Are Men Less Generous to a Smarter Woman in a Non-romantic Setting?

Yuki Takahashi*

Click here for the latest version

March 21, 2023

Abstract

Are men less generous to a woman with a higher IQ in a non-romantic setting? I test this question using a dictator game experiment with a novel design that exogenously varies the receivers' gender and relative IQ. Based on a sufficiently powered analysis, I do not find that men are less generous to a woman whose IQ is higher than theirs; if anything, they are slightly more generous to a higher-IQ woman than a higher-IQ man. In addition, I do not find that women are less generous to a woman with a higher IQ either. The results hold both in mean and distribution, and are not driven by the so-called "beauty premium." These results suggest that although men (and women) care about their IQ very much and dislike smarter women in the marriage market, the men's dislike does not manifest outside the marriage market.

JEL Classification: J16, D91, C91
Keywords: Gender, IQ, generosity, dictator game, laboratory experiment

*Department of Economics, European University Institute. Via dei Roccettini 9, 50014 Fiesole, Italy. Email: yuki.takahashi@eui.eu. I am grateful to Maria Bigoni, Siri Isaksson, and Natalia Montinari whose feedback was essential for this project, and to participants of the experiment for their participation and cooperation. Laura Anderlucci, Tiziano Arduini, Francesca Barigozzi, Enrico Cantoni, Chiara Natalie Focacci, Margherita Fort, Catalina Franco, Fabio Landini, Annalisa Loviglio, Valeria Maggian, Joshua Miller, Monika Pompeo, Eugenio Proto, Tommaso Sonno, Alessandro Tavoni, Bertil Tungodden, ESA Experimental Methods Discussion group, and the University of Bologna's PhD students all provided many helpful comments. This paper also benefited from participants' comments at the Applied Young Economist Webinar, the BEEN Meeting, seminars at Ca' Foscari University, the NHH, and the University of Bologna. Veronica Rattini and oTree help & discussion group kindly answered my questions about oTree programming. Lorenzo Golinelli provided excellent technical and administrative assistance. This study was pre-registered with the OSF registry (https://osf.io/r6d8f/files) and conducted in compliance with the University of Bologna's standardized ethics protocol and the personal data protection policy. The pre-analysis plan is also in Online Appendix E, and the explanation of deviations from the plan is in Online Appendix A.

1 Introduction

Men dislike smarter women in the marriage market (Fisman et al. 2006). Indeed, intelligence is one of the attributes men (and women) care most about that they are even willing to reduce their payoff by distorting their belief updating (Eil and Rao 2011; Zimmermann 2020) and by avoiding information (Castagnetti and Schmacker 2022). Thus, men may exhibit stronger inequality aversion for smarter women than for smarter men and be less generous to them, which is detrimental to women's careers.

This paper studies whether men are less generous to a woman smarter than them in a non-romantic setting. Answering this question using secondary data is difficult due to non-random group formation and that intelligence is correlated with generosity (Falk et al. 2021). Also, a clean measure of one's generosity to others is not readily available in secondary data.

To overcome these challenges, I design a laboratory experiment where participants first work on an incentivized IQ test (Raven matrices) that measures their intelligence. After the test, participants are randomly assigned to a group of six and receive an IQ rank relative to other group members. Then three of the six members are randomly chosen to be dictators and play three rounds of dictator game with the other three members chosen to be receivers, observing the receivers' facial photos and first names, both of which convey information about their gender, and the IQ ranks. I use the dictators' allocation as the measure of generosity, a widely used tool in experimental economics and shown to predict one's generosity outside the laboratory (Franzen and Pointner 2013). The use of photos follows recent literature (Babcock et al. 2017; Coffman 2014; Isaksson 2018) and allows the dictators to infer the receivers' gender without inducing experimenter demand effects.[1] I use dictator IQ fixed effects in the analysis to compare allocations of dictators with the same IQ but assigned different IQ ranks due to random group formation, allowing me to cut the correlation between IQ and generosity.

I first confirm that male and female dictators' behaviors are consistent with the literature: their allocation amount is close to a previous study with a similar level of social distance with receivers and that female dictators allocate more. Yet, I do not find male dictators are not less generous to a female receiver with a higher IQ: the point estimate is quantitatively negligible and statistically indistinguishable from 0, and the confidence interval is tight. If anything, male dictators are slightly more generous to a female receiver with a higher IQ. These results are not driven by the so-called "beauty premium" and hold across the whole distribution of the male dictator allocation. Although statistically insignificant, the belief about receivers' IQ is roughly consistent with the experimental design. I do not find that female dictators are less generous to a female receiver with a higher IQ either. Taken together, these findings suggest that although IQ is one of the attributes men (and women) care most about and that men dislike smarter women in the marriage market, this men's dislike does not manifest outside the marriage market.

This paper's main contribution is to the literature on men's dislike of women who are superior to

1. Yet, I show the robustness of the results to the concern that facial photos may be subject to the so-called "beauty premium."

them by showing that men do not dislike smarter women in a non-romantic setting. In the marriage market, men do not like women superior to them: women who get a promotion are more likely to face divorce (Folke and Rickne 2020), skilled women are less likely to get married, especially in countries with conservative gender norm (Bertrand et al. 2021), and that women smarter than men are less likely to be preferred as a romantic partner (Fisman et al. 2006). Outside the marriage market, men dislike to be led by women: male teachers are more likely to quit when they work under female principals (Husain, Matsa, and Miller 2021), men treat female leaders (Chakraborty and Serra 2022) and female competitors (Datta Gupta, Poulsen, and Villeval 2013) more aggressively than male counterparts, and offer less support to female leaders (Born, Ranehill, and Sandberg 2022). Men are also more likely to lose interest in the work when female managers criticize them than when male managers do the same (Abel 2022). My results show that these men's dislike of superior women outside the marriage market is not driven by those women being smarter than them.

This paper also contributes to the literature on the role of receivers' gender in dictator games by examining the role of IQ interacted with gender. A meta-study of the dictator game by Engel (2011) finds that dictators allocate a higher amount to female receivers across 39 studies regardless of dictators' gender. My results show that results from previous studies are not overturned even by separately examining allocations to higher-IQ and lower-IQ female receivers.

The remainder of the paper proceeds as follows. Section 2 describes the experimental design, procedure, and implementation. Section 3 describes the data. Section 4 discusses the empirical strategy. Section 5 presents the results. Section 6 concludes.

2 Experiment

2.1 Design and procedure

The experiment consists of two parts. Participants receive instructions at the beginning of each part. They earn a participation fee of 2.5€ for their participation.

Pre-experiment: Random desk assignment & photo taking

After registration at the laboratory entrance, participants are randomly assigned to a partitioned computer desk. Afterward, participants have their facial photos taken at a photo booth and enter their first name on their computer. After that, the experimenters go to each participant's desk to check that their photo and first name match them to ensure all participants that other participants' photos and first names are real, following Isaksson (2018).

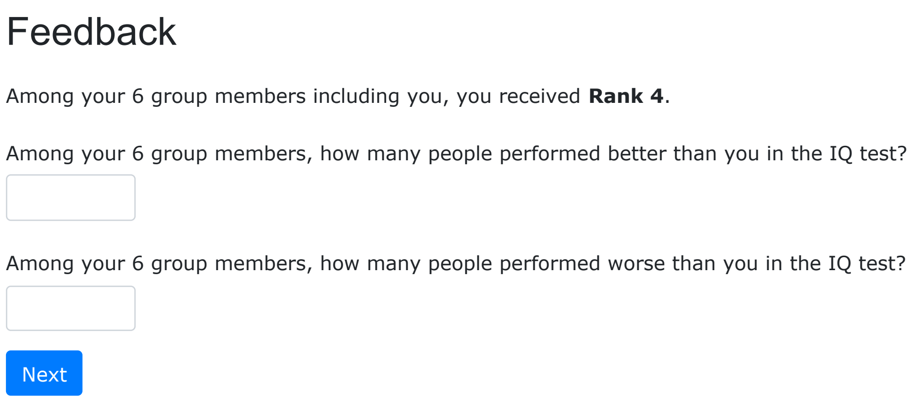
Part 1: IQ test

In part 1, participants work on an incentivized 9 IQ questions for 9 minutes. I use Bilker et al. (2012)'s form A 9-item Raven test that measures one's IQ with accuracy over 90% of the full-length Raven test. Participants receive 0.5€ for each correct answer, and they do not receive information about how many IQ questions they have solved correctly until the end of the experiment. While higher IQ people earn more in this part, I control for this wealth effect in the analysis.

After the IQ test, participants make an incentivized guess on the number of IQ questions they have solved correctly; they receive 0.5€ if their guess is correct. The answer to this question measures their over-confidence level. They do not receive feedback on their guess accuracy until the end of the experiment.

Following Eil and Rao (2011), six participants are randomly grouped and informed of their IQ rank relative to other group members. Ties are broken randomly. They then answer a set of comprehension questions about their IQ rank; see Figure 1 for an example. They cannot proceed to the next part until they answer the questions correctly.

Figure 1: IQ rank comprehension screen



Notes: This figure shows an example of an IQ rank comprehension screen. In this example, the participant's IQ rank is 4.

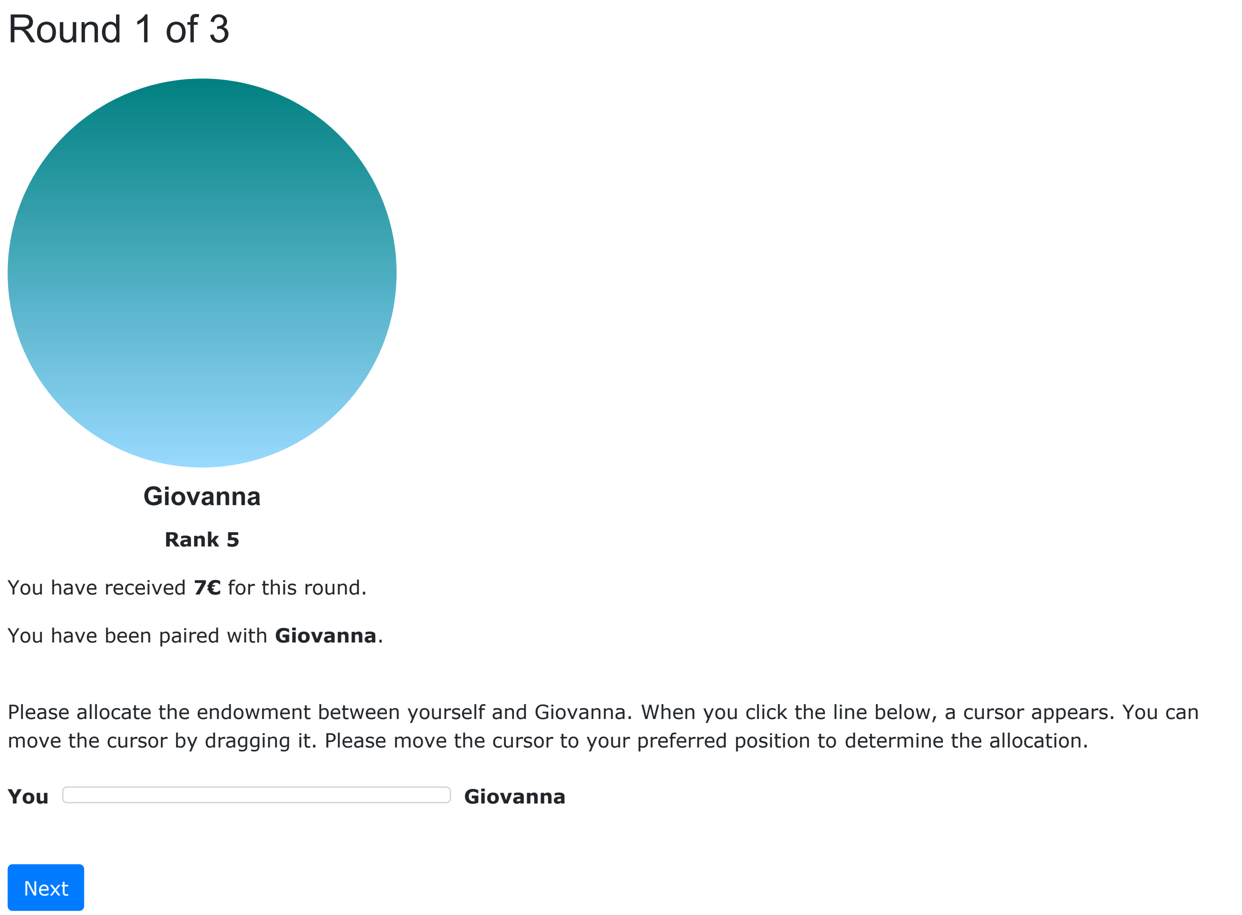
Part 2a: Dictator game (dictators only)

In part 2, three participants in each group are randomly assigned to the role of dictators (which I call "active participant" in the experiment), and the other three participants the role of receivers. Dictators are paired with the three receivers in their group one by one in a random order, receive an endowment, and play a dictator game. Thus, they play a dictator game three times with three different receivers. During the dictator game, dictators observe the receivers' facial photo, first name, and IQ rank; see panel A of figure 2 for an example.[2] The use of photos follows recent literature (Babcock et al. 2017; Coffman 2014; Isaksson 2018) and minimizes experimenter demand effects. While I use photos, I show later that the results are not driven by the so-called "beauty premium."

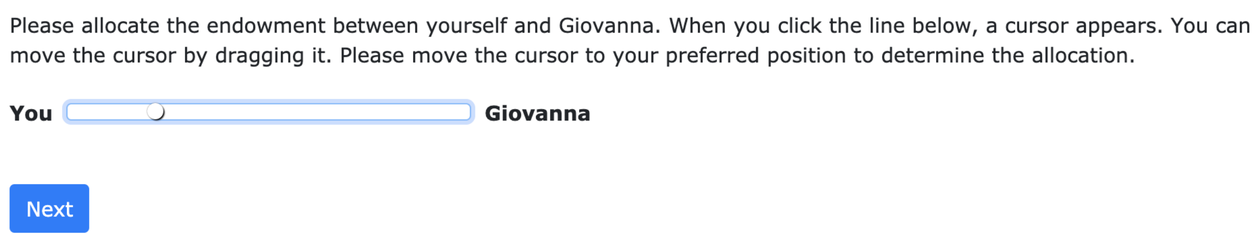
2. One may point out that the photo was too salient on the screen that participants may not have paid enough attention to the IQ rank information. While it may have been the case for some participants, they had to answer the IQ rank comprehension questions (see Figure 1) right after the IQ test and the IQ test is a very ego-relevant task whose performance people really care about (Castagnetti and Schmacker 2022). Thus, I believe it was not a huge concern in my experiment and even if some participants did not pay enough attention to the IQ rank, it does not alter my results because it simply attenuates my results and does not reverse the sign of the coefficients.

