Do competent women receive unfavorable treatment?

Yuki Takahashi*

Latest Version

December 8, 2020

Abstract

Do competent women receive unfavorable treatment than equally competent men? I study this question in a laboratory experiment where unfavorable treatment has material consequences. I find that neither men nor women treat competent women less favorably; if anything, both men and women treat competent women slightly more favorably than equally competent men. The findings provide a piece of evidence that competent women may not necessarily receive unfavorable treatment in settings with material consequences, which may shed new light on hiring and promotion practices in labor markets.

JEL Classification: C91, D91, J16, M51
Keywords: competence, gender bias, labor markets, laboratory experiment

*Department of Economics, University of Bologna. Email: yuki.takahashi2@unibo.it. I am grateful to Maria Bigoni, Natalia Montinari, and Siri Isaksson whose feedback was essential for this project. I am also grateful to participants of the experiment for their participation and cooperation. Ingvild Almås, Laura Anderlucci, Tiziano Arduini, Francesca Barigozzi, Teodora Boneva, Enrico Cantoni, Giovanna d'Adda, Chiara Natalie Focacci, Margherita Fort, Catalina Franco, Astrid Kunze, Fabio Landini, Pascal Langenbach, Annalisa Loviglio, Valeria Maggian, Joshua Miller, Paola Profeta, Eugenio Proto, Tommaso Sonno, Sigrid Suetens, 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. The pre-analysis plan is available at the OSF registry: https://osf.io/ypsmx. The experimental instructions are available in the online appendix: https://drive.google.com/file/d/1uSpKBWSOMaMzK5fQQzxJPKOzhLt5Tmu2.

1 Introduction

A literature argues that people consider competent women as less likable than equally competent men (Heilman 2001; Rudman and Phelan 2008). This is also a view shared by several top female corporate executives.[1] However, it is unclear whether being less likable has practical implications; that is, whether competent women receive unfavorable treatment in decisions such as hiring and promotion. Indeed, this question has been explored mostly by means of questionnaires and hypothetical decisions (Heilman et al. 2004; Phelan, Moss-Racusin, and Rudman 2008; Rudman and Fairchild 2004; Rudman 1998; Rudman et al. 2012).

Evidence from decisions with material consequences mainly comes from audit studies and is mixed: while Quadlin (2018) finds unfavorable treatment, Ceci and Williams (2015) and Williams and Ceci (2015) do not. One possible reason for this mixed evidence is employers' wrong prior belief about competent women's personality which tends to be negative as evidenced by the literature: because employers have to work with their employees for a long period of time, they want to hire people whom they are comfortable to work with. However, their prior must be updated once the employers see the actual job applicants in the interview. Also, in promotion decisions, employers or managers know a potential candidate very well and their prior belief must be irrelevant.

In this paper, I tackle this question by means of a controlled laboratory experiment. I use dictator game allocation as a measure of favorable and unfavorable treatment with clear material consequences and exogenously vary the recipient's gender and competence. I measure competence by an IQ test, an attribute people care most about (Eil and Rao 2011; Zimmermann 2020).[2]

In the experiment, participants first work on an incentivized IQ test. After the test, participants are randomly assigned to a group of six and receive a ranking of their IQ within their group. Once they answer the comprehension questions about their IQ rank, 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 recipients, observing the recipients' facial photos and first names – both of which convey information about gender – and the IQ ranks.[3]

Using dictator IQ fixed effects and exploiting random grouping of participants to address the endogeneity of participants' IQ and recipients' gender, I do not find a significant difference between dictators' allocation to competent women and to competent men; if anything, dictators allocate slightly more to competent women. The point estimate of the difference is positive and statistically indistinguishable from 0. The lower bound of the difference is -3.7% of the dictator endowment, which is quantitatively much smaller (2.4-3.1 times smaller) in absolute

1. In her book *Lean In: Women, Work, and the Will to Lead*, the Facebook's Chief Operating Officer Sheryl Sandberg expresses her view as follows: "If a woman is competent, she does not seem nice enough. If a woman seems really nice, she is considered more nice than competent. Since people want to hire and promote those who are both competent *and* nice, this crates a huge stumbling block for women" (Sandberg 2013).

2. The experimental design, the hypotheses, and the empirical strategy are pre-registered at the OSF registry: `https://osf.io/ypsmx`. However, there are a number of changes to the pre-analysis plan discussed in appendix A.

3. The use of photos follows recent literature and allows the dictators to infer the gender of the recipients in a natural way as they would do in their daily life (Babcock et al. 2017; Coffman 2014; Isaksson 2018), but I address the possibility that recipients' gender-specific characteristics (e.g. women may smile more often in a photo) affect dictators' allocation.

value than the effect sizes of the other dictator game experiments studying the role of social distance. These results hold across the whole distribution of dictators' allocation and even when I separately examine male and female dictators' allocation. All results control for recipients' gender-specific preference for allocation (e.g. women smile more often than men) using the difference in allocation to incompetent women and men as a control group.[4]

Several alternative explanations are inconsistent with the results; most importantly, the results are not due to experimental manipulation failure, ex-post randomization failure, wrong identification assumptions, or lack of statistical power. These findings suggest that competent women do not receive unfavorable treatment in decisions involving material consequences such as hiring and promotion.

This paper mainly relates to two strands of literature. The first focuses on the tradeoff women face between being competent and being likable. The literature finds that people perceive female leaders (Heilman, Block, and Martell 1995; Heilman and Okimoto 2007; Rudman and Kilianski 2000) and competent women (Heilman et al. 2004; Rudman 1998) negatively. It also finds that people evaluate competent women negatively, but these results are obtained in set-ups without real consequences (Phelan, Moss-Racusin, and Rudman 2008; Rudman and Fairchild 2004; Rudman et al. 2012). However, the studies about evaluations towards competent women with real consequences find mixed evidence: while Quadlin (2018) finds top-performing female college students less favorable treatment in hiring than equally qualified male students, Ceci and Williams (2015) and Williams and Ceci (2015) find qualified female applicants for assistant professor positions receive equal or more favorable treatment than equally qualified male applicants. My results suggest that the employers' prior belief about competent women may be driving these mixed findings.

When the consequence of their evaluation is not immediately clear, people seem to evaluate women in traditionally male occupations more critically: Boring (2017) and Mengel, Sauermann, and Zölitz (2019) find that female university instructors receive lower student evaluation. There is also evidence that female economists' work are undervalued (Koffi 2019; Sarsons et al. 2020) and female university faculty are less likely to get promotion (De Paola, Ponzo, and Scoppa 2018). However, these critical evaluations may simply reflect the lack of women in these occupations and thus people do not have enough prior information about women's competence, rather than taste-based discrimination. Sarsons (2019) finds that female surgeons receive a more negative evaluation for their failure and Ditonto (2017) finds that voters care more about female politicians' competence than male politicians' competence. Also, Bohren, Imas, and Rosenberg (2019) find that while women initially receive lower credits than men in their contributions to an online mathematics discussion forum, they receive higher credits than men after they accumulate enough positive evaluations. My findings are compatible with the explanation that people do not have enough prior information about women's competence, and they give fair evaluations to women once they show they are competent.

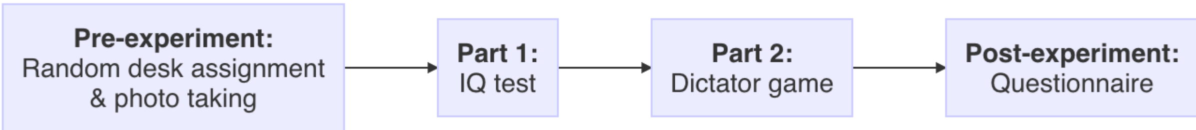
4. While dictators only see the recipients' IQ relative to theirs and thus the competence measure is relative to theirs, dictators do not see their IQ at the time they play dictator games. Indeed, in the real world, we do not have an absolute measure of other people's competence but evaluate relative to some benchmark. Nevertheless, I provide evidence that relative and absolute competence distinction does not matter for my results.

2 Experiment

2.1 Design and procedure

The experiment consists of two parts as shown in figure 1; instructions for each part are only delivered at the end of the previous part. Participants earn a participation fee of 2.5€ for their participation. Experimental instructions are available in the online appendix.

FIGURE 1: OVERVIEW OF THE EXPERIMENT



Notes: This figure shows an overview of the experiment discussed in detail in section 2.1.

Pre-experiment: Random desk assignment & photo taking

After registration at the laboratory entrance, participants are randomly assigned to a desk. Before the start of part 1, participants take their facial photos at a photo booth and enter their first name on their computer. After that, we 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 test questions for 9 minutes. I use Bilker et al. (2012)'s form A 9-item Raven test which predicts one's IQ measured with the full-length Raven test with more than 90% accuracy. Participants receive 0.5€ for each correct answer. They receive information about how many IQ test questions they have solved correctly only at the end of the experiment. I use IQ as the measure of competence because previous studies find it is an attribute people care most about (Eil and Rao 2011; Zimmermann 2020).

