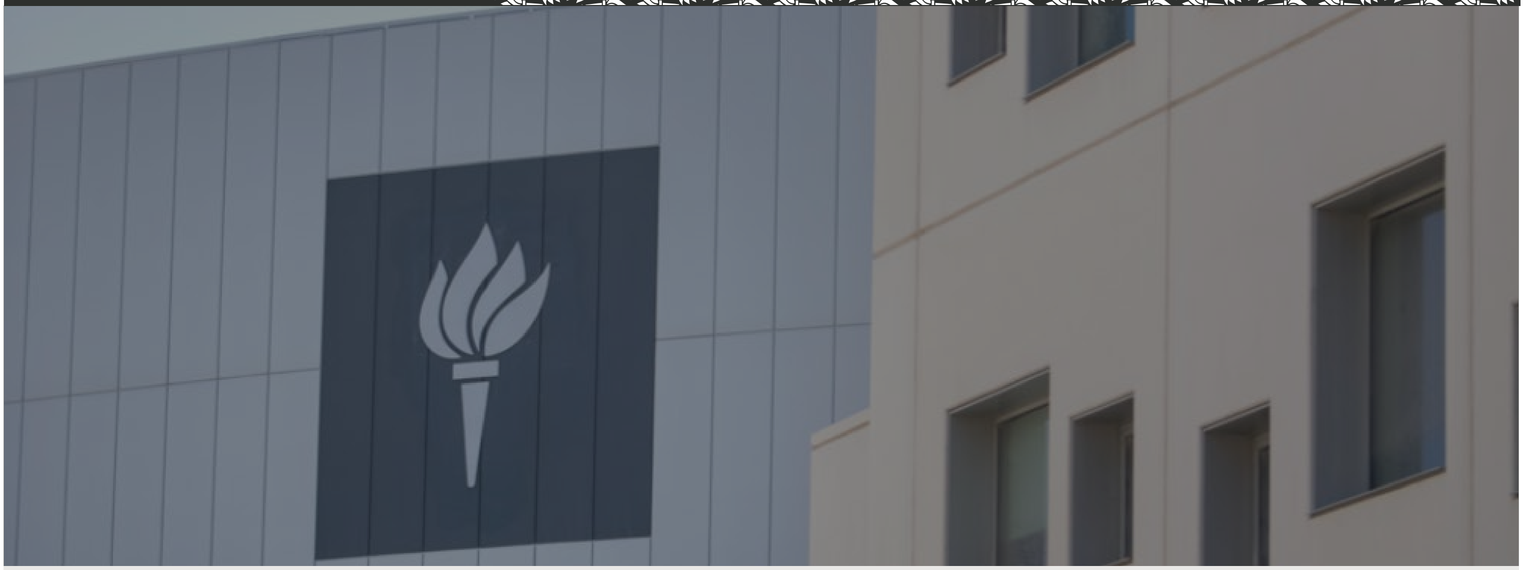
جامعة نيويورك أبوظبي
NYU | ABU DHABI



Taste for competition and the gender gap among young business professionals

Ernesto Reuben, Paola Sapienza, Luigi Zingales

Working Paper # 0031
November 2019

Division of Social Science Working Paper Series

New York University Abu Dhabi, Saadiyat Island P.O Box 129188, Abu Dhabi, UAE

http://nyuad.nyu.edu/en/academics/academic-divisions/social-science.html

Taste for competition and the gender gap among young business professionals

Ernesto Reuben
New York University Abu Dhabi, e-mail: ereuben@nyu.edu

Paola Sapienza
Northwestern University, e-mail: Paola-Sapienza@northwestern.edu

Luigi Zingales
University of Chicago, e-mail: Luigi.Zingales@chicagobooth.edu

ABSTRACT

We study whether and why taste for competition (as measured by Niederle and Vesterlund, 2007) affects MBA salaries and whether this effect can explain the wage gender gap. At graduation, MBAs with higher taste for competition earn $15K (9.3%) more. Over time this effect is mitigated by overconfidence. Seven years after graduation, competitive MBAs with a low degree of overconfidence earn 26% more, while those who are highly overconfident earn 19% less. Taste for competition explains 10% of the gender gap at graduation and none seven years later.

This version: November 2019

JEL Codes: J16, J30, D90, I21

Keywords: gender gap; gender differences; taste for competition; competitiveness; overconfidence; business career.

Acknowledgments: We thank the Templeton Foundation for financial support. We are also thankful for the comments we received from numerous seminars and conferences.

There is a growing literature that attempts to explain differences in career choices and market outcomes (and particularly gender differences) on the basis of differences in taste (e.g., see the discussion in Bertrand 2018). One of the main challenges for this literature is the difficulty of inferring preferences from observed outcomes, since outcomes can be driven not just by taste but also by societal biases and other types of constraints. For example, it is incorrect to conclude that there are differences in preferences regarding academic subjects based merely on the observation that women are less prevalent in STEM fields (Handelsman et al. 2005). Similarly, the choice of working fewer hours may not be necessarily driven by preferences for career interruption, but could be the result of societal biases and obstacles (Blau and Kahn 2017). In this respect, laboratory experiments can provide a powerful tool to measure preferences precisely (Roth 1995).

One measure that has shown a high degree of external validity and typically exhibits substantial gender differences is *taste for competition*, as measured in the lab by Niederle and Vesterlund (2007).[1] In this paper, we use this lab measure to study differences in earnings among high-ability business professionals.

The participants of our study consist of individuals who obtained a master's degree in business administration (MBA) from one of the top-ranked business schools in the United States. This is an interesting group of people to study because many of the jobs these students aspire to are in industries known for being competitive and displaying noticeable gender differences (Bertrand, Goldin, and Katz 2010). Moreover, the recruiting process for these jobs involves a competitive process where recruiting firms meet candidates through social mixers, presentations, and several rounds of formal interviews.

Thanks to an extensive data collection effort, we have access to several incentivized measures of individual traits, earnings, and many demographics. Unique to our study, we measure taste for competition two years before students graduate, and we have information about accepted job offers at graduation, when recruiters set base salaries and bonuses on the basis of expected performance, as well as their compensation seven years later, when reported earnings are based on realized performance (especially bonuses).

To measure taste for competition, we use the experimental design of Niederle and Vesterlund (2007), which consists of giving participants the opportunity to earn money by answering simple arithmetic problems under two different incentive schemes: piece-rate and tournament.[2] In the

[1] See Niederle and Vesterlund (2011) and Dariel et al. (2017) for surveys of published work on gender differences in taste for competition.

[2] Since we study a setting (business) that is stereotypically male, we elicit taste for competition using a task in an area (math) that is typically associated with men (Reuben, Sapienza, and Zingales, 2014; Bohnet, van Geen, and Bazerman 2016). Experiments using this task in various subject pools have consistently found that men choose the tournament more often than women (e.g., Niederle and Vesterlund, 2007; Cason, Masters, and Sheremeta, 2010; Healy and Pate, 2011; Balafoutas and

piece-rate condition, participants do not compete with others and simply earn $4 per correct answer. In the tournament condition, participants compete with three other randomly chosen participants and earn $16 per correct answer if they have the highest performance in their group (and zero otherwise). The participants' taste for competition is assessed by letting them choose between performing under piece-rate and tournament, after controlling for their performance, risk preferences, and degree of overconfidence.

When we look at the accepted job offers at graduation, we find that the total earnings of individuals who exhibited a taste for competition in the experiment two years before are 9 log points higher than those who did not (around $15k more per year), a sizeable effect comparable in magnitude to the effect of gender. In addition, we find that the gender difference in taste for competition accounts for around 10% of the gender difference in total earnings.[3] Earnings at graduation consist of three components: base salary, one-off bonuses (e.g., relocation and tuition benefits), and performance bonuses set in advance based on performance expectations. Both taste for competition and gender explain substantial differences in the performance bonuses, while the differences in base salaries are much smaller. Importantly, the experimental measure of taste for competition is not strongly correlated with the large set of control variables, and therefore, it accounts for variance in earnings and in the gender gap that would otherwise remain unexplained. We explore three sources for the effect of taste for competition: industry selection, more aggressive in bargaining for higher compensation, and the possibility that companies value this attribute. Interestingly, we find that neither industry selection nor bargaining for higher compensation is the primary driver of the effect of taste for competition.

When we repeat the estimation using the 2015 salary and 2014 realized bonus, we find that taste for competition retains a positive impact on total compensation, but this impact is quantitatively smaller and not statistically significant at conventional levels. To explain the differences between salary offers in 2008 and earnings in 2015, we investigate whether taste for competition interacts with other traits that employers are likely to learn over time through the individuals' realized performance.

A company should find taste for competition desirable only when it is not associated with overconfidence because individuals who like to compete but overestimate their probability of winning will enter competitions that they will end up losing. To test this hypothesis, we interact taste for competition with overconfidence. We find that the interaction between taste for competition and overconfidence has a negative effect on total earnings. This effect is small and statistically

---

Sutter, 2012; Niederle, Segal, and Vesterlund, 2013). That being said, gender differences in taste for competition are sometimes diminished when measured with stereotypically-female tasks (e.g., Kamas and Preston 2010; Dreber, von Essen, and Ranehill, 2014; Wozniak, Harbaugh, and Mayr, 2014).

[3] Although explaining 10% of the gender gap might not be considered that much, we should point that the sole measure of taste for competition explains half as much of the gender gap in earnings as a rich set of variables that include demographic characteristics, academic performance, and experimental and survey measures of important psychological attributes.

insignificant in the 2008 data, but it is both large and statistically significant in the 2015 data. In 2015, MBAs with a high taste for competition who have a low degree of overconfidence (one standard deviation below the mean) earn 26% more while those who are highly overconfident (one standard deviation above the mean) earn 19% less. One possible reason for why this result is more pronounced in later in their careers is that employers cannot condition compensation on the degree of overconfidence of recruits because they are unable to detect it at that point, but that can do so over time. An alternative reason is that, unlike salary and bonus at graduation, compensation for business professionals in later years is often based on realized performance. If workers who are competitive and overconfident have lower performance, our results can be explained by performance set bonuses, which is the component of their compensation where the effect of taste for competition is concentrated.

In sum, we find that taste for competition is an important variable in explaining people's compensation. As companies learn their employees' type or are able to set bonuses based on realized performance, they seem to penalize competitive and overconfident workers. However, despite being good predictors of earnings, both taste for competition and overconfidence have a marginal effect in explaining gender differences in compensation, which grew substantially seven years after graduation.

This paper contributes to the growing literature on gender differences in taste for competition and, more specifically, to studies relating incentivized measures of taste for competition to gender differences in labor-market outcomes.[4] The most prominent study in this area is Buser, Niederle, and Oosterbeek (2014), where they use the same measure of taste for competition to predict the educational choices of high school students in the Netherlands. They find that competitive individuals are 20% more likely to select the math and science study track. Moreover, they find that controlling for taste for competition reduces the gender gap in track choice by around 20%. Buser, Peter, and Wolter (2017a) subsequently replicated these findings with high school students in Switzerland. Building on Buser, Niederle, and Oosterbeek (2014), Reuben, Wiswall, and Zafar (2017) and Kamas and Preston (2018) study whether individuals with a high taste for competition major in different fields in private universities in the United States. Buser, Peter, and Wolter (2017b) study the relationship between willingness to compete and educational choices among students of varying abilities. They consistently find that students of both genders who compete make different educational choices than students who do not compete at all points in the ability distribution. Lastly, Zhang (2019) finds that willingness to compete predicts whether middle school students in rural China take a highly demanding high school entrance exam. We extend the findings of these papers

[4] Other related work is that of Flory, Leibbrandt, and List (2015), Leibbrandt and List (2015), and Samek (2019), who demonstrate that jobs perceived as being more competitive affect the willingness of women to apply and negotiate salaries. Unlike the work reviewed here, these studies do not use individual measures of taste for competition.

by demonstrating that taste for competition predicts *actual* labor market outcomes in a considerably different sample of participants (our sample is older, better educated, more diverse, and specialized in business) and for an extended period of time.

A final related study is that of Berge et al. (2015), who find that a high taste for competition is associated with higher investments and profits by small-scale Tanzania's entrepreneurs. Besides studying a very different population, our study has one main advantage vis-à-vis this paper. Namely, we collected the earnings data years after we measured taste of competition and at two points in the careers of the business professionals. First at graduation, after a lengthy competitive recruitment process, but before they had to perform at work. Then, seven years later, when compensation is based on an evaluation of the employee's performance on the job. These longitudinal data allow us to investigate which determinants of earnings at graduation are persistent over time.

The rest of the paper is organized as follows. In Section I, we describe the various sources from which we collect our data. In section II we present descriptive statistics of our sample, including whether there are gender differences in taste for competition and compensation. In section III, we test the relationship between gender, taste for competition, and compensation at graduation. In Section IV we reevaluate these relationships with compensation 7 years after graduation. Finally, we further discuss our findings and conclude in Section V.

I. Study design

Our sample consists of the 2008 MBA cohort at the University of Chicago Booth School of Business. We rely on multiple sources of data of this specific cohort: an experiment and an initial survey conducted at the start of their MBA program, the school's administrative data, and a follow-up survey conducted seven years later.

I.A. Initial survey and experiment

As part of a required core class, all the MBA students of the 2008 cohort completed a survey and participated in an experiment designed to measure several individual-specific characteristics. We conducted both the survey and the experiment in the fall of 2006, during their first month in the business school. Participants completed the survey online before they took part in the experiment. The survey included questions on demographic characteristics as well as standard questionnaires of personality traits.

The experiment consisted of eight distinct parts. Participants were given the instructions for each part before the start of the respective part. They received no feedback concerning the outcome or behavior of others until the experiment had concluded. As compensation, participants received a $20 show-up fee and their earnings in a randomly selected part. On average, participants earned $99 for the 90-minute experiment. In the Online Appendix, we provide a detailed description of the procedures used to conduct the survey and experiment as well as the instructions for the tasks used to measure taste for competition.[5]

To measure taste for competition, we use a variation of the design used by Niederle and Vesterlund (2007). Participants first performed an adding task under both a tournament payment scheme and a piece-rate payment scheme. Subsequently, they performed the task once again under a payment scheme of their choice. Their payment-scheme choice serves as the basis for their taste for competition.

The adding task consisted of computing sums of four two-digit numbers for 150 seconds. The computer randomly drew the two-digit numbers from a uniform distribution with a support of 11 to 99. Calculators were not allowed. After each answer, a new set of numbers appeared on the computer screen along with a message indicating whether their answer was correct or incorrect. Importantly, although participants knew what their own performance was, they did not receive any information about the performance or choices of others during the experiment.

We informed participants that this part of the experiment consists of four periods, one of which would be randomly chosen to determine their earnings. We also informed them that we randomly assigned them to groups of four. Participants read the instructions for each period just before the start of the respective period. In the first two periods, participants performed the addition task once under a piece-rate payment scheme and once under a tournament payment scheme. Under piece-rate, participants earned $4 for every correct answer. Under tournament, participants earned $16 for every correct answer if they had the highest number of correct answers in their group (ties were broken randomly) and earned $0 otherwise. Half the participants performed the addition task first under piece-rate and then under the tournament while the other half performed the tasks in the reverse order.

In the third period, we informed participants that they would perform the addition task once again and asked them to choose one of the two payment schemes to apply in that period. Participants who chose piece-rate earned $4 per correct answer. Participants who chose tournament, earned $16 per correct answer if they had more correct answers than their other group members had when they previously performed the task under the tournament payment scheme. Competing against their group members' past performance has the advantage that the participants' choice and effort in the third period is not affected by the (expected) choices of the other members of the group. The variable "competitive" is a dummy variable equal to 1 when an individual chooses tournament in this period.

