

Reuben, Ernesto; Wiswall, Matthew; Zafar, Basit

Working Paper

Preferences and biases in educational choices and labor market expectations: Shrinking the black box of gender

Staff Report, No. 627

Provided in Cooperation with:

Federal Reserve Bank of New York

Suggested Citation: Reuben, Ernesto; Wiswall, Matthew; Zafar, Basit (2013) : Preferences and biases in educational choices and labor market expectations: Shrinking the black box of gender, Staff Report, No. 627, Federal Reserve Bank of New York, New York, NY

This Version is available at:

<https://hdl.handle.net/10419/93659>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Federal Reserve Bank of New York
Staff Reports

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

Ernesto Reuben
Matthew Wiswall
Basit Zafar

Staff Report No. 627
August 2013



This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the authors and are not necessarily reflective of views at the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

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

Ernesto Reuben, Matthew Wiswall, and Basit Zafar

Federal Reserve Bank of New York Staff Reports, no. 627

August 2013

JEL classification: D81, D84, I21, I23, J10

Abstract

Standard observed characteristics explain only part of the differences between men and women in education choices and labor market trajectories. Using an experiment to derive students' levels of overconfidence, and preferences for competitiveness and risk, this paper investigates whether these behavioral biases and preferences explain gender differences in college major choices and expected future earnings. In a sample of high-ability undergraduates, we find that competitiveness and overconfidence, but not risk aversion, are systematically related with expectations about future earnings: Individuals who are overconfident and overly competitive have significantly higher earnings expectations. Moreover, gender differences in overconfidence and competitiveness explain about 18 percent 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, including test scores and family background, and they are poorly proxied by these same control variables, underscoring that they represent independent variation. While expected earnings are related to college major choices, the experimental measures are not related with college major choice.

Key words: college majors, earnings, gender differences, subjective expectations, risk aversion, overconfidence, competitiveness

Reuben: Columbia University and IZA (e-mail: ereuben@columbia.edu). Wiswall: Arizona State University (e-mail: matt.wiswall@gmail.com). Zafar: Federal Reserve Bank of New York (e-mail: basit.zafar@ny.frb.org). The authors thank the NYU Center for Experimental Social Sciences (CESS) for providing assistance in conducting the survey and experiment. They thank participants at presentations at the 2012 North American ESA Meetings and the University of Essex. Grant Graziani and Victoria Gregory provided outstanding research assistance. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

1 Introduction

While considerable progress has been made to explain gender differences in educational choices and labor market trajectories, residual differences remain unaccounted for by standard variables, such as experience, education, industries, and occupations (Blau and Kahn, 2000; Jarrell and Stanley, 2004; Black et al., 2008). What accounts for the remaining differences? An active area of inquiry documents behavioral or psychological differences between men and women, often using laboratory experiments (for a recent summary see Bertrand, 2011). Whether these documented gender differences correlate with behaviors and beliefs outside the lab remains an open question.

In this paper, we combine experimentally-derived measures of competitiveness, risk aversion, and overconfidence with a survey of labor market expectations and education choices of undergraduate students at New York University, a selective private university. Our analysis follows research documenting that men have a higher level of “competitiveness”—a preference for tasks where rewards are allocated through competition with others (tournaments) rather than through non-competitive means (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).

Replicating previous findings, within our sample of high ability college students, we find substantial gender differences in each experimentally-derived measure. 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. Finally, we find that men are twice as likely as women to choose to participate in the tournament, and 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 our survey of expectations, women and men anticipate future gender gaps in realized earnings: controlling for differences in majors, women expect to earn on average 19% less than men at age 30, and 23% less than men at age 45.¹ In analyzing the combined experi-

¹While there is no direct counterpart to the expected earnings data in realized earnings—the survey data on expectations is about future, unrealized earnings—it is worth noting that expected earnings mirror gender gaps in realized earnings for all US college graduates, with women’s average earnings 17% less than those of men at age 30, and 36% lower than men’s at age 45 (controlling for differences in major composition between genders). Note that our sample of expectations is from a high ability population at an elite private university, and is

ment 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 since we collect data on students’ earnings expectations for *all* majors (as defined by aggregated major categories) and not simply their *chosen* major, 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 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. Furthermore, our results suggest that competitiveness and overconfidence may also explain gender differences in trajectories in the workplace. These findings, based on a sample of high ability students attending an elite university (that is, precisely the kind of students who have a realistic chance of making 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).

Observed gender wage gaps can be the result of gender differences in expected earnings within each major/occupation, as well as gender differences in major/occupational choices. In fact, college major composition has been shown to explain an important part of the earnings differences between men and women (Brown and Corcoran, 1997; Weinberger, 1998; Arcidiacono, 2004). Our results regarding earnings expectations are within major. In the second

therefore not representative of the US population. 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 closely match the true average population earnings.

part of the paper, we investigate the link between the experimental measures and college major choice. We find a gender difference in major choices in our sample, with males more likely to major in business and sciences and women more likely to major in humanities, which mirrors observed gender differences in major choice in nationally representative data of the US (Gemici and Wiswall, 2013). However, 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 three 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.

Our analysis of earnings expectations follows a recent and growing literature which collects and uses subjective expectations data to understand decision-making under uncertainty (for a survey of this literature, see Manski, 2004). In the context of schooling choices, Arcidiacono et al. (2013), Wiswall and Zafar (2013), and Zafar (2013) incorporate subjective expectations into models of college major choices and find that education choices are correlated with earnings expectations—students are more likely to self-select into fields in which they expect relatively higher earnings. While we cannot test the relationship between expected and realized earnings directly—our sample of students is still too young for us to observe their realized earnings—it is important to understand why men and women form such different earnings expectations. In fact, given that realized earnings and other labor market outcomes can be affected by a number of unanticipated events (and may suffer from the problem of reverse causality), we argue that investigating why young men and women form such 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. Furthermore, evidence from different domains—education, investment, labor, retirement, health—shows that expectations tend to be good predictors of choices, above and beyond standard determinants, and that individuals act on their expectations (Wolpin and van der Klaauw, 2008; Jacob and Wilder, 2011; Beaman et al., 2012; Armantier et al., 2012; Stinebrickner and Stinebrickner, 2012; de Paula et al., 2013). Then, it becomes all the more important to understand what determines these expectations, since individuals are likely to act upon them and they can be

come self-fulfilling. For example, in all likelihood, individuals with low earnings expectations are more willing to accept a low-paying job offer because it is in line with their beliefs, and they are also less likely to negotiate for a higher salary since the offer is consistent with their benchmark.