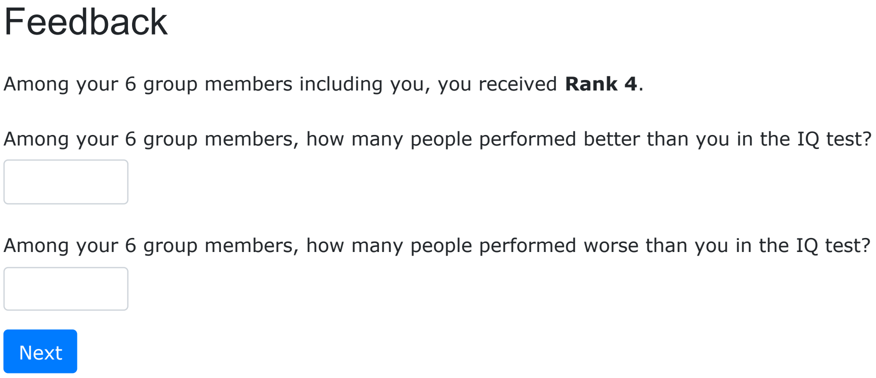
After the IQ test, participants make an incentivized guess on the number of IQ test questions they have solved correctly: they receive 0.5€ if their guess is correct. The answer to this question measures their over-confidence level. They receive feedback on this guess only at the end of the experiment.

Following Eil and Rao (2011), six participants are randomly grouped, and they are informed of the ranking of their IQ relative to other group members. Ties are broken randomly. They then have to answer a set of comprehension questions as shown in figure 2 in order to proceed to the next part.

Part 2: Dictator game

In part 2, three participants in each group are randomly chosen to become dictators and the other three participants become recipients. Dictators are paired with the three recipients in their group one by one in a random order, receive an endowment, and play a dictator game. When they play the dictator game, dictators observe the recipients' facial photo and first name and IQ rank. The

Figure 2: IQ rank assignment and the comprehension questions



Notes: This figure shows an example of the IQ rank assignment and the comprehension questions. In this example, the participant was ranked 4th from the top within a group of 6 participants. Thus, the answer to the first question is 3 (three participants performed better in the IQ test) and the second question is 2 (two participants performed worse in the IQ test).

use of photo allows me to convey information about gender of other participants in a natural way as in the recent literature (Babcock et al. 2017; Coffman 2014; Isaksson 2018).[5] Dictators are also told that their allocation decisions are anonymous except for the experimenters: they are told that their allocation is paid to the recipients 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, as shown in figure 3. I use a cursor to make it more cognitively demanding to figure out fair allocation, which is shown to increase more self-interested decisions (Exley and Kessler 2019). I also vary the endowment across rounds to make each dictator game less repetitive: 7€ for 1st and 3rd rounds, 5€ for 2nd round. At the end of the experiment, one out of three allocations is randomly chosen for each participant as earnings for this part.[6]

I also collect an indirect measure of dictators' beliefs on how many IQ test questions the paired recipients have solved correctly. To prevent the belief elicitation to affect or be affected by the dictator game, I exploit the random assignment of participants to dictators and recipients (derived from the random desk assignment) and use recipients' beliefs as a proxy for dictators' beliefs. Specifically, while dictators are playing the dictator game, recipients are paired with the other two recipients in the same group one by one in random order and make incentivized guesses on how many IQ test questions they have solved correctly, observing the other two recipients' facial photo, first name, and IQ rank. Each correct guess gives them 0.5€.

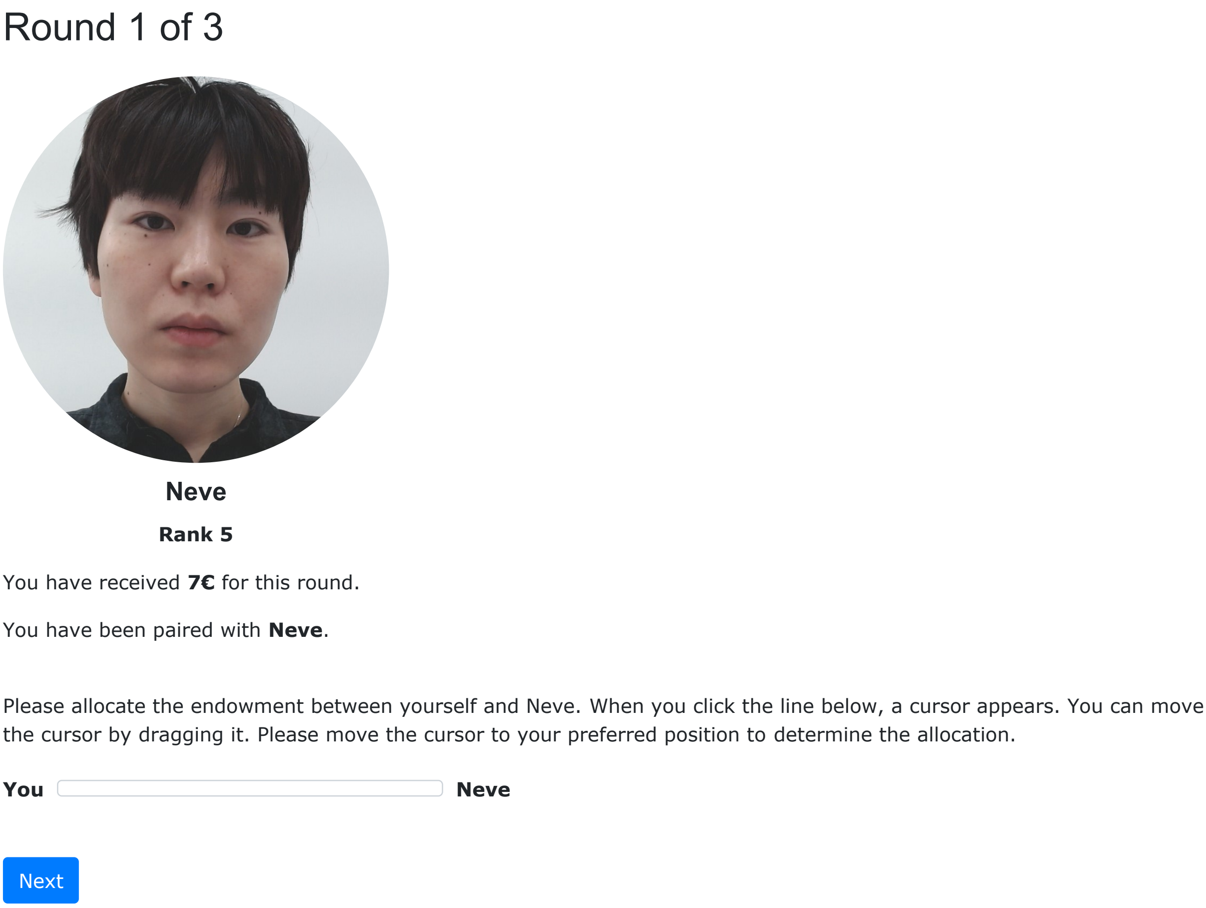
Post-experiment: Questionnaire

After the dictator game and guessing are over, participants are told their earnings from the IQ test, dictator game, and the guesses. Before receiving their earnings, participants answer a short

5. To address the non-anonymity of showing facial photo and first name, I ask participants how well they know the paired participants on a scale of 4 (did not know at all, saw before, knew but not very well, knew very well). I ask this question twice to make sure they do not answer randomly: right after the three dictator games or two guesses and in the post-experimental questionnaire.

6. For each dictator for each round, one of the three recipients in the same group is randomly chosen with replacement and the dictator allocates the endowment between themselves and the recipient. Thus, it is possible that two dictators play dictator game with the same recipient in the same round. At the end of the dictator games, each participant has three allocations, and one of which is randomly chosen for payment.

Figure 3: Dictator's allocation screen



Notes: This figure shows an example of a dictator's allocation screen. In this example, the dictator is playing the first round and paired with a recipient whose first name is Neve with IQ rank 5.

questionnaire about their demographics that are used for balance tests and robustness checks.

2.2 Implementation

The experiment was computerized and 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 of the University of Bologna via ORSEE (Greiner 2015) who (i) were born in Italy, (ii) available to participate in English experiments, and (iii) had not participated in gender-related experiments in the past (as far as I could check). The number of participants was based on the power simulation in the pre-analysis plan to achieve 80% power.[7]

The average length of a session 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 (which I asked for recipients only). I ran 24 sessions in total and the number of participants in each session varied from 12 to 30 and was a multiple of 6.

I limit participants to Italy-born students so that their first name and photo do not signal

7. I exclude the 1st session data because of the problem discussed in appendix A. Nevertheless, the results including the 1st session data give me the same conclusions and are available upon request.

ethnicity, race, or cultural background. As a further attempt to reduce variance, I exclude recipients with non-Italian sounding name,[8] and whom the dictator declared they knew them "very well" at least once.