Figure 2: Dictator's allocation screen

(a) Initial screen



(b) After clicking the slider



Notes: This figure shows an example of a dictator's allocation screen. Panel A shows the screen before clicking the slider bar and panel B after clicking it. In this example, the dictator is playing the first round and paired with a receiver whose first name is Giovanna (a female name) with IQ rank 5. In the actual experiment, dictators see Giovanna's facial photo in place of the green circle.

Dictators are also told that their allocation decisions are anonymous that their allocation will be paid to the receivers as a "top-up" to their earnings. Dictators decide allocation by moving a cursor on a slider where the cursor is initially hidden to prevent anchoring; panel B of figure 2 shows the cursor after clicking the slider. I vary the endowment across rounds to make each dictator game less repetitive: 7€ for the 1st and the 3rd rounds and 5€ for the 2nd round. At the end of the experiment, one of the three allocations is randomly chosen for each participant as earnings for this part.[3]

Part 2b: Belief elicitation (receivers only)

I also collect an indirect measure of dictators' beliefs on how many IQ questions the paired receivers have solved correctly. To prevent the belief elicitation from affecting/being affected by the dictator game, I exploit the random role assignment (derived from the random desk assignment) and use receivers' beliefs as a proxy for dictators' beliefs. This is a valid proxy because (i) both dictators and receivers are exactly in the same experimental environment until the role assignment and (ii) dictators' and receivers' individual characteristics are balanced on average by the random role assignment.[4]

Specifically, while dictators are playing the dictator game, receivers are paired with the other two receivers in the same group one by one in random order and make incentivized guesses on how many IQ questions they have solved correctly, observing the other two receivers' facial photo, first name, and IQ rank, just as dictators do. Each correct guess gives them 0.5€.

To address the non-anonymity of showing facial photos and first names, I ask participants how well they know the paired participants on a scale of 4.[5] I ask this question twice to make sure they do not answer randomly: right after the three dictator games for dictators or two guesses for receivers and in the post-experimental questionnaire.

Post-experiment: Questionnaire

After the dictator game and guessing are over, participants are told their earnings from the IQ test, the dictator game, and the guesses. Before receiving their earnings, participants answer a short questionnaire about their demographics that I use to test ex-post balance and for robustness checks. Receivers are also asked if I could use their photo in another experiment with a gratuity of 1.5€.

2.2 Implementation

The experiment was programmed with oTree (Chen, Schonger, and Wickens 2016) and conducted in English during November-December 2019 at the Bologna Laboratory for Experiments in Social Science (BLESS). I recruited 390 students (195 female and 195 male) of the University of Bologna via ORSEE (Greiner 2015) who (i) were born in Italy, (ii) had not participated in gender-related experiments in the past (as far as I know), and (iii) were available to participate experiments in English. The first condition is to reduce the chance that receivers' photos and first names signal ethnicity, race, or cultural background. The second condition is to reduce experimenter demand

3. For each dictator for each round, one of the three receivers in the same group is randomly chosen *without replacement* and the dictator allocates the endowment between themselves and the receiver. Thus, it is possible that two dictators play the dictator game with the same receiver in the same round. At the end of the dictator games, each participant has three allocations, one of which is randomly chosen for payment.

4. See Appendix Table C3 for the ex-post balance.

5. The answer choices are: "I didn't know him/her at all," "I saw him/her before," "I knew him/her but not very well," and "I knew him/her very well."

effects. The third condition is to run the experiment in English. As a further attempt to make the data cleaner, I exclude receivers with non-Italian sounding names and allocations in which the dictator declared they knew the paired receivers "very well" at least once.[6] These data screenings leave me with 388 participants, 195 dictators, and 558 dictators' allocations.

The number of participants is based on the power simulation in the pre-analysis plan to achieve 80% power.[7] The experiment is pre-registered with the OSF and the pre-analysis plan is in Online Appendix E.[8] Online Appendix A explains deviations from the pre-analysis plan.

I ran 24 sessions in total, and the number of participants in each session was a multiple of 6 (12 to 30). The average session length was 70 minutes, including registration and payment. The average payment per participant was about 10€ including the participation fee and 1.5€ of gratuity for photo use in another experiment (only for those receivers who agreed).

3 Data description

Table 1 describes dictators' own (panel A) and paired receivers' characteristics (panel B) as well as dictators' social distance with paired receivers (panel C) and dictator game allocation (panel D), separately for male and female dictators.

Panel A shows that male dictators solve 0.37 more IQ questions (out of 9) than female dictators, but the difference is quantitatively insignificant. Also, male dictators are less likely to major in humanities and more likely to major in science, technology, engineering, and mathematics (STEM), consistent with the pattern observed in most OECD countries (see, for example, Carrell, Page, and West 2010). In addition, male dictators are more overconfident than female dictators, consistent with other studies (Bertrand 2011; Croson and Gneezy 2009; Niederle and Vesterlund 2011). Further, male dictators are less likely to have completed undergraduate studies than female dictators, consistent with that women are more educated than men in OECD countries (see, for example, Almås et al. 2016; Autor and Wasserman 2013).

Panel B shows that paired receivers' characteristics are roughly balanced, except that male dictators are 10% less likely to be paired with receivers from the Emilia-Romagna region where the experiment was conducted. I control this imbalance in the analysis.

Panel C shows that dictators do not know 95-98% of the paired receivers at all, addressing the concern that dictator game allocation is driven by relationships outside the laboratory. Panel D shows that male dictators allocate six percentage points less than female dictators. To elaborate on these points, Figure 3 plots the empirical cumulative distribution function (CDF) of dictators' allocation separately for male (the solid blue line) and female (the dashed green line) dictators. First, the empirical CDF resembles that of Bohnet and Frey (1999)'s one-way identification with

6. Although it is easy to distinguish Italian and non-Italian sounding names, to make sure not to misclassify, I asked the laboratory manager who was native Italian to check participants' first names after each session.

7. I exclude the 1st session data because of the problem discussed in Online Appendix A.

8. You can find the pre-registration documents registered at the OSF registry at the following URL: https://osf.io/r6d8f/files.

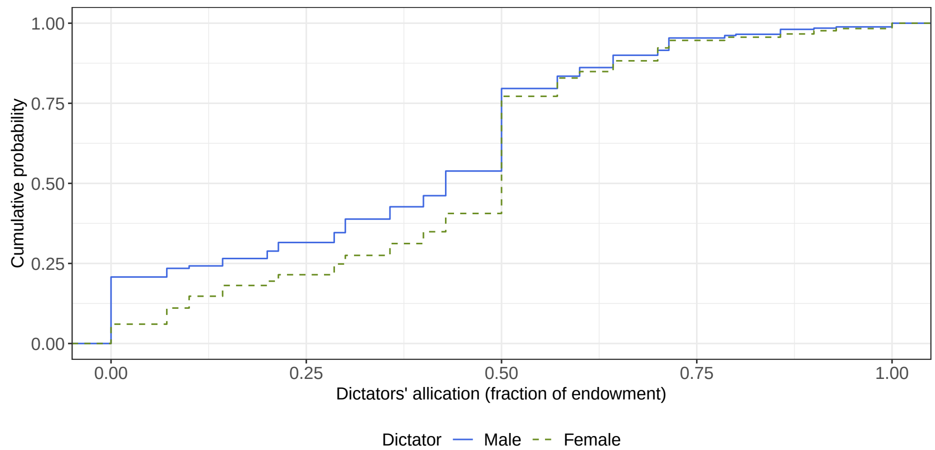
Table 1: Dictators' and paired receivers' characteristics, proximity between dictators and paired receivers, and dictator game allocation

	Male dictators		Female dictators		Difference (Male – Female)		
	Mean	SD	Mean	SD	Mean	SE	P-value
Panel A: Own characteristics							
IQ level	6.89	1.24	6.52	1.20	0.37	0.18	0.04
IQ rank	3.31	1.73	3.83	1.59	-0.52	0.24	0.03
Age	23.23	2.81	23.68	2.62	-0.45	0.39	0.25
From Emilia-Romagna	0.19	0.39	0.18	0.39	0.00	0.06	0.94
Humanities	0.32	0.47	0.58	0.50	-0.26	0.07	0.00
Social sciences	0.24	0.43	0.15	0.36	0.09	0.06	0.13
STEM	0.44	0.50	0.27	0.45	0.17	0.07	0.01
Post bachelor	0.37	0.49	0.53	0.50	-0.16	0.07	0.03
Overconfidence	0.56	0.72	0.31	0.78	0.25	0.11	0.02
Time on feedback (sec.)	107.52	102.26	107.67	89.88	-0.16	13.88	0.99
Observations	91		104				
Panel B: Paired receivers' characteristics							
IQ level	6.91	1.12	6.77	1.19	0.14	0.09	0.11
IQ rank	3.45	1.74	3.39	1.75	0.05	0.10	0.61
Higher IQ	0.48	0.50	0.57	0.50	-0.09	0.05	0.08
Age	23.55	2.98	23.17	2.57	0.37	0.24	0.12
Female	0.43	0.50	0.50	0.50	-0.07	0.04	0.06
From Emilia-Romagna	0.25	0.43	0.15	0.36	0.09	0.04	0.01
Observations	260		298				
Panel C: Social distance with paired receivers							
Did not know at all	0.95	0.23	0.98	0.15	-0.03	0.02	0.14
Knew but not well	0.03	0.18	0.02	0.15	0.01	0.02	0.48
Saw before	0.02	0.14	0.00	0.00	0.02	0.01	0.06
Observations	260		298				
Panel D: Dictator game allocation (fraction of endowment)							
Allocation	0.37	0.25	0.43	0.22	-0.06	0.03	0.04
Allocation (residualized)	-0.03	0.25	0.03	0.22	-0.06	0.03	0.06
Observations	260		298				

Notes: This table shows dictators' own (panel A) and paired receivers' characteristics (panel B) as well as dictators' social distance with paired receivers (panel C) and dictator game allocation (panel D), separately for male and female dictators. Residualized allocation is residual from the regression of the dictator game allocation as a fraction of endowment on IQ fixed effects and shows within dictator IQ variation. P-values for the difference between male and female dictators are calculated with heteroskedasticity-robust standard errors with Bell and McCaffrey (2002)'s small sample bias adjustment for Panel A and with Pustejovsky and Tipton (2018)'s small cluster bias adjustment for Panels B-D.

information treatment where the social distance between dictators and receivers is the closest to my setting. Second, the empirical CDFs show that female dictators allocate slightly more than male dictators below the 50% split but not above that, consistent with the Bilén, Dreber, and

Figure 3: Distribution of the dictators' allocation



Notes: These figures plots empirical cumulative distribution function (CDF) of dictators' allocation separately for male (blue, solid line) and female (green, dashed line) dictators.

Johannesson (2021)'s meta-analysis that female dictators allocate more but the difference is not quantitatively large, and with Klinowski (2018) that female dictators allocate so that the amount between themselves and receivers are equalized, but aside from that, female and male dictators allocate the same amount.

Finally, the standard deviation of the residualized allocation in Panel D shows that there is enough variation in the dictator game allocation after adding dictator IQ fixed effects, which I exploit in the analysis.[9]

4 Empirical strategy

I estimate the following model with OLS with male dictator allocation data:

$$\begin{aligned} Allocate_{ij} =& \beta_1 HigherIQReceiver_{ij} \times FemaleReceiver_j \\ &+ \beta_2 HigherIQReceiver_{ij} + \beta_3 FemaleReceiver_j + X'_{ij}\gamma + \mu_i^{IQ} + \epsilon_{ij} \end{aligned} \tag{1}$$

where each variable is defined as follows:

- $Allocate_{ij} \in [0,1]$: dictator i's allocation to receiver j as a fraction of endowment.
- $HigherIQReceiver_{ij} \in \{0,1\}$: an indicator variable equals 1 if receiver j's IQ is higher than that of dictator i.
- $FemaleReceiver_j \in \{0,1\}$: an indicator variable equals 1 if receiver j is female.
- X_{ij}: a set of additional covariates to increase statistical power and to address the potential ex-post imbalance. Online appendix B provides a full description of the covariates.
- ϵ_{ij}: omitted factors that affect dictator i's allocation to receiver j conditional on covariates.

9. The residualized allocation is the residual from regressing the dictator game allocation on dictator IQ fixed effects.

and $\mu_i^{IQ} \equiv \sum_{k=1}^{9} \theta_k^{IQ} \mathbb{1}[\text{i's IQ} = k]$ is fixed effects for the dictators' IQ (number of IQ questions they have solved correctly), where $\mathbb{1}$ is the indicator variable. Standard errors are clustered at the dictator level with Pustejovsky and Tipton (2018)'s small cluster bias adjustment. Dictator's IQ fixed effects are included following Zimmermann (2020) so that the coefficients in equation 1 capture allocation differences due to the receivers' IQ, not dictators'. Appendix Table C1 provides evidence that dictator IQ rank is uncorrelated with dictator characteristics conditional on dictator IQ fixed effects.[10] There is no order effect because the order dictators play a dictator game with receivers is randomized.

The coefficient of interest is β_1, which captures the difference between dictators' allocation to higher-IQ and lower-IQ female receivers relative to the difference between dictators' allocation to higher-IQ and lower-IQ male receivers. I use the male receiver difference as a comparison group because the higher-IQ receiver would have earned more than the lower-IQ receiver in the part 1 IQ test, which can induce dictators' inequality aversion (Fehr and Schmidt 1999).