In the fourth period, participants did not perform the adding task. In this period, they simply chose whether they wanted their earnings in the fourth period to be calculated based on their past

[5] We provide a detailed description of the other parts of the survey and experiment in Reuben, Sapienza, and Zingales (2008).

performance and either the piece-rate or the tournament payment scheme.[6] Thus, the participants' choice in the fourth period resembled their choice in the third period except that participants who chose the tournament did not perform under the stress (or thrill) one might experience in a competitive environment. The variable "non-competitive tournament" is a dummy variable equal to 1 when an individual chooses tournament in this period.

There are several reasons why participants may prefer a tournament payment scheme. First, they might correctly anticipate that they are a superior performer. Second, they might misperceive their performance and believe they are a superior performer when they are not. Third, they might love risk. Fourth, they might receive a special thrill from performing in a tournament. Following Niederle and Vesterlund (2007), we want to isolate the fourth component. For this reason, we need to construct measures of performance, overconfidence, and risk aversion.

To obtain an individual measure of performance, we compute the participants' average rank in the first and second periods. For this variable to not depend on the specific group matching that occurred in the experiment, we used the number of sums solved by the participants and simulated 1,000,000 matchings to obtain an average rank for each participant. Since average ranks are higher when performance is lower, for ease of interpretation, we define the variable "performance" as the negative of the average ranks.

After the fourth period, we elicited the participants' beliefs concerning their relative performance by asking them to guess how they ranked within their group in each of the first three periods. Participants submitted ranks between 1^{st} and 4^{th} and received $2 for each correct guess.[7] We use the participants' estimated ranks and their actual performance to calculate how overconfident they are. Specifically, the variable "overconfidence" is the difference between the actual average rank of an individual in the first two periods and their expected rank. Note that since a lower rank means higher performance, this variable is indeed greater when participants overestimate their performance.

To measure risk preferences, we gave participants 15 choices between a lottery with an expected value of $100 and a certain amount that ranged from $50 to $120. As is common in the literature, we then use these choices to determine each participant's risk aversion coefficient assuming a CRRA utility function (see Holt and Laury. 2002). The variable "risk aversion" is the CRRA risk aversion coefficient.

I.B. Administrative data

The admission office of the business school supplied us with the gender variable. The career services office of the business school provided us with information regarding the job participants accepted

[6] The participants' choice in the fourth period applied to their performance in the first or second period. Specifically, to the period they completed under the piece-rate payment scheme.

[7] In case of a tie, participants were paid the $2 if their guess corresponded to a rank they could have received when the tie was randomly resolved.

upon graduation. The participants initially reported this information, but the career services office subsequently double-checked it with the respective employers to ensure its accuracy. The information included data on earnings, which include salaries as well as yearly and one-off bonuses (e.g., sign-on, relocation, tuition, and retention at year-end bonuses). Based on this information, we calculated the participants' total earnings in their first year after graduation. The information also included the employers' names, which we used to classify them into three broad industry categories. Specifically, we used two-digit NAICS industry codes to classify each employer into finance (two-digit NAICS code 52), professional services, which we refer to as "consulting" (two-digit NAICS code 54), and "other" (the remaining two-digit codes). We also received from the career services office self-reported information from the participants, which included whether they obtained competing job offers.

I.C. The 2015 follow-up survey

At the end of 2015, we reached out to the same set of MBAs with a follow-up survey. The survey contained questions about their career, work-life balance, and degree of life satisfaction. More importantly, we asked them about their salary and their end of year bonus in 2014. Of the 409 original students who consented to the treatment of their data, 263 (64.3%) answered the follow-up survey.

II. Descriptive statistics

Although participation in some parts of the study was mandatory, participants had the option to opt-out of the study by not consenting to the use of some or all of their data. Out of the 550 students in the cohort, 409 (74%) provided information about their job in 2008 and consented to the analysis of the initial survey, experiment, and administrative data. Note that the decision to consent was made, even for the job placement, in September 2006, two years before the student graduated. Throughout the paper, we concentrate on these participants. However, it is important to understand whether this sample differs systematically from the rest of the cohort. For this reason, in the Online Appendix, we conduct a thorough comparison between the 409 participants in the sample and 129 participants for whom we can analyze data sources other than their job placement data.[8] By and large, we do not find differences between these two populations (see Table A.1 in the Online Appendix). Crucially for this paper, neither the fraction of women nor the fraction of participants who chose the tournament is significantly different (χ^2 tests, $p > 0.388$).[9] Similarly, to understand selection into the sample who responded to the 2015 follow-up survey, in Table A.2 in the Online Appendix, we compare the

[8] Of these 129 participants, we have compensation data for 26 participants who did not consent to the use of their job placement data and 36 who had job offers that were not reviewed by the school's career services office. For the remaining 67 participants, it is unclear whether they failed to report their job placement to the university, or they did not have a job offer.

[9] It is also the case that neither the fraction of men nor the fraction of women who chose the tournament significantly differ between the two populations (χ^2 tests, $p > 0.704$).

Table 1 – Summary statistics by gender

Note: Means, standard deviations, and number of observations for variables used in the paper. The rightmost column displays *p*-values from tests of equality of distributions between men and women (*t*-tests for ordinal variables and χ^2 tests for categorical variables).

	MEN			WOMEN			
	mean	s.d.	N	mean	s.d.	N	*p*-value
Experiment							
Competitive	0.60	0.49	286	0.33	0.47	123	0.000
Performance (rank in sums tasks)	2.39	0.78	286	2.70	0.73	123	0.000
Expected rank in sums tasks	2.11	0.76	286	2.54	0.71	123	0.000
Overconfidence	0.28	0.63	286	0.16	0.65	123	0.095
Risk aversion coefficient	4.22	4.19	286	5.94	4.69	123	0.001
Non-competitive tournament	0.47	0.50	286	0.25	0.44	123	0.000
Jobs data							
Total compensation in 2008	185.84	183.12	286	149.22	36.95	123	0.001
Base salary in 2008	107.71	18.88	286	105.91	15.68	123	0.318
Total bonus in 2008	78.12	176.26	286	43.31	28.45	123	0.001
One-off bonus in 2008	44.16	30.46	286	34.91	22.51	123	0.001
Expected performance bonus in 2008	33.96	168.98	286	8.40	17.33	123	0.012
Number of competing job offers	0.42	0.82	286	0.41	0.81	123	0.882
Fraction with job in finance in 2008	0.58	0.49	286	0.36	0.48	123	0.000
Fraction with job in consulting in 2008	0.20	0.40	286	0.34	0.48	123	0.000
Total compensation in 2015	346.93	231.91	189	228.87	180.59	61	0.000
Base salary in 2015	194.04	108.58	189	160.22	60.18	61	0.002
Performance bonus in 2015	152.89	185.70	189	68.66	143.69	61	0.000
Fraction with job in finance in 2015	0.48	0.50	189	0.18	0.39	61	0.000
Fraction with job in consulting in 2015	0.09	0.29	189	0.21	0.41	61	0.000

characteristics of the 263 respondents and the 146 nonrespondents who had consented to the analysis of their data.

Next, we provide descriptive statistics for participants in our sample and evaluate whether there are gender differences in the experimental, initial survey, and administrative data. Table 1 presents the mean and standard deviation for variables derived from these data sources for both the 286 men and 123 women in the sample. For each variable, the table also displays *p*-values from tests of equality of distributions between men and women based on *t*-tests for ordinal variables and χ^2 tests for categorical variables. In the experiment and initial survey, we replicate many of the gender differences reported in the literature (Croson and Gneezy, 2009). Next, we focus on the difference in taste for competition between male and female MBAs.

II.A. Gender differences in taste for competition

Consistent with the literature on taste for competition, Table 1 shows that 60% of men choose the tournament payment scheme compared to 33% of women. However, on its own, the higher incidence

Table 2 – Determinants of willingness to compete

Note: Regressions of the decision to enter the tournament in the third period of the experiment. Marginal effects from probit regressions and standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III
Woman	–0.268***	–0.223***	–0.148**
	(0.051)	(0.055)	(0.060)
Performance		0.213***	0.351***
		(0.035)	(0.043)
Overconfidence			0.320***
			(0.050)
Risk aversion			–0.019***
			(0.006)
Obs.	409	409	409
χ^2 test	25.069***	63.677***	117.954***

of men choosing the tournament is not enough to conclude that men like to compete more than women do. In particular, Table 1 also reveals that men in our sample outperform women in the adding tasks (the average rank is 2.39 for men and 2.70 for women), and consistent with their higher performance, they expect to be better ranked than women (on average, 2.11 vs. 2.54). These differences, combined with the fact that women are more risk-averse, could explain why men choose the tournament more often than women do.

Do male MBAs like competition more than female MBAs after controlling for their ability, beliefs, and risk preferences? To answer this question, we follow Niederle and Vesterlund (2007) and run a series of probit regressions with the participants' choice of the tournament as the dependent variable. We report the resulting marginal effects in Table 2. In column I, the only independent variable is the participants' gender. Without any controls, the gender gap in choosing the tournament equals 26.8%. In column II, we control for the participants' performance, which reduces the gender gap in choosing the tournament by 4.5 percentage points to 22.3%. In column III, we further control for the participants' beliefs by including the variable overconfidence and for their risk preferences by including their risk-aversion coefficient. Performance, beliefs, and risk preferences are all significant determinants of the choice of tournament. Controlling for these variables still leaves a statistically significant gender gap of 14.8% in the decision to compete (column III). The coefficient of the gender dummy, once we control for performance, beliefs, and risk preferences, can be interpreted as a gender difference in "taste for competition."[10]

[10] This way of measuring taste for competition has recently come under scrutiny because measurement error in the control variables or an incorrectly specified regression can result in the overestimation of the effects of taste for competition (van Veldhuizen 2018; Gillen, Snowberg, and Yariv 2019). In subsection III.A of the Online Appendix, we run a series of robustness checks to test whether this result is susceptible to this problem. We do not find evidence that it is.

II.B. Compensation in 2008

The Booth career office collects data on the base salary and bonuses of all the graduating MBA students. For our analysis, we first consider total compensation, which is composed of base salary and bonus. In 2008, male MBAs received, on average, total compensation of $186K, which is 25% higher than their female graduates ($149K). In Table 1, we also report separate sample statistics for the base salary and the bonus pay. We group the various bonuses into two components: one-off bonuses (relocation, tuition, sign-on, and retention at year-end) and the rest, which are bonuses related to the employers' expected performance of their new hires (stock options, profit sharing, guaranteed performance, and other). We call this component the "expected performance bonus" because firms offer it before the MBAs begin to work, but it is linked to the performance of the firm, and it is likely to incorporate the employees' actual performance in the future. The descriptive statistics reveal that the gender differences are mostly concentrated in the bonus, not in the base salary. For example, the average overall bonus for men is 80% higher than for women. The difference is even starker in the expected performance component, where men's bonus is 404% higher than women's.

The large gender gap in total compensation is partly due to three male outliers, with salaries above $1M. If we ignore those, the average male total compensation drops to $170K, the gender gap to 14%, the average overall bonus for men becomes 44% higher than for women, and the expected performance bonus 222% higher for men compared to women.

II.C. Compensation in 2015

Our 2015 follow-up survey asks for their current salary and year-end performance bonus. We compute total compensation as the sum of the two. We will refer to it as 2015 compensation, even if it is technically the sum of the 2015 salary and 2014 realized bonus.

On average, women make $229K and men $347K (52% more). Unlike in 2008, in 2015, the average base salary of men is significantly higher than the base salary of women by a factor of 1.21. However, the largest difference is once again concentrated in the bonuses, where men's are larger than women's by a factor of 2.23. Eliminating outliers does not change the results much. Nevertheless, to avoid the risk that our results are driven by a few individuals, in our subsequent analysis, we windsorize the compensation data at the 1% and 99% level for both 2008 and 2015.

Although we find both in 2008 and 2015 that the largest gender gap occurs in the bonus component of pay, it is important to highlight that there is a big difference between the 2008 and 2015 bonuses. Bonuses in 2008 are one of the negotiated components during the recruiting process, while bonuses in 2015 are based on realized performance typically related to observable metrics set up in advance. We will discuss these differences in light of our results.

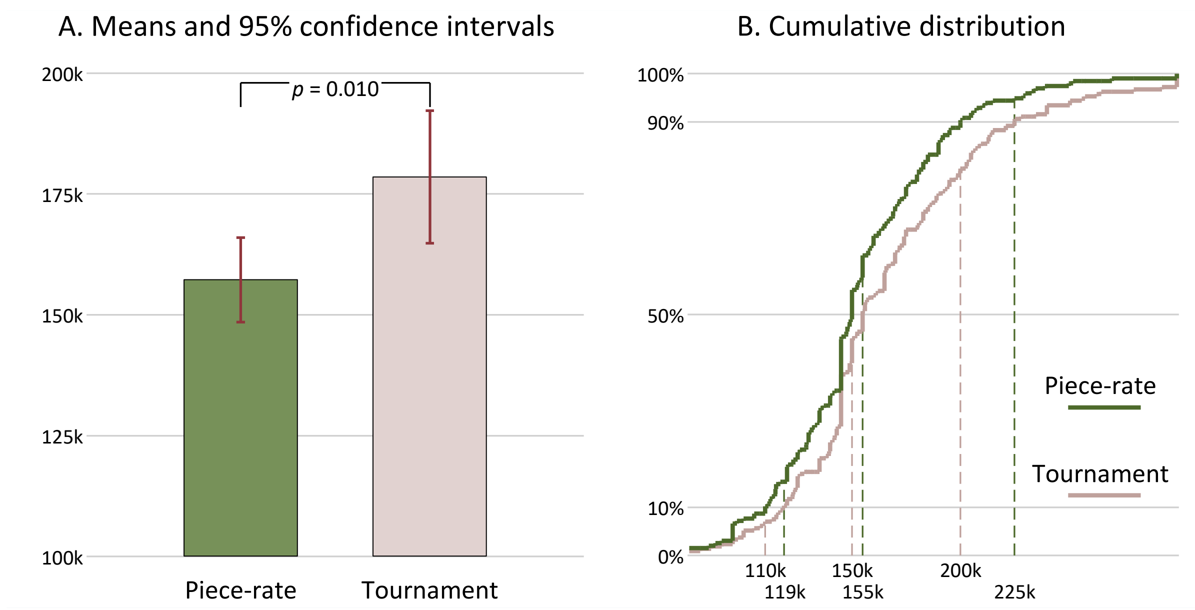


Figure 1 – Total compensation in 2008 depending on tournament choice

III. Taste for competition and compensation in 2008

In the previous section, we have shown that our sample exhibits a gender gap in wages and a gender difference in the taste for competition. In this section, we analyze how the two are related beginning with the compensation at graduation.

III.A. The effect of taste for competition

We start by looking graphically at the association between taste for competition and compensation. As we can see in Figure 1A, choosing the tournament in a laboratory experiment at the beginning of their MBA is associated with higher earnings two years later in the participants' first job. On average, participants who chose the tournament ended up earning $21K more than participants who chose piece-rate (*t*-test, $p = 0.010$). The difference in earnings is larger for the top earners (see Figure 1B).

Does this difference in earnings persist once we control for other determinants of choosing the tournament? To answer this question, in Table 3, we run a series of linear regressions with the log of the participants' total compensation in 2008 as the dependent variable.

In column I, the only explanatory variable is "competitive." Individuals who chose to compete earn approximately 9.3% more. In column II, we add three explanatory variables: overconfidence, risk aversion, and performance in the game, all measured as described in Section I. None of these additional variables is statistically significant, and adding them to the regression does not change the magnitude of the competitive dummy. With the inclusion of these control variables, we follow Niederle and Vesterlund (2007) and interpret the coefficient of the competitive dummy as representing the effect of "taste for competition."