These data screenings leave me 390 participants, 195 dictators, and 558 observations (with dictators' allocation as the unit of observations).

3 Empirical strategy

I estimate the following equation with OLS:

$$Allocate_{ij} = \beta_0 + \beta_1 IQHigher_{ij} + \beta_2 Female_j + \beta_3 IQHigher_{ij} * Female_j + IQFE_i + X'_{ij}\gamma + \epsilon_{ij} \tag{1}$$

where each variable is defined as follows:

- $Allocate_{ij} \in [0, 1]$: dictator i's allocation to recipient j as a fraction of endowment.
- $IQHigher_{ij} \in \{0, 1\}$: an indicator variable equals 1 if recipient j's IQ is higher than dictator i.
- $Female_j \in \{0, 1\}$: an indicator variable equals 1 if recipient j is female.
- $IQFE_i := \sum_{l=2}^{9} \theta^l e_i^l$: fixed effects for the dictators' IQ (number of IQ test questions they have solved correctly), where $e_i^l \in \{0, 1\}$ is an indicator variable equals 1 if dictator i's IQ is l=1,...,9, 0 otherwise.
- X_{ij}: a set of additional covariates to increase statistical power and to address potential imbalance.[9]
- ϵ_{ij}: omitted factors that are correlated with dictator i's allocation to recipient j conditional on covariates.

Dictator's IQ fixed effects is included following Zimmermann (2020) so that the coefficients in equation 1 capture allocation differences due to the recipients' IQ, not the dictators'.[10] I cluster standard error at dictator level (Liang and Zeger 1986) and apply Pustejovsky and Tipton (2018)'s small cluster bias adjustment to address potential inflation of the type I error rate due to moderate cluster size.

Table 1 shows what the coefficients in equation 1 identify. β_1 identifies the allocation difference to male recipients with higher and lower IQ which captures dictators' distributional preference, among other effects. β_2 identifies the allocation difference to female and male recipients with lower IQ, namely every difference due to the recipients being female (e.g. women smile more and dictators like to give more to smiling people due to closer social distance). β_3 identifies the interaction of these two effects. Therefore, the allocation difference between female and male

8. 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.

9. The covariates include dictator characteristics (age, gender dummy, region of origin dummy, social science major dummy, STEM major dummy, post-bachelor dummy, over-confidence level), recipient characteristics (age, region of origin dummy), round fixed effects, and fixed effects for proximity between the dictator and the recipient. The full description of the covariates is in appendix B.

10. This is because people with different IQ (cognitive ability) may have a different distributional preference. For example, Almås et al. (2017) find that people from a low socio-economics family – which can be correlated with their cognitive ability – hold stronger egalitarian views than people from a middle or a high socio-economic family. Fisman et al. (2015) find that students in a top US law school – who presumably are smarter than average US citizens – are more meritocratic and more efficiency-oriented than average US citizens.

recipients with higher IQ is identified by $\beta_2 + \beta_3$, while the main effect of interest, the same allocation difference after controlling for the recipients' any gender-specific effects, is identified by β_3. Note that only the relative IQ matters because dictators only observe the recipients' IQ relative to themselves (and I control for dictators' IQ). Later, I will elaborate on this point more.

Table 1: Dictator's allocation identified by equation 1

		Recipient's gender	
		Female	Male
Recipient's IQ	Higher	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_1$
	Lower	$\beta_0 + \beta_2$	β_0

Notes: This table shows what the coefficients in equation 1 identify. Each cell represents dictator's allocation to recipients with higher (first row) or lower (second row) IQ and whose gender is female (first column) or male (second column).

4 Results

4.1 Data description

Summary statistics Table 2 summarizes the data after excluding participants and observations discussed in subsection 2.2. Looking at panels A and B, participants' average IQ level (number of IQ test questions solved correctly) is about 7 (with a maximum 9) and gender is roughly balanced. Also, dictators took nearly 2 minutes to solve the feedback questions on their IQ rank. Looking at panel C, most dictators did not know the paired recipients (after excluding pairs in which dictator knew the recipient "very well"). Looking at panel D, an average dictator allocated to paired recipients 40% of their endowment and variation in allocation within each IQ is as large as overall variation in allocation.[11] The latter indicates that there is enough variation in allocation I can exploit in my empirical specification (which uses dictator's IQ fixed effects).

Figure C1 shows empirical density (panel A) and empirical distribution (panel B) of dictators' allocation to further elaborate panel D of table 2. First, panel A shows that nearly 45% of dictators have chosen equal allocation. Second, the empirical distribution of giving in panel B resembles the empirical distribution of allocation in Bohnet and Frey (1999)'s one-way identification treatment which also shows recipients' face to the dictators.

Balance tests For coefficients in equation 1 to have causal interpretation, the dictator's IQ rank must be exogenous conditional on the dictator's IQ fixed effects. Table C1 presents evidence for this claim. Also, I have to make sure that randomization was successful ex-post so that dictators face recipients of different gender and IQ in a balanced way conditional on the dictator's IQ fixed effects. Tables C2 and C3 present evidence supporting this claim.

11. Demeaned standard deviation is sample standard deviation of $\widetilde{Allocate}_{ik} = Allocate_{ik} - \overline{Allocate}_k$, where $Allocate_{ik}$ is allocation by dictator i whose IQ is k and $\overline{Allocate}_k = \sum_{i \in k} Allocate_{ik}$ is average allocation by dictators with IQ k.

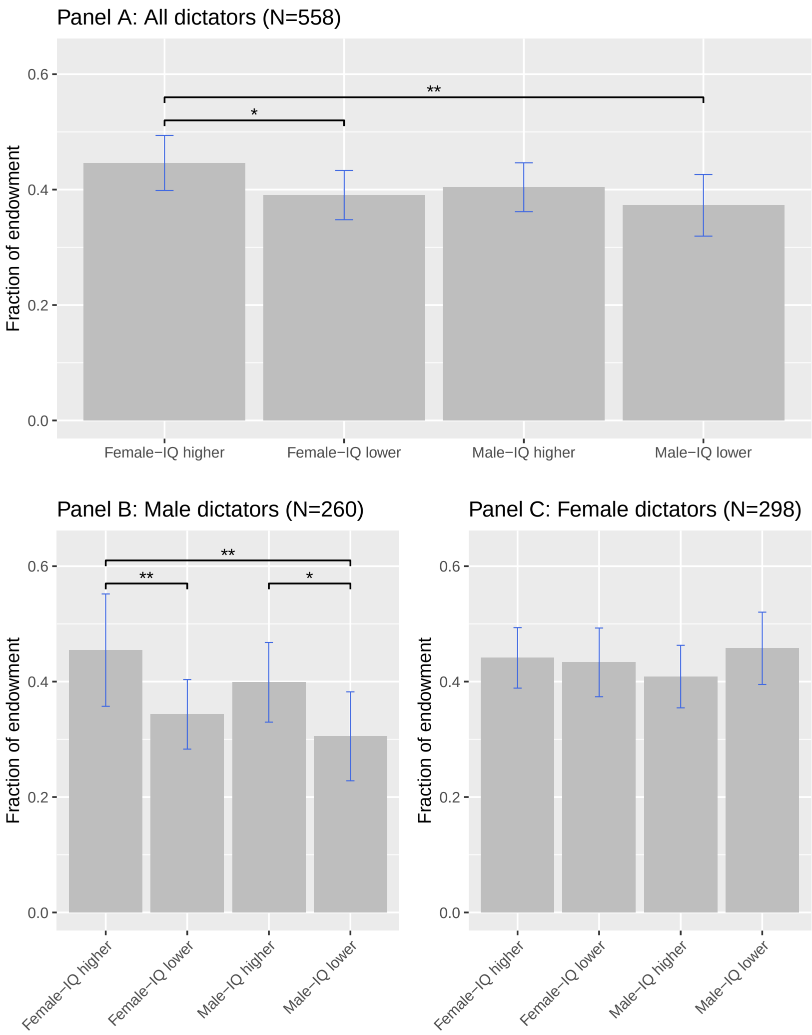
Table 2: Summary statistics: Dictator data

	Mean	SD
Panel A: Dictators		
IQ level	6.69	1.23
IQ rank	3.58	1.67
Age	23.47	2.72
Female	0.53	0.50
From Emilia-Romagna	0.18	0.39
Humanities	0.46	0.50
Social sciences	0.19	0.40
STEM	0.35	0.48
Post bachelor	0.46	0.50
Overconfidence	0.43	0.76
Time on feedback (sec.)	107.60	95.60
Observations	195	
Panel B: Paired recipients		
IQ level	6.84	1.16
IQ rank	3.42	1.74
IQ higher	0.53	0.50
Age	23.35	2.77
Female	0.47	0.50
From Emilia-Romagna	0.20	0.40
Observations	558	
Panel C: Proximity		
Did not know at all	0.96	0.19
Knew but not well	0.03	0.17
Saw before	0.01	0.09
Observations	558	
Panel D: Dictator's allocation (fraction of endowment)		
Allocation	0.40	0.24
Allocation (demeaned)		0.24
Observations	558	

Notes: This table shows summary statistics for the full sample: the dictators' and the paired recipients' characteristics, how well dictators knew the paired recipients, and dictators' allocation. Recipients whose name is non-Italian sounding and whom the dictator declared they knew them "very well" at least one are not included. Standard deviation of demeaned allocation is calculated as sample standard deviation of $\widetilde{Allocate}_{ik} = Allocate_{ik} - \overline{Allocate}_k$, where $Allocate_{ik}$ is allocation by dictator i whose IQ is k and $\overline{Allocate}_k = \sum_{i \in k} Allocate_{ik}$ is average allocation of dictators with IQ k.