The key identification assumption is that conditional on dictator IQ fixed effects, the receiver's gender, IQ rank relative to the dictator's, and their interaction are uncorrelated with factors that affect dictator game allocation. The receiver's gender is exogenous to dictator game allocation by random desk assignment. Receiver's IQ rank is also exogenous to dictator game allocation conditional on dictator's IQ fixed effects by random desk assignment and random matching of dictators and receivers. Appendix Table C2 shows that they are indeed uncorrelated with the dictator or the paired receiver characteristics, dictator game rounds, or social distance between dictators and paired receivers.[11]

5 Results

5.1 Regression results

Columns 1-5 of Table 2 present the regression results of equation 1 for male dictators. Column 1 shows that when we do not control for dictators' IQ, dictators allocate more to higher IQ receivers, which suggests that lower IQ dictators allocate more to higher IQ receivers, although the difference is only marginally statistically significant. Columns 2-5 gradually add more controls and show that coefficient estimates are roughly stable across different specifications, especially after controlling for dictator characteristics some of which are ex-post unbalanced, suggesting irregularities in the data are unlikely to drive the results.

Looking at column 5, my preferred specification, the coefficient estimate on the higher-IQ receiver is positive but statistically insignificant, suggesting that male dictators do not allocate

10. There is some ex-post unbalance that dictators majoring in social sciences are more likely to receive IQ rank 6. I control for dictators' majors in the analysis.

11. There are some ex-post unbalances that dictators from Emilia-Romagna region are less likely to be paired with higher-IQ female receivers and more likely to be paired with higher-IQ male receivers and female receivers and that receivers from Emilia-Romagna region are more likely to be higher IQ. I control dictators' and receivers' regions of origin in the analysis.

Table 2: Male dictators' allocation to higher-IQ female receivers – OLS

Outcome:	Dictator's allocation (fraction of endowment)			
Sample:	Male dictators			
	(1)	(2)	(3)	(4)
Higher IQ receiver x Female receiver	0.018	0.017	0.008	0.035
	(0.060)	(0.060)	(0.062)	(0.060)
	[-0.101, 0.136]	[-0.101, 0.134]	[-0.113, 0.130]	[-0.083, 0.154]
Higher IQ receiver	0.093*	0.054	0.056	0.042
	(0.048)	(0.053)	(0.053)	(0.054)
	[-0.001, 0.188]	[-0.050, 0.159]	[-0.049, 0.161]	[-0.064, 0.147]
Female receiver	0.038	0.031	0.031	0.014
	(0.035)	(0.035)	(0.035)	(0.034)
	[-0.031, 0.107]	[-0.038, 0.100]	[-0.037, 0.100]	[-0.052, 0.081]
Dictator IQ FE	-	✓	✓	✓
Round FE	-	-	✓	✓
Social distance FE	-	-	✓	✓
Dictator controls	-	-	-	✓
Receiver controls	-	-	-	-
H0: Higher IQ receiver x Female receiver +Higher IQ receiver=0 (p-value)	0.043	0.143	0.197	0.139
H0: Higher IQ receiver x Female receiver +Female receiver=0 (p-value)	0.252	0.302	0.414	0.289
Baseline Mean	0.355	0.355	0.355	0.355
Baseline SD	0.262	0.262	0.262	0.262
Adj. R-squared	0.032	0.052	0.051	0.083
Observations	260	260	260	260
Clusters	91	91	91	91
Outcome:	Dictator's allocation (fraction of endowment)			Belief on IQ (fraction of baseline SD)
Sample:	Male dictators	Female dictators	All dictators	Male receivers
	(5)	(6)	(7)	(8)
Higher IQ receiver x Female receiver	0.031	0.057	0.035	0.449
	(0.061)	(0.046)	(0.037)	(0.286)
	[-0.089, 0.151]	[-0.034, 0.147]	[-0.038, 0.107]	[-0.116, 1.014]
Higher IQ receiver	0.048	-0.049	0.008	0.160
	(0.055)	(0.042)	(0.033)	(0.234)
	[-0.060, 0.156]	[-0.131, 0.033]	[-0.058, 0.073]	[-0.301, 0.621]
Female receiver	0.014	-0.014	0.009	-0.139
	(0.034)	(0.037)	(0.026)	(0.199)
	[-0.053, 0.081]	[-0.088, 0.059]	[-0.042, 0.061]	[-0.531, 0.254]
Dictator IQ FE	✓	✓	✓	✓
Round FE	✓	✓	✓	✓
Social distance FE	✓	✓	✓	✓
Dictator controls	✓	✓	✓	✓
Receiver controls	✓	✓	✓	✓
H0: Higher IQ receiver x Female receiver +Higher IQ receiver=0 (p-value)	0.130	0.840	0.151	0.007
H0: Higher IQ receiver x Female receiver +Female receiver=0 (p-value)	0.342	0.150	0.091	0.113
Baseline Mean	0.355	0.359	0.373	3.621
Baseline SD	0.262	0.256	0.261	1.008
Adj. R-squared	0.080	0.021	0.042	0.047
Observations	260	298	558	197
Clusters	91	104	195	104

Notes: This table presents the regression results of equation 1. The standard error (in parenthesis) and the 95% confidence interval (in bracket) are reported below each coefficient estimate. The standard errors are clustered at the dictator level with Pustejovsky and Tipton (2018)'s small cluster bias adjustment. Baseline mean and standard deviation are that of lower-IQ male receivers. Significance levels: * 10%, ** 5%, and *** 1%.

differently to higher-IQ and lower-IQ male receivers. In addition, the coefficient estimate on the female receiver is slightly positive but statistically insignificant, suggesting that male dictators do not discriminate or favor female receivers over male receivers.

The coefficient estimates on the interaction term between the higher-IQ receiver and the female receiver in column 5 is positive but statistically insignificant even at 10%. The 95% confidence interval, presented below the standard error, is relatively tight as well: it is roughly the lower bound of a typical effect size of dictator game experiments that examine the role of social distance with university students (8.9% to 11.4% of the endowment).[12]

Column 6 presents the same specification as column 5 but for female dictators. They allocate less to higher-IQ receivers albeit statistically insignificant, presumably because female dictators prefer equal allocation more than male dictators. Still, the coefficient estimate on the interaction term between the higher-IQ receiver and the female receiver is positive albeit statistically insignificant. The results do not change when we pool male and female dictators, presented in column 7.

Thus, I do not find evidence that men (or women) are less generous to a woman with a higher IQ than them; if anything, they are slightly more generous to a higher-IQ woman than to a higher-IQ men.

5.2 Distribution results

Although OLS only picks up the average effect, Figure 4 shows that the results in section 5.1 also hold in distribution. Panel A of Figure 4 presents empirical CDFs of male dictators' allocation for each receiver type, residualized with the dictator IQ fixed effects to give a causal interpretation to the differences.[13] The figure shows that the CDFs of male dictators' allocation for each receiver type almost coincide; if anything, male dictators allocate less to lower-IQ male receivers (the purple dot-dashed line) and more to higher-IQ female receiver (the solid blue line). The randomization inference (Young 2019) using the Kruskal-Wallis test shows that the p-value of the differences among the CDFs is 0.669, which is far above the conventional 5% cutoff.[14]

Turning to female dictators' allocation, Panel B of Figure 4, all the empirical CDFs almost coincide, again presumably because female dictators prefer equal allocation more than male dictators. The empirical CDF of allocation to the higher-IQ female receivers (the blue solid line) is not on the left of other empirical CDFs, suggesting that female dictators do not allocate less to the higher-IQ female receivers.

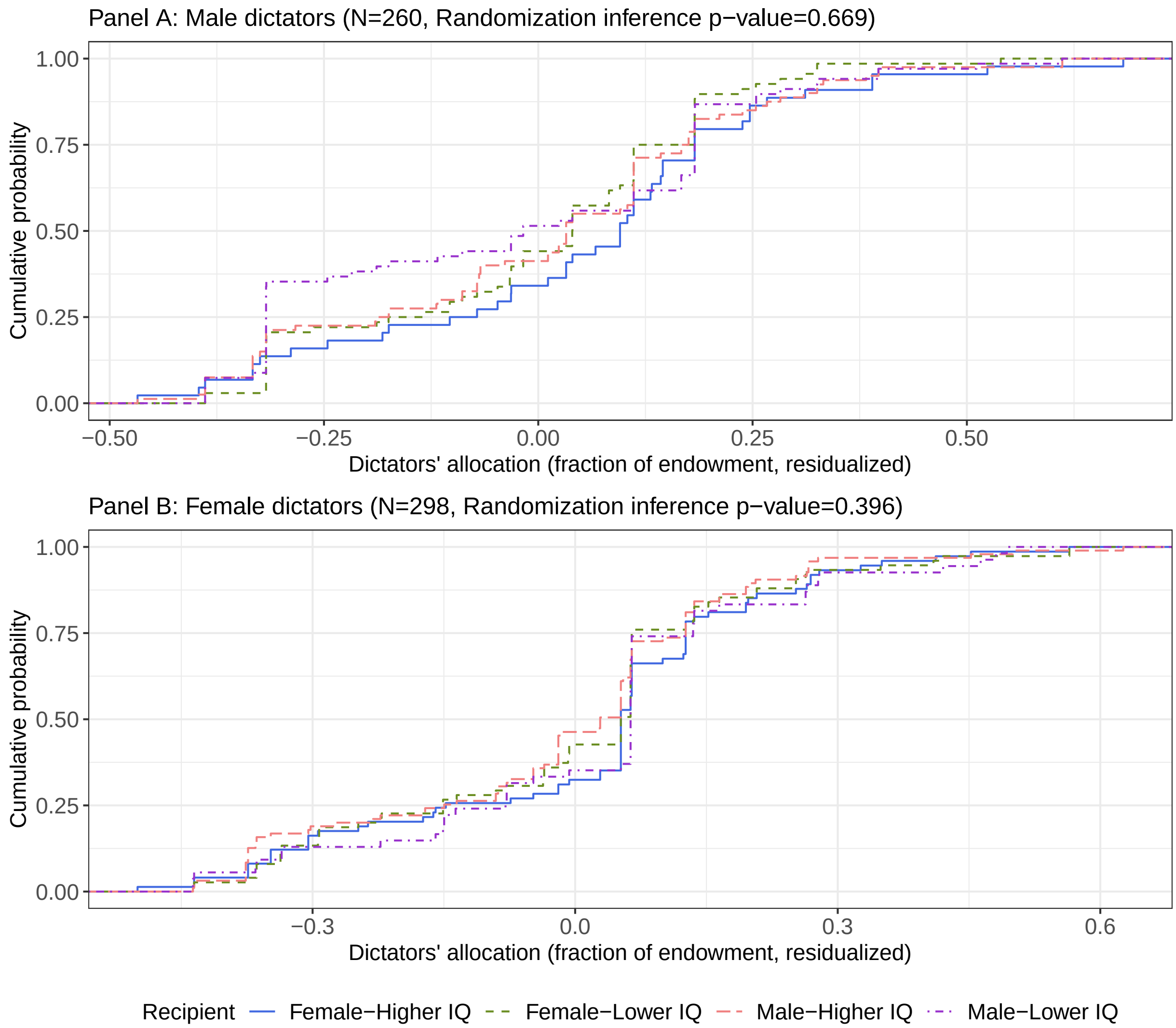
Thus, the results from the OLS hold across the whole distribution of male (and female) dictator

12. For example. Charness and Gneezy (2008) examine how informing the receiver's family name increases the dictators' allocation using a university student sample and find an 8.9 percentage points increase in allocation as a fraction of endowment. Leider et al. (2010) find using a university student sample that dictators increase allocation by 11.4 percentage points as a fraction of endowment for their friends relative to someone living in the same student dormitory. Brañas-Garza et al. (2010) also find using a university student sample that dictators give about 10 percentage points more of their endowment to friends relative to other students in the same class.

13. Residualized allocation is residual from a regression of dictators' allocation on dictator IQ fixed effects.

14. I use randomization inference to address arbitrary dependency among allocations. The null hypothesis is that all CDFs coincide. The number of the random draws is 2000.

Figure 4: Dictators' allocation to higher-IQ female receivers – Distribution



Notes: These figures show the empirical CDF of residualized dictators' allocation by receiver types for male dictators (panel A) and female dictators (Panel B). The randomization inference p-value is calculated with the Kruskal-Wallis test with 2000 random draws.

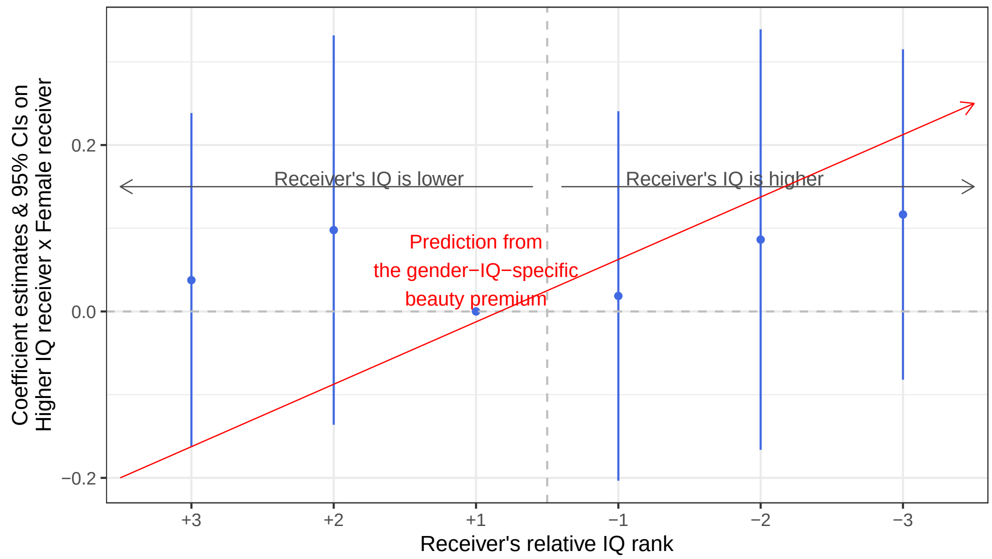
allocation, which suggests that men (or women) are not less generous to a woman than to a man with a higher IQ than them.

