the Tournament round. However, as long as students expect to perform at least as well as in the Tournament round, their beliefs about relative performance in the Tournament round are sufficient to capture the relevant beliefs for the tournament entry decision in the Choice round.

should not and is averse to competition when the converse is true. The remaining “neutral” students make the *correct* choice, that is, enter the tournament when they should (based on utility maximization) and do not enter when they should not.

If we look at the distribution of this measure of competitiveness, we find that 58% of the students make the correct or neutral choice, about 27% are classified as averse to competition, and the remaining 15% are classified as overly competitive. We also see a clear gender difference: 32% of female students compete “too little” versus only 20% of male students and only 12% of females compete “too much” versus 20% of males. Panel B in Table 2 shows the mean and standard deviation of our measure of competitiveness. The mean is negative due to there being more individuals who are averse to competition than individuals who are overly competitive. Wilcoxon rank-sum tests indicate that males are significantly more competitive than females. Thus, consistent with the regression results and the previous literature (see Niederle and Vesterlund, 2011), we find that men are more competitive than women, even after one takes into account differences in ability, performance beliefs, and risk preferences.

3.4 Experimental measures and sample characteristics

Do demographic characteristics explain the variation in the experimentally derived measures of risk preferences, overconfidence, and competitiveness? To test whether there is a relationship between the sample characteristics presented in subsection 2.4 and the experimental measures derived above, we estimate a series of regressions using each of our experimental measures as the dependent variable and including all the demographic variables in Table 1 as regressors. None of these regressions have a significant F-statistic for joint significance of the included demographic variables ($p = 0.143$ for risk aversion, $p = 0.189$ for overconfidence, $p = 0.602$ for competitiveness),¹⁹ indicating that, besides gender, observable characteristics such as age, race, parental income and education, SAT scores, and university grades, etc., are not good predictors of our experimental measures. This is perhaps not surprising given the construction of our key experimental variables: our confidence measure is constructed based on the student’s beliefs about his or her performance net of the student’s true performance; and our competitiveness measure is constructed taking into account heterogeneity in risk preferences and the student’s subjective beliefs. Therefore, our analysis suggests that our experimental measures capture independent variation in individual characteristics that would be otherwise unobservable in standard datasets.

¹⁹Including all the demographic variables in Table 1 as controls in the regression reported in Table 3 has no discernible effect on the coefficients of the experimental variables. In addition, a test of joint significance for the demographic controls does not reject the null hypothesis ($p = 0.195$).

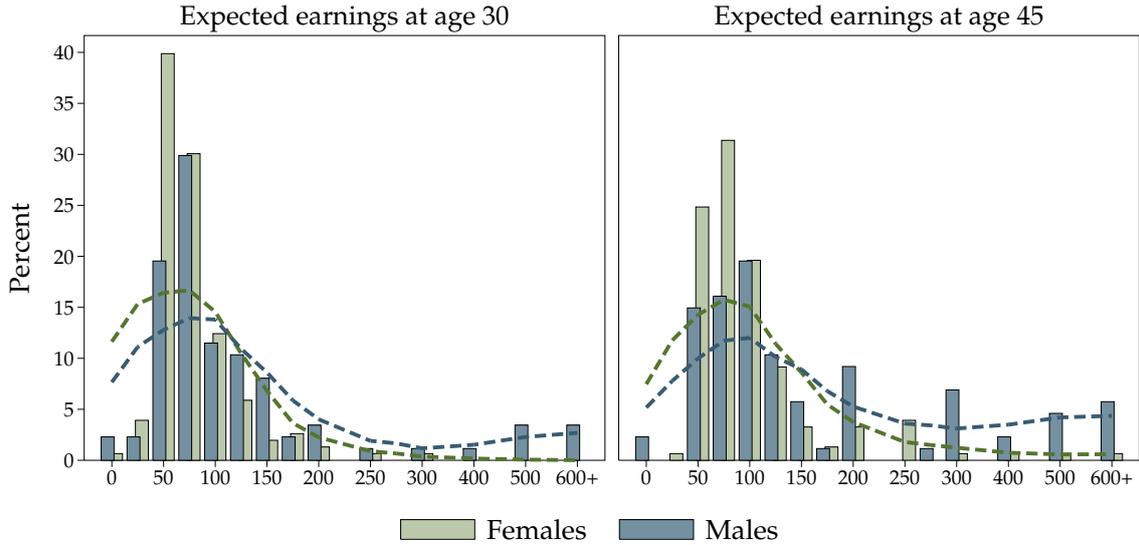


Figure 1: Expected earnings distributions by gender

Note: The bars show the actual distribution of expected earnings while the dashed lines show the same distributions using lowest smoothing. Expected earnings are in \$1000s.

4 Expectations about future earnings

In this section, we first establish that there is an important gender gap in expectations about future earnings, and then, we investigate whether our experimental measures of risk aversion, overconfidence, and competitiveness help explain this gender difference.

4.1 Gender differences in earnings expectations

We elicit the students’ expectations about their own earnings at ages 30 and 45 conditional on graduating in each major category as follows: “*If you received a Bachelor’s degree in each of the following major categories and you were working full time when you are 30 [45] years old, what do you believe is the average amount that you would earn per year?*”. To ensure consistency of the reported expectation across students, we provide a definition of working full time (“working at least 35 hours per week and 45 weeks per year”) and instruct them to ignore the effects of price inflation. We also asked them to incorporate in their response the possibility they might receive an advanced/graduate degree by age 30 (45).²⁰ Given the questions condition on full time/full year labour force participation, our measure of expected earnings is free from biases associated with different labour supply expectations.

We start by analysing the students’ expectations about future earnings for their chosen major, i.e., their actual expected earnings as opposed to the counterfactual expected

²⁰We use a series of practice questions to familiarize students with the format of these questions.

earnings for majors not chosen.²¹ Figure 1 depicts the distributions of these expectations at ages 30 and 45 for both males and females. As is typical for realized earnings distributions, the distributions are positively skewed. It is also clear that the expected earnings distribution of males is shifted to the right and displays a thicker right tail. A Wilcoxon rank-sum test confirms that the distributions of expected earnings differ significantly by gender ($p < 0.001$ both at age 30 and age 45).

The gender differences in expected earnings can also be seen in Panel A of Table 4. In addition to their expected earnings, the table also displays the change in each student’s expected earnings from age 30 to age 45 (labelled “Growth in expected earnings”). For each expectation, the table reports the mean and standard deviation by gender, the difference in means between males and females, and the p -value of testing for equality of means between males and females.²² As we can see, female students clearly expect to earn less than male students and this difference increases with age: on average, females expect to earn around 31% less at age 30, which increases to 39% less by age 45.

While the preceding analysis deals with students’ beliefs about their *own* future earnings, in order to assess how much the students know about the current *population* distribution of earnings, we also asked for each student’s belief about the average earnings of 30-year old individuals of their own gender who graduated with a degree from the same major category as the student (labelled “Expected population earnings”).²³ We compare this to the actual average earnings of the equivalent major \times gender group (labelled “True population earnings”), which we computed from the *National Survey of College Graduates*. Comparing the students’ expectations about their own earnings with their beliefs about population earnings reveals that students believe their earnings will be much higher than the average US college graduate of the same gender and major. This is not surprising given that the students in our sample are drawn from a selective private university and, as revealed by the high average SAT scores and GPA, are of high ability.

One possible reason for the gender difference in earnings expectations is that men and women are misinformed about the distribution of earnings. Table 4 shows that students’ beliefs about the gender gap in average population earnings are quite similar to the true

²¹For younger students, their “chosen” major refers to the major they intend to major in.