Our finding of a lack of relationship between competitiveness and perceived major choice therefore contrasts with the findings in a concurrent study of Buser et al. (2013), which correlates the same type of competitiveness measure to high school tract choice among Dutch students. 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 tract choice, with boys more likely to choose the prestigious science and health tracts over the less prestigious humanities tracts. 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. Our study is 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. In addition, the two settings (US and Europe) may differ in how prestigiousness relates with fields of study (for the European context, they argue that prestigiousness of educational profiles perfectly correlates with their math and science intensity), and how the fields of study map into occupations. Our data are similar to theirs (though our experimental measures are derived somewhat differently), and we additionally have data on earnings expectations. Therefore, our study complements their work by showing that competitiveness and confidence measures strongly relate to earnings expectations, and that these measures can even explain gender differences *within* careers.

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 later, 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 categories, which may hide important sources of heterogeneity. Within the broad science, humanities, and business fields, individuals can choose different majors and antici-

pate working in different occupations.² Therefore, it may 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 behavior within occupations. Therefore, the relationship between earnings expectations and competitiveness and overconfidence can be seen as a kind of summary measure of the anticipated influences of these traits on future labor market choices and outcomes, regardless of the source.

Finally, 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 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 remunerative occupations, even within major categories (Kleinjans, 2009).³ 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, male and female physicians are likely to choose very different specialties (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).⁴ Undercompetitive and underconfident individuals may be less likely to negotiate earnings, which may impact their starting earnings as well as wage trajectories (Babcock and

²Note, however, that this factor also applies to the Buser et al. (2013) context, where the broad high school tracts map into fields of study in college, which then map into labor market occupations. Furthermore, 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. Furthermore, when 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.

³Flory et al. (2010), for example, find that women are less likely to apply to jobs with more competitive payment schemes.

⁴If 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. Since we collect data on students' earnings distribution within each major, we have measures of earnings uncertainty. We find no systematic relationship between earnings uncertainty and overconfidence or competitiveness. This then suggests that, regardless of whether such individuals are more likely to enter tournaments or not, they believe they are more likely to win them.

Laschever, 2003; Rigdon, 2012). Finally, competitiveness and overconfidence, as measured in the lab experiments, may simply proxy for certain psychological traits. For example, Muller and Schwieren (2012) relate competitiveness to the Big Five personality traits, and find that more competitive individuals have lower degrees of neuroticism, and that neuroticism impairs performance.

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 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.⁵ In each round, students first read the instructions for that round. Subsequently, they make the required choices and, if

⁵To 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.

necessary, perform the addition task. In this paper, we analyze the data of the first four rounds, which we describe below.⁶ 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 a 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.

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 two 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

⁶Of 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 actual performance relative to one randomly chosen group member in the piece-rate task and eliciting their updated beliefs about their rank and re-eliciting their choice for the submission task.

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.

Lastly, since risk attitudes may be an important determinant of labor 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 are expected utility maximizers, 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 future earnings that they would earn 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.⁷ 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 students relative

⁷We provided students with a link where they could see how each college major maps into our aggregate categories. Before the official survey began, students first answer a few simple practice questions in order to familiarize themselves with the format of the questions.

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 analyze the results.

2.3 Procedures

The study was administered to New York University (NYU) undergraduate students. Students 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 are available in the supplementary materials.

2.4 Sample characteristics

A total of 257 students participated in the study. However, we decided to drop the 11 students (6 males and 5 females) who major in Engineering and Computer Science because it would be problematic to make robust claims about gender differences in these majors based on so few observations.⁸ This leaves us with a sample of 246 students.

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 (34% of our sample is male and 66% is female). 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

⁸It is not unusual to recruit few students from Engineering and Computer Science since it is a very small major at NYU. The proportion of engineering students in our sample is in line with the distribution of majors among NYU graduates in 2011 according to the *Integrated Post-Secondary Education Data System*.

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 ($n = 246$)		Males ($n = 83$)		Females ($n = 163$)		p-value
Age		21.40	(1.22)	21.43	(1.23)	21.38	(1.21)	$p = 0.733$
Race:	White	29.67%		34.94%		26.99%		$p = 0.320$
	Asian	49.19%		48.19%		49.69%		
	Other	21.14%		16.87%		23.31%		
Parents income (\$1000s)		139.64	(123.39)	144.58	(127.80)	137.13	(121.41)	$p = 0.702$
Mother with B.A. or more		67.89%		73.49%		65.03%		$p = 0.179$
Father with B.A. or more		69.92%		71.08%		69.33%		$p = 0.776$
SAT math score		696.42	(80.77)	718.31	(69.40)	684.44	(84.11)	$p = 0.004$
SAT verbal score		676.83	(75.65)	683.12	(67.85)	673.47	(79.54)	$p = 0.588$
GPA		3.46	(0.32)	3.46	(0.33)	3.47	(0.31)	$p = 0.835$
School year:	Freshman	11.38%		10.84%		11.66%		$p = 0.723$
	Sophomore	10.16%		10.84%		9.82%		
	Junior	36.99%		32.53%		39.26%		
	Senior or more	41.46%		45.78%		39.26%		

significant demographics differences between male and female students except for their SAT math score, where males score significantly higher than females ($p = 0.004$).

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 (additional descriptive statistics are available in the supplementary materials). The first column reports statistics for all students, 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 females (the difference is statistically significant in the Tournament and Piece-rate rounds

Table 2: Descriptive statistics of the experiment

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 Wilcoxon rank-sum tests comparing the distributions of males and females.

