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	Amount	Total income	Base salary	Total bonus Received	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.