²²Since a few outliers may unduly affect our results, all expectations are winsorized at the 2nd and 98th percentiles. Even at these fairly conservative winsorizing values, we still allow for substantial variation in earnings expectations. For earnings expectations at age 30, where the mean is \$84k, the winsorized values range from \$15k to \$300k. For earnings expectations at age 45, where the mean is \$111k, the winsorized values range from \$20k to \$500k.

²³The precise wording of the question is “Among all male [female] college graduates currently aged 30 who work full time and received a Bachelor’s degree in each of the following major categories, what is the average amount that you believe these workers currently earn per year?”.

Table 4: Descriptive statistics for expected earnings

Note: For each expectation, the first two columns report the mean and standard deviation (in parentheses) by gender. The third column reports the difference between males and females and the rightmost column reports the p -value of testing for equality of distributions between males and females based on Wilcoxon rank-sum tests. All expectations are in \$1000s and are winsorized at the 2nd and 98th percentiles.

	Males	Females	Difference	p-value
PANEL A: CONDITIONAL ON THEIR CHOSEN MAJOR				
Expected earnings at age 30	112.00 (74.28)	77.08 (40.85)	34.92	$p < 0.001$
Expected earnings at age 45	165.06 (139.96)	101.06 (71.07)	64.00	$p < 0.001$
Growth in expected earnings	42.79 (66.24)	21.59 (37.64)	21.20	$p = 0.066$
Expected population earnings (age 30)	74.74 (36.17)	61.85 (25.67)	12.88	$p = 0.005$
True population earnings (age 30)	67.36	54.37	12.99	
True population earnings (age 45)	105.87	67.86	38.01	
PANEL B: MEAN OVER ALL MAJOR CATEGORIES				
Expected earnings at age 30	90.12 (62.80)	73.89 (44.77)	16.23	$p = 0.001$
Expected earnings at age 45	120.99 (105.56)	95.28 (72.96)	25.71	$p = 0.006$
Growth in expected earnings	28.82 (52.96)	19.20 (35.61)	9.62	$p = 0.040$
Expected population earnings (age 30)	65.21 (31.65)	61.12 (32.72)	4.19	$p = 0.122$
True population earnings (age 30)	66.05	55.92	10.13	
True population earnings (age 45)	109.88	74.82	35.06	

gender gap.²⁴ Female students believe average earnings for a 30 year old women in their chosen major are 17% less than those of what male students believe average earnings are for men in their chosen major. The students' beliefs about the major-specific gender gap are actually not far from the actual 19% gender gap in the US census data. In other words, we find no evidence that the gender gap in earnings beliefs is mainly driven by systematic misperceptions about population earnings.²⁵

As noted above, an important component of the gender gap in earnings among college graduates is that men and women choose very different fields of study, with men choosing higher paying majors. Therefore, the gender difference in earnings expectations in Panel A of Table 4 may simply be because of the different major composition by gender. An important characteristic of our dataset is that we gathered the students' expected earnings

²⁴Our sample of students is still too young for us to test the relationship between expected and realized earnings directly.

²⁵This is not to say that there are no systematic biases in our students' expected population earnings. We observe that both males and females overestimate the level of population earnings by around \$8k, i.e., the average error (belief - truth) is about \$8k. We simply find a small difference between the perceived gender gap in average earnings and the true gender gap.

for *all* major categories (Business and economics, Engineering and computer science, Humanities and other social sciences, Natural sciences and math, and Never graduate/drop out), not just for the one they have chosen. This allows us to decompose the gender gap in expected earnings using the students' expectations for each major directly rather than make assumptions regarding the counterfactual earnings students would expect in majors not chosen. In contrast to Panel A of Table 4, which computes expectations for the one major chosen, Panel B of Table 4 computes expected earnings for each student by simply averaging each student's expected earnings across all major categories (i.e., weighting each major choice equally). This is equivalent to computing expected earnings by first randomly assigning major choices to the students rather than using the students' self-selected major.

Comparing Panel A and B of Table 4 then allows us to assess how much self-selection affects expected future earnings, and therefore, how much of the gender gap in expected earnings is due to men and women choosing different fields. We find that even if majors are randomly assigned, female students still expect to earn significantly less than male students (Wilcoxon rank-sum tests, $p \leq 0.006$). However, the difference between genders narrows considerably: from \$34.92k (31%) to \$16.23k (18%) at age 30, and from \$64.00k (39%) to \$25.71k (21%) at age 45.²⁶ In other words, differences in major choices account for around one third of the gender gap in expected earnings, which leaves the remaining two thirds to differences in expected earnings *within* each major. Hence, we conduct our subsequent analysis in two steps. First, we examine the relation between the students' expected earnings and their level of risk aversion, overconfidence, and competitiveness, irrespective of their chosen major. Second, we examine the relation between the students' major choice and these experimental measures.

4.2 Experimental measures and expected earnings

To examine whether the students' beliefs about future earnings are systematically correlated with their preferences for risk, overconfidence, and competitiveness, we estimate regressions of the form:

$$Earn_{k,i} = \beta_0 + \beta_1 Male_i + \beta_2 CRRA_i + \beta_3 Overconfidence_i + \beta_4 Competitiveness_i + \gamma X_i + \epsilon_{k,i},$$

where $Earn_{k,i}$ is i 's subjective belief about earnings in major category k , where $k =$ Business, Engineering, Humanities, Natural Sciences, Drop out; $Male_i$ is a dummy that equals one if i is male; $CRRA_i$ is i 's coefficient of relative risk aversion; $Overconfidence_i$ is i 's

²⁶By taking the average across all major categories we are giving each major equal weight. However, other weights lead to a similar result. For instance, if we weight expected earnings based on the observed distribution of chosen majors, the gender gap narrows to \$21.52k (23%) at age 30 and \$38.60k (25%) at age 45.

overestimation of her probability of ranking first; $Competitiveness_i$ is either i 's choice of tournament compensation²⁷ (labelled "Tournament entry") or the measure of competitiveness based on expected utility maximization (labelled "EU competitiveness"); X_i is a vector of control variables; and $\epsilon_{k,i}$ is the error term. Except for our measures of competitiveness, we standardize the continuous independent variables to have a mean of zero and a standard deviation of one to facilitate the interpretation of the coefficients. We use the students' beliefs across all five major categories and cluster standard errors at the individual level.

Table 5 presents the estimates of our regressions. We use two different dependent variables: the students' expected earnings at age 30 and at age 45. To minimize the likelihood of outliers driving our results, we winsorize the dependent variables at the 2nd and 98th percentiles.²⁸ For each dependent variable we run six regressions. In column I, we include only $Male_i$ as an independent variable. As expected, the coefficient of $Male_i$ is positive and statistically significant in both regressions, confirming the existence of a gender gap in expected earnings. In column II, we include the additional demographics control variables described in subsection 2.4.²⁹ The inclusion of these variables, including SAT scores, race, and family background characteristics, has little effect on the gender gap in expectations.

In columns III and IV, we add our experimental measures for risk aversion, overconfidence, and competitiveness (III uses tournament entry and IV uses EU competitiveness). In columns V and VI, we include both the experimental measures and the demographics control variables.