5.3 Robustness of the results

Robustness to the "beauty premium"

Note that the so-called "beauty premium" – that people are more generous to physically attractive people (Landry et al. 2006) and hence allocate more to those people in dictator game (Rosenblat 2008) – is not driving my results. First, my analysis compares higher-IQ and lower-IQ receivers of

Figure 5: Male dictators' allocation to higher-IQ female receivers – Robustness to the "beauty premium"



Notes: This figure plots the coefficient estimate on the interaction between higher-IQ receiver and female receiver using male dictators' allocation, separately for each IQ rank difference between dictator and receiver, along with the 95% confidence intervals. The red arrow is a relationship one should expect in the presence of the gender-IQ-specific beauty premium. The standard errors are clustered at the dictator level with Pustejovsky and Tipton (2018)'s small cluster bias adjustment.

the same gender, so even if male dictators perceive female receivers to be more attractive than male receivers, this effect is differenced out.

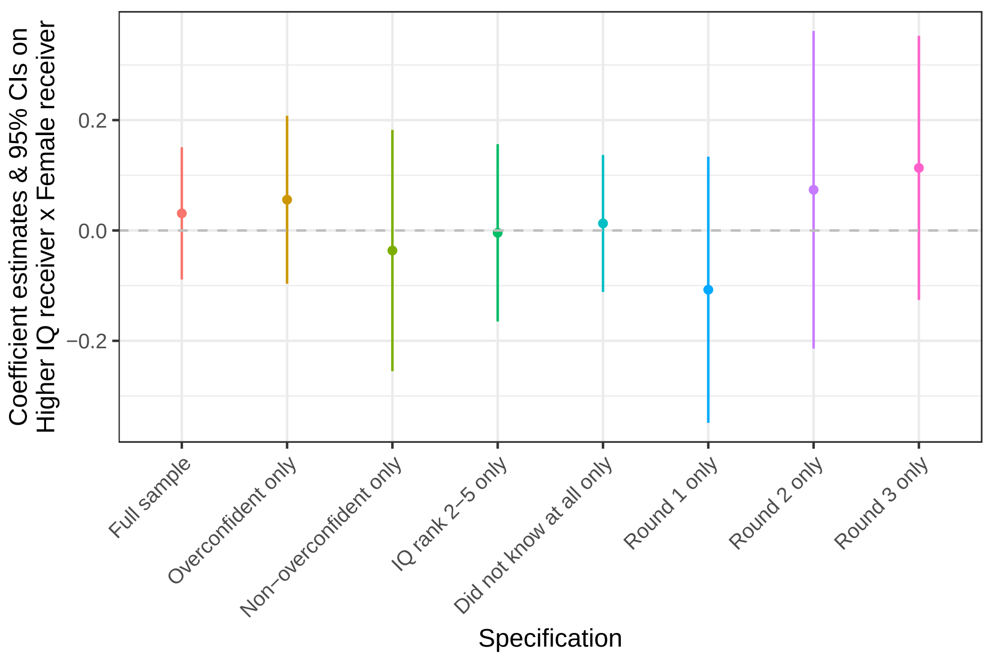
Second, even if higher-IQ people are more physically attractive because they tend to look more confident (Mobius and Rosenblat 2006), it is not driving the results. Figure 5 plots the coefficient estimate on the interaction between higher IQ receiver and female receiver using male dictators' allocation, separately for each IQ rank difference between dictator and receiver (i.e., receiver's IQ rank – dictator's IQ rank), along with the 95% confidence intervals, and the prediction from the gender-IQ-specific "beauty premium" (the red line). The data is inconsistent with the gender-IQ-specific "beauty premium" prediction.

Robustness to male dictators' possible wrong belief about gender differences in IQ

Male dictators allocate no less to female receivers than to male receivers may be because they believe female receivers' IQ is lower than male receivers' IQ conditional on those receivers having higher IQ than them; remember that the more one solves the IQ test questions, the more they earn in part 1 and people may be inequality averse. To address this concern, column 8 of Table 2 presents the regression results of equation 1 but with male receivers' beliefs about paired receivers' IQ as the dependent variable. As discussed in section 2.1, random desk assignment ensures that the receivers' belief proxies the dictators' belief. Appendix Table C3 shows the ex-post balance between dictators and receivers. Column 8 of Table 2 shows that none of the coefficient estimates are statistically significant, which may be because people avoid information that damages their ego-relevant belief even at the cost of reducing their payoff (Castagnetti and Schmacker 2022). However, the coefficient estimate on the higher IQ receiver is positive, consistent with that people correctly believe that male and female receivers with higher IQ solved a larger number of IQ questions. The coefficient estimate on the female receiver is negative albeit statistically insignificant, which might be because men believe women have lower IQ. Yet, the coefficient estimate on the interaction between higher-IQ receiver and female receiver is positive although statistically insignificant, suggesting that the main results are not driven by male dictators' wrong belief about gender differences in IQ.

Robustness to other concerns

Figure 6: Male dictators' allocation to higher-IQ female receivers – Robustness to other concerns



Notes: This figure re-estimates equation 1 with various male dictator sub-samples and plots the coefficient estimates along with their 95% confidence intervals to show the robustness of the findings in Table 2. The standard errors are clustered at the dictator level with Pustejovsky and Tipton (2018)'s small cluster bias adjustment for specifications from "Full sample" to "Did not know at all only" and heteroskedasticity-robust with Bell and McCaffrey (2002)'s small sample bias adjustment for specifications "Round 1 only," "Round 2 only," and "Round 3 only."

In Figure 6, I re-estimate equation 1 with various sub-samples of male dictators and plot the coefficient estimates along with their 95% confidence intervals to show the robustness of the main results to the remaining concerns. I plot the estimate of column 5 of Table 2 as "Full sample" as a reference.

First, overconfident male dictators may dislike higher IQ receivers more and hence allocate less. However, the estimates with overconfident (the specification "Overconfident only") and non-overconfident male dictators (the specification "Non-overconfident only") are very similar to the full sample estimates.

Second, since dictators with IQ rank 1 only face lower IQ receivers and IQ rank 6 only face higher IQ receivers, they may behave differently from other dictators. However, the estimates with male dictators of IQ rank 2-5 only (the specification "IQ rank 2-5 only") provide very similar estimates as the full sample estimates.

Third, although I excluded allocations where dictators knew the paired receivers "very well," knowing the paired receivers even a little may still affect the allocation. However, the estimates with allocations where dictators did not know at all the paired receivers (the specification "Did not know at all only") are very similar to the full sample estimates.

Last, dictators play three rounds of dictator games, and there can be across-round heterogeneity. The specification "Round 1 only" provides estimates with the round 1 data only, the specification "Round 2 only" with the round 2 data only, and the specification "Round 3 only" with the round 3 data only. There is some heterogeneity; especially, in round 1, male dictators allocate less to higher-IQ female receivers, but they are statistically insignificant and male dictators allocate more to higher-IQ female receivers in rounds 2 and 3.

6 Conclusion

This paper provides evidence that men (or women) are not less generous to a woman with IQ higher than them even in a non-romantic setting using a dictator game experiment where the gender and the relative IQ of the receivers are exogenously varied. I show that the results hold across the whole distribution of the dictators' allocation and are robust to various concerns including the so-called "beauty premium" and male dictators' possible wrong belief about women's IQ. Thus, although men (and women) care about their IQ very much and dislike women smarter than them in the marriage market, the men's dislike of smarter women does not manifest outside the marriage market. Although men dislike to be led by women even in non-romantic settings, my results rule out the IQ-related explanation for the men's dislike.

References

Abel, Martin. 2022. “Do Workers Discriminate against Female Bosses?” *Journal of Human Resources.*

Almås, Ingvild, Alexander W. Cappelen, Kjell G. Salvanes, Erik Ø Sørensen, and Bertil Tungodden. 2016. “What Explains the Gender Gap in College Track Dropout? Experimental and Administrative Evidence.” *American Economic Review: Papers & Proceedings* 106 (5): 296–302.

Autor, David, and Melanie Wasserman. 2013. *Wayward Sons: The Emerging Gender Gap in Labor Markets and Education.* Report. Washington, DC: Third Way.

Babcock, Linda, María P. Recalde, Lise Vesterlund, and Laurie Weingart. 2017. “Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability.” *American Economic Review* 107 (3): 714–747.

Bell, Robert M., and Daniel F. McCaffrey. 2002. “Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples.” *Survey Methodology* 28 (2): 169–181.

Bertrand, Marianne. 2011. “New Perspectives on Gender.” In *Handbook of Labor Economics,* edited by David Card and Orley Ashenfelter, 4:1543–1590. Amsterdam, Netherlands: Elsevier.

Bertrand, Marianne, Patricia Cortes, Claudia Olivetti, and Jessica Pan. 2021. “Social Norms, Labour Market Opportunities, and the Marriage Gap Between Skilled and Unskilled Women.” *The Review of Economic Studies* 88 (4): 1936–1978.

Bilén, David, Anna Dreber, and Magnus Johannesson. 2021. “Are Women More Generous than Men? A Meta-Analysis.” *Journal of the Economic Science Association.*

Bilker, Warren B., John A. Hansen, Colleen M. Brensinger, Jan Richard, Raquel E. Gur, and Ruben C. Gur. 2012. “Development of Abbreviated Nine-Item Forms of the Raven’s Standard Progressive Matrices Test.” *Assessment* 19 (3): 354–369.

Bohnet, Iris, and Bruno S. Frey. 1999. “Social Distance and Other-Regarding Behavior in Dictator Games: Comment.” *American Economic Review* 89 (1): 335–339.

Born, Andreas, Eva Ranehill, and Anna Sandberg. 2022. “Gender and Willingness to Lead: Does the Gender Composition of Teams Matter?” *The Review of Economics and Statistics* 104 (2): 259–275.

Brañas-Garza, Pablo, Ramón Cobo-Reyes, María Paz Espinosa, Natalia Jiménez, Jaromír Kovářík, and Giovanni Ponti. 2010. “Altruism and Social Integration.” *Games and Economic Behavior* 69 (2): 249–257.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. “Sex and Science: How Professor Gender Perpetuates the Gender Gap.” *The Quarterly Journal of Economics* 125 (3): 1101–1144.

Castagnetti, Alessandro, and Renke Schmacker. 2022. “Protecting the Ego: Motivated Information Selection and Updating.” *European Economic Review* 142:104007.

Chakraborty, Priyanka, and Danila Serra. 2022. *Gender and Leadership in Organizations: The Threat of Backlash.* Working Paper.

Charness, Gary, and Uri Gneezy. 2008. “What’s in a Name? Anonymity and Social Distance in Dictator and Ultimatum Games.” *Journal of Economic Behavior & Organization* 68 (1): 29–35.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "oTree–An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9:88–97.

Coffman, Katherine Baldiga. 2014. "Evidence on Self-Stereotyping and the Contribution of Ideas." *The Quarterly Journal of Economics* 129 (4): 1625–1660.

Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 47 (2): 448–474.

Datta Gupta, Nabanita, Anders Poulsen, and Marie Claire Villeval. 2013. "Gender Matching and Competitiveness: Experimental Evidence." *Economic Inquiry* 51 (1): 816–835.

Eil, David, and Justin M. Rao. 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself." *American Economic Journal: Microeconomics* 3 (2): 114–138.

Engel, Christoph. 2011. "Dictator Games: A Meta Study." *Experimental Economics* 14 (4): 583–610.

Falk, Armin, Fabian Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers. 2021. "Socioeconomic Status and Inequalities in Children's IQ and Economic Preferences." *Journal of Political Economy* 129 (9): 2504–2545.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics* 114 (3): 817–868.

Fisman, Raymond, Sheena S. Iyengar, Emir Kamenica, and Itamar Simonson. 2006. "Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment." *The Quarterly Journal of Economics* 121 (2): 673–697.

Folke, Olle, and Johanna Rickne. 2020. "All the Single Ladies: Job Promotions and the Durability of Marriage." *American Economic Journal: Applied Economics* 12 (1): 260–287.

Franzen, Axel, and Sonja Pointner. 2013. "The External Validity of Giving in the Dictator Game." *Experimental Economics* 16 (2): 155–169.

Greiner, Ben. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association* 1 (1): 114–125.

Husain, Aliza N., David A. Matsa, and Amalia R. Miller. 2021. *Do Male Workers Prefer Male Leaders? An Analysis of Principals' Effects on Teacher Retention.* Working Paper.

Isaksson, Siri. 2018. *It Takes Two: Gender Differences in Group Work.* Working Paper.

Klinowski, David. 2018. "Gender Differences in Giving in the Dictator Game: The Role of Reluctant Altruism." *Journal of the Economic Science Association* 4 (2): 110–122.

Landry, Craig E., Andreas Lange, John A. List, Michael K. Price, and Nicholas G. Rupp. 2006. "Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment." *The Quarterly Journal of Economics* 121 (2): 747–782.

Leider, Stephen, Tanya Rosenblat, Markus M. Möbius, and Quoc-Anh Do. 2010. "What Do We Expect from Our Friends?" *Journal of the European Economic Association* 8 (1): 120–138.

Mobius, Markus M., and Tanya S. Rosenblat. 2006. "Why Beauty Matters." *American Economic Review* 96 (1): 222–235.

Niederle, Muriel, and Lise Vesterlund. 2011. "Gender and Competition." *Annual Review of Economics* 3 (1): 601–630.

Pustejovsky, James E., and Elizabeth Tipton. 2018. "Small-Sample Methods for Cluster-Robust Variance Estimation and Hypothesis Testing in Fixed Effects Models." *Journal of Business & Economic Statistics* 36 (4): 672–683.

Rosenblat, Tanya S. 2008. "The Beauty Premium: Physical Attractiveness and Gender in Dictator Games." *Negotiation Journal* 24 (4): 465–481.

Young, Alwyn. 2019. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *The Quarterly Journal of Economics* 134 (2): 557–598.

Zimmermann, Florian. 2020. "The Dynamics of Motivated Beliefs." *American Economic Review* 110 (2): 337–361.

Online Appendix

A Deviations from the pre-analysis plan

A.1 Implementation

In the initial design, receivers finished all the tasks except the post-questionnaire and left the laboratory before dictators received their IQ rank so that dictators could play the dictator game without receivers in the same room. The allocation to the receivers was paid electronically as a "participation fee" for the online post-questionnaire, which was sent to receivers via email after the session was over. However, as I ran the 1st session with this initial design with 24 participants, dictators had to wait idly for about 20-30 minutes until receivers left the laboratory, and dictators seemed to have lost concentration during this waiting time: about half of the dictators could not answer the comprehension questions about their IQ rank. Thus, I changed the design and let receivers stay in the laboratory while dictators played the dictator game. I looked at the 1st session data before making this change. I exclude the 1st session data in the analysis, but results are robust to including the 1st session data. The oTree code and instructions used for the 1st session are available upon request.