Manipulation check Figure 4 provides evidence that dictators respond differently to the recipients' gender and IQ information: it shows dictators' average allocation for each category of recipients – female recipients with higher IQ, female recipients with lower IQ, male recipients with higher IQ, and male recipients with lower IQ – along with their 95% confidence intervals. Looking at panel A, we see that dictators allocate most to female recipients with higher IQ, more to male recipients with higher IQ, and slightly more to female recipients with lower IQ – compared to male recipients with lower IQ. In addition, the allocations to female recipients with higher IQ and male recipients with lower IQ are statistically different at 5% level and the allocations to female recipients with higher and lower IQ are marginally statistically different at 10%.

Figure 4: Dictators' allocation by the recipients' category



Notes: This figure shows dictators' allocation as a fraction of endowment by recipients' category along with their 95% confidence intervals for all dictators (panel A), male dictators (panel B), and female dictators (panel C). Confidence intervals are calculated with the standard errors clustered at the dictator level with Pustejovsky and Tipton (2018)'s small cluster bias adjustment. Horizontal lines over categories indicate statistically significant differences. Unit of observation: dictator's allocation. Significance levels: * 10%, ** 5%, and *** 1%.

Panel B, which shows male dictators' average allocation for each category of recipients, presents the same pattern as panel A but the differences are larger. In addition, some differences are more statistically significant despite the smaller sample size: the allocations to female recipients with higher IQ and male recipients with lower IQ are statistically different at 5% so do the allocations to female recipients with higher and lower IQ. Also, the allocations to male recipients with higher and lower IQ are marginally statistically different at 10%.

On the other hand, female dictators' average allocation for each category of recipients presented in panel C shows a rather stark difference between female and male dictators. While male dictators discriminate more based on ability and gender, female dictators do not. Indeed, all the allocation differences are statistically insignificant even at 10%. This observation is consistent with the existing literature that women are more inequality averse (Croson and Gneezy 2009) but inconsistent with Cappelen, Falch, and Tungodden (2019) who find that women dislike male losers more than men. In addition, female dictators' allocation is overall higher than male dictators, consistent with existing dictator game experiments (Engel 2011). The differences in observable characteristics between female and male dictators reported in panel A of table C4 are also consistent with the existing literature.[12]

4.2 Main results

Table 3 presents the results with all dictators. Column 1 presents estimate without controlling for dictator's IQ and shows the direction of the bias without including dictator IQ fixed effects: although statistically insignificant, dictators with lower IQ allocate more to recipients with higher IQ regardless of the recipients' gender as shown by the coefficient estimate on $IQHigher_{ij}$, biasing the estimate upwards. From columns 2 to 5, I gradually increase the number of covariates to check the robustness of my main specification in column 5. They show that the coefficient estimates are stable across 4 columns.

Looking at column 5, the coefficient estimate on $IQHigher_{ij} * Female_j$ is positive and statistically insignificant. To give a statistical claim about the insignificance, I use duality between hypothesis testing and confidence interval (Casella and Berger 2001) and examine what effect size we can reject and whether it is quantitatively important as typically done in epistemology (e.g. Chaisemartin and Chaisemartin 2020). Thus turning to the 95% confidence interval reported below the standard error estimate, the negative end is about -0.037, suggesting that we can reject the effect size lower than -3.7% of the dictator endowment at 5% significance level. This value is very small, about 2.4-3.1 times smaller than the effect size of typical dictator game experiments that examine the role of social distance with university students (e.g., Brañas-Garza et al. 2010; Charness and Gneezy 2008; Leider et al. 2010).[13]

While OLS only picks up the average effect, these observations hold also in distribution. Panel A of figure 5 presents empirical CDFs of dictators' allocation for each recipient category, demeaned by the dictator's IQ fixed effects to give a causal interpretation.[14] The figure shows that the CDF of dictators' allocation to female recipients with higher IQ (solid blue line) almost

12. Table C4 presents the same summary statistics as table 2 but separately for female and male dictators and their differences. It shows that female dictators are more likely to major in humanities, less likely to major in social sciences and STEM, less overconfident, and tend to allocate more to recipients – characteristics consistent with the literature on gender differences.

13. Charness and Gneezy (2008) examine how informing the recipient's family name increases the dictators' giving using a university student sample, and find an 8.9% increase in giving as a fraction of endowment. Leider et al. (2010) find using a university student sample that dictators increase giving by 11.42% 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% more of their endowment to friends relative to other students in the same class.

14. Dictators' allocation is demeaned for dictators' IQ level so that the CDFs correspond to the regression results: $\widetilde{Allocate}_{ik} = Allocate_{ik} - \overline{Allocate}_k + \overline{\overline{Allocate}}$, where $Allocate_{ik}$ is allocation by dictator i whose IQ is k, $\overline{Allocate}_k = \sum_{i \in k} Allocate_{ik}$ is average allocation of dictators with IQ k, and $\overline{\overline{Allocate}} = \sum_i Allocate_{ik}$ is average allocation by all dictators. $\overline{\overline{Allocate}}$ is added to re-center the allocation. Although this re-centering leaves a few observations outside the 0-1 range, they do not alter the results and thus are trimmed for ease of visual inspection.

Table 3: The role of the recipients' gender and IQ in dictators' allocation: All dictators

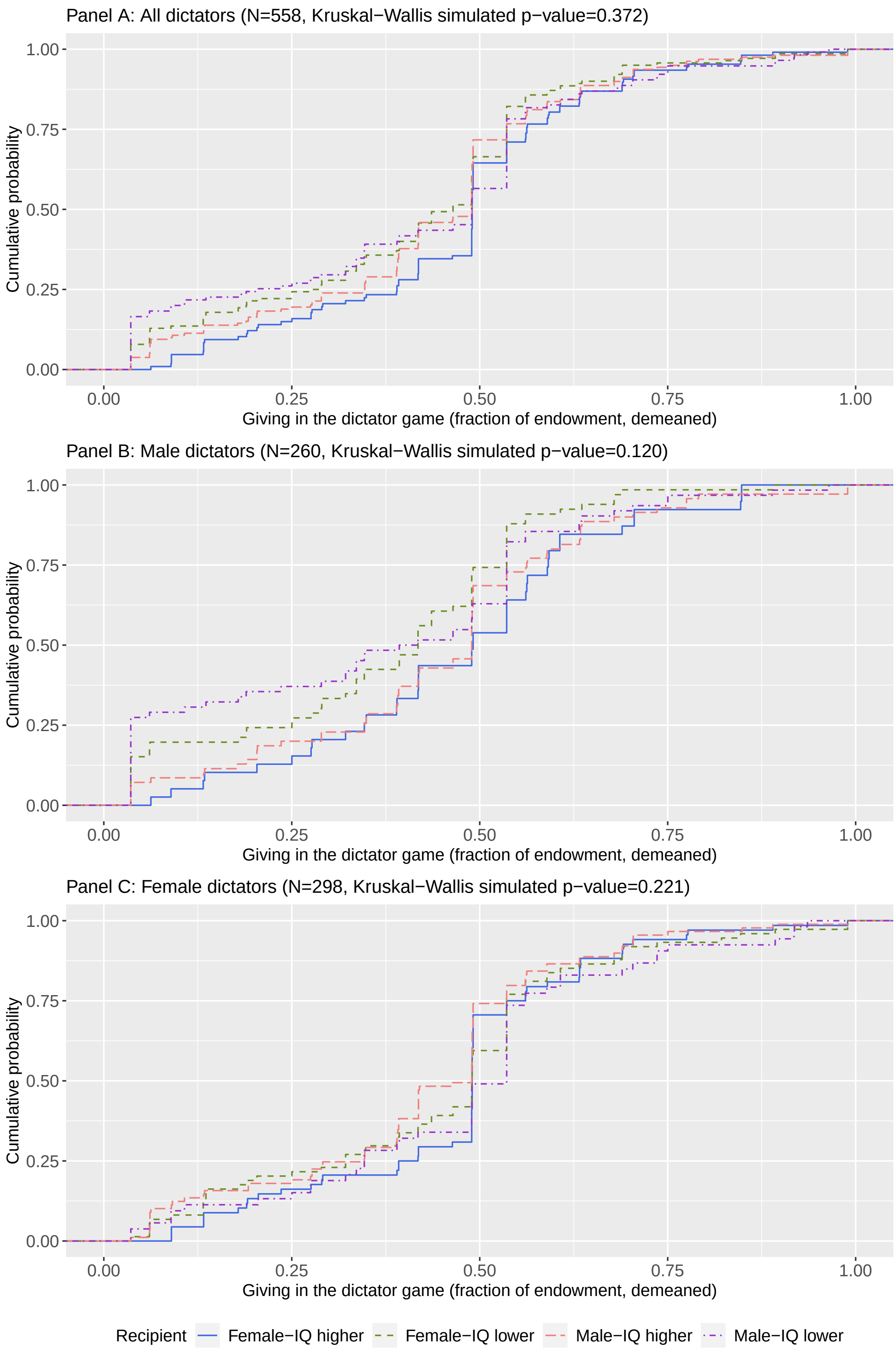
Outcome:	Dictator's allocation (fraction of endowment)				
	(1)	(2)	(3)	(4)	(5)
IQHigher	0.031	0.011	0.013	0.005	0.006
	(0.031)	(0.033)	(0.033)	(0.033)	(0.034)
	[-0.030, 0.093]	[-0.054, 0.075]	[-0.053, 0.078]	[-0.059, 0.070]	[-0.061, 0.072]
Female	0.018	0.014	0.014	0.007	0.006
	(0.027)	(0.027)	(0.027)	(0.026)	(0.026)
	[-0.037, 0.072]	[-0.040, 0.067]	[-0.040, 0.068]	[-0.044, 0.058]	[-0.045, 0.057]
IQHigherxFemale	0.024	0.027	0.026	0.034	0.035
	(0.037)	(0.037)	(0.037)	(0.036)	(0.037)
	[-0.048, 0.097]	[-0.045, 0.100]	[-0.048, 0.099]	[-0.037, 0.105]	[-0.037, 0.107]
Dictator IQ FE	-	✓	✓	✓	✓
Round FE	-	-	✓	✓	✓
Proximity FE	-	-	✓	✓	✓
Dictator controls	-	-	-	✓	✓
Recipient controls	-	-	-	-	✓
Female+IQHigherxFemale	0.042	0.041	0.04	0.041	0.041
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
	[-0.009, 0.093]	[-0.01, 0.092]	[-0.012, 0.091]	[-0.01, 0.092]	[-0.011, 0.093]
Outcome Mean	0.403	0.403	0.403	0.403	0.403
Outcome SD	0.239	0.239	0.239	0.239	0.239
R-squared	0.011	0.025	0.028	0.079	0.086
Observations	558	558	558	558	558
Clusters	195	195	195	195	195