We find a positive relation between competitiveness and expected earnings. With tournament entry as the measure of competitiveness, the effect is stronger for age 30 than age 45 earnings: *ceteris paribus*, competitive individuals expect to earn \$13.4k more at age 30 ($p = 0.007$) and \$7.5k more at age 45 ($p = 0.364$). With our improved measure of competitiveness, the effect is of a similar magnitude and is statistically significant for earnings expectations at both ages: individuals who are overly competitive (averse to competition) expect their age 30 earnings to be \$8.9k higher (lower) and their age 45

²⁷As Buser et al. (2014), when we use i 's choice of tournament compensation as a measure of competitiveness, we also control for i 's probability of ranking first. Note that $Overconfidence_i$ and $CRRRA_i$ already capture i 's beliefs and risk preferences.

²⁸Our results are qualitatively similar if we instead winsorize at the 5th and 95th percentiles.

²⁹Specifically, we include all the variables in Table 1. Moreover, since the students' beliefs in the survey might be affected by their experience in the preceding experiment (e.g. because of changes in their mood, Schwarz and Clore, 1983), we also include all the variables they received feedback on before taking the survey. Namely, their individual performance in all addition tasks and whether they ranked last in the piece-rate round, which was revealed in the additional rounds.

Table 5: The gender gap in expected earnings

Note: OLS estimate with robust standard errors clustered at the individual level. The dependent variables are in \$1000s and are winsorized at the 2nd and 98th percentiles. All regressions have 5 observations for each of the 240 students (one for each of the major categories: Business, Engineering, Humanities, Natural Sciences, and Drop out), resulting in a total of 1200 observations. ***, **, * and * denote statistical significance at 1%, 5%, and 10%, respectively.

Independent variables	Expected earnings at age 30						Expected earnings at age 45					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Male	16.23*** (5.01)	16.77*** (5.04)	12.47** (4.96)	13.91*** (4.85)	12.28** (4.74)	12.91** (4.88)	25.71*** (9.19)	24.81*** (9.05)	20.97** (8.53)	23.11*** (8.57)	18.08** (8.06)	18.85** (8.54)
Tournament entry			12.62** (5.14)		13.44*** (4.95)				7.92 (9.06)		7.48 (8.23)	
EU competitiveness				7.95** (3.54)		8.87** (3.63)				11.86* (6.39)		13.56** (6.26)
Overconfidence			2.15 (3.03)	4.22** (2.01)	2.45 (3.23)	5.27* (2.71)		7.70 (5.05)		6.41* (3.41)	10.52** (5.34)	9.45** (3.94)
CRRRA coefficient			0.72 (1.92)	1.12 (2.03)	0.04 (1.86)	0.52 (1.98)		3.92 (3.23)		4.62 (3.37)	1.68 (3.35)	2.86 (3.31)
Constant	73.89*** (2.30)	73.70*** (2.23)	70.63*** (2.80)	75.76*** (2.49)	70.40*** (2.73)	76.24*** (2.47)	95.28*** (4.09)	95.61*** (3.97)	94.10*** (5.69)	97.76*** (4.44)	95.31*** (5.47)	99.52*** (4.38)
Controls	No	Yes	No	No	Yes	Yes	No	Yes	No	No	Yes	Yes
R ²	0.02	0.05	0.04	0.04	0.07	0.06	0.02	0.08	0.03	0.03	0.08	0.08

earnings to be \$13.6k higher (lower) than competitively “neutral” individuals ($p = 0.015$ for age 30 and $p = 0.031$ for age 45). In the supplementary materials, we show that the effect of competitiveness is driven mostly by the low earnings expectations of students who are averse to competition as opposed to high earning expectations by overly competitive students (see Table SM-1).

Similarly, Table 5 reveals a positive relation between expected earnings and overconfidence. Higher levels of overconfidence are associated with higher expected earnings at ages 30 and 45. The effect is strongest in regressions that use EU competitiveness and control for demographic characteristics. In these regressions, a one-standard deviation increase in overconfidence is associated with a significant increase in expected earnings of \$5.27k at age 30 ($p = 0.053$) and around \$9.45k at age 45 ($p = 0.017$). Although a positive relation between overconfidence and expected earnings might not appear surprising, we do not think it is obvious that overestimating one’s relative rank in a simple arithmetic task correlates with earnings expectations. Moreover, as we discuss next, it is insightful to know that overconfidence helps explain part of the gender difference in earnings expectations. Lastly, given the importance of earnings expectations in the educational and career choices of individuals, overconfident beliefs might nevertheless lead to differences in realized earnings.

Unlike with the other experimental measures, we do not find a consistent relation between earnings expectations and risk aversion. The coefficient of $CRRA_i$ is not significant in any of the regressions ($p > 0.171$).³⁰

Notably, including the experimental measures in the regressions reduces the magnitude of the coefficient of $Male_i$, indicating that part of the gender gap in expected earnings can be accounted for by these variables. Specifically, with the inclusion of these variables, the gender gap narrows by around 23% for age 30 expectations (from a male coefficient of \$16.77k to \$12.91k in regressions with control variables) and around 25% for age 45 expectations (from \$24.81k to \$18.85).

How large are these magnitudes? One way to judge their importance is to compare the impact on the gender gap from the inclusion of our experimental measures to that of the inclusion of the more standard demographic variables (i.e., comparing the changes in the $Male_i$ coefficient from columns I and II vs. columns I and III or IV). Our three experimental measures explain a larger proportion of the gender differences than a rich set of variables capturing ability and family background. This result suggests that these ex-

³⁰See Table SM-2 in the supplementary materials for equivalent regressions including the 17 students whose lottery choices were inconsistent with expected utility maximization. The only perceptible difference of including these students is that the significance of the coefficient for overconfidence weakens slightly in some specifications.

perimental measures are key elements of the gender gap and are capturing individual characteristics that are not otherwise well proxied by standard variables, which is consistent with our previous result of no statistically significant relationship between demographic controls and our experimental measures.

Lastly, note that even though the coefficient of $Male_i$ decreases when we include both experimental and demographic control variables, there is still a significant gender gap in expected earnings that is unaccounted for by these variables. We conclude that although our experimental measures (and additional control variables) are important to our understanding of gender differences in earnings expectations, they are only part of the explanation.

4.3 Competitiveness, overconfidence, and other beliefs

Population earnings

One may argue that differences in the earnings beliefs due to overconfidence or competitiveness are a consequence of differences in the distribution of expected population earnings. In particular, it might be the case that overconfident students expect higher earnings not because they overestimate their own earnings but because they overestimate population earnings. Therefore, it is possible that beliefs about average population earnings are positively associated with competitiveness. To determine whether this is the case, we run regressions with the same specification as the regressions in columns III and IV of Table 5, but as dependent variables, we use the students' expected population earnings (see Table 4) and the difference between expected and true population earnings (i.e. the error in beliefs about the population). The resulting estimates are available in the supplementary materials (see Table SM-5). We find that the coefficients for overconfidence are small and not statistically significant ($p > 0.485$). Thus, overconfident students do not display higher expected earnings because their expected population earnings are higher or more inaccurate. The same result is obtained with both our measures of competitiveness ($p > 0.786$ for tournament entry and $p > 0.151$ for EU competitiveness).

Labour supply

Another possibility is that overconfident and competitive students expect higher earnings because they expect to work more hours. Our survey elicited the average number of hours students expected to be working, conditional on working full time at age 30.³¹

³¹The precise wording of the question is: “If you received a Bachelor’s degree in each of the following major categories and you were working full time when you are 30 years old, what do you believe is the average number of hours you would work per week?”.

To determine whether competitive and overconfident students expect to work more, we run regressions with the same specification as the regressions in columns III and IV of Table 5, but use the students' expected number of work hours as the dependent variable (see the supplementary materials, Table SM-5). We find that the coefficients for overconfidence and both measures of competitiveness are not statistically significant ($p > 0.388$ for overconfidence and $p > 0.404$ for competitiveness). Thus, overconfident and competitive students do not display higher expected earnings because they expect to work more. It should also be pointed out that the results in Table 5 remain qualitatively unaffected, if we add expected number of work hours as a control.