I also made the following changes after the 1st session:

1. I reduced participation fee from 3€ to 2.5€ because participants earned more than I expected in the IQ test.
2. I added more explanation to the instructions on how the IQ rank was assigned and how to allocate endowment in the dictator game.
3. I asked participants' major by simply choosing among the choices from humanities, social sciences, natural sciences/mathematics, medicine, and engineering and letting them type in their degree program name for a check, instead of letting them access to the University of Bologna's degree program website. This is because the computers in the laboratory sometimes did not accept `iframe` or prevented a pop-up to another website due to the security setting.

A.2 Other changes

Interpretation and focus

1. I rephrased smartness as competence to better place my results in the literature.
2. I mainly discussed results for question 3.

Analysis

3. I corrected the definition of $Lower_{ij}$. Consequently, I renamed it as $IQHigher_{ij}$ to make the meaning clearer.

4. I added distributional analysis (in Figure 4) to examine whether the results hold also in distribution.
5. I added round-by-round analysis in Figure 6.
6. I used `lm_robust` instead of `vcovCR` to apply Pustejovsky and Tipton (2018)'s small cluster bias adjustment because `vcovCR` did not make degrees of freedom adjustment.
7. I included in female and male dictator regressions STEM major dummy and Emilia-Romagna dummy because excluding them in regressions where the sample is conditioned by gender made little sense. The results are invariant to the exclusion of these covariates.
8. I divided dictators' allocation by dictator endowment to facilitate the interpretation of the regression results (this does not affect my results because of the round fixed effects).

B Description of covariates

X_{ij} in equation 1 includes the following variables:

Dictator characteristics

- $Age_i \in \mathbb{N}$: dictator i's age.
- $FromEmiliaRomagna_i \in \{0, 1\}$: an indicator variable equals 1 if dictator i is from the Emilia-Romagna region where the experiment was conducted, 0 otherwise.
- $SocialSciences_i \in \{0, 1\}$: an indicator variable equals 1 if dictator i majors in social sciences, 0 otherwise.
- $STEM_i \in \{0, 1\}$: an indicator variable equals 1 if dictator i majors in natural sciences/mathematics, engineering, or medicine; 0 otherwise.
- $PostBachelor_i \in \{0, 1\}$: an indicator variable equals 1 if dictator i is either a master or post-bachelor student, a student in the 4th year or beyond in a bachelor-master combined program (bachelor is a 3 year program in Italy), or PhD student, 0 otherwise.
- $OverConfidence_i \in \{-1, 0, 1\}$: degree of dictator i's overconfidence. It is equal to -1 if dictator i's guess about the number of IQ test questions they have solved correctly is lower than the actual number, 0 if equal to the actual number, and 1 if higher than the actual number.

receiver characteristics

- $Age_j \in \mathbb{N}$: receiver j's age.
- $FromEmiliaRomagna_j \in \{0, 1\}$: an indicator variable equals 1 if receiver j is from the Emilia-Romagna region where the experiment was conducted, 0 otherwise.

Fixed effects

- $\sum_{k=2}^{3} \theta_k^{round} \mathbb{1}[\text{round}_{ij} = k]$: fixed effects for dictator game or belief elicitation round. $\mathbb{1}$ is the indicator variable.
- $\sum_{k=2}^{3} \theta_k^{social\ distance} \mathbb{1}[\text{social distance}_{ij} = k]$: fixed effects for social distance between dictator i and receiver j. social distance$_{ij} = 1$ means dictator i did not know receiver j at all, $= 2$ knew but not well, and $= 3$ saw before. $\mathbb{1}$ is the indicator variable.

C Additional tables

Table C1: Exogeneity of male dictator IQ rank conditional on dictator IQ fixed effects

Outcome:	Age	From Emilia-Romagna	Human-ities	Social sciences	STEM	Post bachelor	Over-confidence
Sample:				Male dictators			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
IQ rank = 2	-0.744	0.024	-0.316**	0.099	0.217	0.038	0.304
	(0.950)	(0.128)	(0.153)	(0.145)	(0.177)	(0.174)	(0.242)
IQ rank = 3	-0.449	0.055	-0.123	0.155	-0.032	0.141	0.350
	(0.883)	(0.148)	(0.198)	(0.180)	(0.192)	(0.192)	(0.320)
IQ rank = 4	-0.059	0.051	-0.232	0.175	0.058	0.163	0.339
	(1.167)	(0.145)	(0.169)	(0.176)	(0.193)	(0.188)	(0.321)
IQ rank = 5	-0.056	0.300	-0.133	0.165	-0.031	-0.100	-0.057
	(1.231)	(0.184)	(0.230)	(0.192)	(0.217)	(0.190)	(0.367)
IQ rank = 6	-1.563	0.495*	-0.476*	0.638***	-0.162	-0.232	0.227
	(1.886)	(0.251)	(0.271)	(0.220)	(0.296)	(0.318)	(0.629)
Dictator IQ FE	✓	✓	✓	✓	✓	✓	✓
F statistic	0.354	0.951	1.575	2.219*	0.540	0.667	0.802
Adj. R-squared	-0.023	-0.023	-0.027	0.027	-0.047	-0.023	-0.046
Observations	91	91	91	91	91	91	91

Notes: This table shows male dictator IQ rank is uncorrelated with their characteristics conditional on dictator IQ fixed effects. The F statistic shows the joint significance of IQ rank = 2 to IQ rank = 6 dummies. Heteroskedasticity-robust standard errors with Bell and McCaffrey (2002)'s small sample bias adjustment are reported below each coefficient estimate. Significance levels: * 10%, ** 5%, and *** 1%

Table C2: Exogeneity of the main regression's covariates conditional on dictator IQ fixed effects: Male dictators

Outcome:	Age	From Emilia-Romagna	Human-ities	Social sciences	STEM	Post bachelor	Over-confidence
Sample:	Male dictators						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Higher IQ receiver x Female receiver	0.452	-0.342***	-0.098	-0.073	0.171	0.155	0.021
	(0.678)	(0.091)	(0.112)	(0.097)	(0.124)	(0.120)	(0.156)
Higher IQ receiver	-0.337	0.216***	-0.064	0.163*	-0.099	-0.102	-0.082
	(0.513)	(0.075)	(0.094)	(0.086)	(0.092)	(0.092)	(0.148)
Female receiver	-0.581	0.156***	0.032	0.048	-0.080	-0.099	-0.001
	(0.415)	(0.058)	(0.072)	(0.063)	(0.079)	(0.078)	(0.091)
Dictator IQ FE	✓	✓	✓	✓	✓	✓	✓
F statistic	0.707	4.780***	1.171	1.299	0.680	0.678	0.132
Adj. R-squared	0.048	0.044	-0.002	0.056	0.018	0.040	-0.010
Observations	260	260	260	260	260	260	260
Clusters	91	91	91	91	91	91	91

Outcome:	Age (receiver)	From Emilia-Romagna (receiver)	Dictator game round 1	Dictator game round 2	Dictator game round 3	Did not know at all	Saw before	Knew but not very well
Sample:	Male dictators							
	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Higher IQ receiver x Female receiver	0.224	-0.064	0.183	-0.245**	0.062	-0.012	-0.013	0.025
	(0.773)	(0.103)	(0.128)	(0.104)	(0.131)	(0.034)	(0.040)	(0.041)
Higher IQ receiver	-0.079	0.194**	-0.158*	0.057	0.101	0.034	-0.029	-0.006
	(0.560)	(0.077)	(0.086)	(0.085)	(0.082)	(0.036)	(0.030)	(0.034)
Female receiver	-0.263	-0.002	-0.135	0.058	0.077	0.049	-0.032	-0.017
	(0.518)	(0.053)	(0.087)	(0.080)	(0.083)	(0.032)	(0.028)	(0.017)
Dictator IQ FE	✓	✓	✓	✓	✓	✓	✓	✓
F statistic	0.092	2.983**	1.262	2.216*	1.739	1.163	1.627	1.175
Adj. R-squared	-0.006	0.012	-0.015	-0.011	-0.009	0.178	0.060	0.134
Observations	260	260	260	260	260	260	260	260
Clusters	91	91	91	91	91	91	91	91

Notes: This table shows receiver gender, receiver's IQ rank relative to dictator's, and their interaction are uncorrelated with male dictator or paired receiver characteristics, dictator game rounds, or social distance between dictators and paired receivers. The F statistic shows the joint significance of all covariates. Cluster-robust standard errors with Pustejovsky and Tipton (2018)'s small cluster bias adjustment are reported below each coefficient estimate. Significance levels: * 10%, ** 5%, and *** 1%.

Table C3: Balance between male dictators and male receivers

	Male receivers		Male dictators		Difference (Receivers – Dictators)		
	Mean	SD	Mean	SD	Mean	SE	P-value
Panel A: Own characteristics							
IQ level	7.12	1.05	6.89	1.24	0.23	0.17	0.18
IQ rank	2.98	1.65	3.31	1.73	-0.33	0.24	0.18
Age	23.30	2.82	23.23	2.81	0.07	0.40	0.87
From Emilia-Romagna	0.22	0.42	0.19	0.39	0.03	0.06	0.55
Humanities	0.25	0.44	0.32	0.47	-0.07	0.07	0.29
Social sciences	0.26	0.44	0.24	0.43	0.02	0.06	0.78
STEM	0.49	0.50	0.44	0.50	0.05	0.07	0.48
Post bachelor	0.49	0.50	0.37	0.49	0.12	0.07	0.10
Overconfidence	0.63	0.62	0.56	0.72	0.07	0.10	0.45
Time on feedback (sec.)	88.95	82.00	107.52	102.26	-18.56	13.40	0.17
Observations	104		91				
Panel B: Paired receivers' characteristics							
IQ level	6.77	1.18	6.91	1.12	-0.15	0.10	0.15
IQ rank	3.59	1.73	3.45	1.74	0.15	0.12	0.23
Higher IQ	0.38	0.49	0.48	0.50	-0.10	0.05	0.08
Age	23.44	2.76	23.55	2.98	-0.11	0.28	0.69
Female	0.54	0.50	0.43	0.50	0.11	0.04	0.01
From Emilia-Romagna	0.18	0.38	0.25	0.43	-0.07	0.04	0.10
Observations	197		260				
Panel C: Social distance with paired receivers							
Did not know at all	0.98	0.14	0.95	0.23	0.03	0.02	0.13
Knew but not well	0.02	0.14	0.03	0.18	-0.01	0.02	0.41
Saw before	0.00	0.00	0.02	0.14	-0.02	0.01	0.06
Observations	197		260				
Panel D: Belief on paired receiver's IQ level (fraction of baseline SD)							
Belief on IQ level	3.50	1.01					
Belief on IQ level (residualized)	0.00	1.00					
Observations	197		260				

Notes: This table shows that male receivers and male dictators are comparable also ex-post. P-values for the difference between male receivers and male dictators are calculated with heteroskedasticity-robust standard errors with Bell and McCaffrey (2002)'s small sample bias adjustment for Panel A and with Pustejovsky and Tipton (2018)'s small cluster bias adjustment for Panels B-D.

D Experimental instructions

To the experimenter:

- *Before subjects arrive:*
- *Clear image cache from the browser.*
- *Put on each desk (i) a scratch paper and (ii) a pencil.*
- *Have a printed instructions ready.*
- *Set up photo booth. The brightness of the camera should be 172 and resolution 0.7 mb with 4:3 aspect ratio.*
- *Leave a paper in which participants write down their desk number on the photo booth.*
- *After registration:*
- *Give them photo taking instructions.*
- *Ask them to take photo at the photo booth, then take seat.*
- *After subjects took photo:*
- *Check that all the participants' photos are neutral: they must not signal nothing other than their gender.*
- *Make sure that the photos are saved as Pxx.jpg where xx is participant's desk number.*
- *After reserve participants left the room:*
- *Rename the photo name to the new desk number's for those who moved to new desks.*
- *Store photos in _static/photo folder.*
- *Startup Chrome & oTree*

App: personal_info

Page: DeskNumber

Please enter your desk number and click "Next"

[Your desk number:]

Page: PersonalInfo

Please check that the photo is yours

[Participant's photo]

The photo you took is displayed above. Please check that the photo is yours. Please also enter your first name. We will come to each desk and check the photo and the first name.

[Your first name:]

[Digital signature (please wait for us to sign you in):]

To the experimenter: before type in the password, do the followings:

- *Check that the photo and the first name correspond to the participant.*

Then click "Next" to let participants to proceed.

Page: Introduction

To the experimenter: read the instructions aloud.

Welcome!

You are participating in a study of the BLESS. For your participation, you will receive a fixed amount of [Participation fee]€. There are 2 parts in which you can earn additional earnings. The expected length is 1 hour.

During the study, we use your photo and first name to identify you. Your photo and the first name will only be used in this session and deleted immediately afterwards. However, we may ask some of you to allow us to use their photo in another study, which you can opt out.

The study is computerized, meaning that the computer program will give you precise instructions in each task. In the following you will find general instructions of the study, which you can always find in the bottom of the screen.

General instructions

- Please turn off your mobile phone.
- Please do not communicate with other participants.
- Please only use paper and pencil.
- Once you understand the instructions or enter your decisions, please click "Next" to proceed unless instructed otherwise.
- If you have any questions, please raise your hand at any time.

If there is no question, we will start the study.

To the experimenter:

- *Confirm that everyone turned off their mobile phone.*
- *Then, if there is no question, click "Advance slowest user(s)."*

After that, just sit in the experimenter area unless someone raises her or his hand. Do not read instructions aloud unless this document says to do so.

App: iqtest

Page: Introduction

Part 1: Instructions

In part 1, you will work on an IQ test, which is frequently used to measure intelligence. The IQ test you will work on is the Raven's Standardized Progressive Matrices Test.

You will solve the IQ test as follows: for each question, you will see an image in which a piece is missing. Below the image there will be several options. Choose the correct option among them to complete the image. There will be only one correct option.