Table 3 – Determinants of total compensation in 2008

Note: Regressions of the log of total compensation in 2008. OLS estimates and standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III	IV	V	VI	VII	VIII
Competitive	0.093***	0.092**		0.079**	0.074**	0.075**		0.062*
	(0.031)	(0.036)		(0.036)	(0.031)	(0.036)		(0.036)
Woman			–0.117***	–0.107***			–0.112***	–0.104***
			(0.036)	(0.036)			(0.036)	(0.036)
Overconfidence		0.014	0.024	0.003		0.008	0.014	–0.002
		(0.029)	(0.027)	(0.029)		(0.029)	(0.027)	(0.029)
Risk aversion		0.000	0.000	0.002		0.000	0.001	0.002
		(0.004)	(0.004)	(0.004)		(0.004)	(0.004)	(0.004)
Performance		–0.004	0.006	–0.016		–0.004	0.004	–0.015
		(0.025)	(0.023)	(0.025)		(0.025)	(0.023)	(0.025)
Finance					0.128***	0.128***	0.125***	0.118***
					(0.038)	(0.039)	(0.038)	(0.038)
Consulting					0.127***	0.125***	0.144***	0.133***
					(0.044)	(0.045)	(0.044)	(0.044)
Obs.	409	409	409	409	409	409	409	409
R^2	0.021	0.022	0.032	0.044	0.050	0.051	0.064	0.071
F test	8.853***	2.312*	3.361***	3.686***	7.172***	3.598***	4.564***	4.359***

In column III, we repeat specification II with the gender dummy instead of the competitive dummy. Women make $18K (11.7%) less, and this difference is statistically significant. Given the tendency to wage compression at this stage of an MBA's career and that most companies have predetermined wages for newly hired MBAs, we find this result to be notable.

In column IV, we include both the competitive and gender dummies (as well as all the controls). Once we control for the taste for competition, the magnitude of the gender coefficient drops by around $2K (10% of the gender gap) but remains statistically significant. The coefficient for taste for competition remains positive and statistically significant: individuals who like to compete, earn $13K (7.9%) more than the rest.

Some industries tend to pay MBAs significantly more (Oyer, 2008). Thus, compensation can vary because of differences in the industry chosen by MBAs at graduation. Since one of the effects of taste for competition could be different sorting across industries, we initially chose not to control for industry to estimate the full effect of taste for competition on compensation. It is interesting, however, to check how the results change if we control for the industry chosen by the MBAs at graduation. This is what we do in columns V to VIII, where we repeat the specifications I to IV including industry fixed effects. As explained in Section I, we classify employers into three industries: finance, consulting, and the rest. Each industry was chosen by roughly a third of the MBAs. The coefficients of both the competitive and gender dummies are only slightly smaller. Thus, industry sorting does not seem to be the main driver of the results.[11]

To better understand the relationship between compensation and taste for competition, we separate the base salary and the bonuses in Table 4. In column I, the dependent variable is the log of the base salary in 2008. The explanatory variables are the same as in Table 3. The results show that the competitive dummy does not affect the MBAs' base salary, nor does the gender dummy. Since not all MBAs receive a bonus, and we are estimating the regressions in logs, we opted for a two-step hurdle model to estimate first the impact of the independent variables on probability of getting the bonus (column II) and then on the magnitude of the bonus received (column III) (Cragg, 1971). Neither taste for competition nor gender predicts the probability of receiving a bonus, which is hardly surprising since almost everyone (92.9%) receives some form of bonus. By contrast, both the competitive and gender dummies correlate with the size of the bonus. MBAs with a higher taste for competition receive an $8K (15.8%) higher bonus and women receive an $18K (37.8%) lower bonus.

Next, we study the two bonus components separately.[12] The dependent variable in column IV is the probability of getting the one-off component of the bonus. Most MBAs (91.4%) received some one-off bonus, which might be why the selection equation does not show any significant coefficient. In column V, the dependent variable is the log of the one-off bonus. Interestingly, we find that the competitive dummy is not significantly associated with the one-off bonus. One possible explanation is that this bonus does not vary much across individuals. However, this is not true: the one-off bonus is $25K at the 25th percentile and $55K at the 75th percentile, an interquartile range of $30K, which is similar to that of the expected performance bonus ($37K). Furthermore, we do observe a sizeable negative effect for women: on average, they receive a $9K (24.4%) lower one-off bonus.

In the next two columns, we repeat the hurdle model for the expected performance bonus. Only 37.4% of the MBAs received this bonus. Nevertheless, the probability of receiving it is not correlated with any of our explanatory variables (Column V). By contrast, the magnitude of this bonus is highly correlated with taste for competition and gender (Column VI). MBA students with a higher taste for competition receive a $13K higher expected performance bonus, while women receive a $10K lower bonus.

---

[11] In the Online Appendix, we report the results of two other robustness checks. In Table A.5, we evaluate whether we are overestimating the effect of taste for competition due to potential measurement errors in the control variables (van Veldhuizen 2018; Gillen, Snowberg, and Yariv 2019). In Table A.6, we repeat our basic specification adding a large set of individual controls to the regression (following Bertrand, Goldin, and Katz 2010). In both cases, we find very similar results.

[12] As previously described, we group the various bonuses into two components: one-off bonuses (relocation, tuition, sign-on, and retention) and expected performance bonuses (stock options, profit sharing, guaranteed performance, and other). As a robustness test, we dropped other bonuses from the latter category. The results are largely unchanged (see Table A.8 in the Online Appendix).

Table 4 – Taste for competition and different 2008 compensation measures

Note: Regression of the log of the base salary in 2008 in column I. Hurdle model of the likelihood of receiving a bonus in column II and its magnitude in column III. Hurdle model of the likelihood of receiving a one-off bonus in column IV and its magnitude in column V. Hurdle model of the likelihood of receiving an expected performance bonus in column VI and its magnitude in column VII. Linear estimates in columns I, III, V, and VII. Marginal effects in columns II, IV, and VI. Regressions in Panel A do not include industry fixed effects while those in Panel B do. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Base salary	Total bonus Received	Total bonus Amount	One-off bonus Received	One-off bonus Amount	Exp. perform bonus Received	Exp. perform bonus Amount
	PANEL A: WITHOUT INDUSTRY FIXED EFFECTS						
	I	II	III	IV	V	VI	VII
Competitive	0.022	0.015	0.158***	0.024	0.044	0.010	0.571***
	(0.017)	(0.033)	(0.079)	(0.036)	(0.073)	(0.056)	(0.211)
Woman	–0.010	0.032	–0.378***	0.035	–0.244***	–0.073	–0.499**
	(0.017)	(0.027)	(0.079)	(0.029)	(0.072)	(0.056)	(0.231)
Overconfidence	0.003	0.024	–0.051	0.013	–0.043	–0.023	0.176
	(0.013)	(0.026)	(0.064)	(0.027)	(0.058)	(0.046)	(0.173)
Risk aversion	0.003*	–0.003	–0.000	–0.003	–0.013*	–0.004	0.028
	(0.002)	(0.003)	(0.008)	(0.003)	(0.007)	(0.006)	(0.023)
Performance	–0.005	0.007	–0.076	–0.016	–0.064	0.010	–0.065
	(0.012)	(0.022)	(0.055)	(0.022)	(0.050)	(0.039)	(0.148)
Obs.	409	409	380	409	374	409	153
F test / χ^2 test	0.975	3.705	31.981***	3.916	18.800***	3.540	16.286***
	PANEL B: WITH INDUSTRY FIXED EFFECTS						
	I	II	III	IV	V	VI	VII
Competitive	0.009	0.013	0.124*	0.019	0.007	0.035	0.478**
	(0.015)	(0.033)	(0.075)	(0.036)	(0.066)	(0.058)	(0.215)
Woman	–0.035**	0.027	–0.300***	0.033	–0.147**	–0.115**	–0.473**
	(0.015)	(0.027)	(0.076)	(0.030)	(0.066)	(0.058)	(0.231)
Overconfidence	–0.002	0.023	–0.058	0.012	–0.048	–0.021	0.159
	(0.012)	(0.026)	(0.060)	(0.027)	(0.052)	(0.047)	(0.173)
Risk aversion	0.001*	–0.004	0.005	–0.004	–0.007	–0.007	0.029
	(0.002)	(0.003)	(0.008)	(0.003)	(0.007)	(0.006)	(0.023)
Performance	0.000	0.008	–0.081	–0.015	–0.074*	0.013	–0.060
	(0.010)	(0.022)	(0.051)	(0.022)	(0.045)	(0.040)	(0.146)
Finance	–0.031*	–0.017	0.540***	0.006	0.620***	–0.296***	0.389*
	(0.016)	(0.031)	(0.081)	(0.033)	(0.071)	(0.064)	(0.231)
Consulting	0.140***	0.022	0.172*	0.040	0.137*	–0.104	0.350
	(0.019)	(0.040)	(0.093)	(0.043)	(0.082)	(0.072)	(0.249)
Obs.	409	409	380	409	374	409	153
F test / χ^2 test	17.282***	4.992	86.274***	5.033	106.100***	28.058***	19.912***

In the bottom panel of Table 4, we re-estimate all the specifications, controlling for industry fixed effects. The effects of taste for competition remain largely the same. For gender, we find a somewhat lower but still highly statistically significant effect on the magnitude of the one-off bonus. Moreover, after controlling for industry, we see that women are 11.5% less likely to receive an expected performance bonus.

III.B. Robustness: Taste for competition or taste for high rewards

A clever feature of the experimental design of Niederle and Vesterlund (2007) is that participants make two choices between tournament and piece-rate. In the third period, participants perform under the chosen payment scheme while in the fourth period, the payment scheme is simply applied to their past performance (see Section I). Because it does not include performing in a competitive environment, Niederle and Vesterlund (2007) argue that this latter choice between piece-rate and tournament is unaffected by the participants' taste for competition, and is determined by preferences for highly non-linear payoffs that reward high performers. In Table A.7 in the Online Appendix, we replicate the analysis in Tables 3 and 4 adding as an explanatory variable a dummy equal to one if an individual chooses tournament in the fourth period. Adding this variable enables us to test whether the effect of the "competitive" variable in Tables 3 and 4 is driven by a taste for competition or a "taste for high rewards."

The taste-for-high-reward variable always has a small coefficient that is statistically not different from zero. By contrast, the coefficients of the competitive dummy remain substantially unchanged in all specifications. These results provide compelling evidence that the association between tournament and compensation is indeed driven by the participants' taste for competition and is not related to the choice of tournament pay *per se*.

III.C. Why does taste for competition matter?

In this section, we study why a higher taste for competition is associated with higher compensation. One possibility is that MBAs with a high taste for competition are simply better at generating higher salary offers through negotiation. Another possibility is that firms expect MBAs with a higher taste for competition to add more value. We test the first hypothesis empirically in Table 5.

The average MBA receives only 0.4 competing offers at graduation, and 73.1% of MBAs receive none. It is reasonable to expect that MBAs who receive multiple offers can extract more rents through negotiation. Given the highly competitive nature of the MBA recruiting process, it also seems plausible that MBAs with a higher taste for competition can generate more offers. We test these predictions in Table 5.

To test the effect of multiple offers on compensation, in columns I and II, we rerun regressions IV and VIII of Table 3 of the log of total compensation in 2008, adding as an explanatory variable a dummy equal to one if the MBA received at least one competing job offer. MBAs who generate competing offers do indeed earn around $16K (10.1%) more. The coefficients of the competitive dummy, however, slightly increase with the inclusion of the competing offers dummy. We also check

Table 5 – Taste for competition and competing job offers

Note: OLS regressions of the log of total compensation in 2008 in columns I and II. Marginal effects from negative binomial regressions of the number of competing job offers in 2008 in columns III and IV. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Total compensation		# competing offers	
	I	II	III	IV
Competitive	0.083**	0.066*	–0.004	0.015
	(0.035)	(0.035)	(0.098)	(0.092)
Woman	–0.103***	–0.095***	–0.059	–0.111
	(0.035)	(0.035)	(0.094)	(0.087)
Overconfidence	0.012	0.009	–0.098	–0.094
	(0.029)	(0.028)	(0.082)	(0.078)
Risk aversion	0.002	0.002	0.009	0.006
	(0.004)	(0.004)	(0.010)	(0.009)
Performance	–0.014	–0.013	–0.051	–0.034
	(0.025)	(0.024)	(0.074)	(0.069)
Competing job offer	0.101***	0.128***		
	(0.035)	(0.035)		
Finance		0.142***		–0.371***
		(0.044)		(0.115)
Consulting		0.128***		–0.002
		(0.035)		(0.106)
Obs.	409	409	409	409
F test / χ^2 test	4.532***	5.571***	2.875	21.559***

whether individuals with a taste for competition generate more competing job offers. In columns III and IV of Table 5, we run a negative binomial regression of the number of competing offers each MBA got on the same explanatory variables we have been using so far. The results show that neither taste for competition nor gender predict the number of competing offers. Thus, we can exclude that generating multiple job offers is the source of the association between taste for competition and higher earnings.

Hence, the evidence so far is more consistent with firms expecting a higher average performance from employees with a higher taste for competition and setting total compensation based on this expectation. There are various plausible reasons why employers in business might value employees who have a taste for competition. For instance, these employees might perform better in competitive or stressful situations (Gneezy, Niederle, and Rustichini 2003; Apesteguia and Palacios-Huerta 2010) or might be easier to retain in competitive industries. However, the experimental evidence of Niederle and Vesterlund (2007), as well as in this paper, suggest that rewarding individuals only based on their willingness to compete could be a mistake. While individuals who choose the tournament payment scheme are likely to have a taste for competition, they are also likely to be overconfident, which is, by definition, a suboptimal trait.

The MBA recruiting process is a long courtship that lasts six months, involving corporate conversations, networking nights, pre-interview one-to-one meetings, and several rounds of formal interviews.[13] During this process, it is plausible that recruiters can observe competitive behavior. However, since recruiters do not have access to our data, it is unlikely that they can separate those who compete because they are overconfident from those who compete because they have a high taste for competition. If employers value a high taste for competition but are also aware that they cannot separate it from a trait like overconfidence, they should expect that hiring individuals based on their competitive behavior will also result in a higher variance in performance. These conjectures are consistent with our results so far since overconfidence is not a significant determinant of compensation in 2008, and the reward for having a high taste for competition is concentrated on the expected performance bonus rather than on the base salary.

This line of reasoning also suggests that the association between overconfidence, taste for competition, and compensation might change over time. As employees build a track record, it is more likely that employers can discriminate these two traits and reward them differently. Also, over time, bonuses are typically based on *realized* rather than expected performance. Thus, if realized performance is affected negatively by overconfidence, the realized bonus will be showing a correlation with this trait, even if the company cannot assess this and other characteristics. We can explore these hypotheses by studying the relationship between taste for competition and compensation in 2015, which includes the 2014 realized bonuses.

IV. Taste for competition and compensation in 2015

In Table 6, we study the relationship between taste for competition and overall compensation, base salary, and bonuses in 2015. In columns I to IV, we do not control for industry, while in columns V to VIII, we repeat the same specifications controlling for industry fixed effects.

In columns I and V, we regress the log of total compensation in 2015 on the competitive and gender dummies and the controls. Compared to the coefficients from 2008, the coefficient of the competitive dummy in 2015 drops by half and is no longer statistically different from zero. It is even smaller when we control for industry fixed effects. To check whether this change is due to sample size, given that only 61% of the sample answered to the 2015 follow-up survey, in the Online Appendix, we re-estimate the regressions for the compensation in 2008 solely for the sample of MBAs for whom we have 2015 data (see Table A.11). The results show that the coefficient of the taste-for-competition variable is roughly the same as in the full sample, but it is not statistically different

[13] The structure of this recruiting process is similar to the one adopted at the undergraduate level to recruit business professionals in prestigious consulting firms, investment banks, and technology firms.