Earnings uncertainty

Finally, competitive individuals may have higher earnings expectations if they expect to enter more tournaments. However, if they over-enter tournaments, they are also likely to have higher earnings uncertainty. In addition to their expected mean earnings, our survey also asked students about the probability that their earnings will exceed \$35k and \$85k in each major category.³² A student's answers to these questions provide some information on beliefs about the expected variance in her future earnings. To provide a direct measure of variance, we calculate each student's interquartile range of their future earnings distribution assuming the earnings expectations of student i for major category k follow a log-normal distribution with mean $\mu_{i,k}$ and variance $\sigma_{i,k}^2$. We compute the value of $\sigma_{i,k}^2$ that best fits with the three data points that we elicit from each student and for each major and then use it to calculate the distribution interquartile range (see the supplementary materials for details). To determine whether competitive and overconfident students perceive higher earnings uncertainty, we run regressions with the same specification as the regressions in columns V and VI of Table 5, but use the students' interquartile range as the dependent variable (see the supplementary materials, Table SM-5).

The coefficients of both measures of competitiveness are positive and close to statistical significance ($p = 0.127$ for tournament entry and $p = 0.082$ for EU competitiveness), which is consistent with competitive students anticipating that they will enter tournaments at work and thus face higher variation in earnings. Similarly, in some specifications, we find that the coefficient for overconfidence is positive ($p > 0.068$), suggesting that overconfident students expect both higher earnings and higher earnings uncertainty.

³²The precise wording of the questions is: "What do you believe is the percent chance that you would earn: (1) at least \$85,000 per year, (2) at least \$35,000 per year, when you are 30 years old if you worked full time and you received a Bachelor's degree in each of the following major categories?"

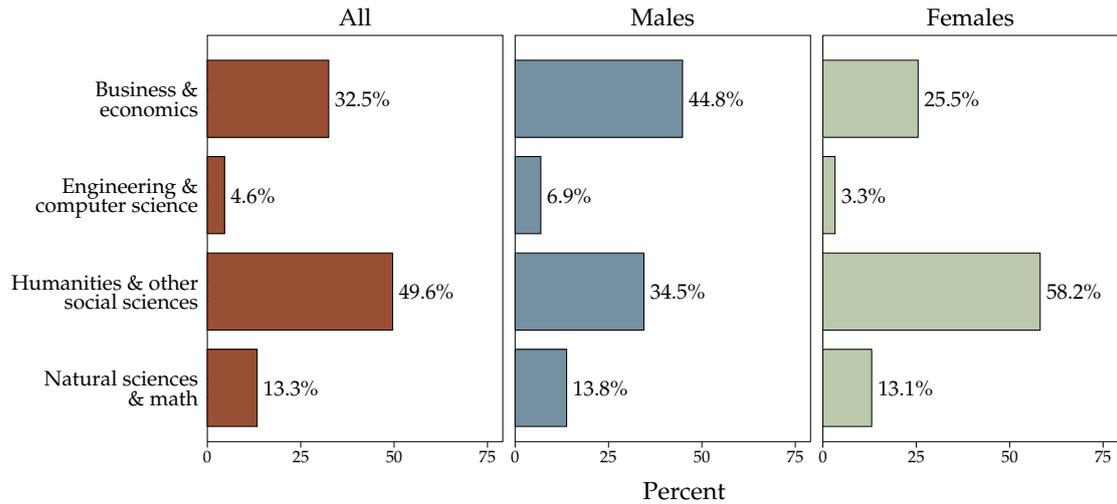


Figure 2: Major category distributions by gender

5 Major choice

We turn to the second part of our analysis, and examine whether the students’ levels of risk aversion, overconfidence, and competitiveness help explain gender differences in major choice. Figure 2 depicts the distribution of the students’ major choice. Most students choose a major in the “Humanities and other social sciences” (henceforth *humanities*), followed by “Business and economics” (henceforth *business*), “Natural sciences and math” (henceforth *natural sciences*), and then “Engineering and computer science” (henceforth *engineering*). Moreover, there is a strong and significant gender difference in their choice of a college major (χ^2 test, $p = 0.003$): while 44.8% of the male students major in business and only 34.5% major in humanities, 58.2% of females major in humanities and only 25.5% major in business.³³

5.1 Student perceptions of college majors

Before analysing their major choice, we use questions from the survey to look at how students perceive the riskiness, difficulty, returns, and competitiveness of jobs in each major category. Descriptive statistics for these questions are shown in Table 6.

The first variable in the table serves as a measure of difficulty. Specifically, it is

³³Our sample is weighted more toward business and economics majors than in the actual NYU population graduating in 2010, possibly because the experimental laboratory is located in the building of the faculty of economics. For the NYU population of students who graduated in 2010 (IPEDS), the fraction of students completing degrees in each field are as follows: For women, 14.1 percent graduated in economics or business, 71.7 percent in humanities, arts, or other social sciences, and 13.7 percent in natural sciences, engineering, or computer science. For men, 31.1 percent graduated in economics or business, 61.2 percent in humanities, arts, or other social sciences, and 7.8 percent in natural sciences, engineering, or computer sciences. Engineering and computer science were less than 2 percent of graduates.

the expected number of study hours students need to graduate with a GPA of 4.0 in a major category.³⁴ According to this measure, both males and females consider the natural sciences and engineering the most difficult, followed by business, which leaves humanities as the least difficult major category. Given that overconfident students consider themselves as more capable than others, if overconfidence plays a role in their major choice then we ought to see that students from the natural sciences and engineering are more overconfident.

Our survey design also included a number of variables to measure the students' perceptions about the level of competition in jobs within a major category. The next three rows of Table 6 describe various measures of a major's competitiveness, namely: (1) the importance of relative performance for job compensation, (2) the probability of being fired, and (3) the fraction of male employees.³⁵ Table 6 shows that both male and female students expect jobs in business to be consistently competitive. According to the ratio of bonus pay and the fraction of male employees, jobs in engineering and natural sciences are more competitive than jobs in the humanities. This ordering reverses for the probability of being fired, where jobs in engineering and natural sciences are considered safer than jobs in the humanities. Hence, if competitiveness matters for major choice, we ought to see a higher fraction of underconfident students in the humanities compared to business and to a lesser extent in engineering and the natural sciences.

The second to last variable in Table 6 gives us an indication of the variability of the earnings expectations of each student in each major category, and reports the interquartile range of the earnings distribution (which, as explained in subsection 4.3, is obtained from fitting the three points on the students' earnings beliefs distribution to a log-normal distribution). Compared to humanities, both males and females consider business to have more variable earnings and females think the same is true for the natural sciences and engineering. Hence, if risk aversion plays a role in major choice then we ought to see that risk averse students self-select themselves into the humanities.