An example is provided below. In the image, there are 9 large white squares each containing a small black square. In the first column, the small black square is located on the left; in the second column, in the middle; in the third column, on the right. In the first row, it is located on the top; in the second row, in the middle; in the third row, in the bottom. Thus, in the third column of the third row, the small black square must be located in the right bottom, thus the correct option is 5.

[Raven matrix no. 31 here]

There are 9 questions in total and you have 9 minutes. Once the time is over, you will automatically be directed to the next page. You will earn [Payoff per IQ test]€ for each correct answer. There is no penalty for wrong answers. You can use paper and pencil on your desk.

Page: IQTest (9 minutes)

Please complete the image by choosing the correct option

[Raven IQ test]

Page: Guess

Guess the number of questions you solved

The IQ test is over.

We have randomly formed a group of 6 participants including you in this room and constructed a ranking among the 6 group members based on their IQ test performance.

A group member with rank 1 performed the best in the IQ test, followed by a group member with rank 2, 3, 4, 5, and 6. In case of a tie between group members, the computer randomly decided who receives the higher rank.

How many questions do you think you have solved correctly? If your guess is correct, you will additionally earn [Earnings from guess]€.

[Guess]

[Dictator] ***Page: Feedback***

Feedback

Among your 6 group members including you, you received **Rank [Participant's rank]**.

[Among your 6 group members, how many people performed better than you in the IQ test?:]

[Among your 6 group members, how many people performed worse than you in the IQ test?:]

App: dictator

[Dictator] ***Page: IntroductionDict***

Part 2: Instructions

In this part, half of you will be active participants who will work on the task described in the next page, and the remaining half will be passive participants who will NOT work on the task described in the next page.

[Dictator] ***Page: IntroductionDictCont***

Part 2: Instructions

You are assigned to a role of **active participant**.

Part 2 consist of 3 rounds. In each round, you will first receive an endowment (money). After that, you will be paired with a passive participant in your group.

Your task in this part is to allocate the endowment to yourself and the paired passive participant. The passive participants, other active participants, or anyone else other than us will never know who allocated how much.

At the end of the study, the computer will randomly select 1 out of 3 rounds and the amount you allocated to you in that round will be your earnings in this part.

The computer will also randomly select 1 out of 3 rounds for the paired passive participants and the amount you allocated to him or her in that round will be his or her earnings in this part.

[Recipient] ***Page: IntroductionRecip***

Part 2: Instructions

In part 2 consists of 2 rounds. In each round, you will be paired with another participant in your group.

Your task in this part is to guess how many questions the paired participant has solved correctly in the IQ test. For each correct guess, you will earn [Earning from guess other]€.

[Dictator] ***Page: PrepEndow***

Round [Round number] of 3

Please wait.

[Dictator] ***Page: OfferDict1-3***

Round [Round number] of 3

[Paired participant's photo]

[Paired participant's first name]

Rank [Paired participant's rank]

You have received **[7/5/7]€** for this round.

You have been paired with **[Paired participant's first name].**

Please allocate the endowment between yourself and [Paired participant's first name]. When you click the line below, a cursor appears. You can move the cursor by dragging it. Please move the cursor to your preferred position to determine the allocation.

[Slider from 0 to endowment that moves with increment of 0.5]

[Recipient] ***Page: GuessOther1-3***

Round [Round number] of 2

[Paired participant's photo]

[Paired participant's first name]

Rank [Paired participant's rank]

You have been paired with **[Paired participant's first name].**

How many questions do you think [Paired participant's first name] has solved correctly?

[Guess]

[Dictator] **Page: AnonymityCheckDict**

Round 3 of 3

Below we display the participants whom you were paired with. How well did you know him/her before participating in this study?

[Paired participant 1's photo] [Paired participant 1's first name] [I didn't know him/her at all, I saw him/her before, I knew him/her but not very well, I knew him/her very well]	[Paired participant 2's photo] [Paired participant 2's first name] [I didn't know him/her at all, I saw him/her before, I knew him/her but not very well, I knew him/her very well]	[Paired participant 3's photo] [Paired participant 3's first name] [I didn't know him/her at all, I saw him/her before, I knew him/her but not very well, I knew him/her very well]

[Recip] **Page: AnonymityCheckRecip**

Round 2 of 2

Below we display the participants whom you were paired with. How well did you know him/her before participating in this study?

[Paired participant 1's photo] [Paired participant 1's first name] [I didn't know him/her at all, I saw him/her before, I knew him/her but not very well, I knew him/her very well]	[Paired participant 2's photo] [Paired participant 2's first name] [I didn't know him/her at all, I saw him/her before, I knew him/her but not very well, I knew him/her very well]

Page: ShowResults

Results

The study is over. The results are provided below.

- In part 1, you solved [**Number of IQ test questions solved**] questions and earned [**Earnings from IQ test]€.** [If guess is correct] You have additionally earned **[Earnings from guess]€** because your guess about the number of questions solved was correct.
- [Dictator] In part 2, computer selected **round [1/2/3]** in which you allocated **[Allocation to self]€** to yourself.
- [Recipient] In part 2, you made **[Number of correct guesses on others]** guesses correct. So you earned **[Earnings from guesses other]€**.
- [Recipient] You additionally earned a top-up of **[Allocation from dictator]€**.

So, your total earnings are **[Participant's earnings]€** including [Participation fee]€ of participation fee.

Thank you for participating in this study! We will prepare your payment soon. Meanwhile, please answer a short questionnaire by clicking "Next." Your answer will be kept anonymous and will not affect your payment.

Page: Questionnaire1

Questionnaire 1 of 3

[Your study program: Agricultural and Food Sciences; Economics and Management; Education; Engineering and Architecture; Humanities; Languages and Literatures, Interpreting and Translation; Law; Medicine; Pharmacy and Biotechnology; Political Sciences; Psychology; Sciences; Sociology; Sport Sciences; Statistics; Veterinary Medicine]

[Please also type your full study program name in Italian:]

If you are enrolled in a specialized or professional program, please choose the closest study program. If you are enrolled in a post-bachelor vocational program, please choose the study program of your bachelor's degree. If you are an exchange student, please choose the study field closest to the one in your home university.

[Your degree program: Bachelor, Master/Post-bachelor, Bachelor-master combined (ciclo unico), Doctor]

[Your year in the degree program: 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, 8th]

[Your age:]

[Your gender: Male, Female]

[Are you from Emilia-Romagna region?: Yes, No]

[Recipient] In another study, we'd like to use your photo. We will show your photo to some people in the University of Bologna only in this room, but no other people except us will see your photo. Your photo will be deleted immediately after we finish another study. For your cooperation, we will pay you gratuity of [Gratuity for photo use]€. May we use your photo in another study?

[Yes, I allow the researcher to use my photo in another study; No, I do NOT allow the researcher to use my photo in another study]

[What do you think the study you participated was about?]

[Was there anything unclear or confusing about the study you participated?]

[Do you have any other comments? (optional)]

Page: Questionnaire2

To the experimenter:

- *Prepare payment.*

Questionnaire 2 of 3

Below we display the participants whom you were paired with. How well did you know him/her before participating in this study?

[Dictator]

[Paired participant 3's photo]	[Paired participant 1's photo]	[Paired participant 2's photo]
[Paired participant 3's first name]	[Paired participant 1's first name]	[Paired participant 2's first name]
[I didn't know him/her at all, I saw him/her before, I knew him/her but not very well, I knew him/her very well]	[I didn't know him/her at all, I saw him/her before, I knew him/her but not very well, I knew him/her very well]	[I didn't know him/her at all, I saw him/her before, I knew him/her but not very well, I knew him/her very well]

[Recipient]

[Paired participant 2's photo]	[Paired participant 1's photo]
[Paired participant 2's first name]	[Paired participant 1's first name]
[I didn't know him/her at all, I saw him/her before, I knew him/her but not very well, I knew him/her very well]	[I didn't know him/her at all, I saw him/her before, I knew him/her but not very well, I knew him/her very well]

Page: Questionnaire3

Questionnaire 3 of 3

[What do you think this study was about?]

[Was there anything unclear or confusing about this study?]

[Do you have any other comments? (optional)]

[Participants with payment less than 5€] ***Page: ExtraTask***

Extra task

Please solve the additions below and click next to earn [5€ – Participant's earnings]€.

84	33	64

[Sum of the above numbers:]

19	65	97

[Sum of the above numbers:]

E Pre-analysis plan

Conditionally gentlemen? Capped gender equality ideal

Yuki Takahashi

University of Bologna

November 22, 2019

1. **Study summary**

In this study, I test the following questions:

- Are men and women equally or more generous to a woman who is less smart than them?
- Are men and women less generous to a woman who is smarter than them?
- Is the difference in men's and women's generosity to a person who is smarter than them and to another person who is less smart than them more negative when these persons are women?

I use IQ test as a measure of smartness and giving in dictator game as a measure of generosity. I also test alternative explanation and effect heterogeneities.

2. **Experimental procedure and design**

The main experiment is conducted during November-December 2019 at the Bologna Laboratory for Experiments in Social Science (BLESS) – a physical laboratory – on computer and the online survey is conducted with Qualtrics. I recruit students at the University of Bologna who were born in Italy and declared they could participate in a session in English via ORSEE (Greiner 2015)[1] and program the experiment with oTree (Chen, Schonger, and Wickens 2016). The expected length is between 40 minutes to 1 hour depending on whether a subject is a dictator or a recipient. Each session consists of a multiple of 6 subjects. I will recruit about 400 subjects to achieve sufficient power (see power simulation in appendix A).

The experiment proceed as follows:

- Upon arrival to the laboratory, each subject draws a plastic coin from a bag and is assigned a desk corresponding to the number written on the coin. This is the source of randomization and the rest are prespecified for each oTree participant ID which is assigned by the server for each desk. I then ask subjects to take their photo at a

[1] I do not recruit subjects who participated in experiments related to gender in the past (as far as I could check).

dedicated photo booth and upload it to the server.[2][3] I confirm their photo does not signal things other than gender, for example their favorite soccer team or their department.[4] After taking photo, subjects go to their desk and enter their desk number and first name.[5] I use subjects' photo and first name to tell their gender to other subjects without making their gender explicit. Then, the laboratory manager and I go to each subject's desk and check that subjects enter their true first name and that the photo corresponds to the actual subject before proceed. This process ensures subjects that the other subjects' photo and first name are real.[6]

- After the photo and first name are checked, subjects work on 9 IQ test questions displayed on their computer screen for 9 minutes. I tell subjects that it is frequently used to measure intelligence. The 9 IQ test questions are the form A 9-item Raven test proposed by Bilker et al. (2012). Subjects earn 0.5€ for each correct answer. The total number of correct answers as well as how much they earned are told to subjects with odd numbered oTree participant ID only at the end of the experiment and to subjects with even numbered oTree participant ID right after the next task.
- After the IQ test, I split the subjects into a group of 6 based on their oTree participant ID and assign them their rank within their group based on the number of questions they solved in the IQ test. I break ties with a random draw (if any). I also ask subjects to guess the number of questions they solved correctly, which is used as a control. Subjects who guessed correctly earn 0.5€ and whether the guess is correct is only told at the end of the experiment to subjects with odd numbered oTree participant ID and right after the guess to subjects with even numbered oTree participant ID. For subjects with even numbered oTree ID, I tell them their rank, ask them to answer how many subjects in their group are ranked higher and lower than them to make sure they understand their rank, and ask to also guess the number of questions other 2 recipients in the same group have solved from which they earn 0.5€ for each correct guess.
- Then I assign subjects with odd numbered oTree participant ID to the dictator role and with even numbered oTree participant ID to the recipient role.[7] I tell those who are assigned the recipient role will participate in part 2 via online (which is simply an online survey) from which they will earn additional participation fee and ask them to provide their email address (participation is still optional, although I tell them it is part 2 of the experiment), to answer a short questionnaire, and whether they give us the right to use their photo in another experiment.[8] Then I ask them to leave the laboratory.

[2] I inform subjects in the invitation email that the experiment uses their facial photo and first name.

[3] I use the same photo booth for all subjects to standardize photo. I also provide written instructions as an additional effort for standardization.

[4] However, different from Isaksson (2018), I allow subjects to wear their glasses.

[5] Desk number is only used to associate their photo with their oTree participant ID.

[6] I do not use photo or first name of subjects assigned to the dictator role, but letting them go through the same process ensures them that the recipients' photo and first name are real.

[7] I mention in the beginning of the study that who will participate in part 2 will be randomly determined so that the information does not affect subjects' IQ test performance.

[8] I ask the following questions to the recipients: (i) age, (ii) gender, (iii) whether from Emilia-Romagna region (the region where the University of Bologna is located), (iv) how well they know the 2 other

The laboratory manager pays them outside the laboratory the participation fee and their earnings from the IQ test and from the guesses, as well as gratuity for their photo use in another experiment (if they agree).

- After the recipients leave the laboratory, I first tell dictators their rank, and ask them to answer how many subjects in their group are ranked higher and lower than them. I then pair each dictator with all the three recipients in the same group – who have already left the laboratory – one by one in an order pre-determined by their oTree participant ID. Each time a dictator is paired with a recipient, the dictator receives an endowment and the computer screen displays the recipient's photo, first name, and rank. The dictator then allocates the endowment between herself or himself and the recipient using slider without tick mark, label, or initial cursor position.[9] I vary the endowment across rounds – 7€ in the 1st and the 3rd rounds and 5€ in the 2nd round – and do not tell how they are determined or whether they are the same for other dictators. I make it clear that the paired recipients, other subjects in the laboratory, or anyone else – except us – cannot identify who allocated how much and that the allocation will be paid to recipients as participation fee for the online part of part 2 in which recipients will be participating. To prevent dictators to choose similar allocations across 3 rounds, I set 5 seconds of waiting time before each dictator game during which they see a countdown timer and a sentence "Preparing your endowment and pairing you with another participant."
- Right after the 3 rounds of dictator games, I show 3 recipients' photo and first name with whom dictators were paired with and ask them how well they know paired participants. They select from 4 choices: (a) do not know at all, (b) have seen before, (c) know but not very well, (d) know very well. I ask this question again in the questionnaire at the end of the study to verify their response.
- After the dictator game is over, 1 out of the 3 rounds is randomly chosen for each subject. Then I pay dictators the allocation from that round's dictator game along with participation fee, earnings from the IQ test and earnings from the guess. Then, while the laboratory manager and I prepare the payment, each dictator answers a short questionnaire which is used as controls and to identify subject's gender.[10][11]

recipients on whom they made guesses ((a) do not know at all, (b) have seen before, (c) know but not very well, (d) know very well).