		All		Males		Females		<i>p</i> -value
PANEL A: EXPERIMENTAL OUTCOMES								
Earnings		39.66	(10.78)	40.83	(12.45)	39.07	(9.80)	<i>p</i> = 0.611
Correct answers:	Tournament	11.85	(3.80)	12.83	(4.75)	11.34	(3.12)	<i>p</i> = 0.032
	Choice	12.59	(4.08)	13.42	(5.06)	12.17	(3.43)	<i>p</i> = 0.196
	Piece-rate	12.88	(4.32)	14.00	(5.01)	12.31	(3.80)	<i>p</i> = 0.022
Subjective probability of ranking 1 st		0.34	(0.26)	0.45	(0.32)	0.28	(0.21)	<i>p</i> < 0.001
Proportion choosing Tournament		35.77%		54.22%		26.38%		<i>p</i> < 0.001
Number of lottery choices		6.76	(2.07)	7.23	(1.80)	6.52	(2.16)	<i>p</i> = 0.004
PANEL B: INDIVIDUAL-SPECIFIC MEASURES								
CRRA coefficient		0.62	(0.99)	0.41	(0.68)	0.73	(1.11)	<i>p</i> = 0.008
Overconfidence		0.09	(0.27)	0.14	(0.27)	0.06	(0.26)	<i>p</i> = 0.041
Competitiveness		-0.15	(0.63)	-0.04	(0.62)	-0.21	(0.63)	<i>p</i> = 0.047
Competitiveness rank		-7.78	(22.19)	-3.14	(18.39)	-10.32	(23.69)	<i>p</i> = 0.030
Residual competitiveness		0.00	(0.41)	0.08	(0.45)	-0.05	(0.39)	<i>p</i> = 0.022

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 26% of females students ($p < 0.001$).⁹ 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

⁹This 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; Andersen et al., 2012; Booth and Nolen, 2012; Gupta et al., 2013).

risk preferences, overconfidence, and competitiveness.

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 .¹⁰ Note that, 17 ($\approx 7\%$) of our students had choice patterns that are inconsistent with expected utility maximization.¹¹ Given that our analysis calls for an accurate measure of risk preferences, we decided that the most prudent step is to drop these students from all subsequent data analysis. However, the results in the paper are robust to including these students and using the first instance of them switching to the certain payoff as their switching point.

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 (0.73 vs. 0.41, Wilcoxon rank-sum test $p = 0.008$). 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, 49% of females and 30% of males exhibit choices that are consistent with risk averse preferences, and 11% of females and 15% of males exhibit choices that are consistent with risk loving preferences.

¹⁰We 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.

¹¹It is a commonly found in the literature that a small fraction of participants, typically around 10%, either switch multiple times or switch once from the certain payoff to the lottery (see Holt and Laury, 2002).

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). 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.¹² Mirroring the gender difference in number of correct sums, men have a significantly higher true probability of being ranked first than women (33% vs. 21%, Wilcoxon rank-sum test, $p = 0.032$).

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.¹³ 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), the mean level of overconfidence is larger for men than for women (14 percentage points vs. 6 percentage points, Wilcoxon rank-sum test, $p = 0.047$).

3.3 Competitiveness

Following Niederle and Vesterlund (2007), we obtain measures of competitiveness using the students’ decision to enter the tournament in the second round. However, in addition to individual differences in competitiveness, it is to be expected that this decision will also be affected by ability, beliefs about relative performance, and risk preferences.¹⁴ Thus, we use

¹²We 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.

¹³Both males and females significantly overestimate their probability of ranking first according to Wilcoxon signed-rank tests ($p < 0.001$)

¹⁴There is evidence of positive selection into the tournament for each of these variables. Students who choose the tournament compensation scheme have a higher performance in the tournament round, a higher belief of

additional data about the students' beliefs and characteristics to construct a series of residual competitiveness measures which net out each student's ability, performance beliefs, and risk preferences. To test the robustness of our competitiveness measures, we also construct an alternative measure based on Buser et al. (2013).

Since risk preferences differ systematically in the population, we use CRRA utility to incorporate the data on heterogeneous risk preferences.¹⁵ 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.¹⁶ Utility-maximizing students 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 first 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 words, a student is overly competitive if she enters the tournament when she 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.

being ranked first, and a lower CRRA coefficient (Wilcoxon rank-sum tests, $p < 0.005$). See the supplementary materials for further analysis of the tournament entry decision.

¹⁵We 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.

¹⁶Technically, 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.

Our second measure of competitiveness follows the same logic as the first, but it uses the additional information contained in the actual difference in utilities between the two compensation schemes. Note that, among students classified as overly competitive, the difference $U_i^P - EU_i^T$ equals the amount of utility that i gives up by choosing the tournament compensation scheme, and therefore it serves as an indication of how competitive i is. Similarly, among students classified as averse to competition, $EU_i^T - U_i^P$ serves as an indication of how averse to competition i is. Based on this observation, we construct the following variable:

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

where $R_i^+(\cdot)$ ranks the competitiveness of i among the overly competitive students (the least competitive gets rank 1) and $R_i^-(\cdot)$ ranks i 's aversion to competition among the students who are averse to competition (the least averse gets rank 1). In other words, a student obtains a high (positive) competitiveness rank if she enters the tournament and the difference in utilities $U_i^P - EU_i^T$ is large compared to others, and a low (negative) competitiveness rank if she does not enter the tournament and the difference in utilities $U_i^P - EU_i^T$ is low compared to others. Students who make the correct entry choice obtain a competitiveness rank of zero. We use ranks as opposed to the actual differences in utilities because the nonlinear nature of the CRRA functional form produces outliers in the distribution of competitiveness.

If we look at the distribution of competitiveness, we find that 58% of the students make the correct or neutral choice, about 28% are classified as averse to competition, and the remaining 14% are classified as overly competitive. We also see a clear gender difference: 32% of female students compete “too little” versus only 21% of male students and only 11% of females compete “too much” versus 17% of males. Panel B in Table 2 shows the mean and standard deviation of our two measures of competitiveness. The means are 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 in both measures. Thus, consistent with previous literature (see Niederle and Vesterlund, 2011), we find that men are more competitive than women, even after ones takes into account differences in ability, performance beliefs, and risk preferences.

An alternative measure of competitiveness

As an alternative to our measures of competitiveness, we consider a measure based on Buser et al. (2013). It consists of first regressing the tournament entry choice (τ_i) on the number of correct sums in the Tournament round (q_i), the subjective probability of being ranked first in the Tournament round (p_i^{1st}), and the CRRA coefficient (ρ_i). The residual of this regression for each student is then used as an indication of how competitive that student is. This measure of competitiveness implicitly assumes that each of these determinants affect tournament entry in a linear and separable way. This is in contrast to the measure described above, which incorporates beliefs and risk preferences in a non-separable way through maximization of expected utility. The mean and standard deviation of the alternative competitiveness variable, which we call “Residual competitiveness”, is available in Table 2. It is positive for males, indicating that the average male student is competitive, and negative for females, indicating the opposite for the average female student. Like the measures of competitiveness constructed above, this measure also differs significantly by gender with a Wilcoxon rank-sum test ($p = 0.022$).