Finally, Table 6 also reports the student's beliefs about the average population earnings

³⁴The wording of the question is “*How many hours per week do you think you would need to spend studying (excluding class time) in each of the following major categories in order to achieve an average GPA in that major category of 4.0?*”

³⁵The wording of the questions is: (1) “*What do you believe would be the average amount of bonus pay based on relative performance (as a percent of your annual base pay) among the job offers you receive at age 30 if you received a Bachelor's degree in each of the following major categories?*” (2) “*What do you believe would be the percent chance of being fired or laid off in the next year from positions similar to those from which you would receive job offers at age 30 if you received a Bachelor's degree in each of the following major categories?*” and (3) “*What do you believe would be the proportion of men in positions similar to those from which you would receive job offers at age 30 if you received a Bachelor's degree in each of the following major categories?*”

Table 6: Student perceptions of majors

Note: The table reports mean and standard deviations (in parentheses) separately for male (M) and female (F) students. Earnings expectations are in \$1000s and are winsorized at the 2nd and 98th percentiles. For each variable and gender, the last column reports the statistical significance of pairwise Wilcoxon rank-sum tests comparing the three major categories: \ggg , \gg , and $>$ indicate a significant difference at 1%, 5%, and 10% respectively; \approx indicates there is no significant difference at 10%; major categories are identified by their initial.

		Business	Engineer- ing	Humani- ties	Natural sciences	Statistical comparisons
Study hours needed for a 4.0 GPA	M	23.54 (17.20)	28.76 (18.55)	20.05 (16.97)	28.51 (18.39)	E \approx N \ggg B \ggg H
	F	25.40 (12.61)	29.73 (16.31)	19.59 (10.38)	28.04 (13.68)	E \ggg N \ggg B \ggg H
Fraction of salary based on performance pay	M	0.46 (0.54)	0.19 (0.22)	0.13 (0.27)	0.14 (0.19)	B \ggg E \ggg N \ggg H
	F	0.39 (0.49)	0.26 (0.39)	0.15 (0.24)	0.18 (0.26)	B \ggg E \ggg N \ggg H
Probability of being fired	M	0.15 (0.18)	0.09 (0.11)	0.10 (0.10)	0.08 (0.08)	B \ggg H \approx E \approx N
	F	0.18 (0.17)	0.14 (0.15)	0.18 (0.17)	0.13 (0.14)	B \approx H \ggg E \approx N
Fraction of male employees	M	0.61 (0.16)	0.61 (0.19)	0.42 (0.14)	0.53 (0.18)	E $>$ B \ggg N \ggg H
	F	0.62 (0.16)	0.66 (0.18)	0.42 (0.13)	0.56 (0.17)	E \ggg B \ggg N \ggg H
Expected earnings uncertainty	M	50.54 (51.49)	40.93 (36.01)	37.19 (25.15)	40.85 (35.63)	B \gg E \approx N \approx H
	F	44.91 (37.82)	46.54 (35.14)	34.71 (23.68)	43.62 (31.72)	E \approx B \approx N \ggg H
Expected population earnings	M	89.31 (41.64)	81.58 (30.30)	54.74 (16.30)	68.48 (22.29)	B \ggg E \ggg N \ggg H
	F	81.67 (35.34)	78.29 (30.78)	52.56 (21.64)	65.93 (31.57)	B \approx E \ggg N \ggg H

for each major. Both males and females believe average earnings for business majors are the highest, followed by engineering, the natural sciences, and in last place the humanities. Female beliefs about the average earnings of female workers are quite similar. While it is difficult to conclude that these beliefs uniquely reflect beliefs about the difficulty or competitiveness of the major, the ordering of majors is similar as for other major characteristics.

5.2 Experimental measures and major choice

To evaluate whether major choice is systematically correlated with our experimental measures of individual attributes, we estimate alternative-specific conditional logit regressions (McFadden, 1974), where we allow the latent utility of each major choice to depend on characteristics of the major, characteristics of the student, and interactions of major and student characteristics. The latent utility to individual i from completing major k is

$$V_{k,i} = \gamma_k + \beta_k X_i + \alpha Y_{i,k} + \epsilon_{k,i}, \quad (1)$$

where γ_k is a major-specific fixed effect; X_i is a set of variables that vary only across individuals (e.g., gender); $Y_{i,k}$ is a vector of variables that vary across major categories

within the same individual (e.g., each student’s expected future earnings in each major); and $\epsilon_{k,i}$ is the error term, assumed to have an extreme value distribution that gives rise to the logit form. By allowing the coefficients β_k to vary across major categories, we allow for the individual attributes in X_i , including our experimentally derived measures of risk, competitiveness, and confidence, to have differential effects on the utility for each major.³⁶ Given the extreme value distribution assumption, the probability of completing major k is given by $p_{k,i} = \exp(\bar{V}_{k,i}) / \sum_j \exp(\bar{V}_{j,i})$, where $\bar{V}_{k,i}$ denotes $V_{k,i}$ net of $\epsilon_{k,i}$. We estimate the marginal effects $\partial p_{k,i} / \partial X_i$ and $\partial p_{k,i} / \partial Y_{k,i}$.

Table 7 presents the estimated marginal effects from four conditional logit regressions. For convenience, each regression is presented in four columns, each containing the estimates for one major. In all regressions, X_i contains a dummy indicating the students’ gender ($Male_i$), the experimentally derived variables that measure risk aversion ($CRRA_i$), overconfidence ($Overconfidence_i$), either the choice of tournament compensation (regressions I and III) or the measure of competitiveness based on expected utility maximization (regressions II and IV), and the additional control variables described in subsection 2.4 and used in Table 5. In regressions II and III, $Y_{i,k}$ is empty while in regressions III and IV we explore the impact of earnings expectations on major choice by including i ’s earnings expectations in major k in $Y_{i,k}$. We use earnings expectations at age 45 since long-run earnings ought to be the more important determinant of major choice. Except for our measures of competitiveness, we standardize the continuous independent variables to have a mean of zero and a standard deviation of one to facilitate the interpretation of the marginal effects.

Our findings are as follows. First, we find that, as hypothesized, overconfident students are more likely to choose a major in the natural sciences (from $p = 0.035$ in III to $p = 0.077$ in II). Albeit, we do not find evidence of a statistically significant effect for engineering or business. For the humanities, the size of the marginal effect of overconfidence is considerable and in the hypothesised direction but it does not reach statistical significance (from $p = 0.105$ in III to $p = 0.152$ in II). Second, we do not find support for our hypotheses concerning competitiveness and major choice. That is, we do not find that more competitive students are significantly over-represented in business ($p > 0.121$) or under-represented in the humanities ($p > 0.447$). In fact, the marginal effect for business is negative and the one for humanities is positive, which is the converse of what one would expect to find since the humanities should be the least competitive major category and business the most competitive.³⁷ Third, consistent with the literature on major choice,

³⁶Note that specifications in which the vector of major-specific variables $Y_{i,k}$ is empty are equivalent to a standard multinomial logit regression.

³⁷In Table SM-3 of the supplementary materials, we separate the EU competitiveness measure into one

Table 7: The gender gap in major choice

Note: Marginal effects of logit estimates. Robust standard errors for the marginal effects clustered at the individual level are reported in parentheses. All regressions have major and individual fixed effects, and 4 observations for each of the 240 students (one for each of the major categories: Business, Engineering, Humanities, and Natural Sciences), resulting in a total of 960 observations. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