Table 6 – Determinants of compensation in 2015

Note: Regressions of the log of total compensation in 2015 in columns I and V and of the log of base salary in 2015 in columns II and VI. Hurdle model of the likelihood of receiving a bonus in columns III and VII, and of its magnitude (in logs) in columns IV and VIII. Linear estimates in columns I, II, IV, V, VI, and VIII. Marginal effects in columns III and VII. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Total income	Base salary	Bonus		Total income	Base salary	Bonus	
			Received	Amount			Received	Amount
	I	II	III	IV	V	VI	VII	VIII
Competitive	0.035	0.003	0.003	0.085	0.001	–0.008	–0.007	0.030
	(0.098)	(0.066)	(0.047)	(0.195)	(0.089)	(0.064)	(0.048)	(0.172)
Woman	–0.384***	–0.194***	–0.002	–0.986***	–0.218**	–0.142**	0.026	–0.623***
	(0.103)	(0.069)	(0.049)	(0.205)	(0.097)	(0.070)	(0.051)	(0.188)
Overconfidence	–0.128	–0.082	–0.101***	–0.027	–0.133*	–0.087*	–0.104***	–0.083
	(0.081)	(0.054)	(0.037)	(0.161)	(0.073)	(0.053)	(0.038)	(0.143)
Risk aversion	–0.024**	–0.012*	–0.005	–0.026	–0.028***	–0.013**	–0.007	–0.035**
	(0.010)	(0.007)	(0.005)	(0.020)	(0.009)	(0.006)	(0.005)	(0.018)
Performance	–0.048	–0.056	–0.008	–0.050	–0.012	–0.046	–0.001	0.038
	(0.068)	(0.046)	(0.032)	(0.133)	(0.062)	(0.045)	(0.035)	(0.118)
Finance					0.632***	0.256***	0.110**	1.290***
					(0.084)	(0.060)	(0.053)	(0.162)
Consulting					0.284***	0.221**	0.016	0.395
					(0.123)	(0.089)	(0.064)	(0.241)
Obs.	250	250	250	218	250	250	250	218
F test / χ^2 test	5.205***	2.958**	10.051*	30.871***	12.708***	5.109***	28.058***	104.120***

from zero. Thus, in the 2015 sample, we certainly have a power issue, yet this is not the only reason why taste for competition is not statistically significant with 2015 compensation. Keeping the same sample, taste for competition seems to have less of an impact on compensation in 2015 than in 2008.[14]

Another difference with the results from 2008 is the coefficient on overconfidence. In 2008, the size of the coefficient was basically equal to zero (less than 1%). In 2015, the absolute size of the coefficient increases to 12.8% and 13.3%, depending on whether we control for industry fixed effects, and it is marginally significant. This change is consistent with our hypothesis that employers were not able to link salaries to overconfidence in 2008 but might be able to so in 2015.

[14] One might still worry that these results are driven by selection of respondents in the 2015 survey relative to the 2008 sample. To study this possibility, we also exploit the fact that some individual characteristics of these students predict their attachment to the University (and the willingness to spend time filling a survey) and their tendency to value time. Thus, in the Online Appendix, we present estimates of a Heckman selection model where we first estimate the marginal effect of the probability of answering the survey and then a linear model of the total compensation in 2015. The results remain substantially unchanged.

The gender dummy is now much larger than in 2008: the coefficient more than tripled, from 10.7% to 38.4% (or from 10.4% to 21.8%, controlling for industry fixed effects), implying an increase in the gender gap from $17K to $89K (or from $16K to $53K with industry fixed effects).

In columns II and VI, we repeat the same specifications for the log of the base salary. The results show a statistically and economically insignificant effect of taste for competition. Overconfidence has a borderline negative and significant effect. The gender gap is $31K (19.4%) or $23K (14.2%) with industry fixed effects, which is much higher than the 2008 gender gap in base salary of around $2K.

In the remaining columns, we run a two-step hurdle model to estimate, in the first stage, the probability of getting a bonus (columns III and VII of Table 6), and in the second stage, the log of the bonus received (columns IV and VIII). Note that, on average, the bonus was $132K, and 87% of the sample received at least some bonus. Neither taste for competition nor gender predicts the probability of receiving a bonus, while overconfidence has a negative and strongly significant effect. When we look at the magnitude of the bonus, we find that neither taste for competition nor overconfidence is economically or statistically significant. By contrast, we find that gender has a powerful effect: on average, women receive $60K less. In part, this effect is due to industry selection. When we control for industry, the effect drops to $41K.

How is it possible that taste for competition affects MBA earnings at graduation, but not seven years later? As mentioned before, taste for competition might be a valuable trait only when it is not combined with overconfidence. Overconfident employees who are eager to compete because they mistakenly think they are going to win will likely end up losing money for their employers. Thus, employers would like to hire and reward only MBAs who like competition but are not overconfident.[15] Unfortunately, employers are unlikely to observe the degree of overconfidence of potential employees: they must see them in action. Therefore, if they want to attract employees who like competition, employers have to offer a higher *expected* bonus to all "competitive" MBAs at the beginning of their career. Over time, employers could learn the degree of overconfidence of their competitive employees and start to reward only those who are not overconfident, penalizing the others. What is more, even if overconfidence remains unobservable, as long as employers can condition bonuses on observed performance and performance is negatively affected by overconfidence, we would observe a negative correlation between the realized bonus and overconfidence. This hypothesis is consistent with the fact that overconfidence (by itself) has no impact on compensation in 2008 but has a negative effect in 2015.

[15] Overconfidence may be less of an issue for employees who do not have a taste for competition. Their aversion to competition means that they will take fewer risks and therefore the overestimation of their abilities is bound to be less costly to employers.

Table 7 – Interaction with overconfidence

Note: Regressions of the log of total compensation in 2015 in columns I and V and of the log of base salary in 2015 in columns II and VI. Hurdle model of the likelihood of receiving a bonus in columns III and VII, and of its magnitude (in logs) in columns IV and VIII. Linear estimates in columns I, II, IV, V, VI, and VIII. Marginal effects in columns III and VII. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Total income	Base salary	Bonus Received	Bonus Amount	Total income	Base salary	Bonus Received	Bonus Amount
	I	II	III	IV	V	VI	VII	VIII
Competitive	0.120	0.040	0.025	0.214	0.064	0.018	0.013	0.125
	(0.103)	(0.070)	(0.049)	(0.200)	(0.094)	(0.068)	(0.048)	(0.177)
Woman	–0.391***	–0.197***	–0.004	–1.008***	–0.226**	–0.146**	0.022	–0.645***
	(0.102)	(0.069)	(0.048)	(0.203)	(0.097)	(0.070)	(0.044)	(0.187)
Overconfidence	0.017	–0.018	–0.063	0.245	–0.027	–0.042	–0.064	0.117
	(0.098)	(0.067)	(0.054)	(0.194)	(0.089)	(0.065)	(0.052)	(0.173)
Overconfidence	–0.350**	–0.152	–0.080	–0.678**	–0.254**	–0.108	–0.074	–0.494**
× competitive	(0.137)	(0.093)	(0.064)	(0.278)	(0.125)	(0.091)	(0.063)	(0.247)
Risk aversion	–0.023**	–0.012*	–0.005	–0.025	–0.027***	–0.013**	–0.006	–0.035**
	(0.010)	(0.007)	(0.004)	(0.020)	(0.009)	(0.006)	(0.004)	(0.017)
Performance	–0.077	–0.069	–0.018	–0.097	–0.034	–0.055	–0.010	0.003
	(0.068)	(0.046)	(0.033)	(0.133)	(0.062)	(0.045)	(0.032)	(0.118)
Finance					0.614***	0.248***	0.091**	1.258***
					(0.084)	(0.061)	(0.041)	(0.161)
Consulting					0.267**	0.213**	0.014	0.363
					(0.123)	(0.089)	(0.071)	(0.239)
Obs.	250	250	250	218	250	250	250	218
F test / χ^2 test	5.519***	2.923***	12.150*	37.653***	11.774***	4.655***	17.100**	110.015***

To test this hypothesis, in Table 7, we re-estimate the specifications of Table 6, adding an interaction between taste for competition and overconfidence. Consistent with our hypothesis, we find that the interaction between taste for competition and overconfidence has a negative and significant effect on salary. This effect is economically large and is visualized in Figure 2, where we report the estimated marginal effect of taste for competition on total compensation (in thousands of USD) and the 90% confidence intervals at different levels of overconfidence. Points to the right of the dotted line correspond to overconfident individuals. Among MBAs whose degree of overconfidence is one standard deviation below the mean, having a high taste for competition implies $74K (25.8%) higher earnings. By contrast, among MBAs whose degree of overconfidence is one standard deviation above the mean, having a high taste for competition implies $44K (18.6%) lower earnings. Table 7 and Figure 2 also show that the interaction between taste for competition and overconfidence is more prominent in the bonus component of the MBAs' earnings.[16]

[16] Like before, in the Online Appendix, we test whether these results are robust to adding a large set of individual controls (Table A.9) and to the use of a Heckman selection model to correct for the selection of respondents into the 2015 follow-up survey (Table A.10). The results remain substantially unchanged.

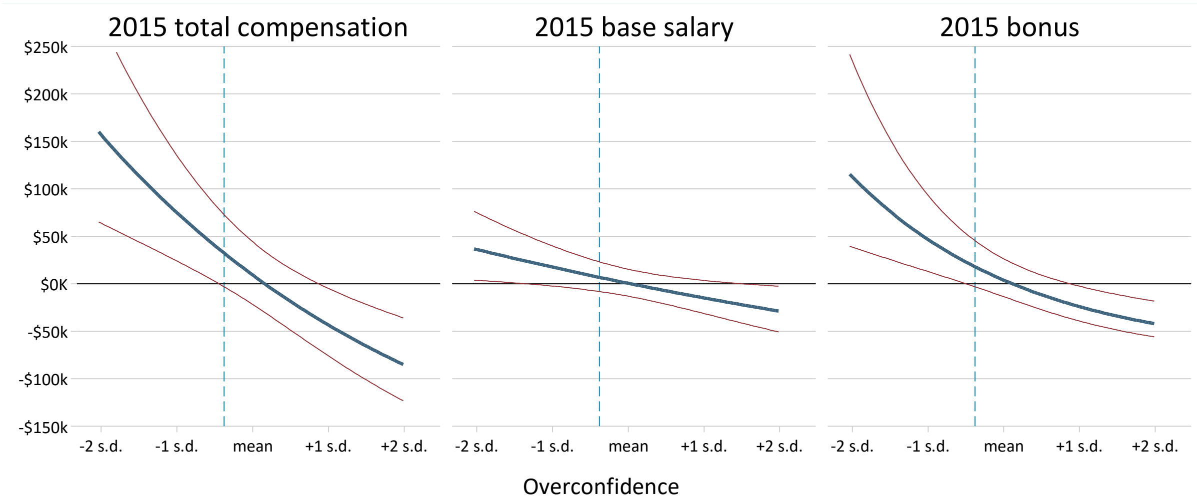


Figure 2 – Estimated marginal effect of taste for competition on total compensation in 2015 (in thousands of USD) and 90% confidence intervals at different levels of overconfidence. Points to the right of the dotted line correspond to overconfident individuals.

In unreported regressions, we repeated the same specifications with the 2008 data. The interaction between taste for competition and overconfidence is both economically and statistically indistinguishable from zero, suggesting that our interpretation that companies learn only over time the degree of overconfidence of their employees or that performance indirectly reveals overconfidence is a plausible one.

In conclusion, the evidence is most consistent with a higher taste for competition being a positive characteristic only in the absence of overconfidence. When employers are unable to observe overconfidence, they are willing to compensate all "competitive" MBAs with a higher pay-for-performance. Once bonuses are paid based on realized performance, only competitive MBAs who are not overconfident are rewarded.[17]

V. Conclusions

In this paper, we study whether gender differences in taste for competition can explain the observed gender gap in wages among MBAs. We find that taste for competition, as measured by a laboratory experiment, positively predicts wages. We also find that, on average, men exhibit a higher taste for competition. Nevertheless, we find that taste for competition explains only 10% of the gender gap at graduation and none seven years later.

[17] One can interpret our results regarding risk aversion similarly. It could be that, given the recruitment process, employers cannot observe attitudes toward risk early on, resulting in a nonexistent correlation between risk aversion and compensation in 2008. Over time, irrespective of whether employers directly observe risk aversion, the negative association between risk aversion and compensation can be the result of a negative relationship between risk aversion and performance. That being said, for risk aversion, it is also plausible that the lower compensation of risk-averse individuals is driven by self-selection into lower-paying but less risky jobs over time.

We explore possible explanations for the correlation between taste for competition and earnings. We do not find evidence that "competitive" MBAs are better at generating multiple offers, which can boost their salaries. Our evidence is consistent with the hypothesis that employers consider taste for competition to be a valuable trait because it can boost performance, but only when it is not associated with overconfidence. We find that the taste for competition is linked to the variable component of earnings, bonuses. At graduation, bonuses are set in advance, without knowledge of effective performance. Recruiters, unable to observe overconfidence before hiring, reward all competitive MBAs at graduation with higher expected bonuses. Seven years later, bonuses are based on realized performance. Hence, over time, only the non-overconfident MBAs are rewarded for being competitive through higher realized bonuses, while the overconfident MBAs are penalized.

These results could be explained in two different ways: either employers learn over time the characteristics of workers and tie compensation to those characteristics, or the correlation emerges simply because these characteristics affect performance, which is used to set bonuses ex-post. Unfortunately, with our current data, we are unable to distinguish between these two explanations.

These results can also explain why taste for competition explains less of the gender gap as we move from compensation based on expected performance to compensation based on realized performance. Men like to compete more, but they are also more overconfident. Thus, initially, men get rewarded more as firms cannot observe overconfidence, and workers have yet to perform. Over time, as overconfidence and performance are revealed, men get penalized more, eliminating this component of the gender gap. Indeed, one of the main advantages of our framework is that we observe two "types" of compensations: early on, when compensation is more likely to be based on expected performance, and later on, when salaries and bonuses are more likely to be based on realized performance (particularly bonuses). This result is especially important to evaluate potentially naïve policy implications that suggest that women should "change" and become more competitive (on this point, see also Niederle, 2017).

A few words of caution are warranted when thinking about the external validity of our results. In this paper, we study young professional careers. The recruiting process for MBAs involves a complicated competitive process where recruiting firms meet candidates through social mixers, presentations, and several rounds of formal interviews. Since this recruiting process is particularly competitive, our results may have limited external validity in other professions. Because our sample is non-representative, future research is needed to understand whether these results extend to the general population. However, while we focus on MBAs, many sought-after jobs share a protracted recruitment process that is very similar (especially for jobs in consulting, banking, and some of the STEM fields).

Our paper contributes to the growing literature linking measurable characteristics in the lab with relevant labor-market outcomes. Compared to other studies, our work highlights the importance of measuring the effect of lab-generated measures over time as their relationship with labor-market outcomes can vary over people's careers. In our sample of young business professionals, the lab-generated measure of taste for competition plays an important role during the recruiting phase, presumably, when few other characteristics are observable to recruiters. To the extent that employers can observe some of these traits over time or that performance is correlated with these traits, performance-based compensation is more likely to correlate with characteristics previously measured in the lab that are unobservable to recruiters.

References

Apesteguia, Jose and Ignacio Palacios-Huerta. 2010. "Psychological Pressure in Competitive Environments: Evidence from a Randomized Natural Experiment." *American Economic Review* 100(5): 2548–64.