Notes: This table shows OLS estimates of the role of the recipients' gender and IQ in dictators' allocation. The outcome variable is dictators' allocation as a fraction of endowment. The main specification is column 5 which includes all covariates (see the main text for detail). Columns 2-4 provide robustness of the main specification by excluding some covariates and column 1 shows bias of not including dictator IQ fixed effects. Joint statistical significance of coefficient estimate on Female+IQHigherxFemale is calculated using t-test. 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. R-squared is net of the dictator IQ fixed effects. Unit of observation: dictator's allocation. Significance levels: * 10%, ** 5%, and *** 1%.

always lies to the right of the other CDFs (although all CDFs are statistically indistinguishable from each other at 5% significance level), suggesting people do not treat competent women unfavorably than competent men.

The results also hold separately for male and female dictators. Column 1 of table 4 presents results with male dictators only and column 2 results with female dictators only, both will full control. First, the coefficient estimate on $IQHigher_{ij} * Female_j$ is positive and statistically insignificant both for male and female dictators. Second, while the 95% confidence interval is wider due to the reduction of sample size by about half, we can still reject at 5% significance level the effect size lower than -9.0% for male dictators and -3.5% for female dictators. -9.0% is still the magnitude of the effect size of typical dictator game experiments. As with the full sample estimate, these observations also hold in distribution as reported in panel B (male dictators) and in panel C (female dictators) of figure 5. For both male dictators and female dictators, the CDF of dictators' allocation to female recipients with higher IQ (solid blue line) almost always lies on the right of the other CDFs (although all CDFs are statistically indistinguishable from each other at 5% significance level), suggesting that neither men nor women do not treat competent

Figure 5: CDFs of dictators' allocation by the recipients' category (demeaned)



Notes: These figures show the empirical distribution of demeaned dictators' allocation by recipients' category for all dictators (panel A), male dictators (panel B), and female dictators (panel C). Demeaning was done with respect to the dictators' IQ so that the CDFs have causal interpretation: $\widetilde{Allocate}_{ik} = Allocate_{ik} - \overline{Allocate}_k + \overline{\overline{Allocate}}$, where $Allocate_{ik}$ is allocation by dictator i whose IQ is k, $\overline{Allocate}_k = \sum_{i \in k} Allocate_{ik}$ is average allocation of dictators with IQ k, and $\overline{\overline{Allocate}} = \sum_i Allocate_{ik}$ is average allocation by all dictators. $\overline{\overline{Allocate}}$ is added to re-center the allocation. Values below 0 and above 1 are trimmed for ease of visual inspection, but including those observations does not alter my results (there are only a few observations outside 0-1 range). Kruskal-Wallis simulated p-values are calculated using randomization inference (Young 2019) to address arbitrary dependency among observations with 2000 draws under the null hypothesis of no location difference (i.e. all CDFs coincide). Unit of observation: dictator's allocation.

women unfavorably than competent men.

Table 4: The role of the recipients' gender and IQ in dictators' allocation: Robustness checks

Outcome:	Dictator's allocation (fraction of endowment)				Belief on IQ
Sample:	Male	Female	Over-confident	Non-over-confident	Evaluator
	(1)	(2)	(3)	(4)	(5)
IQHigher	0.048	-0.049	0.032	-0.032	0.232
	(0.055)	(0.042)	(0.048)	(0.049)	(0.303)
	[-0.062, 0.158]	[-0.132, 0.034]	[-0.065, 0.128]	[-0.130, 0.067]	[-0.371, 0.834]
Female	0.014	-0.014	-0.007	0.013	-0.352
	(0.034)	(0.037)	(0.033)	(0.042)	(0.292)
	[-0.054, 0.082]	[-0.089, 0.061]	[-0.073, 0.060]	[-0.072, 0.098]	[-0.931, 0.226]
IQHigherxFemale	0.031	0.057	0.038	0.046	0.512
	(0.061)	(0.046)	(0.050)	(0.054)	(0.392)
	[-0.090, 0.152]	[-0.035, 0.148]	[-0.060, 0.136]	[-0.063, 0.154]	[-0.261, 1.286]
Female+IQHigherxFemale	0.045	0.042	0.031	0.059	0.16
	(0.047)	(0.029)	(0.035)	(0.038)	(0.257)
	[-0.048, 0.138]	[-0.015, 0.1]	[-0.037, 0.1]	[-0.016, 0.133]	[-0.346, 0.666]
Outcome Mean	0.369	0.432	0.385	0.427	6.342
Outcome SD	0.253	0.223	0.241	0.235	1.89
R-squared	0.151	0.084	0.112	0.151	0.097
Observations	260	298	325	233	368
Clusters	91	104	115	80	193

Notes: This table shows OLS estimates of the role of the recipients' gender and IQ in dictators' allocation for male and female dictators (columns 1-2), overconfident and non-overconfident dictators (columns 3-4), and evaluators' belief on the recipients' IQ (column 5). The outcome variable is dictators' allocation as a fraction of endowment in columns 1-2 and evaluators' belief on the recipients' IQ level in column 5. All specifications include dictator IQ fixed effects, round fixed effects, proximity fixed effects, dictator (or evaluator) controls, and recipient controls, except columns 1 and 2 where dictator's gender dummy is excluded and columns 3-4 where dictator's overconfidence measure is excluded. Joint statistical significance of coefficient estimate on Female+IQHigherxFemale is calculated using t-test. 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 or the evaluator level with Pustejovsky and Tipton (2018)'s small cluster bias adjustment. R-squared is net of the dictator IQ fixed effects. Unit of observation: dictator's allocation (columns 1-2) and evaluator's belief (column 3). Significance levels: * 10%, ** 5%, and *** 1%.

While dictators only see the recipients' IQ relative to theirs and thus the competence measure is relative to theirs, dictators do not see their IQ at the time they play dictator games. Indeed, in the real world, we do not have an absolute measure of other people's competence but evaluate relative to some benchmark. Yet, if anything, overconfident people is likely to consider people whose competence is higher than themselves as more competent in absolute terms than non-overconfident people do, after controlling for their actual competence.