Independent variables	Business				Engineering				Humanities				Natural Sciences			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
Male	0.20** (0.09)	0.20** (0.08)	0.15* (0.08)	0.14* (0.08)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	-0.20** (0.09)	-0.22** (0.09)	-0.16* (0.09)	-0.18* (0.09)	-0.02 (0.05)	0.00 (0.05)	-0.01 (0.05)	0.01 (0.05)
Tournament entry	-0.07 (0.08)	-0.09 (0.08)	-0.09 (0.08)		0.03 (0.03)	0.03 (0.03)	0.03 (0.03)		0.02 (0.09)		0.03 (0.10)		0.02 (0.05)		0.02 (0.05)	
EU competitiveness		-0.08 (0.06)		-0.09 (0.06)		0.02 (0.02)		0.02 (0.02)		0.03 (0.07)		0.05 (0.07)		0.03 (0.04)		0.03 (0.04)
Overconfidence	0.05 (0.05)	0.03 (0.04)	0.06 (0.05)	0.03 (0.05)	0.03 (0.02)	0.01 (0.01)	0.02 (0.01)	0.01 (0.01)	-0.10 (0.06)	-0.08 (0.05)	-0.11 (0.07)	-0.08 (0.06)	0.07** (0.03)	0.06* (0.03)	0.07** (0.04)	0.06* (0.03)
CRRRA coefficient	0.03 (0.04)	0.02 (0.04)	0.03 (0.03)	0.02 (0.03)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)	0.02 (0.04)	-0.03 (0.03)	-0.03 (0.03)	-0.03 (0.03)	-0.03 (0.03)
Expected earnings			0.11*** (0.04)	0.11*** (0.04)		0.01 (0.01)	0.01 (0.01)	0.01 (0.01)			0.13*** (0.04)	0.13*** (0.04)			0.05** (0.02)	0.05*** (0.02)

columns III and IV both show that students select into majors that they believe will provide them with relatively higher earnings (see Arcidiacono, 2004; Arcidiacono et al., 2013; Wiswall and Zafar, 2015). This effect is strongest for business and the humanities.

One concern with these results is that students from different majors could have differentially selected to participate in our study. For example, if humanities students who are willing to participate in incentivized laboratory experiments are more competitive than the average humanities student and business students who have time to participate in experiments are less competitive than the average business student, then the impact of competitiveness on major choice would be biased downwards. While we cannot rule out this possibility or test it (since these measures do not exist for the entire population of NYU students), we believe it is unlikely. First, the NYU CESS laboratory runs many different types of experiments, which attract a variety of students. However, the content of individual experiments are not conveyed through the recruitment materials, and therefore, students with a particular trait cannot self-select into specific experiments. Second, although there are no studies evaluating self-selection into laboratory experiments depending on competitiveness, several studies report that the social and risk preferences of participants of laboratory experiments do not vary from those of other students (e.g., Cleave et al., 2013; Abeler and Nosenzo, 2015). Evidence of differential selection across majors is scarcer. However, both Krawczyk (2011) and Falk et al. (2013) find that social preferences do not predict selection into laboratory experiments among students majoring in business and economics as well as among students majoring in other fields.

Lastly, we should also note that, in contrast to what we see in the regressions in Table 5, our experimental measures and additional control variables do not help explain the large gender difference in major choice. For example, the 19 percentage point difference between men and women majoring in business is comparable to the estimated 20 percentage point difference given by the marginal effect for $Male_i$ in regressions I and II.

Reverse Causality?

Our experimental measures are collected from students after they are in college, and when they have potentially been exposed to different experiences in the various majors. A potential concern could then be reverse causality. For example, if competitiveness is

for students who are averse to competition and one for students that are overly competitive. We find that, compared to neutral students, students who are overly competitive or averse to competition are more likely to major in natural sciences and are less likely to major in humanities. In Table SM-4 we include the 17 students whose lottery choices were inconsistent with expected utility maximization. The only discernible effect of including these students is that the effect of overconfidence on majoring in the humanities is now significantly negative ($p < 0.043$).

taught in certain majors, such as business, then the interpretation of estimates in Table 7 is not clear. However, if this concern were taken at its face value (that certain majors “teach” competitiveness), we would expect to find results biased in the direction of finding a systematic relationship between competitiveness and major choice. Instead, we find no evidence of that in Table 7. Nonetheless, as a further robustness test, we reran the estimations excluding students who are beyond their third year in college. Arguably, younger students have more similar coursework experiences, and their choice of college major is still reversible. Estimates based on this restricted sample are very similar to those presented in Table 7, suggesting that such concerns cannot explain our results. It should also be pointed out that sunk investments in particular forms of human capital are intrinsic to the sequential nature of educational investments. Administering experiments along the lines done in this study to individuals before they attend college does not rule out the concern that individuals may have different classroom experiences in earlier grades.

Why are competitiveness and overconfidence not related to major choice?

Regressions III and IV of Table 7 show that earnings expectations are a significant determinant of major choice. Therefore, at first, it may seem puzzling that the positive and significant relationships between earnings expectations, competitiveness, and overconfidence (documented in Section 4) do not have a stronger effect on major choice. However, closer examination reveals that the associations between the experimental measures and earnings expectations exist in *each* major category and not only for, say, their chosen major. This can be seen in Figure 3, which depicts the students’ expected earnings in each major category depending on their EU competitiveness and on whether they are overconfident or underconfident. To better observe the effect of competitiveness and overconfidence, expected earnings are standardized to have a mean of zero and a standard deviation of one *within each major category*. Since, we observe the same pattern in *all* majors, it is conceivable for competitiveness and overconfidence to affect earnings expectations and at the same time have a muted impact on major choice in spite of relative earnings affecting the latter decision. These findings are consistent with competitiveness and overconfidence having an impact on the expected workplace trajectories of individuals conditional on major choice.

In summary, competitiveness and overconfidence help explain the gender gap in expected earnings within majors and thus might help explain gender differences within a given career (such as the glass ceiling phenomenon, Bertrand and Hallock, 2001; Albrecht et al., 2003), but they do not help explain the gender gap in major choice and thus might not be good candidates to explain gender differences in career choice.

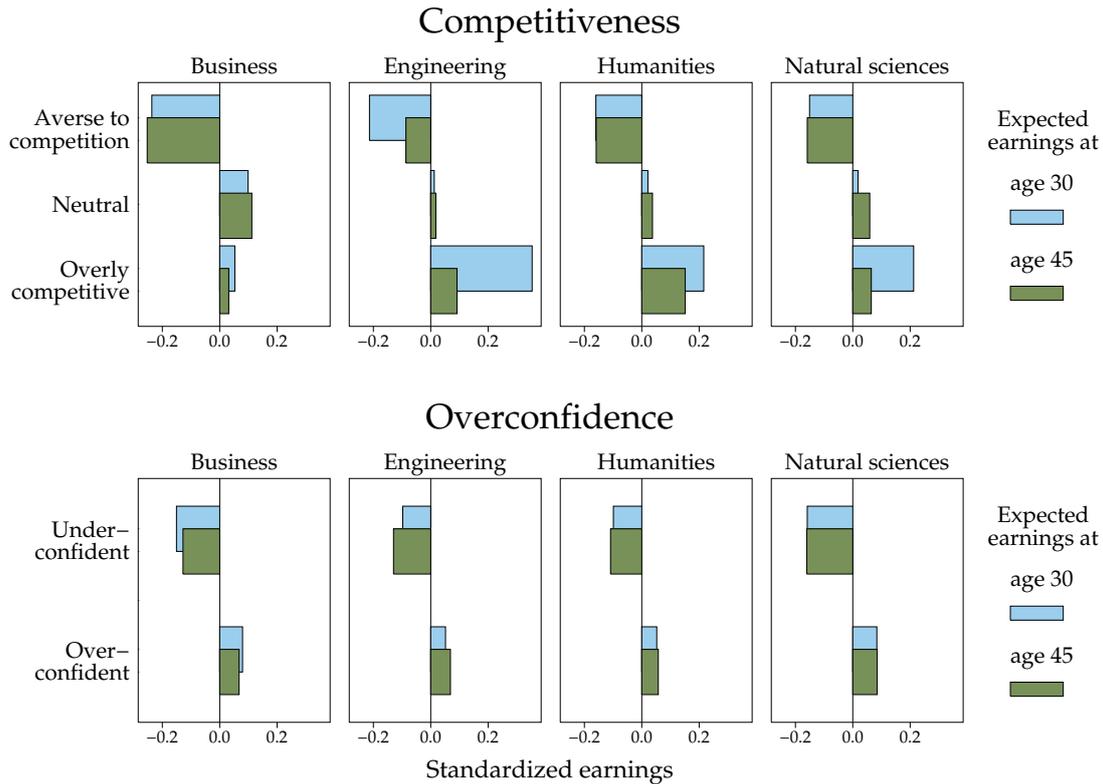


Figure 3: Earnings, competitiveness, and overconfidence by major category
Note: Expected earnings are standardized within each major category.