Balafoutas, Loukas and Matthias Sutter. 2012. "Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory." *Science* 335(6068): 579–82.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Armando Jose Garcia Pires, and Bertil Tungodden. 2015. "Competitive in the Lab, Successful in the Field?" *Journal of Economic Behavior & Organization* 118: 303–17.

Bertrand, Marianne. 2018. "Coase Lecture - The Glass Ceiling." *Economica* 85(338): 205–31.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz. 2010. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics* 2(3): 228–55.

Blau, Francine D. and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55(3): 789–865.

Bohnet, Iris, Alexandra van Geen, and Max Bazerman. 2016. "When Performance Trumps Gender Bias: Joint vs. Separate Evaluation." *Management Science* 62(5): 1225–34.

Buser, Thomas, Muriel Niederle, and H. Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." *The Quarterly Journal of Economics* 129(3): 1409–47.

Buser, Thomas, Noemi Peter, and Stefan C. Wolter. 2017a. "Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland." *American Economic Review* 107(5): 125–30.

Buser, Thomas, Noemi Peter, and Stefan C. Wolter. 2017b. "Gender, Willingness to Compete and Career Choices along the Whole Ability Distribution." Working Paper.

Cason, Timothy N., William A. Masters, and Roman M. Sheremeta. 2010. "Entry into Winner-Take-All and Proportional-Prize Contests: An Experimental Study." *Journal of Public Economics* 94(9–10): 604–11.

Cragg, John G. 1971. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica* 39(5): 829.

Croson, Rachel and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 47(2): 448–74.

Dariel, Aurelie, Curtis Kephart, Nikos Nikiforakis, and Christina Zenker. 2017. "Emirati Women Do Not Shy Away from Competition: Evidence from a Patriarchal Society in Transition." *Journal of the Economic Science Association* 3(2): 121–36.

Dreber, Anna, Emma von Essen, and Eva Ranehill. 2014. "Gender and Competition in Adolescence: Task Matters." *Experimental Economics* 17(1): 154–72.

Flory, Jeffrey A., Andreas Leibbrandt, and John A. List. 2015. "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions." *Review of Economic Studies* 82(1): 122–55.

Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study." *Journal of Political Economy* 127(4): 1826–63.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *The Quarterly Journal of Economics* 118(3): 1049–74.

Handelsman, Jo, Nancy Cantor, Molly Carnes, Denice Denton, Eve Fine, Barbara Grosz, Virginia Hinshaw, Cora Marrett, Sue Rosser, Donna Shalala, and Jennifer Sheridan. 2005. "More Women in Science." *Science* 309(5738): 1190–91.

Healy, Andrew and Jennifer Pate. 2011. "Can Teams Help to Close the Gender Competition Gap?" *The Economic Journal* 121(555): 1192–1204.

Holt, Charles A. and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92(5): 1644–55.

Kamas, Linda and Anne Preston. 2010. Are Women Really Less Competitive than Men? Santa Clara University Working Paper.

Kamas, Linda and Anne Preston. 2018. "Competing with Confidence: The Ticket to Labor Market Success for College-Educated Women." *Journal of Economic Behavior & Organization* 155: 231–52.

Leibbrandt, Andreas and John A. List. 2015. "Do Women Avoid Salary Negotiations? Evidence from a Large-Scale Natural Field Experiment." *Management Science* 61(9): 2016–24.

Niederle, Muriel. 2017. "A Gender Agenda: A Progress Report on Competitiveness." *American Economic Review* 107(5): 115–19.

Niederle, Muriel, Carmit Segal, and Lise Vesterlund. 2013. "How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness." *Management Science* 59(1): 1–16.

Niederle, Muriel and Lise Vesterlund. 2007. "Do Women Shy Away From Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122(3): 1067–1101.

Niederle, Muriel and Lise Vesterlund. 2011. "Gender and Competition." *Annual Review of Economics* 3(1): 601–30.

Oyer, Paul. 2008. "The Making of an Investment Banker: Stock Market Shocks, Career Choice, and Lifetime Income." *The Journal of Finance* 63(6): 2601–28.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2008. A Description of the Templeton Chicago MBAs Longitudinal Study. University of Chicago.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. "How Stereotypes Impair Women's Careers in Science." *Proceedings of the National Academy of Sciences* 111(12): 4403–8.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. 2017. "Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender." *The Economic Journal* 127(604): 2153–86.

Roth, Alvin E. 1995. "Introduction to Experimental Economics." Pp. 3–109 in *The Handbook of Experimental Economics*, edited by J. H. Kagel and A. E. Roth. Princeton, NJ: Princeton University Press.

Samek, Anya. 2019. "Gender Differences in Job Entry Decisions: A University-Wide Field Experiment." *Management Science* 65(7): 3272–81.

van Veldhuizen, Roel. 2018. "Gender Differences in Tournament Choices: Risk Preferences, Overconfidence or Competitiveness?" WZB Working Paper.

Wozniak, David, William T. Harbaugh, and Ulrich Mayr. 2014. "The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices." *Journal of Labor Economics* 32(1): 161–98.

Zhang, Y. Jane. 2019. "Culture, Institutions and the Gender Gap in Competitive Inclination: Evidence from the Communist Experiment in China." *The Economic Journal* 129(617): 509–52.

Online Appendix for the paper: Taste for competition and the gender gap among young business professionals

This document contains supplementary materials for the paper Reuben, Sapienza, and Zingales (2019). Section I presents a detailed analysis of whether MBAs who consented to the study in 2008 vary systematically from MBAs who did not. Section II presents a similar analysis of whether MBAs who answered the survey in 2015 vary systematically from those who did not. Section III describes in detail the numerous robustness checks reported in the paper but not fully described there due to space constraints. Section IV describes the procedures used to conduct the experiment and survey, including a sample of the instructions used to elicit taste for competition. Lastly, Section V describes the additional variables used in the robustness checks.

I. Selection into the sample in 2006

In this section, we evaluate whether the 409 participants who consented to the analysis of all their data (including their earnings) differ from the 129 participants who consented to the analysis of only some of their data. In the top part of Table A.1, we present the means and standard deviations of variables related to taste for competition plus the fraction of women. For each variable, the table also displays the *p*-value obtained when we test whether the two groups of participants are significantly different from each other. Specifically, we use simple *t*-tests for the continuous variables and χ^2 tests for categorical variables. In the bottom part of Table A.1, we present the same information for the control variables that we will use for the robustness checks in Section III. We describe these variables and how we collected them in Section V.

By and large, we find no significant differences between the participants who fully consented to the study and those who did not. If we use an unadjusted significance threshold of 5%, we find a significant difference in four out of twenty-five variables (age, GMAT verbal percentile, GPA, and the survey measure of overconfidence). However, if we adjust *p*-values with the Benjamini-Hochberg method to account for multiple comparisons (Benjamini and Hochberg 1995), then we find a significant difference in only one variable (GPA). Importantly for this paper, neither the fraction of women nor the fraction of participants who chose the tournament are significantly different.

Table A.1 – Summary statistics depending on consenting to all parts of the study in 2006

Note: Means, standard deviations, and number of observations for variables of interest. The rightmost column displays *p*-values from tests of equality of distributions between people who consented to the analysis of all their data and those who did not (*t*-tests for ordinal variables and χ^2 tests for categorical variables).

	CONSENTED			DID NOT CONSENT			
	mean	s.d.	N	mean	s.d.	N	*p*-value
Competitive	0.52	0.50	409	0.51	0.50	123	0.867
Performance (rank in sums tasks)	2.48	0.77	409	2.52	0.77	123	0.649
Expected rank in sums tasks	2.24	0.77	409	2.31	0.82	123	0.430
Overconfidence	0.24	0.63	409	0.21	0.76	123	0.697
Risk aversion coefficient	4.74	4.41	409	3.87	4.79	123	0.074
Taste for high reward	0.41	0.49	409	0.44	0.50	123	0.513
Fraction of women	0.3	0.46	409	0.34	0.48	129	0.388
Additional control variables							
Age	28.22	2.44	409	28.93	2.72	129	0.009
Fraction non-white	0.55	0.50	409	0.64	0.48	129	0.062
Fraction US residents	0.77	0.42	409	0.74	0.44	129	0.584
Fraction married before MBA	0.26	0.44	409	0.22	0.41	129	0.362
Fraction religious	0.47	0.50	409	0.42	0.50	129	0.312
GMAT Quantitative percentile	81.91	12.81	406	80.84	16.06	129	0.489
GMAT Verbal percentile	88.02	11.45	406	85.31	12.75	129	0.033
GMAT Analytic percentile	71.91	21.75	383	68.7	22.63	112	0.184
GPA	3.33	0.34	391	3.18	0.42	99	0.002
CRT score	2.44	1.33	409	2.43	1.35	129	0.979
RMET score	0.75	0.10	409	0.74	0.10	129	0.469
Discount rate	0.05	0.04	376	0.05	0.04	108	0.718
Trust	0.38	0.30	409	0.34	0.29	123	0.183
Reciprocity	0.36	0.20	409	0.33	0.20	123	0.151
Cooperation	0.33	0.47	409	0.29	0.46	123	0.436
Survey overconfidence	0.90	4.56	391	1.92	4.42	99	0.044
Survey risk aversion (general)	6.44	1.89	409	6.57	2.19	129	0.541
Survey risk aversion (monetary)	1.49	1.01	409	1.62	0.95	129	0.182

Moreover, if we test for each gender whether the fraction of individuals who chose the tournament differs between those who fully consented and those who did not, we do not find a statistically significant difference for men ($p = 0.794$) or women ($p = 0.704$).

Finally, to test whether taste for competition differs between participants who fully consented to the study and those who did not, we run a probit regression with the participants' choice of the tournament as the dependent variable. In line with the regressions in Table 2, as independent variables, we include the participants' gender, performance, overconfidence, and risk aversion coefficient. In addition, we also add a dummy variable equal to one for participants who did not fully consent. We find that the estimated marginal effect of the dummy variable is minimal (0.003) and is not statistically significant (p = 0.957).

II. Attrition in the 2015 follow-up survey

Of the 409 participants who consented to the analysis of their data in 2006, 263 (64.3%) answered the follow-up survey in 2015. To evaluate whether the 263 survey respondents differ from the nonresponding 146 participants, in the top part of Table A.2 we present the means and standard deviations of the variables in Table 1 for which we have data for both samples. For each variable, the table also displays the p-value obtained when we test whether the two samples are significantly different from each other (t-tests for continuous variables and χ^2 tests for categorical variables). In the bottom part of Table A.2, we present the same information for the control variables that we will use for the robustness checks in Section III. We describe these variables and how we collected them in Section V.

For most variables, there are no statistically significant differences between the participants who answered the survey and those who did not. If we use an unadjusted significance threshold of 5%, then we find a significant difference in three out of the fifteen variables in the top part of Table A.2 (overconfidence, one-off bonuses, and gender) and in six out of the nineteen variables in the bottom part of Table A.2 (donations to University of Chicago, discount rate, fraction of US residents, fraction of white individuals, GMAT verbal percentile, and CRT scores). However, accounting for multiple comparisons by adjusting p-values with the Benjamini-Hochberg method, reduces the number of variables with significant differences to one out of thirty-four (donations to the University of Chicago). The fact that we find significant differences for this variable is not very surprising. It is to be expected that alumni who donated money to the University of Chicago are more likely to respond to a survey sent by professors from that university.

We also test whether taste for competition differs between participants who consented to the analysis of all their data and those who did not. To do so, we run a probit regression with the participants' choice of the tournament as the dependent variable and the participants' gender, performance, overconfidence, and risk aversion coefficient as independent variables. In addition, we also included a dummy variable equal to one for participants who did not respond to the follow-up survey. We find that the estimated marginal effect of the dummy variable is small (0.022) and is not statistically significant (p = 0.690).

In conclusion, although there is no clear-cut evidence that there are strong selection effects into responding to the follow-up survey, there might be reasons for worry. In particular, if we do not correct p-values for multiple testing, we do see significant differences in three important variables for the paper: overconfidence, one-off bonuses, and gender. Moreover, although the difference is not statistically significant, the total compensation in 2008 of respondents to the survey is noticeably

Table A.2 – Summary statistics depending on responding to the follow-up survey in 2015

Note: Means, standard deviations, and number of observations for variables of interest. The rightmost column displays *p*-values from tests of equality of distributions between people who responded to the follow-up survey in 2015 and those who did not (*t*-tests for ordinal variables and χ^2 tests for categorical variables).

	RESPONDENT			NON-RESPONDENT			
	mean	s.d.	N	mean	s.d.	N	*p*-value
Competitive	0.52	0.50	263	0.52	0.50	146	0.994
Performance (rank in sums tasks)	2.43	0.78	263	2.58	0.76	146	0.063
Expected rank in sums tasks	2.25	0.78	263	2.23	0.76	146	0.735
Overconfidence	0.18	0.63	263	0.35	0.64	146	0.008
Risk aversion coefficient	4.52	4.40	263	5.13	4.42	146	0.182
Taste for high reward	0.41	0.49	263	0.40	0.49	146	0.957
Total compensation in 2008	180.77	161.02	263	164.11	144.32	146	0.284
Base salary in 2008	108.15	18.65	263	105.40	16.62	146	0.127
Total bonus in 2008	72.63	153.66	263	58.70	140.26	146	0.354
One-off bonus in 2008	43.65	30.50	263	37.30	24.37	146	0.022
Expected performance bonus in 2008	28.98	144.84	263	21.40	137.19	146	0.600
Number of competing job offers	0.47	0.87	263	0.32	0.71	146	0.068
Fraction working in finance in 2008	0.50	0.50	263	0.53	0.50	146	0.674
Fraction working in consulting in 2008	0.26	0.44	263	0.22	0.42	146	0.674
Fraction of women	0.27	0.44	263	0.36	0.48	146	0.041
Additional control variables							
Age	28.04	2.30	263	28.54	2.65	146	0.057
Fraction non-white	0.50	0.50	263	0.64	0.48	146	0.005
Fraction US residents	0.81	0.39	263	0.68	0.47	146	0.003
Fraction married before MBA	0.25	0.43	263	0.27	0.45	146	0.552
Fraction religious	0.48	0.50	263	0.45	0.50	146	0.600
GMAT Quantitative percentile	81.34	12.61	262	82.96	13.14	144	0.228
GMAT Verbal percentile	89.34	9.80	262	85.61	13.67	144	0.004
GMAT Analytic percentile	73.47	21.34	245	69.13	22.27	138	0.064
GPA	3.35	0.34	255	3.29	0.35	136	0.072
CRT score	2.54	1.33	263	2.26	1.32	146	0.045
RMET score	0.75	0.10	263	0.74	0.10	146	0.120
Discount rate	0.05	0.04	244	0.06	0.05	132	0.002
Trust	0.38	0.30	263	0.40	0.30	146	0.475
Reciprocity	0.36	0.20	263	0.37	0.21	146	0.828
Cooperation	0.35	0.48	263	0.30	0.46	146	0.358
Survey overconfidence	0.93	4.51	255	0.85	4.66	136	0.858
Survey risk aversion (general)	6.54	1.83	263	6.26	1.98	146	0.155
Survey risk aversion (monetary)	1.46	1.02	263	1.53	1.01	146	0.526
Donations to University of Chicago	0.87	1.24	260	0.48	0.94	145	0.000

higher than that of non-respondents (17K or 10% more). For this reason, in Section III, we perform a series of robustness checks where we account for selection into the follow-up survey.