3.4 Experimental measures and sample characteristics

Do demographic characteristics explain the variation in the experimentally derived measures of risk preferences, overconfidence, and competitiveness? Our experimental measures become less important in some sense if student background characteristics are good proxies for them. 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 ($p = 0.249$ for risk aversion, $p = 0.132$ for overconfidence, $p = 0.263$ for competitiveness, and $p = 0.131$ for competitiveness rank), 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 unsurprising 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.

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).¹⁷ Given the questions condition on full time/full year labor force participation, our measure of expected earnings is free from biases associated with different labor supply expectations.

We start by analyzing the students' expectations about future earnings for their chosen major, i.e., their actual expected earnings as opposed to the counterfactual expected earnings for majors not chosen.¹⁸ 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 3. In addition to their expected earnings, the table also displays the change in each student's expected earnings from age 30 to age 45 (labeled “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

¹⁷We use a series of practice questions to familiarize the students with the format of these types of questions.

¹⁸For younger students, their “chosen” major refers to the major that they report they intend to major in.

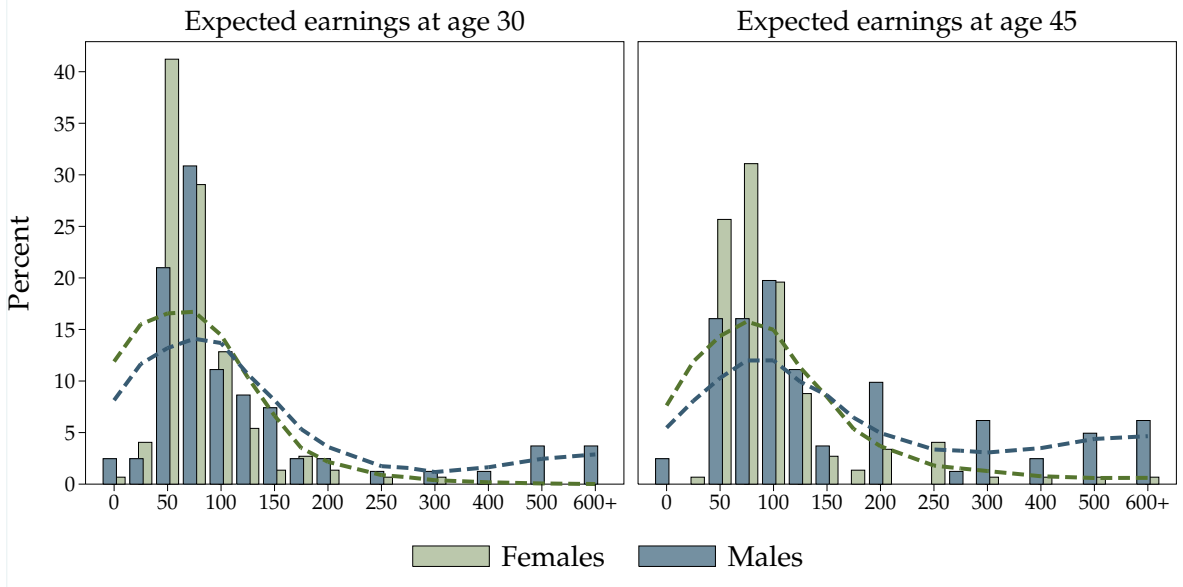


Figure 1: Expected earnings distributions by gender

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

males and females.¹⁹ 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 (labeled “Expected population earnings”).²⁰ We compare this to the actual average earnings of the equivalent major \times gender group (labeled “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

¹⁹Since a few outliers may unduly affect our results, all expectations are winsorized at the 2nd and 98th percentiles. Results are qualitatively similar and gender differences are in fact stronger if we do not winsorize.

²⁰The 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 3: 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	110.79 (76.21)	76.32 (40.96)	34.47	$p < 0.001$
Expected earnings at age 45	165.49 (143.76)	100.89 (72.08)	64.61	$p < 0.001$
Growth in expected earnings	45.86 (72.55)	22.86 (41.40)	23.01	$p = 0.114$
Expected population earnings (age 30)	73.11 (36.10)	61.27 (25.69)	11.84	$p = 0.016$
True population earnings (age 30)	66.25	53.67	12.58	
True population earnings (age 45)	105.40	65.29	40.11	
PANEL B: MEAN OVER ALL MAJOR CATEGORIES				
Expected earnings at age 30	85.32 (40.92)	69.21 (27.50)	16.11	$p = 0.002$
Expected earnings at age 45	117.48 (78.78)	89.88 (47.89)	27.61	$p = 0.007$
Growth in expected earnings	31.62 (41.27)	19.37 (27.64)	12.25	$p = 0.072$
Expected population earnings (age 30)	59.53 (17.45)	56.87 (19.97)	2.67	$p = 0.075$
True population earnings (age 30)	61.97	51.13	10.84	
True population earnings (age 45)	109.13	69.60	39.53	

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 3 shows that students' beliefs about the gender gap in average population earnings are quite similar to the true gender gap. Female students believe average earnings for 30 year old women in their chosen major are 16% less than those of what male students believe average earnings are for men in their chosen major. The student's 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.²¹

²¹This 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 \$7k, i.e., the average error (belief - truth) is about \$7k. We simply find a small difference between the perceived gender gap in average earnings and the true gender gap.

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 3 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 for *all* major categories (Business and economics, 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 student’s expectations for each major directly rather than make assumptions regarding the counterfactual earnings a student would expect in majors not chosen. In contrast to Panel A of Table 3, which computes expectations for the one major chosen, Panel B of Table 3 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 3 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, $p \leq 0.007$). However, the difference between genders narrows considerably: from \$34.47k (31%) to \$16.11k (19%) at age 30, and from \$64.61k (39%) to \$27.61k (23%) at age 45.²² 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

²²By 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 \$24.00k (23%) at age 30 and \$38.60k (27%) at age 45.