[9] Dictators can choose any number between 0 and the endowment with increment of 0.5 as an amount to allocate to the recipient.

[10] I ask the following questions to dictators: (i) study program, (ii) degree program (bachelor, master/post-bachelor, bachelor-master combined, doctor), (iii) years in the program, (iv) age, (v) gender (male, female), (vi) whether from Emilia-Romagna region (the region where the University of Bologna is located), (vii) how well they know the paired recipients ((a) do not know at all, (b) have seen before, (c) know but not very well, (d) know very well). Classification of study program is in appendix C.

[11] If a subject's earning is less than 5€, which is the minimum payment I must guarantee in the BLESS, I ask them to work on a short addition task – which I do not need for my analysis – and pay additional euros. I do the same for recipients. However, I expect this would not happen because IQ test is multiple choice.

- After the dictators left the room, I send a link to the online survey to recipients who agreed to participate. They first enter their email address so that I can associate the survey data with the laboratory data. Then they answer the same questionnaire that the dictator answered (except those they already answered in the laboratory) as well as how they find the experiment.[12] Then I pay to those who completed the survey the allocation from the chosen round's dictator game as participation fee.[13]

3. Previous experiment

In June 2019, I run an experiment with 52 subjects at the same experimental laboratory with different but closely related research question and design, which I call previous experiment.[14] The insights from the previous experiment informed the current experiment in several ways. First, I reframed the research question so that it has more relevant real-world implications. Second, I changed the design so that subjects' rank is exogenous. Third, I used IQ test to induce dictators' stronger bias. Fourth, I introduced actual photo and first name to naturally inform subjects' gender. Fifth, I changed the generosity elicitation task to dictator game to make the decision problem simple. Sixth, I let passive subjects (recipients in the current experiment) to leave the laboratory before generosity elicitation so that active subjects (dictators in the current experiment) could choose more self-interested decision. Seventh, I dropped most heterogeneity analyses to focus on the main research question. The main differences of the previous experiment are listed in appendix B. The working paper, codes, data, and instructions of the previous experiment are available upon request.[15]

4. Questions and testing

I use 5% (two-tailed) as my significance level. I use OLS with cluster-robust covariance matrix estimator with Pustejovsky and Tipton (2018)'s small cluster bias adjustment. Standard errors are clustered at the dictator level. Since the power simulation in appendix A suggests that the analysis may be underpowered, I report type M error ratio and type S error probability for all statistically significant results. I use R's plm and vcovCR to estimate the model and retrodesign for calculating type M ratio and type S error probability. I exclude

[12] I ask (i) what they think the study was about, (ii) whether there was anything unclear or confusing about the study, (iii) whether they have any other comments.
[13] They can choose how they will be paid from (i) Amazon Italia digital gift card or (ii) bank transfer.
[14] Also, because I was initially planning to use GRE quantitative test for the task, I run a pilot session with 14 subjects to choose 10 out of 20 GRE questions of appropriate level. Pilot consisted of photo taking, 20 GRE questions, and post-questionnaire asked about subjects' gender, age, field of study, degree program, years in the degree program, how hard they found each of the 20 GRE questions, whether they prefer to take their photo by themselves or us to take their photo, whether they could find their field of study and degree program from the choices, and free comments.
[15] In the previous experiment the task performance was not displayed on the subjects' computer screen due to computer program error, and subjects could not see the payoff consequence of their decision in the generosity elicitation.

observations in which recipient has non-Italian sounding first name.[16] Also, I exclude observations in which dictators or evaluators answer that they knew the recipients "very well" at least once.

4.1. Main tests

Question 1: *Are men and women equally or more generous to a woman who is less smart than them?*

Question 2: *Are men and women less generous to a woman who is smarter than them?*

Question 3: *Is the difference in men's and women's generosity to a person who is smarter than them and to another person who is less smart than them more negative when these persons are women?*

To answer these questions, I estimate the following equation:

$$\text{Giving in DG}_{\text{i,j}} = \beta_0 + \beta_1 Lower_{i,j} + \beta_2 Female_j + \beta_3 Lower_{i,j} * Female_j + X'_{i,j}\gamma + \sum_{l=1}^{9} \theta^l d_i^l + \epsilon_{i,j} \tag{1}$$

Where

- $\text{Giving in DG}_{\text{i,j}} \in \{0,0.5, \dots, Endowment\}$ is the amount dictator i allocates to recipient j. Endowment is either 5 or 7.
- $Lower_{i,j} \in \{0,1\}$ is an indicator variable equals 1 if dictator i's rank is lower than recipient j, 0 otherwise.
- $Female_j \in \{0,1\}$ is an indicator variable equals 1 when recipient j is female, and 0 when recipient j is male.
- $d_i^l \in \{0,1\}$ is an indicator variable equals 1 if dictator i solved l=0,1,…,9 questions correctly, and 0 otherwise.
- $X_{i,j}$ is a vector of dictator and recipient characteristics as well as sets of fixed effects. The main specification is to include full control and I rely on statistical significance from that specification. However, for robustness check, I also estimate specification only with number of questions solved fixed effects, with fixed effects only, and with fixed effects and dictator controls to see if the direction and magnitude of the estimates are similar across different specifications.
- $\epsilon_{i,j}$ is error term that can be autocorrelated within dictator i and heteroskedastic across dictators.

In $X_{i,j}$, I include the following variables:

<u>Dictator characteristics:</u>

- $Age_i \in \mathbb{N}$ is dictator i's age.
- $Male_i \in \{0,1\}$ is an indicator variable equals 1 if dictator i is male, 0 otherwise.

[16] Ideally, I should also exclude recipients with non-Italian face, but it is difficult to objectively distinguish Italian face and non-Italian face. On the other hand, distinguishing Italian and non-Italian first name is pretty straight forward.

- $FromEmiliaRomagna_i \in \{0,1\}$ is an indicator variable equals 1 if dictator i is from Emilia-Romagna region, 0 otherwise.
- $SocialSciences_i \in \{0,1\}$ is an indicator variable equals 1 if dictator i's study program is social sciences, 0 otherwise.
- $STEM_i \in \{0,1\}$ is an indicator variable equals 1 if dictator i's study program is natural sciences/mathematics, engineering, or medicine, 0 otherwise.
- $PostBachelor_i \in \{0,1\}$ is an indicator variable equals 1 if dictator i's degree program is either master/post-bachelor, in the 4th year or higher of bachelor-master combined program, or doctor, 0 otherwise.
- $OverConfidence_i \in \{-1,0,1\}$ is degree of dictator i's overconfidence. It is equal to -1 if dictator i's guess on the number of questions solved is lower than the actual number, 0 if the dictator i's guess is equal to the actual number, and 1 if the dictator i's guess is higher than the actual number.

<u>*Recipient characteristics:*</u>

- $Age_j \in \mathbb{N}$ is recipient j's age.
- $FromEmiliaRomagna_j \in \{0,1\}$ is an indicator variable equals 1 if recipient j is from Emilia-Romagna region, 0 otherwise.

<u>*Fixed effects:*</u>

- $\sum_{l=2}^{3} r_{i,j}^{l}$, where $r_{i,j}^{l} \in \{0,1\}$ is an indicator variable equals 1 if the round dictator i is paired with recipient j is l=1,2,3, 0 otherwise.
- $\sum_{l=2}^{3} q_{i,j}^{l}$, where $q_{i,j}^{l} \in \{0,1\}$ is an indicator variable showing proximity between dictator i and recipient j, and equals 1 if does not know at all (l=1), has seen before (l=2), knows but not very well (l=3).

Note that this is conceptually a 2x2 design:

<u>Table 1: Graphical representation of equation 1</u>

		Recipient	
		Female	Male
Dictator	Ranked higher than recipient	A $(\approx \beta_0 + \beta_2)$	B $(\approx \beta_0)$
	Ranked lower than recipient	C $(\approx \beta_0 + \beta_1 + \beta_2 + \beta_3)$	D $(\approx \beta_0 + \beta_1)$

Note: Each cell represents generosity of corresponding dictator (in rows) to recipient (in columns).

Where each cell represents generosity – allocation to the recipient in dictator game – by dictators (in rows) to recipients (in columns). Assuming that dictator characteristics who are ranked higher than and lower than recipient is balanced, $\beta_2 \approx A - B$, $\beta_2 + \beta_3 \approx C - D$, $\beta_3 \approx C - A - (D - B)$ conditional on the dictators' number of questions solved. Thus,

- β_2 answers Question 1. In particular, $\beta_2 \geq 0$ indicates that the answer is yes.
- $\beta_2 + \beta_3$ answers Question 2. In particular, $\beta_2 + \beta_3 < 0$ indicates that the answer is yes.
- β_3 answers Question 3. In particular, $\beta_3 < 0$ indicates that the answer is yes.

4.2. Alternative explanation

I test the following alternative explanation for Question 1-3:

Explanation 1: *Dictators believe that recipients of one gender whose rank is higher/lower than them solved more/less questions than recipients of other gender with equivalent rank.*

To test this possible explanation, I estimate the following equation:

$$\widehat{QsSolved}_{k,j} = \beta_0 + \beta_1 Lower_{k,j} + \beta_2 Female_j + \beta_3 Lower_{k,j} * Female_j + X'_{k,j}\gamma + \sum_{l=1}^{9} \theta^l d_k^l + \epsilon_{k,j} \tag{2}$$

Where

- $\widehat{QsSolved}_{k,j} \in \{0,1,\dots,9\}$ is evaluator k's guess on recipient j's number of questions solved.
- $Lower_{k,j} \in \{0,1\}$ is an indicator variable equals 1 if evaluator k's rank is lower than recipient j, and 0 otherwise.
- $d_k^l \in \{0,1\}$ is an indicator variable equals 1 if evaluator k solved l=0,1,…,9 questions correctly, and 0 otherwise.
- $X_{k,j}$ is a vector of evaluator and recipient characteristics as well as sets of fixed effects. I estimate the model with full control (main specification), with number of questions solved fixed effects only, with fixed effects only, with fixed effects and evaluator controls. Evaluator characteristics are the same as dictator characteristics used in the main test.
- $\epsilon_{k,j}$ is error term that can be autocorrelated within evaluator k and heteroskedastic across evaluators.

Other variables are as defined above.

If the explanation is plausible, I should observe:

- $\beta_2 \geq 0$ for Question 1.
- $\beta_2 + \beta_3 < 0$ for Question 2.
- $\beta_3 < 0$ for Question 3.

4.3. Heterogeneity analyses

Dictators' gender

I test whether gender of the dictator matters for Question 1-3. I do this by separately estimating equation 1 for male and female dictators. I exclude from $X_{i,j}$ $STEM_i$, $Male_i$, and $FromEmiliaRomagna_i$.

Dictators' overconfidence

I test whether overconfidence of the dictator matters for Question 1-3. I do this by separately estimating equation 1 for dictators with $OverConfidence_i = -1{,}0$ and with $OverConfidence_i = 1$. I exclude from $X_{i,j}$ $OverConfidence_i$.

Rank differences

I test whether rank difference matters for dictators' allocation. For example, dictators who face a recipient whose rank is higher by 1 and by 2 may choose different allocation. To see this, I estimate the following equation:

$$\text{Giving in DG}_{\text{i,j}} = \beta_0 + \sum_{l\in\{-3,-2,1,2,3\}} \beta_1^l Lower_{i,j}^l + \beta_2 Female_j + \sum_{l\in\{-3,-2,1,2,3\}} \beta_3^l Lower_{i,j}^l * Female_j + X_{i,j}'\gamma + \sum_{l=1}^{9} \theta^l d_i^l + \epsilon_{i,j} \tag{3}$$

Where

- $Lower_{i,j}^l \in \{0,1\}$ is an indicator variable equals 1 if dictator i's rank is l=-3,-2,-1,1,2,3 points lower than recipient j's rank, and 0 otherwise. 3 includes 4 and 5 and -3 includes -4 and -5.

Other variables are as defined above.

4.4. Robustness check

Excluding dictators/evaluators who always face recipients with higher or lower rank

Dictators/evaluators with rank 1 will never face recipients with higher rank. Similarly, dictators/evaluators with rank 6 will never face recipients with lower rank. This might introduce potential unbalance across cells in table 1. To address this possibility, I estimate equations 1-3 without dictators/evaluators whose rank is either 1 or 6 to show robustness of my parameter estimates.

Further excluding observations in which dictators/evaluators know the recipients

In the main specifications, I excluded observations in which dictators/evaluators answer that they knew the recipients "very well." To check the robustness of my analysis, I also estimate equations 1-3 without observations in which dictators/evaluators answer that they knew the recipients "not very well." I also estimate equation 1-3 without observations in which dictators/evaluators answer that they "saw before".

4.5. Complementary experiments

Once I finish this experiment and analyze the data to answer the above questions, I conduct a complementary experiment to investigate the following points as well as to show robustness of my results after controlling for attractiveness. Their details will be written in separate pre-analysis plans.

Appearance

As Isaksson (2018, p.10, footnote 6; pre-analysis plan p. 9), I examine how a person's appearance matters in the dictators' generosity. Specifically, I examine how recipients looking feminine, looking masculine, looking more attractive, looking more approachable, and looking selfish matter in the dictators' generosity. I will recruit students of the University of Bologna satisfying the same requirements outlined in the procedure section but have not participated in this experiment, and let them rate the photo in terms of (i) feminineness, (ii) masculineness, (iii) attractiveness from the same gender's view, (iv) attractiveness from the different gender's view, (v) approachability, and (vi) appearing selfish.

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Bilker, Warren B., John A. Hansen, Colleen M. Brensinger, Jan Richard, Raquel E. Gur, and Ruben C. Gur. 2012. "Development of Abbreviated Nine-Item Forms of the Raven's Standard Progressive Matrices Test." *Assessment* 19 (3): 354–69. https://doi.org/10.1177/1073191112446655.