III. Robustness checks

III.A. Gender differences in taste for competition

Measurement error and misspecification

Our first robustness check addresses concerns about using a residual measure for taste for competition (Gillen, Snowberg, and Yariv 2019; van Veldhuizen 2018). As Niederle and Vesterlund (2007), we measure taste for competition by looking at whether individuals choose a tournament payment scheme. However, since there are several reasons why an individual might choose tournament pay. We interpret the choice of tournament pay as indicating a higher taste for competition only after we control for the individual's performance, overconfidence, and risk aversion. This way of measuring taste for competition has recently come under scrutiny because it is not a direct measure of the trait of interest. More precisely, if there is measurement error in the control variables or they are incorrectly specified in the regression, it is possible for there to be a bias in the estimated effects of taste for competition.

If the concerns raised by Gillen et al. (2019) are valid in our data, they could steer us towards incorrect inferences. First, in regressions where we evaluate whether there are gender differences in taste for competition (Table 2), a significant coefficient for the gender dummy might be due to (residual) gender differences in risk aversion or overconfidence and might not be due to differences in taste for competition. Second, in regressions where we test the effect of taste of competition on compensation (Tables 3, 4, 6, and 7), a significant coefficient for the competitive dummy might again be due to (residual) effects of risk aversion or overconfidence. In this subsection, we take a closer look at the identification of gender differences in taste for competition. In subsection III.B, we do the same for the effect of taste of competition on compensation.

To evaluate whether there is bias in the identification of gender differences in taste for competition, we reran the probit regressions in Table 2 with additional control variables. We report the resulting marginal effects in Table A.3. In column I, we simply reproduce the last regression of Table 2, where we regress the participants' choice between tournament and piece-rate pay on the participants' gender, performance, overconfidence, and risk preferences. In this regression, there is a significant gender gap in the choice of tournament pay of 14.8% (p = 0.013), which we interpret in the paper as a gender difference in taste for competition.

In column II, we include the square of the original control variables. Introducing these variables allows us to capture non-linear relations between choosing tournament pay and performance, risk aversion, and overconfidence, which can be a source of bias in the identification of gender differences in the taste for competition. As can be seen in Table A.3, the coefficient of the gender dummy is mostly unaffected by the introduction of all these control variables: it increases by 0.4 percentage points to 15.2% (p = 0.012), which is less than 0.07 standard deviations.

Table A.3 – Robustness of the gender gap in taste for competition to measurement error and misspecification

Note: Regressions of the choice of tournament pay. Marginal effects from probit regressions and standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III	IV
Woman	–0.148**	–0.152**	–0.147**	–0.156**
	(0.060)	(0.061)	(0.064)	(0.065)
Performance	0.351***	1.093***	0.352***	1.086***
	(0.043)	(0.253)	(0.044)	(0.254)
Overconfidence	0.320***	0.354***	0.316***	0.349***
	(0.050)	(0.061)	(0.050)	(0.061)
Risk aversion	–0.019***	–0.025	–0.019***	–0.027
	(0.006)	(0.017)	(0.007)	(0.017)
Performance²		0.149***		0.147***
		(0.049)		(0.050)
Overconfidence²		–0.057		–0.055
		(0.051)		(0.051)
Risk aversion²		0.001		0.001
		(0.001)		(0.001)
Survey risk aversion (general)			0.022	0.020
			(0.015)	(0.016)
Survey risk aversion (monetary)			0.032	0.036
			(0.031)	(0.031)
Overconfidence in GPA			–0.005	–0.005
			(0.006)	(0.006)
Obs.	409	409	409	409
χ^2 test	117.954***	128.365***	121.059***	131.256***

In column III, we use answers to the initial survey to include additional measures of the participants' risk aversion and overconfidence. These variables are bound to be noisier than our lab measures since we did not elicit them with incentive-compatible methods. However, as demonstrated by Gillen et al. (2019), they can capture some of the potential measurement error of the incentive-compatible variables. For risk aversion, we use a commonly-used survey measure of general attitudes towards risk (Falk et al. 2018) and another self-reported measure of risk attitudes in the monetary domain. For overconfidence, we use the participants' expected GPA decile (estimated in 2006) minus their actual GPA decile (in 2008). We provide descriptive statistics for these variables in Tables A.1 and A.2 and a detailed description in Section V. Once again, the coefficient of the gender dummy is robust to the introduction of all these variables: it decreases by 0.1 percentage points to 14.7% (i.e., less than 0.02 standard deviations, p = 0.023).

Finally, in column IV, we include all the additional control variables from the regressions in columns II and III. Compared to the regression in column I, we find a slight increase in the gender coefficient

Table A.4 – Robustness of the gender gap in taste for competition to additional controls

Note: Regressions of the choice of tournament pay. Marginal effects from probit regressions and standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	I	II	III
Woman	–0.148**	–0.119*	–0.125*
	(0.060)	(0.067)	(0.073)
Performance	0.351***	0.321***	1.122***
	(0.043)	(0.047)	(0.265)
Overconfidence	0.320***	0.300***	0.326***
	(0.050)	(0.053)	(0.065)
Risk aversion	–0.019***	–0.019***	–0.030
	(0.006)	(0.007)	(0.019)
Additional controls	No	Yes	Yes
Measurement error controls	No	No	Yes
Obs.	409	409	409
χ^2 test	117.954***	140.918***	154.838***

of 0.8 percentage points to 15.6% (p = 0.016), which is only around 0.13 standard deviations. Hence, overall, we do not find evidence that the gender difference in taste for competition is due to measurement error or misspecification.

Additional control variables

Here, we test the robustness of the gender difference in taste for competition to the inclusion of a large set of control variables. Specifically, we include all the fifteen additional control variables seen in Tables A.1 and A.2. These variables include demographic characteristics (e.g., age, race, residence, marital status, religiosity), measures of different kinds of abilities (e.g., mathematical, verbal, and analytical skills, capacity for cognitive reflection, and emotional intelligence), and other standard experimental measures (e.g., willingness to trust, reciprocate, and cooperate with others). We describe all these variables in detail in Section V. This exercise allows us to evaluate whether taste for competition describes variation between individuals that is not captured by typically measured observables. In addition, if risk aversion and overconfidence affect variables such as GMAT scores, trust, and cooperation, then the inclusion of these variables ought to reduce the effects of measurement error (as reasoned above).

The results of including all these control variables are seen in Table A.4. As above, in column I, we simply reproduce the last regression of Table 2. In column II, we include the fifteen additional control variables, while in column III, we also add the six variables used in Table A.3 to reduce bias due to measurement errors and misspecification. We can see that the inclusion of all these control variables has a moderate effect on the magnitude of the gender gap in tournament pay. In column II, it shrinks by 2.9 percentage points (around 0.47 standard deviations) to 11.9% (p = 0.078), and in column III, by 2.3 percentage points (around 0.36 standard deviations) to 12.5% (p = 0.089).

We think that this is compelling evidence that, by and large, competitiveness captures individual variation that would otherwise remain unobserved. Moreover, even though the shrinking of the gender gap (and the increase in its p-value) might suggest that there is a bias due to measurement error, we should point out that these regressions are not as appropriate to test measurement error as the regressions in Table A.3 (where the coefficient does not shrink). The reason being that the variables Table A.3 are measures of risk preferences and overconfidence that are not related to taste for competition, while the additional control variables in these regressions could be related to this trait. For instance, it is conceivable that taste for competition, which has been shown to affect educational choices (Buser, Niederle, and Oosterbeek 2014; Reuben, Wiswall, and Zafar 2017), has a direct effect on variables meant to measure abilities such as GMAT test scores, which would also explain the attenuation of the coefficient.

III.B. Effect of taste for competition on compensation in 2008

Next, we provide a series of robustness checks for the effect of taste for competition on compensation in 2008 (reported in Tables 3 and 4). Throughout this subsection, we focus solely on regressions without industry fixed effects to keep the number and size of the tables at a reasonable level. In line with the main body of the paper, our results do not change much if we control for industry. We can provide the regressions with industry fixed effects upon request.

Measurement error and misspecification

We start by looking at whether the effect of taste of competition on compensation in 2008 is overestimated due to measurement error or misspecification of the control variables (Gillen et al. 2019). Similarly to subsection III.A, in Table A.5, we reran the more interesting regressions from Tables 3 and 4 adding the following independent variables: the squares of performance, risk aversion, and overconfidence, the two survey measures of risk aversion, and the survey measure of overconfidence (see Tables A.1 and A.2 for descriptive statistics of these variables and Section V for a detailed description).

In column I, the dependent variable is the log of total compensation in 2008. The inclusion of the control variables slightly reduces the coefficient of the competitive dummy from 0.079 to 0.074 (i.e., form $13K to $12K or 0.14 standard deviations), but it remains both economically and statistically significant. In column II, the dependent variable is the log of base salaries in 2008. The additional control variables do not change the result of no relationship between base salaries and taste for competition (the coefficient of the competitive dummy shrinks from 0.022 to 0.015 and remains statistically insignificant). In columns III and IV, we reran the two-step hurdle model used to estimate both the probability of getting a bonus (column III) and the magnitude of the bonus received (column IV). In columns V and VI, we repeat the same regression but solely for expected performance bonuses.

Table A.5 – Robustness of the effect of taste for competition on compensation in 2008 to measurement error and misspecification

Note: Regressions of the log of total compensation in 2008 in column I and of the log of base salary in 2008 in column II. Hurdle model of the likelihood of receiving a bonus in column III and its magnitude in column IV. Hurdle model of the likelihood of receiving an expected performance bonus in column V and its magnitude in column VI. Linear estimates in columns I, II, III, and V. Marginal effects in columns IV and VI. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Total income	Base salary	Total bonus		Exp. perform bonus	
			Received	Amount	Received	Amount
	I	II	III	IV	V	VI
Competitive	0.074**	0.015	0.007	0.163**	0.001	0.591***
	(0.036)	(0.017)	(0.032)	(0.080)	(0.057)	(0.213)
Woman	–0.103***	–0.011	0.028	–0.353***	–0.084	–0.452*
	(0.038)	(0.018)	(0.029)	(0.084)	(0.060)	(0.252)
Overconfidence	0.024	0.008	0.021	0.001	–0.005	0.156
	(0.032)	(0.015)	(0.028)	(0.069)	(0.051)	(0.190)
Risk aversion	0.003	–0.004	–0.001	0.023	0.011	–0.004
	(0.009)	(0.004)	(0.008)	(0.021)	(0.015)	(0.069)
Performance	–0.078	0.053	0.198	–0.387	0.071	–0.855
	(0.130)	(0.060)	(0.131)	(0.282)	(0.205)	(0.763)
Obs.	409	409	409	380	409	153
F test / χ^2 test	2.289***	1.370	7.835	39.783***	9.948	20.031**

Once again, we do not find that the inclusion of the control variables has an important effect on the coefficient of the competitive dummy. It slightly increases from 0.158 to 0.163 (less than 0.07 standard deviations) when considering the magnitude of all bonuses and from 0.571 to 0.591 (form $13K to $14K or less than 0.10 standard deviations) when considering the magnitude of expected performance bonuses.

In summary, we do not find evidence that the relationship between the various forms of compensation in 2008 and the competitive dummy is due to measurement error or misspecification, which suggest that it is indeed driven by taste for competition.

Additional control variables

Here, we test the robustness of the effect of taste of competition on compensation in 2008 to the inclusion of a large set of control variables. Like in subsection III.A, in Table A.6, we reran a selection of the regressions from Tables 3 and 4 including the fifteen additional control variables seen in Tables A.1 and A.2 (described in Section V). These variables include demographic characteristics (e.g., age, race, residence, marital status, religiosity), measures of different kinds of abilities (e.g., mathematical, verbal, and analytical skills, capacity for cognitive reflection, and emotional intelligence), and other common experimental measures (e.g., willingness to trust, reciprocate, and

Table A.6 – Robustness of the effect of taste for competition on compensation in 2008 to additional controls

Note: Regressions of the log of total compensation in 2008 in column I and of the log of base salary in 2008 in column II. Hurdle model of the likelihood of receiving a bonus in column III and its magnitude in column IV. Hurdle model of the likelihood of receiving an expected performance bonus in column V and its magnitude in column VI. Linear estimates in columns I, II, III, and V. Marginal effects in columns IV and VI. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Total income	Base salary	Total bonus Received	Total bonus Amount	Exp. perform bonus Received	Exp. perform bonus Amount
	I	II	III	IV	VII	VIII
Competitive	0.078**	0.024	0.007	0.152*	0.038	0.528**
	(0.036)	(0.017)	(0.029)	(0.079)	(0.059)	(0.212)
Woman	–0.083**	0.008	0.034	–0.345***	–0.044	–0.584**
	(0.038)	(0.018)	(0.025)	(0.085)	(0.063)	(0.247)
Overconfidence	0.011	0.008	0.024	–0.043	–0.017	0.267
	(0.029)	(0.014)	(0.024)	(0.064)	(0.048)	(0.184)
Risk aversion	–0.001	0.003	–0.004	–0.004	–0.005	0.015
	(0.004)	(0.002)	(0.003)	(0.008)	(0.006)	(0.025)
Performance	–0.030	–0.009	–0.005	–0.099*	0.010	–0.064
	(0.026)	(0.012)	(0.020)	(0.057)	(0.043)	(0.155)
Obs.	409	409	409	380	409	153
F test / χ^2 test	2.657***	1.769**	32.123*	52.725***	38.558**	34.361**

cooperate with others). This robustness check lets us evaluate whether taste for competition describes variation in compensation that is not captured by typically measured observables.

In column I, the dependent variable is the log of total compensation in 2008. The inclusion of the additional control variables does not affect the coefficient of the competitive dummy (it changes from 0.079 to 0.078, which is less than 0.03 standard deviations). In column II, the dependent variable is the log of base salaries in 2008. The additional control variables do not modify the result of no relationship between base salaries and taste for competition (the coefficient changes from 0.022 to 0.024 and remains statistically insignificant). In columns III and IV, we reran the two-step hurdle model used to estimate both the probability of getting a bonus (column III) and the magnitude of the bonus received (column IV). In columns V and VI, we repeat the same regression but solely for expected performance bonuses. Once again, we do not find that the inclusion of the control variables has a large effect on the coefficient of the competitive dummy. It slightly decreases from 0.158 to 0.152 (less than 0.08 standard deviations) when considering the magnitude of all bonuses and from 0.571 to 0.528 (form $13K to $12K, around 0.20 standard deviations) when considering the magnitude of expected performance bonuses.

Table A.7 –Taste for competition or taste for high rewards

Note: Regressions of the log of total compensation in 2008 in columns I and V and of the log of base salary in 2008 in columns II and VI. Hurdle model of the likelihood of receiving a bonus in columns III and VII, and of its magnitude (in logs) in columns IV and VIII. Linear estimates in columns I, II, IV, V, VI, and VIII. Marginal effects in columns III and VII. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Total income	Base salary	Total bonus Received	Total bonus Amount	Total income	Base salary	Total bonus Received	Total bonus Amount
	I	II	III	IV	V	VI	VII	VIII
Non-competitive	0.036	0.027	–0.016	0.019	0.014	0.023	–0.020	–0.032
tournament	(0.036)	(0.017)	(0.032)	(0.081)	(0.038)	(0.018)	(0.029)	(0.085)
Competitive					0.075**	0.015	0.024	0.167**
					(0.038)	(0.017)	(0.041)	(0.083)
Woman	–0.114***	–0.010	0.034	–0.397***	–0.106***	–0.009	0.034	–0.379***
	(0.036)	(0.016)	(0.032)	(0.079)	(0.036)	(0.017)	(0.030)	(0.079)
Overconfidence	0.014	0.001	0.036	–0.015	0.000	–0.002	0.030	–0.044
	(0.029)	(0.014)	(0.032)	(0.065)	(0.030)	(0.014)	(0.031)	(0.066)
Risk aversion	0.001	0.003*	–0.004	–0.002	0.002	0.003*	–0.003	0.000
	(0.004)	(0.002)	(0.004)	(0.008)	(0.004)	(0.002)	(0.003)	(0.008)
Performance	–0.004	–0.007	0.018	–0.037	–0.019	–0.010	0.012	–0.069
	(0.025)	(0.012)	(0.026)	(0.055)	(0.026)	(0.012)	(0.026)	(0.057)
Obs.	409	409	409	380	409	409	409	380
F test / χ^2 test	2.886**	1.161	3.769	27.786***	3.088***	1.092	4.163	32.141***

In summary, we do not find that the relationship between taste for competition and compensation in 2008 is affected by the inclusion of a large set of control variables. Therefore, it seems likely that taste for competition is explaining variance in earnings that would otherwise remain unexplained.