In columns 3-4 of table 4, I present the results separately for overconfident dictators (dictators who guess their IQ higher than their actual IQ, column 3) and non-overconfident dictators (dictators who guess their IQ equal to or lower than their actual IQ, column 4). For both types of dictators, the coefficient estimate on $IQHigher_{ij} * Female_j$ is positive and statistically insignificant and the lower bound of the estimate at 5% significance level is almost identical (-6.0% for over-confident dictators and -6.3% for non-overconfident dictators). Thus, relative or absolute does not matter for my main results.

4.3 Alternative explanations

Female dictators' in-group preference One competing explanation is female dictators' favoritism towards people who belong to the same social group, or in-group preference (Tajfel and Turner 1979), which biases my β_3 estimates upward. However, this explanation is inconsistent with the data. First, I use the difference in allocation between lower IQ female and male recipients as a control group, which eliminates the recipients' gender-specific allocation preference for analysis with female dictators. Second, the results with male dictators who do not have an in-group preference towards female recipients still reject the effect size lower than that of a typical dictator game experiment studying the effect of social distance using a university student sample.

Distaste against lower IQ male recipients Although I use the difference in allocation to lower IQ female and male recipients to control for any recipient gender-specific allocation preference, this may not be a clean control because people may have a negative bias against under-performing men (Cappelen, Falch, and Tungodden 2019; Moss-Racusin, Phelan, and Rudman 2010). This explanation is also inconsistent with the data. First, the allocation difference between higher IQ female and male recipients without using lower IQ female-male allocation differences, estimate of $\beta_2 + \beta_3$, still suggests the same conclusion: we can reject at the 5% significance level the effect size lower than -1.1% of dictator endowment for all dictators (table 3, column 5), lower than -4.8% for male dictators (table 4, column 1), lower than -1.5% for female dictators (table 4). Second, while these single-difference estimates do not control for the recipients' gender-specific allocation preference, Cappelen, Falch, and Tungodden (2019) find that the distaste mostly comes from women and the results with male dictators only should not be affected by this distaste.

A wrong belief that female recipients are less competent My empirical specification compares female and male recipients with higher IQ. The identification fails if dictators consider female recipients as less competent than male recipients even if they have a higher IQ than dictators. Although this is unlikely, Fiske et al. (2002) find that people consider women as less competent than men where the competence measures include intelligence.

This explanation indeed does not apply to my sample. Column 5 of table 4 presents results from a regression where I replace dictators' allocation with recipients' belief (whom I call evaluator) about the other recipients' IQ level which proxies dictators' belief. Recipients' belief is a valid proxy for dictators' belief by the random assignment of participants to dictators and recipients and that both dictators and recipients face the same environment until the start of the dictator game.[15] The estimate of β_3 is positive albeit statistically insignificant, suggesting dictators do not believe that higher IQ female recipients are less competent than higher IQ male recipients.

This belief analysis, however, points to a potentially interesting difference in people's belief updating process about women's and men's competence: people may update women's competence

15. Table C5 presents evidence that recipients and dictators do not differ in their observable characteristics and characteristics of paired recipients.

more than men's competence once they see an objective measure of women's competence as shown by the positive but statistically insignificant estimate of β_3.

Experimental manipulation failure The effect size becomes null if dictators do not respond to the recipients' gender and IQ information. However, dictators in my sample *do* respond to the recipients' gender and IQ information in statistically significant ways as we already see in figure 4.

Ex-post randomization failure My empirical specification cannot detect causal effects if either (i) dictators' IQ rank is endogenous even conditional on the dictators' IQ fixed effects or (ii) dictators of specific characteristics face recipients with a specific gender or/and with lower or higher IQ. However, both concerns are addressed by a random desk assignment. Also, we see that even ex-post, the random assignment was successful in tables C1, C2, and C3. While the recipients' region of origin is unbalanced (table C3, column 10) – which can happen by the definition of type I error – I include recipients' region of origin dummy in my main specification which controls the imbalance nonparametrically. Table C6 presents results for various subsamples and we can still reject the effect size lower than -4.3% to -8.7% at 5% significance level, which further addresses concerns for ex-post imbalance.[16] Last, while I pool all the higher and lower IQ recipients despite that dictators can also see the IQ rank differences, figure C2 shows that taking into account the IQ rank differences does not alter the results.[17]

Wrong identification assumptions Any causal inference relies on several assumptions, so failure to reject the null hypothesis of no effect can be no effect, but can also be that some identification assumptions are wrong. However, aside from those discussed thus far, I do not make any significant assumptions because my empirical specification is a simple double difference-in-means. I also apply Pustejovsky and Tipton (2018)'s small cluster bias adjustment to address the finite-sample bias of the standard error. Thus, it is unlikely that the failure to reject the null can be attributed to some implausible identification assumptions. Note that I also show I can reject a very small effect size using confidence intervals.

Lack of statistical power When the power is low (type II error rate is high), the confidence interval becomes wider. However, my confidence interval can reject a very small effect size at a 5% significance level. Also, while there is an ex-post minimum detectable effect estimate, it is simply 2.8 times the standard error and mostly useful for cross-study comparison (McKenzie and Ozier 2019); the information used in the confidence interval is strictly larger than the information used in the ex-post minimum detectable effect.

16. In table C6, column 1 excludes dictators with IQ rank 1 and 6 who never face recipients with lower / higher IQ. Column 2 excludes dictator-recipient pairs in which the dictator knows the recipients even a little and column 3 pairs in which the dictator saw the recipients before.

17. Figure C2 shows OLS estimates of equation 1 but splitting $IQHigher_{ij}$ into 6 separate dummies indicating the recipients' IQ rank differences relative to the dictators'. The lower/higher the recipient's IQ, the more negative/positive their IQ rank difference. For brevity, the figure only plots the coefficient estimates on the interaction terms between the 6 separate $IQHigher_{ij}$ and $Female_j$, $\hat{\beta}_3$ along with their 95% confidence intervals.

5 Conclusion

This paper examines whether competent women receive unfavorable treatment compared to competent men. Using dictator game giving as a measure of favorable and unfavorable treatment and exogenously varying gender and competence measured by an IQ test, I show that people treat competent women no less favorably than competent men; if anything, people treat competent women slightly more favorably. The lower bound of my estimate is -3.7% of dictator endowment, which is much smaller than the effect size of dictator game experiments studying the role of social distance. I also show that experimental manipulation is successful, randomization was successful even ex-post, identification assumptions are plausible, and the experiment has sufficient statistical power.

This paper contributes to the literature in two ways. First, I provide evidence that, in the stylized environment where unfavorable treatment has material consequences, the argument that women face a tradeoff between being competent and being likable does not hold. This suggests that competent women may receive fair treatment in hiring and promotion if the results are externally valid. Second, while several studies show that women are more critically evaluated in traditionally male occupations, my results indicate that a plausible explanation for this evidence is people's lack of enough prior about women's competence in these occupations rather than taste-based gender discrimination.

Indeed, there is ample evidence that female leaders (Chakraborty and Serra 2019; Håkansson 2020) and competitors (Datta Gupta, Poulsen, and Villeval 2013) receive more aggressive treatments and receive less support by men (Born, Ranehill, and Sandberg 2020). My study is silent to gender discrimination where there are intense interactions and competition among women and men; there is evidence that men hold motivated gender bias (Sinclair and Kunda 2000) and it is a topic of future research. Still, my results apply to vertical relationships such as workers vs. managers and employees vs. employers and provide a piece of evidence in considering hiring and promotion practices in labor markets.

References

Almås, Ingvild, Alexander W Cappelen, Kjell G Salvanes, Erik Ø Sørensen, and Bertil Tungodden. 2017. "Fairness and family background." *Politics, Philosophy & Economics* 16 (2): 117–131.

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.

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.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg. 2019. "The Dynamics of Discrimination: Theory and Evidence." *American Economic Review* 109 (10): 3395–3436.

Boring, Anne. 2017. "Gender biases in student evaluations of teaching." *Journal of Public Economics* 145:27–41.

Born, Andreas, Eva Ranehill, and Anna Sandberg. 2020. "Gender and Willingness to Lead: Does the Gender Composition of Teams Matter?" *The Review of Economics and Statistics.*

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.

Cappelen, Alexander, Ranveig Falch, and Bertil Tungodden. 2019. *The Boy Crisis: Experimental Evidence on the Acceptance of Males Falling Behind.* HCEO Working Paper 2019-014.

Casella, George, and Roger L. Berger. 2001. *Statistical Inference.* 2nd ed. Pacific Grove, CA: Duxbury Press.

Ceci, Stephen J., and Wendy M. Williams. 2015. "Women have substantial advantage in STEM faculty hiring, except when competing against more-accomplished men." *Frontiers in Psychology* 6:1532.

Chaisemartin, Clément de, and Luc de Chaisemartin. 2020. "BCG vaccination in infancy does not protect against COVID-19. Evidence from a natural experiment in Sweden." *Clinical Infectious Diseases.*

Chakraborty, Priyanka, and Danila Serra. 2019. *Gender differences in top leadership roles: Does worker backlash matter?* 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.