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 ; $Male_i$ 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 overestimation of her probability of ranking first; $Competitiveness_i$ is i 's level of competitiveness according to either our first or second measure; 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. Thus, the constant can be interpreted as the earnings belief of an average female student who is neither overly competitive nor averse to competition. We use the students' beliefs across all four major categories and cluster standard errors at the individual level.

Table 4 presents the estimates of our regressions. We use two different dependent variables: the students' expected earnings at age 30 and at age 45. 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 2.4.²³ The inclusion of these variables, including SAT scores, race, and family background characteristics, reduces the gender gap in expectations by about 11 and 13 percent (for age 30 and 45 expectations, respectively).

In columns III and IV, we add our experimental measures for risk aversion, overconfidence, and competitiveness (III uses the first measure of competitiveness and IV uses the second). These regressions show a systematic relation between expected earnings and both overconfidence and competitiveness. Higher levels of overconfidence are associated with higher expected earnings at ages 30 and 45. Specifically, a one-standard deviation increase in overconfidence is associated with a significant increase in expected earnings of around \$4.20k at age 30 (about 6% more than the baseline) and around \$6.60k at age 45 (about 7% more).

Similarly, we find a positive relation between competitiveness and expected earnings. With our first measure of competitiveness, the effect has a higher level of significance for age 45 than age 30 earnings ($p = 0.115$ for age 30 and $p = 0.057$ for age 45). With our second measure

²³Specifically, we include all the variables in Table 1 except for GPA, which suffers from obvious causality problems. 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 their experimental earnings.

Table 4: 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 4 observations for each of the 229 students, resulting in a total of 916 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	16.11*** (5.06)	14.33*** (5.29)	14.34*** (5.02)	14.22*** (4.93)	12.23** (5.28)	11.94** (5.19)	27.61*** (9.57)	24.07*** (9.45)	25.01*** (8.98)	25.05*** (8.95)	20.05** (8.98)	19.58** (8.84)
Competitiveness			5.67 (3.58)		5.43 (3.62)				12.90* (6.75)		13.85** (6.92)	
Competitiveness rank				0.24** (0.11)		0.24* (0.12)				0.46** (0.23)		0.55** (0.25)
Overconfidence			4.15** (2.05)	4.37*** (2.06)	4.20* (2.23)	4.44** (2.25)		6.48* (3.44)		6.77* (3.48)	7.61** (3.26)	8.07** (3.31)
CRRRA coefficient			1.20 (2.27)	3.22 (2.71)	0.54 (2.16)	2.61 (2.73)		4.54 (3.53)		8.23* (4.67)	2.47 (3.46)	7.05 (4.68)
Constant	69.21*** (2.26)	68.48*** (3.95)	70.68*** (2.48)	71.71*** (2.65)	71.39*** (2.50)	72.53*** (2.69)	89.88*** (3.93)	88.53*** (7.79)	92.71*** (4.35)	94.35*** (4.72)	94.60*** (4.31)	96.97*** (4.84)
Controls	No	Yes	No	No	Yes	Yes	No	Yes	No	No	Yes	Yes
R ²	0.02	0.05	0.03	0.03	0.05	0.06	0.02	0.08	0.03	0.04	0.10	0.10

of competitiveness, the significance of the coefficient improves in both the age 30 and age 45 regressions (to $p = 0.041$ and $p = 0.048$, respectively).²⁴ The sign of the estimates for both measures of competitiveness indicates that competitiveness is positively related to earnings expectations. The interpretation of the coefficients on the first measure of competitiveness is that individuals who are overly competitive (averse to competition) expect age 30 earnings to be about \$5.5k higher (lower) and age 45 earnings to be \$13k higher (lower) than competitively “neutral” individuals. 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.

In columns V and VI, we include both the experimental measures and demographics control variables. These regressions show that the positive and statistically significant effects of overconfidence and competitiveness are unaffected by the inclusion of a large set of control variables, with the effect of overconfidence increasing somewhat, especially for age 45 earnings expectations. This robustness to the inclusion of control variables is consistent with our previous results of no statistically significant relationship between demographic controls and our experimental measures.

By and large, we do not find a significant relation between earnings expectations and risk aversion. The coefficient of $CRRA_i$ is significant only in column IV for earnings at age 45, but the significance disappears once we add controls in columns V and VI.

Lastly, note that 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 16.7% for age 30 expectations (from a male coefficient of \$14.33k to \$11.94k in models with control variables) and around 18.7% for age 45 expectations (from \$24.07k to \$19.58).

How large are these magnitudes? One way to judge their importance is to compare the relative magnitude of the reduction in the gender gap from our experimental measures to that from the inclusion of the more standard demographic variables. Comparing the reduction in the $Male_i$ coefficient in columns I and II vs. columns I and III (or IV) indicates that our three experimental measures reduce the gender gap by about as much as the demographic variables for age 30 expectations, and about as much as the demographic variables for age

²⁴We obtain results in the same direction with the residuals measure of competitiveness. Specifically, we obtain a positive coefficient in the regression for earnings at age 30 as well as the regression for earnings at age 45, although the coefficient is statistically significant only in the former ($p = 0.060$ and $p = 0.375$, respectively).

45 expectations. That our three experimental measures explain a similar proportion of the gender differences as a rich set of variables capturing ability and family background, variables including SAT scores and family income, is suggestive that these experimental measures are key elements of the gender gap. That these experimental measures are uncorrelated with these same demographic variables suggests further that the experimental measures are capturing individual characteristics that are not otherwise well proxied by standard variables. Note however, that even though the coefficient of $Male_i$ decreases further 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 Relationship of competitiveness and overconfidence with 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 4, but as dependent variables, we use the students' expected population earnings (see Table 3) 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. We find that the coefficients for overconfidence are small and not statistically significant ($p > 0.295$ for expected population earnings and $p > 0.308$ for the error in beliefs about the population). 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 ($p > 0.183$ for expected population earnings and $p > 0.160$ for the error in beliefs about the population).