Taste for high rewards

In this subsection, we analyze the effect of choosing tournament pay without having to perform under competitive conditions. Recall that, like in the third period of the experiment, in the fourth period, participants had to choose whether they wanted to be compensated for their past performance according to the piece-rate or tournament payment schemes. Unlike in the third period, however, they did not have to perform the adding task again as their decision applied to their past piece-rate performance. Niederle and Vesterlund (2007) argue that this decision is akin to a choice between a certain payoff and a lottery with ambiguous probabilities and is not affected by the participants' attitudes towards competition. If this is the case, it is interesting to analyze whether this variable is also a good predictor of the participants' compensation in 2008.

To test the effect of a 'taste for high rewards' on compensation in 2008, in Table A.7 we reran a selection of the regressions from Tables 3 and 4 including a dummy variable that equals one if the participant chooses tournament pay in the fourth period. We label this variable "non-competitive tournament." In columns I through IV, we use the non-competitive tournament dummy instead of the competitive dummy. In columns V through VIII, we use both the non-competitive tournament and competitive dummies.

In columns I and V, the dependent variable is the log of total compensation in 2008. When the non-competitive tournament dummy is included alone (in column I), its coefficient is positive but is less than half the coefficient of competitive in Table 3 (0.036 vs. 0.079, which is a difference of around one standard deviation), and it is not statistically different from zero (p = 0.322). When both the non-competitive tournament and competitive dummies are included, the coefficient for competitive is economically and statistically significant (around $12K, p = 0.046) whereas the coefficient for the non-competitive tournament is small and far from statistical significance (around $2K, p = 0.709). In columns II and VI, the dependent variable is the log of base salaries in 2008. Consistent with the results reported in the main body of the paper, neither the coefficient of non-competitive tournament nor of competitive display a significant association with base salaries. Finally, in columns III and IV as well as VII and VIII, we reran the two-step hurdle model used to estimate both the probability of getting a bonus (columns III and VII) and the magnitude of the bonus received (columns IV and VIII). Once again, we do not find that the non-competitive tournament dummy is significantly associated with the magnitude of the bonus while the competitive dummy is.

In summary, we find compelling evidence that the coefficient of the competitive dummy variable is indeed capturing a relationship between the participants' taste for competition and compensation that is not related to the choice of a tournament payment scheme *per se*.

Expected performance bonus

In this subsection, we test the robustness of the results for the expected performance bonus component. Recall that we group the various bonuses into two components: the one-off bonus component, which includes relocation, tuition, sign-on, and retention bonuses, and the expected performance bonus component, which includes stock options, profit sharing, guaranteed performance, and other bonuses. Since it is not entirely clear what bonuses are classified as "other," in Table A.8, we reran the two-step hurdle model used to estimate both the probability of getting some expected performance bonus (column VI in Table 4) and the magnitude of the expected performance bonus (column VII in Table 4) excluding "other" bonuses. The exclusion of other bonuses decreases the coefficient of the competitive dummy from 0.571 to 0.461 when considering the magnitude of the expected performance bonus, but it remains large and statistically significant. The exclusion of other bonuses also leaves unchanged the lack of a significant correlation between receiving some expected performance bonus and competitive.

III.C. Effect of taste for competition on compensation in 2015

Next, we provide a series of robustness checks for the effect of taste for competition on compensation in 2015 (reported in Tables 6 and 7). Once again, we focus on regressions without

Table A.8 –Robustness of the expected performance bonus to the exclusion of bonuses classified as "other"

Note: Hurdle model of the likelihood of receiving an expected performance bonus in column I and of its magnitude (in logs) in column II. Marginal effects in column I and linear estimates in column II. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Received	Amount
	I	II
Competitive	0.043	0.461**
	(0.053)	(0.219)
Woman	–0.030	–0.474**
	(0.054)	(0.238)
Overconfidence	–0.007	0.126
	(0.044)	(0.176)
Risk aversion	0.001	0.006
	(0.006)	(0.024)
Performance	–0.028	0.070
	(0.039)	(0.157)
Obs.	409	128
F test / χ^2 test	1.321	13.512**

industry fixed effects since results do not change much by controlling for industry (regressions with industry fixed effects are available upon request).[1]

Additional control variables

Next, we test the robustness of the effect of taste of competition on compensation in 2015 to the inclusion of a large set of control variables. Like in subsection III.A, in Table A.9, we reran the regressions from Tables 6 and 7 including the fifteen additional control variables seen in Tables A.1 and A.2 (described in Section V). These variables include demographic characteristics (e.g., age, race, residence, marital status, religiosity), measures of different kinds of abilities (e.g., mathematical, verbal, and analytical skills, capacity for cognitive reflection, and emotional intelligence), and other standard experimental measures (e.g., willingness to trust, reciprocate, and cooperate with others).

In columns I and V, the dependent variable is the log of total compensation in 2015. The inclusion of the additional control variables marginally decreases the magnitude of the coefficient of the competitive dummy by around two log points (less than 0.13 standard deviations) and that of the

[1] We do not report robustness checks of whether the coefficients of the competitive dummy in regressions for compensation in 2015 are affected by measurement errors or misspecification of the control variables. The reason is that the important result in those regressions are based on the interaction between the competitive dummy and overconfidence (see Table 7), and it would be unclear how to interpret this coefficient once we include the square of overconfidence. We did run this robustness check for the regressions without the interaction (i.e., those in Table 6) and found that the additional control variables do not change the lack of an association between the competitive dummy and compensation in 2015.

Table A.9 – Robustness of the effect of taste for competition on compensation in 2015 to additional controls

Note: Regressions of the log of total compensation in 2015 in columns I and V and of the log of base salary in 2015 in columns II and VI. Hurdle model of the likelihood of receiving a bonus in columns III and VII, and of its magnitude (in logs) in columns IV and VIII. Linear estimates in columns I, II, IV, V, VI, and VIII. Marginal effects in columns III and VII. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Total income	Base salary	Bonus		Total income	Base salary	Bonus	
			Received	Amount			Received	Amount
	I	II	III	IV	V	VI	VII	VIII
Competitive	0.022	0.003	–0.010	0.045	0.108	0.047	0.006	0.186
	(0.101)	(0.069)	(0.037)	(0.193)	(0.106)	(0.073)	(0.042)	(0.198)
Woman	–0.354***	–0.207**	–0.040	–0.709***	–0.362***	–0.211**	–0.041	–0.740***
	(0.121)	(0.083)	(0.052)	(0.231)	(0.120)	(0.082)	(0.056)	(0.228)
Overconfidence	–0.114	–0.082	–0.084***	–0.052	0.027	–0.011	-0.057	0.219
	(0.084)	(0.057)	(0.031)	(0.161)	(0.101)	(0.069)	(0.044)	(0.192)
Overconfidence					–0.347**	–0.176*	-0.064	–0.693**
× competitive					(0.141)	(0.098)	(0.056)	(0.274)
Risk aversion	–0.031***	–0.017**	–0.008*	–0.035*	–0.031***	–0.017**	–0.008**	–0.036*
	(0.010)	(0.007)	(0.005)	(0.020)	(0.010)	(0.007)	(0.004)	(0.020)
Performance	–0.043	–0.055	0.000	–0.085	–0.068	–0.068	–0.008	–0.125
	(0.072)	(0.049)	(0.028)	(0.134)	(0.072)	(0.049)	(0.028)	(0.133)
Obs.	250	250	250	218	250	250	250	218
F test / χ^2 test	2.080***	1.155	38.771**	57.476***	2.320***	1.261	40.344**	65.580***

interaction between competitive and overconfidence, which shifts from –0.350 to –0.347 (less than 0.12 standard deviations). In columns II and VI, the dependent variable is the log of base salaries in 2015. The additional control variables did not affect the coefficient of competitive in column II, and they marginally increased it in column VI. In columns III, IV, VII, and VIII, we reran the two-step hurdle model used to estimate both the probability of getting a bonus (columns III and VII) and the magnitude of the bonus received (columns IV and VIII). Once again, we do not find that the inclusion of the control variables has an important qualitative or quantitative effect on the results reported in the paper. The coefficients of the competitive dummy are very similar in all regressions, overconfidence has a negative effect on the probability of getting a bonus, and there is a strong negative interaction between the competitive dummy and overconfidence.

In summary, like for compensation in 2008, we find that the relationship between taste for competition and compensation in 2015 is unaffected by the inclusion of a large set of control variables. Therefore, once again, it seems likely that taste for competition in combination with overconfidence is explaining variance in earnings that would otherwise remain unexplained.

Attrition in the 2015 follow-up survey

As we reported in Section II, there is no evidence of strong selection effects for those responding to the 2015 follow-up survey. However, with uncorrected *p*-values, we find significant differences in

Table A.10 – Robustness of the effect of taste for competition on compensation in 2015 to selection into the follow-up survey

Note: Regressions of the log of total compensation in 2015 in columns I and IV, the log of base salary in 2015 in columns II and V, and the log of the performance bonus in 2014 in columns III and VI. Panel A contains linear estimates corrected for selection into the follow-up survey using Heckman's two-step procedure. Panel B reports the marginal effects of the selection equation. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Total income	Base salary	Bonus Amount	Total income	Base salary	Bonus Amount
	PANEL A: LOG OF INCOME					
	I	II	III	IV	V	VI
Competitive	0.046	0.006	0.137	0.130	0.043	0.256
	(0.099)	(0.066)	(0.206)	(0.104)	(0.070)	(0.208)
Woman	−0.345***	−0.184**	−0.840***	−0.352***	−0.187**	−0.877***
	(0.119)	(0.079)	(0.242)	(0.117)	(0.079)	(0.238)
Overconfidence	−0.103	−0.075	0.091	0.042	−0.012	0.342
	(0.089)	(0.060)	(0.192)	(0.104)	(0.071)	(0.216)
Overconfidence × competitive				−0.350***	−0.152*	−0.659**
				(0.135)	(0.092)	(0.278)
Risk aversion	−0.022**	−0.012*	−0.019	−0.021**	−0.012*	−0.019
	(0.010)	(0.007)	(0.021)	(0.010)	(0.007)	(0.021)
Performance	−0.052	−0.057	−0.069	−0.082	−0.070	−0.112
	(0.068)	(0.046)	(0.138)	(0.068)	(0.046)	(0.137)
	PANEL B: SELECTION INTO THE FOLLOW-UP SURVEY					
	I	II	III	IV	V	VI
Competitive	−0.036	−0.036	−0.042	−0.036	−0.036	−0.052
	(0.054)	(0.054)	(0.056)	(0.054)	(0.054)	(0.056)
Woman	−0.126**	−0.126**	−0.140**	−0.126**	−0.126**	−0.166***
	(0.052)	(0.052)	(0.054)	(0.052)	(0.052)	(0.053)
Overconfidence	−0.069	−0.069	−0.102**	−0.069	−0.069	−0.091**
	(0.044)	(0.044)	(0.046)	(0.044)	(0.044)	(0.045)
Risk aversion	−0.007	−0.007	−0.008	−0.007	−0.007	−0.009
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)
Performance	0.038	0.038	0.039	0.038	0.038	0.043
	(0.037)	(0.037)	(0.039)	(0.037)	(0.037)	(0.038)
Discount rate	−1.593***	−1.593***	−1.786***	−1.593***	−1.593***	−1.565***
	(0.525)	(0.525)	(0.557)	(0.525)	(0.525)	(0.563)
Donations to University of Chicago	0.066***	0.066***	0.071***	0.066***	0.066***	0.059***
	(0.022)	(0.022)	(0.023)	(0.022)	(0.022)	(0.021)
Uncensored obs.	250	250	218	250	250	218
Censored obs.	146	146	146	146	146	146
χ^2 test	15.153**	9.921*	21.075***	22.233***	12.751**	27.427***

three important variables: overconfidence, one-off bonuses, and gender. For this reason, in Table A.10, we reestimated the coefficients reported in Tables 6 and 7 correcting for selection into the follow-up survey using Heckman's two-step procedure (Heckman 1979).[2]

In the first stage, we include the same independent variables as in the second stage (i.e., competitive and gender dummies, overconfidence, risk aversion, and performance). In addition, to limit potential problems caused by collinearity between the correction term and the independent variables, we include two exclusion restrictions (Puhani 2000). The first exclusion restriction is the participants' donations to the University of Chicago. As we saw in Section II, the more money a student donated to their class gift for the University of Chicago, the more likely they are to respond to the follow-up survey (a one standard deviation increase in donations predicts an 8.3% higher chance of responding to the follow-up survey). This effect is most likely due to how much individuals identify with the university and not due to their compensation or taste for competition. The second exclusion restriction is the participants' elicited discount rate. Higher discount rates are strongly associated with a lower likelihood of responding to the follow-up survey (a one standard deviation increase in the discount rate predicts a 7.2% lower chance of responding to the follow-up survey). This association is almost certainly an independent effect of discount rates, which have been linked theoretically and empirically to procrastination in filling out questionnaires (Ariely and Wertenbroch 2002; O'Donoghue and Rabin 1999; Reuben, Sapienza, and Zingales 2015), that is unrelated to compensation in 2015 or taste for competition.[3]

In columns I and IV, the dependent variable of the second stage is the log of total compensation in 2015. Correcting for selection into the follow-up survey slightly increases the magnitude of the coefficient of the competitive dummy (by around one log point or 0.10 standard deviations in both columns I and IV). The interaction between the competitive dummy and overconfidence remains unchanged at 0.350. Hence, selection into the follow-up survey had little effect on the estimated effects of taste for competition. It is important to note that this is not the result of a weak first stage. Both exclusion restrictions are strong predictors of answering the survey, and we do see somewhat larger differences in the estimated effect of gender. For example, in column IV, the magnitude of the coefficient of the gender dummy is four log points smaller once we correct for selection effects (falls from –0.391 to –0.352 or about 0.35 standard deviations). In columns II and V, the dependent variable of the second stage is the log of base salaries in 2015. Correcting for selection into the follow-up survey had a negligible effect on the coefficient of the competitive dummy in both columns II and V as well as on the coefficient of the interaction term between the competitive dummy and overconfidence. In columns III and VI, the dependent variable of the second stage is the log of performance bonus. Note that, to run these regressions, we dropped the survey respondents who did

[2] We obtain very similar results if we use maximum likelihood to do the estimation. Also, note that 263 individuals answered the 2015 follow-up survey. However, among the respondents there were 13 individuals who were not employed at that time. Since our dependent variable is their income form employment, we dropped these 13 individuals from the analysis. However, if we rerun the first stage of the two-step procedure including these 13 individuals, we obtain very similar coefficients.

[3] The correlation coefficients between donations to the University of Chicago or discount rates with either the competitive dummy or total compensation in 2008 are low (less than 0.073 for donations and 0.020 for discount rates) and are not statistically significant.