De Paola, Maria, Michela Ponzo, and Vincenzo Scoppa. 2018. “Are Men Given Priority for Top Jobs? Investigating the Glass Ceiling in Italian Academia.” *Journal of Human Capital* 12 (3): 475–503.

Ditonto, Tessa. 2017. “A High Bar or a Double Standard? Gender, Competence, and Information in Political Campaigns.” *Political Behavior* 39 (2): 301–325.

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.

Exley, Christine L., and Judd B. Kessler. 2019. *Motivated Errors.* Working Paper.

Fiske, Susan T., Amy J. C. Cuddy, Peter Glick, and Jun Xu. 2002. “A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition.” *Journal of personality and social psychology* 82 (6): 878–902.

Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Daniel Markovits. 2015. “The distributional preferences of an elite.” *Science* 349 (6254): aab0096.

Greiner, Ben. 2015. “Subject pool recruitment procedures: organizing experiments with ORSEE.” *Journal of the Economic Science Association* 1 (1): 114–125.

Håkansson, Sandra. 2020. “Do women pay a higher price for power? Gender bias in political violence in Sweden.” *Journal of Politics.*

Heilman, Madeline E. 2001. “Description and Prescription: How Gender Stereotypes Prevent Women’s Ascent Up the Organizational Ladder.” *Journal of Social Issues* 57 (4): 657–674.

Heilman, Madeline E., Caryn J. Block, and Richard F. Martell. 1995. “Sex stereotypes: Do they influence perceptions of managers?” *Journal of Social Behavior and Personality* 10 (6): 237–252.

Heilman, Madeline E., and Tyler G. Okimoto. 2007. “Why are women penalized for success at male tasks?: the implied communality deficit.” *The Journal of Applied Psychology* 92 (1): 81–92.

Heilman, Madeline E., Aaron S. Wallen, Daniella Fuchs, and Melinda M. Tamkins. 2004. “Penalties for success: reactions to women who succeed at male gender-typed tasks.” *The Journal of Applied Psychology* 89 (3): 416–427.

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

Koffi, Marlène. 2019. *Innovative Ideas and Gender Inequality.* Working Paper.

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.

Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal data analysis using generalized linear models." *Biometrika* 73 (1): 13–22.

MacKinnon, James G, and Halbert White. 1985. "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties." *Journal of Econometrics* 29 (3): 305–325.

McKenzie, David, and Owen Ozier. 2019. *Why ex-post power using estimated effect sizes is bad, but an ex-post MDE is not.* Development Impact. The World Bank.

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2019. "Gender Bias in Teaching Evaluations." *Journal of the European Economic Association.*

Moss-Racusin, Corinne A., Julie E. Phelan, and Laurie A. Rudman. 2010. "When men break the gender rules: Status incongruity and backlash against modest men." *Psychology of Men & Masculinity* 11 (2): 140–151.

Phelan, Julie E., Corinne A. Moss-Racusin, and Laurie A. Rudman. 2008. "Competent yet Out in the Cold: Shifting Criteria for Hiring Reflect Backlash Toward Agentic Women." *Psychology of Women Quarterly* 32 (4): 406–413.

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.

Quadlin, Natasha. 2018. "The Mark of a Woman's Record: Gender and Academic Performance in Hiring." *American Sociological Review* 83 (2): 331–360.

Rudman, Laurie A. 1998. "Self-promotion as a risk factor for women: The costs and benefits of counterstereotypical impression management." *Journal of Personality and Social Psychology* 74 (3): 629–645.

Rudman, Laurie A., and Kimberly Fairchild. 2004. "Reactions to counterstereotypic behavior: the role of backlash in cultural stereotype maintenance." *Journal of Personality and Social Psychology* 87 (2): 157–176.

Rudman, Laurie A., and Stephen E. Kilianski. 2000. "Implicit and Explicit Attitudes Toward Female Authority." *Personality and Social Psychology Bulletin* 26 (11): 1315–1328.

Rudman, Laurie A., Corinne A. Moss-Racusin, Julie E. Phelan, and Sanne Nauts. 2012. "Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders." *Journal of Experimental Social Psychology* 48 (1): 165–179.

Rudman, Laurie A., and Julie E. Phelan. 2008. "Backlash effects for disconfirming gender stereotypes in organizations." *Research in Organizational Behavior* 28:61–79.

Sandberg, Sheryl. 2013. *Lean In: Women, Work, and the Will to Lead.* New York, NY: Knopf.

Sarsons, Heather. 2019. *Interpreting Signals in the Labor Market: Evidence from Medical Referrals.* Working Paper.

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram. 2020. "Gender Differences in Recognition for Group Work." *Journal of Political Economy.*

Sinclair, Lisa, and Ziva Kunda. 2000. "Motivated Stereotyping of Women: She's Fine if She Praised Me but Incompetent if She Criticized Me." *Personality and Social Psychology Bulletin* 26 (11): 1329–1342.

Tajfel, Henry, and John Turner. 1979. "An integrative theory of intergroup conflict." In *The social psychology of intergroup relations,* edited by William G. Austin and Stephen Worchel, 33–47. Monterey, CA: Brooks Cole Publishing.

Williams, Wendy M., and Stephen J. Ceci. 2015. "National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track." *Proceedings of the National Academy of Sciences* 112 (17): 5360–5365.

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.

Appendix A Changes to the pre-analysis plan

In the initial design, recipients finished all the tasks except the post-questionnaire and left the laboratory before dictators receive their IQ rank, so that dictators could play dictator game without recipients in the same room. The allocation to the recipients was paid electronically as a "participation fee" for the online post-questionnaire which was sent to recipients 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 recipients left the laboratory and dictators seemed to have lost concentration during this period: about half of the dictators could not answer the comprehension questions about their IQ rank. Thus, I changed the design and let recipients 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 including the 1st session data delivers the same conclusion and available upon request. Also, the oTree code and instructions used for the 1st session are available upon request.

I also made the following minor 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.

Other changes are the following:

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 5) to also examine whether the results hold also in distribution.
5. 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.
6. 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.
7. 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).

Appendix B Description of covariates

X_{ij} in the main specification (equation 1) includes the following variables:

Dictator characteristics

- $Age_i \in \mathbb{N}$: dictator i's age.
- $Female_i \in \{0, 1\}$: an indicator variable equals 1 if dictator i is female, 0 otherwise.
- $FromEmiliaRomagna_i \in \{0, 1\}$: an indicator variable equals 1 if dictator i is from Emilia-Romagna region (where the University of Bologna is located), 0 otherwise.
- $SocialSciences_i \in \{0, 1\}$: an indicator variable equals 1 if dictator i's major is social sciences, 0 otherwise.
- $STEM_i \in \{0, 1\}$: an indicator variable equals 1 if dictator i's major is natural sciences/mathematics, engineering, or medicine; 0 otherwise.
- $PostBachelor_i \in \{0, 1\}$: an indicator variable equals 1 if dictator i's degree program is either master/post-bachelor, in the 4th year or beyond of bachelor-master combined program, or PhD, 0 otherwise.[18]
- $OverConfidence_i \in \{-1, 0, 1\}$: degree of dictator i's overconfidence. It is equal to -1 if dictator i's guess about the number of IQ test questions they correctly solved is lower than the actual number, 0 if equal to the actual number, and 1 if higher than the actual number.