6 Conclusion

Our research combines an experiment and survey of expectations to investigate the gender gap in education choices and labour market earnings expectations. Our analysis reveals two key findings. First, we extend the prior research by showing that confidence and competitiveness, but not risk preferences, are related systematically to students' expectations about future earnings and help explain an important proportion of the gender gap in earnings expectations. Second, we show that while earnings expectations are related to major choice, there is no direct relationship between college field of study (aggregated up to broad major categories) and the experimental measures. These findings provide important insights into the underlying reasons behind the observed gender differences in the labour market outcomes such as persistent differences in occupational choice and the glass ceiling phenomenon.

At first, it may seem puzzling that earnings expectations—which are significant determinants of major choice—are positively and significantly related with competitiveness and overconfidence, yet these measures do not have a direct effect on major choice. As we show, the associations between the experimental measures and earnings expectations exist in each major category and not only for, say, the chosen major of the student. It is

then conceivable for competitiveness and overconfidence to affect earnings expectations, and at the same time have a muted impact on major choice. This does raise the question of why these measures are not independently related to major choice? One possible factor is that our survey lumps majors in broad science, humanities, and business categories, which may hide important sources of heterogeneity. Within the broad fields, individuals can choose different majors and anticipate working in different occupations.³⁸ However, given that males and females may choose very different occupations even within very fine occupations/majors (Goldin and Katz, 2011), it is not clear to what extent our findings would change if the categorization of majors were finer.³⁹ It may, therefore, be easier to observe an association between the experimentally-measured individual attributes and future earnings expectations because expectations incorporate beliefs about individual-specific decisions such as pursuing a graduate degree, training investments, occupational choices, and negotiating and bargaining behaviour within occupations. Our findings show that students have already internalized their level of competitiveness and confidence, and this has affected their beliefs about future labour market outcomes 10-25 years later. Therefore, the relationship between earnings expectations and the experimental measures can be seen as a kind of summary measure of the anticipated influences of these traits on future labour market choices and outcomes, regardless of the source.

Our findings also underscore the importance of combining experimental measures of individual traits with more traditional surveys of labour market behaviour and beliefs.⁴⁰ We find that our experimental measures explain nearly the same proportion of the gender gap in earnings expectations as do traditional demographic variables, such as test scores and family background. In addition, we find that these same traditional demographic variables are weakly correlated with the experimental measures and therefore poor proxies, which indicates that the experimental measures provide real added value to the analysis of gender in the labour market.

Why do competitiveness and overconfidence positively relate to earnings expectations? This is an open question to which our data cannot provide a clear answer. Individuals with different levels of confidence and competitiveness may expect to pursue different occupations on the extensive margin and more aggressively negotiate for salary on the intensive margin. Undercompetitive and underconfident individuals may anticipate choosing less

³⁸Note, however, that this factor also applies to the Buser et al. (2014) context, where the broad high school tracks map into fields of study in college, which then map into labour market occupations.

³⁹When we looked at the precise major that students are pursuing, we did not find any notable differences in the specific majors that the two genders are choosing within our broad major categories.

⁴⁰A small and growing literature studies the link between experimental measures and actual behaviour in the field (e.g., Karlan, 2005; Ashraf et al., 2006; Benz and Meier, 2008; Sapienza et al., 2009; Fehr and Leibbrandt, 2011; Hopfensitz and Miquel-Florensa, 2013; Zhang, 2013; Buser et al., 2014).

remunerative occupations, even within major categories (Kleinjans, 2009).⁴¹ While the occupational distribution conditional on major can explain a large part of the earnings differences across majors (Phipps and Ransom, 2010), the mapping of majors to occupations is far from one-to-one. For example, within medicine, the proportion of female physicians differs substantially across specialties, ranging from almost 70% to less than 10% percent (Goldin and Katz, 2011). Even conditional on choosing the same occupation, undercompetitive and underconfident individuals may have different earnings trajectories because they believe they are less likely to enter and/or win tournaments (i.e., promotions in the workplace).⁴² Undercompetitive and underconfident individuals may be less likely to negotiate earnings, which may impact their starting earnings as well as wage trajectories (Babcock and Laschever, 2003; Rigdon, 2012). Finally, it is possible that competitive and overconfident individuals are simply overly optimistic when predicting their future earnings, even if they expect to work in similar occupations and have comparable success in promotions and salary negotiations. In this regard, future studies that follow students over several years would be valuable to determine whether the accuracy of individuals' earnings expectations is related to traits such as competitiveness, overconfidence, or risk aversion.

To conclude, we would like to stress that our findings are based on a sample of students with relatively high ability and socioeconomic status. The composition of our sample might be important given the results of Almås et al. (2015), who find that gender differences in competitiveness are stronger among individuals with high socioeconomic status. This suggests that competitiveness is potentially a better explanation for gender differences in the labour market outcomes among the well-off (e.g., for the glass ceiling phenomenon Bertrand et al., 2010). Hence, an interesting topic for future research is to explore the relationship between competitiveness and earnings expectations among different socioeconomic strata. Moreover, given that students may have selectively chosen to participate in the study, and this selection may vary by field of study, it would be useful to replicate our study with more representative samples.

⁴¹Flory et al. (2015), for example, find that women are less likely to apply to jobs with more competitive payment schemes.

⁴²If competitive individuals are more likely to enter tournaments, it would make sense for their expected earnings to be higher. However, by entering more tournaments, their earnings uncertainty should increase. Using our measure of earnings uncertainty, we find a positive relationship between uncertainty and competitiveness (but not between uncertainty and overconfidence).

References

- Abeler, J. and Nosenzo, D. (2015). Self-selection into laboratory experiments: pro-social motives versus monetary incentives. *Experimental Economics*, 18:195–214.
- Albrecht, J., Björklund, A., and Vroman, S. (2003). Is there a glass ceiling in Sweden? *Journal of Labor Economics*, 11:145–177.
- Almås, I., Cappelen, A. W., Salvanes, K. G., Sorensen, E., and Tungodden, B. (2015). Willingness to compete: Family matters. *Management Science*, forthcoming.
- Andersen, S., Ertac, S., Gneezy, U., List, J. A., and Maximiano, S. (2013). Gender, competitiveness and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics*, 95:1438–1443.
- Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121:343–375.
- Arcidiacono, P., Hotz, J., and Kang, S. (2013). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, 166:3–16.
- Armantier, O., van der Klaauw, W., Nelson, S., Topa, G., and Zafar, B. (2015). Inflation expectations and behavior: Do survey respondents act on their beliefs. *International Economic Review*, 56:505–536.
- Ashraf, N., Karlan, D., and Yin, W. (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *The Quarterly Journal of Economics*, 121:635–672.
- Babcock, L. and Laschever, S. (2003). *Women Don't Ask: Negotiation and the Gender Divide*. Princeton University Press, Princeton, NJ.
- Balafoutas, L. and Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science*, 335:579–582.
- Beaman, L., Duflo, E., Pande, R., and Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science*, 335:582–586.
- Belzil, C. and Leonardi, M. (2007). Can risk aversion explain schooling attainments? evidence from Italy. *Labour Economics*, 14:957–970.
- Bénabou, R. and Tirole, J. (2015). Bonus culture: Competitive pay, screening, and multitasking. *Journal of Political Economy*, forthcoming.
- Bengtsson, C., Persson, M., and Willenhag, P. (2005). Gender and overconfidence. *Economics Letters*, 86:199–203.
- Benz, M. and Meier, S. (2008). Do people behave in experiments as in the field?—evidence from donations. *Experimental Economics*, 11:268–281.
- Bertrand, M. (2011). New perspectives on gender. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, pages 1543–1590. Elsevier, Amsterdam.
- Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2:228–255.
- Bertrand, M. and Hallock, K. (2001). The gender gap in top corporate jobs. *Industrial and Labor Relations Review*, 55:3–21.

- Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59:960–970.
- Black, D. A., Haviland, A. M., Sanders, S. G., and Taylor, L. (2008). Gender wage disparities among the highly educated. *Journal of Human Resources*, 43:630–659.
- Blau, F. and Kahn, L. (2000). Gender differences in pay. *Journal of Economic Perspectives*, 14:75–99.
- Blau, F. and Kahn, L. (2012). The sources of the gender pay gap. In Grusky, D. B. and Kricheli-Katz, T., editors, *The New Gilded Age: The Critical Inequality Debates of our Time*, pages 843–860. Stanford University Press, Stanford, CA.
- Booth, A. and Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81:542–555.
- Brown, C. and Corcoran, M. (1997). Sex-based differences in school content and the male-female wage gap. *Journal of Labor Economics*, 15:431–465.
- Buser, T., Niederle, M., and Osterbeek, H. (2014). Gender, competitiveness and career choices. *The Quarterly Journal of Economics*, 129:1409–1447.
- Cason, T. N., Masters, W. A., and Sheremeta, R. M. (2010). Entry into winner-take-all and proportional-prize contests: An experimental study. *Journal of Public Economics*, 94:604–611.
- Cleave, B. L., Nikiforakis, N., and Slonim, R. (2013). Is there selection bias in laboratory experiments? the case of social and risk preferences. *Experimental Economics*, 16:372–382.
- Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47:448–474.
- de Paula, A., Shapira, G., and Todd, P. E. (2014). How beliefs about HIV status affect risky behaviors: Evidence from Malawi. *Journal of Applied Econometrics*, 29:944–964.
- Dohmen, T. and Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, 101:556–590.
- Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100:1238–1260.
- Eckel, C. C. and Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. In Plott, C. R. and Smith, V. L., editors, *Handbook of Experimental Economics Results, volume 1*, pages 1061–1073. Elsevier, Amsterdam.
- Falk, A., Meier, S., and Zehnder, C. (2013). Do lab experiments misrepresent social preferences? the case of self-selected student samples. *Journal of the European Economic Association*, 11:839–852.
- Fehr, E. and Leibbrandt, A. (2011). A field study on cooperativeness and impatience in the tragedy of the commons. *Journal of Public Economics*, 95:1144–1155.
- Flory, J. A., Leibbrandt, A., and List, J. A. (2015). Do competitive workplaces deter female workers? a large-scale natural field experiment on gender differences in job-entry decisions. *Review of Economic Studies*, 82:122–155.
- Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77:1637–1664.
- Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104:1091–1119.

- Goldin, C. and Katz, L. (2011). The cost of workplace flexibility for high-powered professionals. *The Annals of the American Academy of Political and Social Science*, 638:45–67.
- Gupta, N. D., Poulsen, A., and Villeval, M.-C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, 51:816–835.
- Healy, A. and Pate, J. (2011). Can teams help to close the gender competition gap? *The Economic Journal*, 121:1192–1204.
- Hegewisch, A., Williams, C., and Harbin, V. (2013). The gender wage gap by occupation. Fact sheet c350a, Institute for Women’s Policy Research.
- Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92:1644–1655.
- Hopfensitz, A. and Miquel-Florensa, P. (2013). Public good contributions among coffee farmers in Costa Rica: Cooperativists and private market participants. Working paper 13-395, Toulouse School of Economics.
- Hyde, J., Fennema, E., and Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107:139–155.
- Jacob, B. and Wilder, T. (2011). Educational expectations and attainment. In Duncan, G. J. and Murnane, R. J., editors, *Whither Opportunity? Rising Inequality, Schools, and Children’s Life Chances*, pages 133–162. Russell Sage Press, New York, NY.
- Jarrell, S. B. and Stanley, T. (2004). Declining bias and gender wage discrimination? a meta-regression analysis. *Journal of Human Resources*, 39:828–838.
- Karlan, D. (2005). Using experimental economics to measure social capital and predict financial decisions. *The American Economic Review*, 95:1688–1699.
- Kleinjans, K. (2009). Do gender differences in preferences for competition matter for occupational expectations? *Journal of Economic Psychology*, 30:701–710.
- Krawczyk, M. (2011). What brings your subjects to the lab? a field experiment. *Experimental Economics*, 14:482–489.
- Lundeberg, M. A., Fox, P. W., and Punóchař, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, 86:114–121.
- Malmendier, U. and Tate, G. (2008). Who makes acquisitions? ceo overconfidence and the market’s reaction. *Journal of Financial Economics*, 89:20–43.
- Manski, C. (2004). Measuring expectations. *Econometrica*, 72:1329–1376.
- McFadden, D. L. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York, NY.
- Moen, E. R. and Rosén, A. (2005). Performance pay and adverse selection. *Scandinavian Journal of Economics*, 107:279–298.
- Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115:502–517.
- Niederle, M., Segal, C., and Vesterlund, L. (2012). How costly is diversity? affirmative action in light of gender differences in competitiveness. *Management Science*, 59:1–16.
- Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122:1067–1101.

- Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annual Review in Economics*, 3:601–630.
- Nielsen, H. S. and Vissing-Jorgensen, A. (2006). The impact of labor income risk on educational choices: Estimates and implied risk aversion. Working paper, Northwestern University.
- Phipps, A. and Ransom, M. (2010). Career and occupational implications of college majors: Evidence from the national survey of college graduates. Working paper, Brigham Young University.
- Reuben, E., Rey-Biel, P., Sapienza, P., and Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organisation*, 83:111–117.
- Reuben, E., Sapienza, P., and Zingales, L. (2014). How stereotypes impair women’s careers in science. *Proceedings of the National Academy of Sciences of the United States of America*, 111:4403–4408.
- Rigdon, M. (2012). An experimental investigation into gender differences in wage negotiations. Working paper, Rutgers University.
- Sapienza, P., Zingales, L., and Maestripieri, D. (2009). Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences of the United States of America*, 106:15268–15273.
- Schwarz, N. and Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45:513–523.
- Stinebrickner, T. and Stinebrickner, R. (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics*, 30:707–748.
- Weinberger, C. (1998). Race and gender wage gaps in the market for recent college graduates. *Industrial Relations*, 37:67–84.
- Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *Review of Economic Studies*, 82:791–824.
- Wolpin, K. I. and van der Klaauw, W. (2008). Social security and the savings and retirement behavior of low income households. *Journal of Econometrics*, 145:21–42.
- Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, 48:545–595.
- Zhang, Y. J. (2013). Can experimental economics explain competitive behavior outside the lab? Working paper 2292929, SSRN.