Carpenter, Thomas, Ruth Pogacar, Chris Pullig, Michal Kouril, Stephen J Aguilar, Jordan P. LaBouff, Naomi Isenberg, and Alek Chakroff. 2018. "Survey-Based Implicit Association Tests: A Methodological and Empirical Analysis." Working Paper. https://osf.io/hgy3z.

Charness, Gary, and Uri Gneezy. 2008. "What's in a Name? Anonymity and Social Distance in Dictator and Ultimatum Games." *Journal of Economic Behavior & Organization* 68 (1): 29–35. https://doi.org/10.1016/j.jebo.2008.03.001.

Chen, Daniel L., Martin Schonger, and Chris Wickens. 2016. "OTree—An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9 (March): 88–97. https://doi.org/10.1016/j.jbef.2015.12.001.

Coppock, Alexander. 2019. "Ordered Probit with Control." DeclareDesign Community Discussion. http://discuss.declaredesign.org/t/ordered-probit-with-control/68.

Gelman, Andrew, and John Carlin. 2014. "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science* 9 (6): 641–51. https://doi.org/10.1177/1745691614551642.

Greenwald, Anthony G., Brian A. Nosek, and Mahzarin R. Banaji. 2003. "Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm." *Journal of Personality and Social Psychology* 85 (2): 197–216. https://doi.org/10.1037/h0087889.

Greiner, Ben. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association* 1 (1): 114–25. https://doi.org/10.1007/s40881-015-0004-4.

Isaksson, Siri. 2018. "It Takes Two: Gender Differences in Group Work." Working Paper.

Lu, Jiannan, Yixuan Qiu, and Alex Deng. 2019. "A Note on Type S/M Errors in Hypothesis Testing." *British Journal of Mathematical and Statistical Psychology* 72 (1): 1–17. https://doi.org/10.1111/bmsp.12132.

Maggian, Valeria, and Natalia Montinari. 2017. "The Spillover Effects of Gender Quotas on Dishonesty." *Economics Letters* 159 (October): 33–36. https://doi.org/10.1016/j.econlet.2017.06.045.

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. https://doi.org/10.1162/qjec.122.3.1067.

Pustejovsky, James E., and Elizabeth Tipton. 2018. "Small-Sample Methods for Cluster-Robust Variance Estimation and Hypothesis Testing in Fixed Effects Models." *Journal of Business & Economic Statistics* 36 (4): 672–83. https://doi.org/10.1080/07350015.2016.1247004.

Appendix A: Power simulation

I simulate the power and bias I would get for answering Question 3 with 200 subjects as well as 300 and 400 subjects (N=100, N=150, N=200, respectively, because half of the subjects are recipients) with type I error probability kept at 5% (two-tailed). Power is defined as 1 minus type II error probability. While my experiment has earnings stage (letter-number typing task), dictators see the recipients' photos and first name which would reduce social distance and increase allocation. Thus, I refer to Charness and Gneezy (2008)'s single blind dictator experiment to determine parameters: I use their treatment effect (difference in average allocation between the name and no name treatment) and use average allocation in the name treatment to be the allocation to the male recipients whose rank is lower than the dictators' and the average allocation in the no name treatment to be the allocation to the male recipients whose rank is higher than the dictators'. The R code for simulation is available on the OSF page (PowerSim.R).

The allocation to the female recipients are determined somewhat arbitrarily: the allocation for female recipients whose rank is lower than the dictators' is 1/2 of the treatment effect higher than the allocation for male recipients with equivalent rank and the allocation for female recipients whose rank is higher than the dictators' is 1/4 of the treatment effect lower than the allocation for male recipients with equivalent rank. These are because the effect of gender bias must be lower than the effect of rank differences especially in negative domain.[17]

I consider the baseline estimation without covariates: while covariates increase power, modelling the effect of potential unbalance across cells is complicated. Instead, I assume unobserved individual heterogeneity does not predict outcome. I also assume that the endowment is 7€ for all rounds for simplicity.

I assume the following data generating process. I add i to subscript for female recipient dummy to make variable notations easier to interpret (so j here simply indicates round):

$$Y^*_{i,j} = \left(b_0 + \tilde{b}^i_0\right) + \left(b_1 + \tilde{b}^i_1\right)Lower_{i,j} + \left(b_2 + \tilde{b}^i_2\right)Female_{i,j} + \left(b_3 + \tilde{b}^i_3\right)Lower_{i,j} * Female_{i,j} + \#Questions'_i\gamma + e_{i,j} \tag{A1}$$

with $i = 1, \ldots, N$ and $j = 1,2,3$. Each variable is defined as follows:

- $Lower_{i,j} \sim^{iid} Bernoulli(0.5)$
- $Female_{i,j} \sim^{iid} Bernoulli(0.5)$
- $\#Questions_i = [0, \ldots ,1, \ldots ,0]'_{10\times1}$ where location of 1 is determined by $v_i \sim^{iid} Bin(10,0.5)$
- $e_{i,j} \sim^{iid} N(0, \sigma^2)$
- $\tilde{b}^i_0 \sim^{iid} N(0, \sigma^2_0)$
- $\tilde{b}^i_1 \sim^{iid} N(0, \sigma^2_1)$

[17] I calculate these values as well as standard deviation from the data point of name treatment dictator game in table 1 of Charness and Gneezy (2008). Since their total endowment is 100, I divide the averages and standard deviation by 100 and multiply them by 7 to make them consistent with my endowment.

- $\tilde{b}_2^i \sim^{iid} N(0, \sigma_2^2)$
- $\tilde{b}_3^i \sim^{iid} N(0, \sigma_3^2)$

And each parameter is set as follows:

- $b_0 = 1.90$ so that male recipients whose rank is lower than the dictators receive 1.90€ on average.
- $b_1 = -0.62$ so that male recipients whose rank is higher than the dictators receive 1.28€ on average.
- $b_2 = 0.32$ so that female recipients whose rank is lower than the dictators receive 2.22€ on average.
- $b_3 = -0.47$ so that female recipients whose rank is higher than the dictators receive 1.13€ on average.
- $\gamma = [u]_{10\times 1}$ where $u \sim^{iid} Unif(-1,1)$ so that dictators' giving who solved different number of questions varies by 2€.
- $\sigma = 1.56$: this is sample standard deviation of Charness and Gneezy (2008)'s name treatment and means that 95% of dictators' allocation for male recipients whose rank is lower than the dictators' fall somewhere between 0€ to 4.96€. Since Charness and Gneezy (2008) have no covariates whereas I include covariates, I also consider $\sigma = 1.56/2$ and $\sigma = 1.56/3$.
- $\sigma_0 = \frac{b_0}{4}, \sigma_1 = \frac{|b_1|}{4}, \sigma_2 = \frac{b_2}{4}, \sigma_3 = \frac{|b_3|}{4}$: these are somewhat arbitrary but I set them so that almost all dictators react in a same direction.

The allocations obtained from these parameterizations are summarized below:

Table A1: Assumed allocations and their approximate 95% range

		Recipient: Female	Recipient: Male
Dictator	Ranked higher than recipient	2.22€ [-0.90€, 5.34€] (32%) [-13%, 76%]	1.90€ [-1.22€, 5.02€] (27%) [-28%, 61%]
	Ranked lower than recipient	1.13€ [-1.99€, 4.25€] (16%) [-28%, 61%]	1.28€ [-1.84€, 4.25€] (18%) [-26%, 63%]

Note: Each cell represents assumed allocations of corresponding dictator (in rows) to recipient (in columns). Approximate 95% range (+-2 standard deviation) is in bracket. Negative values are truncated in the realized values. The allocations as a percent of endowment, 7€, are shown in parentheses.

From the above data generating process, I determine the realized value of $Y_{i,j}^*$ as follows:

$$Y_{i,j} = \begin{cases} 0\ if\ Y_{i,j}^* \leq 0.25 \\ 0.5\ if\ 0.25 < Y_{i,j}^* \leq 0.75 \\ 1\ if\ 0.75 < Y_{i,j}^* \leq 1.25 \\ \vdots \\ 6.5\ if\ 6.25 < Y_{i,j}^* \leq 6.75 \\ 7\ if\ 6.75 < Y_{i,j}^* \end{cases}$$

(A2)

I draw 1000 independent random samples from equation A1, and for each sample, I estimate the following equation via OLS with cluster-robust covariance matrix estimator with Pustejovsky and Tipton (2018)'s small cluster adjustment and run two-tailed t-test for $\beta_3 = 0\ vs.\beta_3 \neq 0$:

$$Y_{i,j} = \beta_0 + \beta_1 Lower_{i,j} + \beta_2 Female_{i,j} + \beta_3 Lower_{i,j} * Female_{i,j} + \sum_{l=1}^{9} \theta^l d_i^l + \epsilon_{i,j}$$

(A3)

I then calculate power as the number of times the t-test rejects $\beta_3 = 0$ divided by the number of independent sample draws (1000 in my case):

$$Power(N, \sigma) = \frac{\#\ Rejections(N, \sigma)}{\#\ Draws}$$

(A4)

I also calculate average β_3 estimate:

$$\bar{\hat{\beta}}_3(N, \sigma) = \frac{1}{\#\ Draws} \sum_{l=1}^{\#\ Draws} \hat{\beta}_{3,l}\,(N, \sigma)$$

(A5)

As well as average standard error of β_3 estimate:

$$\bar{\hat{\sigma}}_3(N, \sigma) = \frac{1}{\#\ Draws} \sum_{l=1}^{\#\ Draws} \hat{\sigma}_{3,l}\,(N, \sigma)$$

(A6)

The simulation results are reported in table A2 (power), table A3 (average β_3 estimate as well as its average standard error).

Table A2: Estimated power

	$\sigma = 1.56$	$\sigma = 1.56/2$	$\sigma = 1.56/3$
$N = 100$	20.9%	55.8%	76.4%
$N = 150$	33.1%	75.6%	92.3%
$N = 200$	42.7%	85.9%	97.1%

Note: Each cell presents estimated power for a given sample size (in row) and standard deviation (in column). Power is estimated via Monte Carlo simulation with 1000 draws.

First, looking at table A2, we see that it is hard to achieve 80% power with 200 subjects (N=100) unless the standard deviation is very small. A likely scenario with controls is $\sigma = 1.56/2$, at which we can achieve 80% power with 400 subjects (N=200). However, it is still optimistic because the data generating process assumptions can be wrong. Thus, I report type M error ratio and type S error probability for all statistically significant estimates because with underpowered estimation statistically significant estimates are likely to be results of picking up extreme values to because the standard error is very large – they tend to be highly exaggerated relative to their actual values (which we cannot know) and their sign can even be reversed (Gelman and Carlin 2014). To calculate type M error ratio and type S error probability, I use Lu, Qiu, and Deng (2019)'s closed-form expressions with the hypothesized actual value -0.47.

Table A3: Estimated average $\hat{\beta}_3$ and its standard error

	$\sigma = 1.56$	$\sigma = 1.56/2$	$\sigma = 1.56/3$
$N = 100$	-0.40 (0.329)	-0.43 (0.204)	-0.45 (0.162)
$N = 150$	-0.40 (0.266)	-0.44 (0.167)	-0.45 (0.133)
$N = 200$	-0.40 (0.230)	-0.44 (0.143)	-0.45 (0.115)

Note: Each cell presents estimated average $\hat{\beta}_3$ as well as its standard error (in parenthesis) for a given sample size (in row) and standard deviation (in column). Average $\hat{\beta}_3$ as well as its standard error are estimated via Monte Carlo simulation with 1000 draws. The true b_3 in the data generating process is -0.47.

Second, looking at table A3, while the underlying process is ordered probit and there is truncation from below as well as from above (but no selection), OLS approximates the underlying true value pretty well, as shown by Angrist and Pischke (2009) for truncated data and by Coppock (2019) for ordered data. In particular, the OLS estimate is biased downward, presumably because the data is truncated mostly from below, which works against finding my expected results. In addition, it is difficult to address small cluster bias adjustment or serial correlation in the covariance matrix with Tobit or ordered probit. Thus, I use OLS.

Appendix B: Main differences from the previous experiment

The main differences of the experiment run in June 2019 are the followings:

- The research questions of the previous experiment were (i) whether one's performance in a male task matters in her or his perception of others' gender norm incongruity (*both* women and men) and (ii) whether gender matters in her or his perception of gender norm incongruity.
- The previous experiment used a version of Niederle and Vesterlund (2007)'s addition task – used in Maggian and Montinari (2017) – as the task to rank subjects.
- The previous experiment assigned subjects fake first names that correspond to their gender with which they identified each other's gender.
- The previous experiment split subjects into managers and workers (to frame the setting as a real world workplace decision), matched up managers and workers, and asked managers whether to appoint the matched worker as their collaborators or not (appointing the matched worker was always payoff maximizing in expectation but increased payoff of the matched worker more).
- The rank in the previous experiment was not randomly assigned.
- The passive subjects (workers in the previous experiment and recipients in the current experiment) stayed the laboratory while the active subjects were making decisions.
- The previous experiment asked subjects to work on form A of the 9-item Raven's progressive matrices test (Bilker et al. 2012) and gender-career version of the Implicit Association Test to investigate heterogeneity with respect to intelligence and implicit gender bias. The Implicit Association Test was conducted on Qualtrics (Carpenter et al. 2018) and the raw data were processed according to Greenwald, Nosek, and Banaji (2003)'s method.

Appendix C: Classification of study program

The followings are classification of study program I use in the analysis. Each subcategory is from the University of Bologna's list of degree program website (https://www.unibo.it/en/teaching/degree-programmes) from which subjects choose their field and enter it in the questionnaire.

- Humanities: Languages and Literatures, Interpreting and Translation; Education; Law; Humanities
- Social Sciences: Economics and Management; Psychology; Political Sciences; Sociology
- Natural Sciences/Mathematics: Sciences; Statistics
- Medicine: Pharmacy and Biotechnology; Medicine; Veterinary Medicine
- Engineering: Engineering and Architecture; Agricultural and Food Sciences; Sport Sciences

If they enroll in a specialized or professional program, I ask them to choose the closest study program. If they are enrolled in post-bachelor vocational program, I ask them to choose the study program of their bachelor's degree. If they are exchange students, I ask them to choose the study program closest to the one in their home university.