Labor 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 30.²⁵ 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 4, but use the students' expected number of work hours as the dependent variable (see the supplementary materials). We find that the coefficient for overconfidence is negative and is not statistically significant ($p > 0.297$). Similarly, both our measures of competitiveness are unrelated with expected work hours ($p > 0.788$ for both measures). 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 4 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.²⁶ 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 standard deviation of future earnings assuming the earnings expectations of student i for major category k follows a log-normal distribution with mean $\mu_{i,k}$ and variance $\sigma_{i,k}^2$, and 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 (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 III and IV of Table 4, but use the student's earnings' standard deviations $\sigma_{i,k}$ as the dependent variable (see the supplementary materials). We find that the coefficient for overconfidence is positive but is not statistically significant ($p > 0.182$). In the same way, the coefficients for both measures of competitiveness are small and not distinguishable from zero ($p > 0.560$ for both measures). Thus, while overconfident and competitive students expect higher earnings,

²⁵The 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?"

²⁶The 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?"

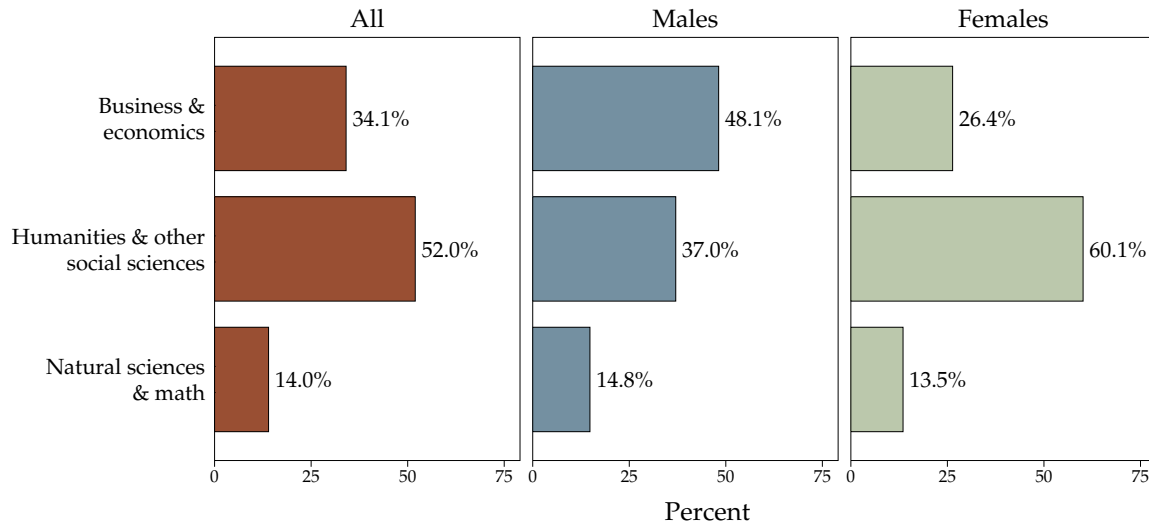


Figure 2: Major category distributions by gender

they do not expect higher earnings uncertainty. This would suggest that, if such individuals are more likely to enter tournaments at work then they must also expect to win more of them.

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*), and then "Natural sciences and math" (henceforth *natural sciences*). However, there is a strong and significant gender difference in their choice of a college major (χ^2 test, $p = 0.002$): while 48.1% of the male students major in business and only 37.0% major in humanities, 60.1% of females major in humanities and only 26.4% major in business.

5.1 Student perceptions of college majors

Before analyzing 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 5.

The first variable in the table serves as a measure of difficulty. Specifically, it is the expected number of study hours students need to graduate with a GPA of 4.0 in a major

category.²⁷ According to this measure, both males and females consider the natural sciences 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 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 5 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.²⁸ Table 5 shows that both male and female students expect jobs in business to be the most competitive. According to the ratio of bonus pay and the fraction of male employees, jobs in natural sciences are the second most competitive and jobs in the humanities are the least competitive. This ordering reverses for the probability of being fired, where jobs in natural sciences are considered the safest and jobs in the humanities the second safest. 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 the natural sciences.

The second to last variable in Table 5 gives us an indication of the variability of the earnings expectations of each student in each major category, and reports the standard deviation in earnings (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. 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 5 also reports the student's beliefs about the average population earnings

²⁷The 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?”. The mean and standard deviation of this variable are 32.53 and 19.44 hours.

²⁸The precise 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 5: Student Perceptions about Majors

Note: The table reports mean and standard deviations (in parentheses). 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	Humani- ties	Natural sciences	Statistical comparisons
Study hours needed for a 4.0 GPA	Males	21.10 (14.44)	17.65 (14.04)	26.59 (16.14)	N \ggg B \ggg H
	Females	25.57 (12.73)	19.74 (10.52)	28.27 (13.77)	N \ggg B \ggg H
Fraction of salary based on relative performance	Males	0.47 (0.55)	0.13 (0.27)	0.14 (0.20)	B \ggg N \ggg H
	Females	0.39 (0.50)	0.16 (0.25)	0.18 (0.27)	B \ggg N \ggg H
Probability of being fired	Males	0.15 (0.19)	0.10 (0.10)	0.08 (0.09)	B \gg H \ggg N
	Females	0.18 (0.17)	0.18 (0.17)	0.13 (0.14)	B \approx H \ggg N
Fraction of male employees	Males	0.62 (0.15)	0.43 (0.15)	0.54 (0.18)	B \ggg N \ggg H
	Females	0.62 (0.16)	0.41 (0.13)	0.55 (0.17)	B \ggg N \ggg H
Expected earnings uncertainty	Males	54.45 (46.05)	38.90 (30.52)	42.51 (35.09)	B \gg N \approx H
	Females	45.39 (36.30)	36.12 (33.07)	43.74 (34.29)	B \approx N \ggg H
Expected population earnings	Males	85.49 (36.28)	52.74 (15.11)	67.30 (22.45)	B \ggg N \ggg H
	Females	81.32 (33.79)	52.18 (21.97)	65.21 (30.29)	B \ggg N \ggg H

for each major. Both males and females believe average earnings for business majors are the highest, with males reporting average earnings for a 30-year-old male full time worker of about \$85k, compared to natural sciences with \$67k and humanities with \$53k. 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 the same 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 models (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 given by

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

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.²⁹ 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 $\epsilon_{k,i}$. Normalizing the model relative to a base major category \tilde{k} , we set the parameters $\gamma_{\tilde{k}} = 0$ and $\beta_{\tilde{k}} = 0$. Given equation (1), the log odds of student i completing major k relative to the baseline major \tilde{k} is then given by

$$\ln \left(\frac{p_{k,i}}{p_{\tilde{k},i}} \right) = \gamma_k + \beta_k X_i + \alpha (Y_{i,k} - Y_{i,\tilde{k}}). \quad (2)$$

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.