Table A.11 – Effect of taste for competition on compensation in 2008 restricting the regressions to the respondents of the 2015 follow-up survey

Note: Regressions of the log of total compensation in 2008 in column I and of the log of base salary in 2008 in column II. Hurdle model of the likelihood of receiving a bonus in column III and its magnitude in column IV. Hurdle model of the likelihood of receiving an expected performance bonus in column V and its magnitude in column VI. Linear estimates in columns I, II, III, and V. Marginal effects in columns IV and VI. Standard errors in parenthesis. ***, **, and * indicate statistical significance at 0.01, 0.05, and 0.10.

	Total income	Base salary	Total bonus		Exp. perform bonus	
			Received	Amount	Received	Amount
	I	II	III	IV	VII	VIII
Competitive	0.062	0.001	0.020	0.151	–0.052	0.458*
	(0.048)	(0.022)	(0.032)	(0.103)	(0.074)	(0.272)
Woman	–0.126**	0.001	–0.003	–0.367***	–0.081	–0.744**
	(0.050)	(0.023)	(0.026)	(0.108)	(0.075)	(0.311)
Overconfidence	0.039	0.031*	0.020	–0.050	–0.030	0.299
	(0.039)	(0.018)	(0.024)	(0.084)	(0.060)	(0.226)
Risk aversion	–0.003	–0.001	–0.002	–0.005	–0.005	0.014
	(0.005)	(0.002)	(0.003)	(0.010)	(0.007)	(0.031)
Performance	–0.014	0.020	–0.006	–0.103	0.010	–0.096
	(0.033)	(0.015)	(0.018)	(0.070)	(0.050)	(0.185)
Obs.	250	250	250	237	250	101
F test / χ^2 test	2.947**	0.920	4.097	16.593***	2.791	12.317**

not receive a bonus. For this reason, the coefficients in Table A.10 are not entirely comparable to the coefficients in Tables 6 and 7, which are based on a two-step hurdle model. Nonetheless, it is telling to see that the coefficients for taste for competition are similar in both regressions. In particular, the strong negative interaction between the competitive dummy and overconfidence.

In summary, we find that the relationship between taste for competition and compensation in 2015 is mostly unchanged once we control for selection into responding to the 2015 follow-up survey using Heckman's two-step correction.

Another way of checking whether attrition in the 2015 follow-up survey affects the results reported in the paper is to reestimate the regressions for compensation in 2008 solely for the sample of MBAs for whom we have 2015 data. This exercise allows us to see how much standard errors increase due to the reduction in the sample size from 409 to 250 observations, and also to observe whether there are selection effects by comparing the reestimated coefficients to those estimated with the full 2008 sample. We report the reestimated regressions in Table A.11.

In column I, the dependent variable is the log of total compensation in 2008. Restricting the sample to the follow-up survey respondents reduces the coefficient of the competitive dummy slightly (from 0.079 to 0.062) and increases its standard error (from 0.036 to 0.048), resulting in a nonsignificant effect. Despite this decrease, the estimated coefficient for the competitive dummy is about twice as large for total compensation in 2008 compared to 2015 (0.062 vs. 0.035).[4] We observe a similar effect for the base salaries (column II), the magnitude of the bonus received (column IV), and the magnitude of the expected performance bonus (column VIII). In other words, there seems to be a power issue in the 2015 sample, yet this is not the main reason why taste for competition is not statistically significant in 2015. On average, taste for competition seems to have less of an impact on compensation in 2015 compared to 2008.

IV. Procedures for the initial survey and experiment

This section describes the procedures used to conduct the initial survey and the experiment. We concentrate on the parts of the survey and experiment that are relevant to the paper. Further details can be found in Reuben, Sapienza, and Zingales (2008), including all survey questions and experimental instructions.

IV.A. The initial survey

Participants completed the online survey in the fall of 2006. The deadline to complete the survey was the day participants took part in the experiment. Completing the survey was a requirement to pass one of the MBA core courses and took approximately one hour. The survey included questions on demographic characteristics as well as standard questionnaires of personality traits. We do not use the survey's variables in the main body of the paper, but we do use them in the robustness checks. In Section V, we describe the variables used in these checks.

VI.B. The experiment

We ran the experiment in October 2006 in four sessions of around 140 participants. It lasted for about 90 minutes. Participation in the experiment was a requirement of one of the MBAs' core courses. The experiment was programmed and run using zTree (Fischbacher 2007).

The experiment consisted of eight parts: three decision problems and five games. Participants played the eight parts in the following order: lottery with losses, asset market game, trust game, taste for competition game, chocolate auction, social dilemma game, lottery without losses, and discount rate elicitation task. We gave the instructions for each part before the start of the respective part (the only exception being the instructions of the asset market game, which they received before their arrival). Importantly, participants received no information about the outcome of the games or lotteries during the experiment. Instead, they received feedback on their performance in specific games and on the behavior of other participants a few days later through an email.

[4] These results are consistent with the results for 2015 compensation using Heckman's correction, which slightly increase the coefficient of competitive compared to the OLS regression (from 0.035 to 0.046) but leave it around half as large as the coefficient for 2008 compensaiton (0.046 vs. 0.079).

Participants received a $20 show-up fee, which could be used to cover potential losses during the experiment. Also, we paid participants the amount they earned in one randomly chosen part (we did the randomization only among six parts since we always paid the lottery with losses and discount rate elicitation tasks). We paid participants who earned more than the show-up fee with a check delivered to their mailbox. Including the show-up fee, participants earned $99 on average (the standard deviation was $63).

In the main body of the paper, we describe the parts of the experiment used to measure taste for competition. Below, we provide the instructions of these parts of the experiment. Moreover, in Section V, we describe the parts of the experiment used to measure the additional control variables utilized in the robustness checks.

Instructions of the sums tasks

This game is divided into 4 periods. At the beginning of the game, you will be divided into groups of four. The participants in your group will be the same throughout the 4 periods.

In each of the first 3 periods, you will be given a series of *addition tasks* (sums of four 2-digits numbers like the one below). You will have 150 seconds to answer as many questions as you want. The computer will record the number of sums that you answer correctly. You may use paper and pencil, but you *cannot* use a calculator. In each period, the rules for the payment are different and will be explained in detail before the start of the respective period.

One of the 4 periods will be randomly selected by the computer to determine your earnings for Game 3. In addition, after period 4 there will be a bonus section consisting of four questions. Any money earned in the bonus section will be added to this experiment's earnings.

Instructions for the piece-rate period

In this period, you will be paid $4 for each correct answer you give.

Example: If you answer 6 questions correctly, your earnings for period 1 equal $24. Remember, you can write down the numbers on a piece of paper, but you *cannot* use a calculator.

Instructions for the tournament period

In this period, you will compete against the other *three participants* in your group. Your payment is contingent on you having the highest number of correct answers. You will be paid *$16* for each correct answer if you have the *highest* number of correct answers in your group. If you do not have the highest number of correct answers, you will earn *$0* in this period. If there are two or more group members tied in first place, one of them will be randomly selected to be paid $16 for each correct answer (the others are paid $0). Note that all group members will face the same difficulty. That is, everyone will face the same sequence of numbers.

Example: Suppose that the other three participants in your group answer 5, 9, and 12 questions correctly. If you answer 11 questions correctly, your earnings in this period would equal $0. If you answer 13 questions correctly, your earnings in this period would equal $208. Remember, you can write down the numbers on a piece of paper, but you *cannot* use a calculator.

Instructions for the choice period

In this period, you will replay the same game, but you choose the rule according to which you will be paid. You can be paid with Rule 4 or with Rule 16:

Rule 4: If you choose this rule, you will be paid $4 for each correct answer regardless of what others do.

Rule 16: If you choose this rule, you will be paid according to your performance relative to the performance of the other three group members. You will earn $16 for each correct answer if you have more correct answers than the other group members had in period 2. If you do not have more correct answers than the other group members had, you will earn $0 in this period. If you tie in first place, a random draw will determine whether you are paid $16 for each correct answer or $0.

Remember, you can write down the numbers on a piece of paper, but you *cannot* use a calculator.

Instructions for the uncompetitive choice period

In this period, you do not have to repeat the addition task, but you have the choice to be paid *again* for your period 1 performance in two ways. You can choose to be paid according to Rule 4 or Rule 16.

Rule 4: If you choose this rule, you will be paid $4 for each question answered correctly in period 1 regardless of what others did.

Rule 16: If you choose this rule, you will be paid $16 for each correct answer in period 1 if you have more correct answers than the other three group members had in period 1. If you did not have more correct answers than the other group members had, you will earn $0 in this period. If you tie in first place, a random draw will determine whether you are paid $16 for each correct answer or $0.

Recall that in period 1, you correctly answered *XX questions*. Note that this choice determines your period 4 earnings; it does not affect your earnings from period 1.

Instructions to elicit the participants' expected rank in each period

In this screen, we would like you to estimate your performance relative to that of the other three players. For each of the first three periods, indicate whether you think you ranked first, second, third or fourth. You will receive $2 for every period in which you correctly estimate your rank. In case of a tie, you will receive the $2 if there is a way of resolving the tie that makes your estimate correct.

Example: Suppose that in period 1 you had 8 correct answers and the other three group members had 6, 8, and 11 correct answers. You will receive $2 if you guess that your rank is second or third in period 1.

V. Description of additional control variables

This section describes the additional control variables used in the robustness checks. We divide them depending on the source of the data: administrative data from the University of Chicago, the initial survey, or the experiment.

V.A. Administrative data

In addition to gender, the business school supplied us with the following variables:

- Age (in months).
- Race, which we used to construct a dummy variable indicating non-white individuals.
- Visa status, which we used to construct a dummy variable indicating whether an individual is a US resident (citizens and legal residents).
- Marital status, which we used to construct a dummy variable indicating married individual.
- GMAT percentile scores. Both the aggregate score and the score of each of the three components: quantitative, verbal, and analytic.
- GPA at graduation.
- Pledged donations to the class gift to the University of Chicago, coded into the following bins: $0 (0), $1 to $50 (1), $51 to $100 (2), $101 to $500 (3), $501 to $1000 (4), and $1000 or more (5).

Note that all these variables, except for GPA and donations to the university, were collected in 2006, before the students started their MBA. We collected the last two variables in 2008 at graduation.

V.B. Initial survey

We use the following variables from the initial survey in the robustness checks.

- Religiosity, which we measured with the yes/no question: “Are you religious now?”
- A self-reported measure of the participants’ general attitude towards risk that has been shown to correlate with incentivized measures and is commonly used in the literature (Falk et al. 2018). It consists of the question “Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please select a number between 0 and 10 where 0 means unwilling to take risks and 10 means fully prepared to take risk.” For this variable to measure risk aversion, we reverse coded it so that higher numbers imply more aversion to risk.
- Another self-reported measure of the participants’ attitude towards risk, which was elicited in the monetary domain. Specifically, we asked participants to indicate “What is the maximum price you are willing to pay for a ticket in a lottery that pays you $5K with 50% probability and nothing with 50% probability?” For this variable to measure risk aversion, we use $2.5K minus their answer to the question so that higher values imply more aversion to risk.
- We elicited the participants estimated academic performance by asking them: “In your future

exams at the University of Chicago, in which decile of the GPA distribution do you expect yourself to be?" We then used their answer to this question minus their actual GPA decile at graduation to create the non-incentivized survey measure of overconfidence.

- We measured the participants' tendency to suppress intuitive responses using the Cognitive Reflection Test or CRT (Frederick 2005). We simplified the original test to the following four questions: (i) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? (ii) If you flipped a fair coin 3 times, what is the probability that it would land "heads" at least once? (iii) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (iv) Two cars are on a collision course, traveling towards each other in the same lane. Car A is traveling 70 miles an hour. Car B is traveling 80 miles an hour. How far apart are the cars one minute before they collide? The CRT score consists of the number of correct answers.
- To measure the participants' ability to detect emotions (a key component of emotional intelligence), we use the "reading the mind in the eyes" test or RMET (Baron-Cohen et al. 2001). It consists of correctly recognizing the emotions of various individuals by looking at pictures of their eyes. The RMET score consists of the fraction of correct answers.

V.C. Experiment

From the experiment, we use the following measures important individual characteristics.

Discount rate

To measure time preferences, we gave participants a series of choices of the following form: receive *x* dollars today or receive (1 + *y*)*x* dollars in two weeks, where x equaled their earnings in the experiment. Each subject answered thirteen such questions where *y* varied from 0 to 0.12 in steps of 0.01. After that, one of the questions was randomly selected and paid. We always paid participants by dropping a check into their mail folder during a day in which they had to attend class.

Trust and reciprocity

We measure the participants' propensity to trust and reciprocate by having them play a trust game (Berg, Dickhaut, and McCabe 1995). In the game, a first-mover is endowed with $50 and decides how much to send to a second-mover (in multiples of $5). Any amount sent is multiplied by three. The second mover then decides how much to return to the first mover.

Each participant played two trust games. First, they played as the first mover and then as the second mover. Participants made their second-mover decision using the strategy method. That is, they indicated how much to return for each possible sent amount without knowing how much the first mover actually sent. They received no feedback in-between decisions and knew they were not playing with the same participant. We use the fraction of the $50 sent as first movers as the participants' measure of trust and the fraction they returned conditional on receiving $150 as their measure of reciprocity.

Cooperation

To measure their willingness to cooperate, participants played a variation of the design used by Fischbacher, Gächter, and Fehr (2001). Specifically, participants were randomly assigned to groups of eight, given an endowment of $50, and asked to make two contribution decisions to a linear public good game: an "unconditional" decision and a "conditional" decision. For their unconditional decision, each participant i indicated whether he/she is willing to contribute $c_i \in \{0, 50\}$ to the public good. For their conditional decision, each participant i indicated whether he/she is willing to contribute $c_i(x) \in \{0, 50\}$ given that $x \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ other group members contribute. To determine the final contributions to the public good, seven unconditional decisions were selected at random and were used to determine the conditional decision of the remaining group member. Participant i's earnings equaled $50 - c_i + 0.3 \times \sum_j c_j$. We use the unconditional contribution as a participant's willingness to cooperate.

References

Ariely, Dan and Klaus Wertenbroch. 2002. "Procrastination, Deadlines, and Performance: Self-Control by Precommitment." *Psychological Science* 13:219–24.

Baron-Cohen, Simon, Sally Wheelwright, Jacqueline Hill, Yogini Raste, and Ian Plumb. 2001. "The 'Reading the Mind in the Eyes' Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-Functioning Autism." *Journal of Child Psychology and Psychiatry* 42(2):241–51.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society Series B (Methodological)* 57:289–300.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10(1):122–42.

Buser, Thomas, Muriel Niederle, and H. Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." *The Quarterly Journal of Economics* 129(3):1409–47.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133(4):1645–92.

Fischbacher, Urs. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10(2):171–78.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters* 71(3):397–404.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19(4):25–42.

Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2019. "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study." *Journal of Political Economy* 127(4):1826–63.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1):153.

Niederle, Muriel and Lise Vesterlund. 2007. "Do Women Shy Away From Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122(3):1067–1101.

O'Donoghue, Ted and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review* 89(1):103–24.

Puhani, Patrick. 2000. "The Heckman Correction for Sample Selection and Its Critique." *Journal of Economic Surveys* 14(1):53–68.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2008. *A Description of the Templeton Chicago MBAs Longitudinal Study*. University of Chicago.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2015. "Procrastination and Impatience." *Journal of Behavioral and Experimental Economics* 58:63–76.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2019. "Taste for Competition and the Gender Gap among Young Business Professionals." Working paper, New York University Abu Dhabi.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. 2017. "Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender." *The Economic Journal* 127(604):2153–86.

van Veldhuizen, Roel. 2018. "Gender Differences in Tournament Choices: Risk Preferences, Overconfidence or Competitiveness?" WZB Working Paper.