Table 6 presents estimates and robust standard errors clustered at the individual level from seven logit models. In model I, X_i contains only a dummy indicating the students' gender ($Male_i$) while $Y_{i,k}$ is empty. In models II and III, in addition to gender, X_i contains the experimentally derived variables that measure risk aversion ($CRRA_i$), overconfidence ($Overconfidence_i$), and either the first (model II) or second (model III) measure of competitiveness. In models IV and V, X_i also includes the additional control variables described in subsection 2.4 and used in Table 4. Finally, in models VI and VII, we explore the impact of earnings expectations on major choice. These models use the same specification as models IV and V except that $Y_{i,k}$ now contains i 's earnings expectations in major k . We use earnings expectations at age 45 since they show the strongest association with competitiveness (see Table 4). To facilitate the interpretation of the coefficient, we standardize expectations to have a mean of zero and a standard deviation of one.

Table 6 presents odds ratios of the estimated coefficients using humanities as the omitted major. Our findings are as follows. First, we find that, as hypothesized, overconfident students are relatively more common in the natural sciences than in business or the humanities ($p < 0.045$ in VI and $p < 0.050$ in VII). Albeit, we do not find evidence of a statistically significant

²⁹Note that models in which the vector of major-specific variables $Y_{i,k}$ is empty is equivalent to a standard multinomial logit regression.

Table 6: The gender gap in major choice

Note: Odd ratios of logit estimates with robust standard errors clustered at the individual level. All regressions have major and individual fixed effects, and 3 observations for each of the 229 students, resulting in a total of 687 observations. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

Independent variables	I	II	III	IV	V	VI	VII
BUSINESS							
Male	2.97*** (0.92)	3.39*** (1.10)	3.40*** (1.10)	3.89*** (1.50)	3.98*** (1.55)	3.31*** (1.33)	3.39*** (1.37)
Competitiveness		0.67 (0.17)		0.78 (0.22)		0.75 (0.22)	
Competitiveness rank			0.98* (0.01)		0.99 (0.01)		0.99 (0.01)
Overconfidence		0.77* (0.12)	0.76* (0.12)	0.88 (0.16)	0.87 (0.17)	0.85 (0.16)	0.85 (0.16)
CRRA coefficient		0.92 (0.14)	0.82 (0.15)	1.04 (0.17)	0.97 (0.19)	1.03 (0.16)	0.93 (0.17)
NATURAL SCIENCES							
Male	1.78 (0.75)	1.43 (0.66)	1.49 (0.68)	1.68 (0.89)	1.75 (0.92)	1.67 (0.89)	1.73 (0.91)
Competitiveness		1.16 (0.45)		1.32 (0.53)		1.30 (0.52)	
Competitiveness rank			1.00 (0.02)		1.00 (0.01)		1.00 (0.01)
Overconfidence		1.42 (0.31)	1.39 (0.30)	1.67* (0.44)	1.63* (0.43)	1.74** (0.47)	1.70** (0.46)
CRRA coefficient		0.78 (0.17)	0.77 (0.18)	0.79 (0.20)	0.80 (0.19)	0.77 (0.21)	0.78 (0.20)
MAJOR k							
Expected earnings						1.52** (0.26)	1.54** (0.27)
Controls	No	No	No	Yes	Yes	Yes	Yes
Wald χ^2	12.32	22.32	22.12	77.71	75.49	78.86	76.28

difference between business and the humanities ($p = 0.399$ in VI and $p = 0.387$ in VII). 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

compared to the humanities ($p = 0.337$ in VI and $p = 0.225$ in VII) or the natural sciences ($p = 0.337$ in VI and $p = 0.225$ in VII). In fact, the odds ratio for business is less than one, 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.³⁰ Third, the estimated odds ratios for the coefficient of CRRA are generally below one, which is consistent with risk averse students being less common in business and the natural sciences than in the humanities, but they are not statistically significant at conventional levels ($p > 0.333$ in VI and $p > 0.330$ in VII). Fourth, consistent with the literature on major choice, columns VI and VII 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, 2013). The estimates imply that a one standard deviation increase in expected earnings in a major relative to the baseline major increases the odds of majoring in that field by about 1.5.

Lastly, we should also note that, in contrast to what we see in the regressions in Table 4, including our experimental measures and additional control variables does not help explain the large gender difference in major choice (i.e., the lack of women majoring in business compared to humanities). In fact, the coefficient for $male_i$ increases as we add more independent variables.

Why are competitiveness and overconfidence not related to major choice?

Models VI and VII of Table 6 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 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

³⁰In the supplementary materials, we separate each competitiveness measure into one for students who are averse to competition and one for students that are overly competitive. We find that compared to both business and the humanities, the natural sciences display a higher proportion of *both* overly competitive students and students who are adverse to competition. If we use the residuals measure of competitiveness, we obtain a positive coefficient for competitiveness in the natural sciences and a negative coefficient for business, with neither one being statistically significant ($p = 0.620$ and $p = 0.574$, respectively).

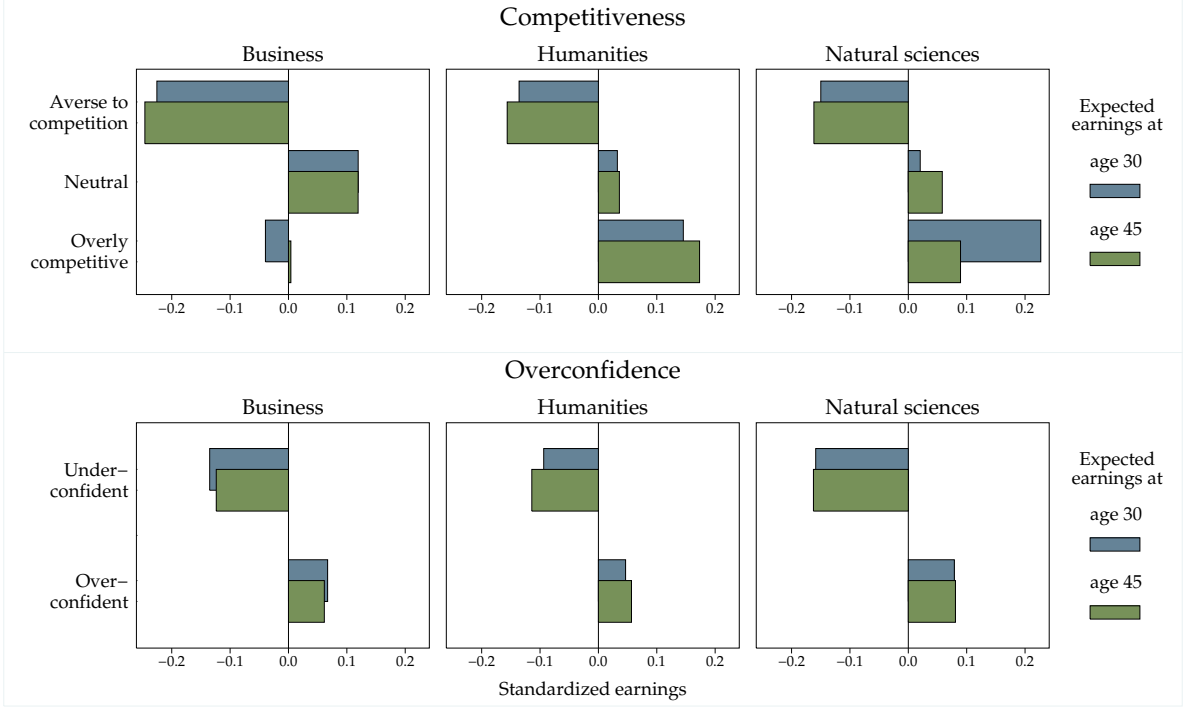


Figure 3: Earnings, competitiveness, and overconfidence by major category

Note: Expected earnings are standardized within each major category.

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 terms of the specific logit model estimated above, while α in equation (2) is significant and positive (indicating that students are more likely to choose majors which they believe have higher future earnings), the competitiveness trait reduces expectations by a roughly proportional amount for all majors, hence the $Y_{i,k} - Y_{i,\tilde{k}}$ term from equation (2) does not change.

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.

6 Conclusion

Our research, combining an experiment and survey of expectations, reveals three key findings. First, our experimental results confirm past findings of substantial gender differences in levels of confidence, competitiveness, and preferences for risk. Second, 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. Third, we show that while earnings expectations are related to major choice, there is no direct relationship between major choice and the experimental measures, at least for choices aggregated up to our broad major categories.

Our findings underscore the importance of combining experimental measures of individual traits with more traditional surveys of labor market behavior and beliefs.³¹ 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 labor market.

The finding of a relationship between overconfidence and competitiveness and earnings expectations provides important insight into the mechanisms underlying gender differences in the labor 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. Our data do not allow us to distinguish the two explanations. 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 earning 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.³² These results may form at least a

³¹A small and growing literature studies the link between experimental measures and actual behavior in the field (e.g., Karlan, 2005; Ashraf et al., 2006; Benz and Meier, 2008; Sapienza et al., 2009; Fehr and Leibbrandt, 2011; Buser et al., 2013; Hopfensitz and Miquel-Florensa, 2013).

³²There are various reasons why occupations with tournament-based pay might have higher expected earn-

partial explanation for the glass ceiling phenomenon (Bertrand and Hallock, 2001; Albrecht et al., 2003), whereby higher earning and higher prestige positions require aggressive negotiation and compensation is based on relative performance. This is consistent with, for example, Goldin and Katz (2011) who find that the proportion of female physicians differs substantially across specialties within medicine, ranging from almost 70% to less than 10% percent.

References

- Albrecht, J., Björklund, A., and Vroman, S. (2003). Is there a glass ceiling in sweden? *Journal of Labor Economics*, 11:145–177.
- Andersen, S., Ertac, S., Gneezy, U., List, J. A., and Maximiano, S. (2012). Gender, Competitiveness and Socialization at a Young Age: Evidence from a Matrilineal and a Patriarchal Society. *Review of Economics and Statistics*, forthcoming.
- 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. (2012). Inflation expectations and behavior: Do survey respondents act on their beliefs. mimeo, Federal Reserve Bank of New York.
- 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 Magazine*, 335:582–586.
- Belzil, C. and Leonardi, M. (2007). Can risk aversion explain schooling attainments? evidence from italy. *Science Magazine*, 14:957–970.
- Bénabou, R. and Tirole, J. (2013). Bonus culture: Competitive pay, screening, and multitasking. Working paper 18936, NBER.

ings. For example, if performance pay is used in markets with averse selection to differentiate employees according to their ability, it can lead to overincentivization of the most talented workers (Moen and Rosén, 2005; Bénabou and Tirole, 2013). 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.

- 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 Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, page 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.
- 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. *The Quarterly Journal of Economics*, 99:31–44.
- Buser, T., Niederle, M., and Osterbeek, H. (2013). Gender, competitiveness and career choices. Working paper, Stanford University.
- 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.
- 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. (2013). How beliefs about hiv status affect risky behaviors: Evidence from malawi. *Journal of Applied Econometrics*, Forthcoming.
- 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 B.V., Amsterdam, NL.
- 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. (2010). Do competitive workplaces deter female workers? a large-scale natural field experiment on gender differences in job-entry decisions. Working paper 16546, NBER.

- Gemici, A. and Wiswall, M. (2013). Evolution of gender differences in post-secondary human capital investments: College majors at the intensive margin. *International Economic Review*, forthcoming.
- 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. 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.
- 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 expectation and attainment. In Duncan, G. J. and Murnane, R. J., editors, *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children*. 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.
- Lundeberg, M. A., Fox, P. W., and Puncóhař, 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.
- Muller, J. and Schwieren, C. (2012). Can personality explain what is underlying women’s unwillingness to compete? *Journal of Economic Psychology*, 33:448–460.
- 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.
- 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. (2013). Determinants of college major choice: Identification using an information experiment. Staff report 500, Federal Reserve Bank of New York.
- 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*, forthcoming.