Recipient characteristics

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

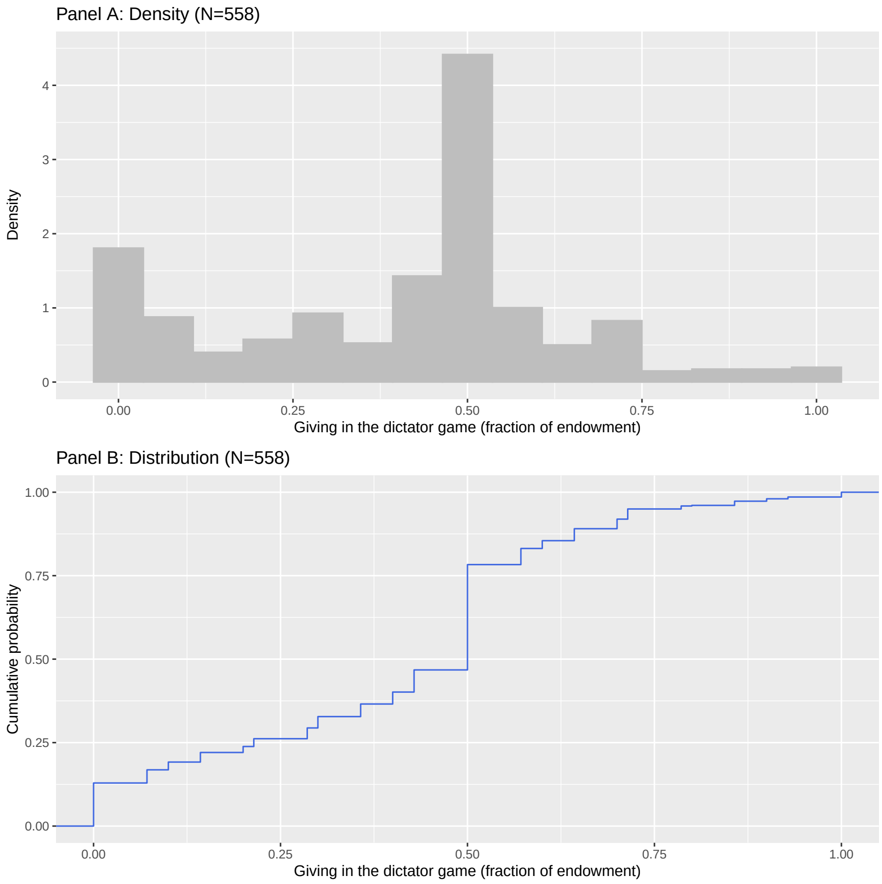
Fixed effects

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

18. In Italy, bachelor is a 3 year program.

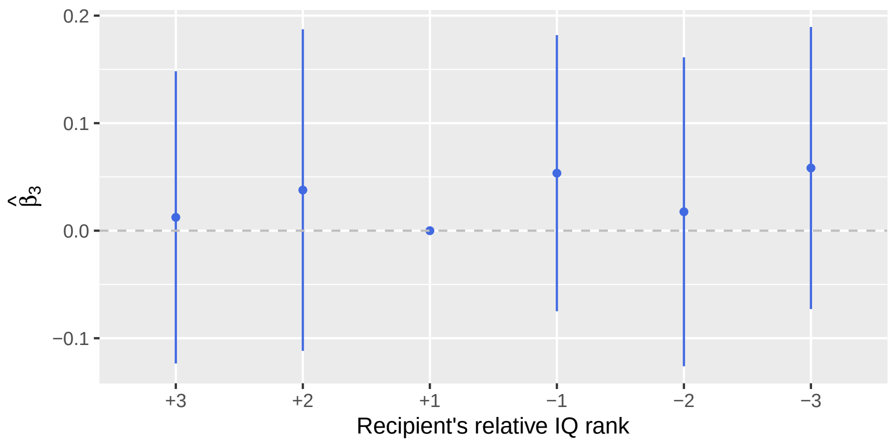
Appendix C Additional figures and tables

Figure C1: Density and distribution of the dictators' allocation



Notes: These figures show the empirical density (panel A) and the empirical distribution (panel B) of the dictators' allocation as a fraction of endowment. Recipients whose name is non-Italian sounding and whom the dictator declared they knew them "very well" at least once are excluded. Unit of observation: dictator's allocation.

Figure C2: The role of the recipients' IQ and gender in dictators' allocation: Taking into account for IQ rank differences



Notes: This figure shows OLS estimates of the role of recipient's gender and IQ in dictators' allocation that takes into account for the IQ rank differences dictators observe by splitting $IQHigher_{ij}$ into 6 separate dummies indicating the recipients' IQ rank differences relative to the dictators'. The lower/higher the recipient's IQ, the more negative/positive their IQ rank difference. The specification includes dictator IQ fixed effects, round fixed effects, proximity fixed effects, dictator controls, and recipient controls. The outcome variable is dictators' allocation as a fraction of endowment. For brevity, the figure only plots the coefficient estimates on the interaction terms between the 6 separate $IQHigher_{ij}$ and $Female_j$, $\hat{\beta}_3$, along with their 95% confidence intervals, which is calculated with standard errors clustered at dictator level with Pustejovsky and Tipton (2018)'s small cluster bias adjustment. Unit of observation: dictator's allocation.

Table C1: Balance test: IQ rank

Outcome:	Age	Female	From Emilia-Romagna	Human-ities	Social sciences	STEM	Post bachelor	Over-confidence
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IQ rank = 2	0.010	0.221*	0.074	-0.095	0.034	0.061	0.151	0.146
	(0.796)	(0.128)	(0.104)	(0.130)	(0.088)	(0.130)	(0.127)	(0.200)
IQ rank = 3	-0.300	0.139	-0.007	-0.101	0.183	-0.081	0.183	0.160
	(0.776)	(0.143)	(0.103)	(0.142)	(0.120)	(0.137)	(0.137)	(0.241)
IQ rank = 4	-0.536	0.094	0.138	-0.146	0.101	0.045	0.187	0.430*
	(0.894)	(0.148)	(0.116)	(0.148)	(0.123)	(0.148)	(0.145)	(0.258)
IQ rank = 5	0.534	0.092	0.062	-0.220	0.166	0.054	0.061	0.158
	(0.959)	(0.165)	(0.128)	(0.175)	(0.128)	(0.165)	(0.156)	(0.271)
IQ rank = 6	-0.040	0.070	0.021	-0.368*	0.442***	-0.074	0.013	0.346
	(1.093)	(0.191)	(0.147)	(0.201)	(0.162)	(0.173)	(0.191)	(0.306)
Dictator IQ FE	✓	✓	✓	✓	✓	✓	✓	✓
F statistic	0.571	0.634	0.704	0.697	1.91*	0.626	0.739	0.83
R-squared	0.040	0.067	0.040	0.042	0.074	0.062	0.027	0.032
Observations	195	195	195	195	195	195	195	195

Notes: This table shows balance across dictators with different IQ ranks. The estimates are obtained by running OLS regression of various dictator characteristics on IQ rank dummies with dictator IQ fixed effects. The F statistic shows the joint significance of IQ rank = 2 to IQ rank = 6 dummies. HC2 heteroskedasticity-robust standard errors (MacKinnon and White 1985) with Bell and McCaffrey (2002)'s small sample bias adjustment are reported below each coefficient estimate. R-squared is net of dictator IQ fixed effects. Unit of observation: dictator. Significance levels: * 10%, ** 5%, and *** 1%.

Table C2: Balance test: Recipient's category

Outcome:	Age	Female	From Emilia-Romagna	Human-ities	Social sciences	STEM	Post bachelor	Over-confidence
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IQHigher	-0.429	0.001	0.105**	-0.065	0.106**	-0.041	-0.071	0.063
	(0.350)	(0.064)	(0.048)	(0.065)	(0.051)	(0.060)	(0.063)	(0.107)
Female	-0.228	0.060	0.080*	-0.026	0.015	0.011	-0.043	0.040
	(0.336)	(0.059)	(0.048)	(0.057)	(0.046)	(0.057)	(0.060)	(0.090)
IQHigherxFemale	0.431	0.010	-0.148**	0.014	-0.063	0.049	0.069	-0.051
	(0.458)	(0.082)	(0.064)	(0.081)	(0.062)	(0.079)	(0.084)	(0.129)
Dictator IQ FE	✓	✓	✓	✓	✓	✓	✓	✓
F statistic	0.522	1.078	2.074	0.505	1.731	0.661	0.417	0.119
R-squared	0.029	0.052	0.034	0.025	0.028	0.050	0.014	0.007
Observations	558	558	558	558	558	558	558	558
Clusters	195	195	195	195	195	195	195	195

Notes: This table shows that dictators were matched recipients of different gender and IQ in a balanced way even ex-post. The estimates are obtained by running OLS regression of various dictator characteristics on covariates of interest with dictator IQ fixed effects. The F statistic shows the joint significance of all covariates. The standard errors are clustered at the dictator level with Pustejovsky and Tipton (2018)'s small cluster bias adjustment are reported below each coefficient estimate. R-squared is net of dictator IQ fixed effects. Unit of observation: dictator-recipient pair. Significance levels: * 10%, ** 5%, and *** 1%.

Table C3: Balance test: Recipient's category (cont.)

Outcome:	Age (recipient)	From Emilia-Romagna (recipient)	Dictator game round 1	Dictator game round 2	Dictator game round 3	Did not know at all	Saw before	Knew but not very well
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
IQHigher	-0.792**	0.188***	-0.084	-0.026	0.110*	-0.002	0.008	-0.006
	(0.374)	(0.050)	(0.065)	(0.064)	(0.061)	(0.026)	(0.022)	(0.018)
Female	-0.284	0.025	-0.084	0.037	0.047	0.020	-0.011	-0.009
	(0.344)	(0.038)	(0.062)	(0.058)	(0.059)	(0.020)	(0.017)	(0.010)
IQHigherxFemale	0.626	-0.100	0.137	-0.084	-0.053	-0.020	0.005	0.014
	(0.462)	(0.062)	(0.084)	(0.079)	(0.084)	(0.026)	(0.025)	(0.020)
Dictator IQ FE	✓	✓	✓	✓	✓	✓	✓	✓
F statistic	1.537	5.51***	0.941	0.89	1.207	0.666	0.415	1.071
R-squared	0.013	0.041	0.006	0.006	0.007	0.047	0.014	0.074
Observations	558	558	558	558	558	558	558	558
Clusters	195	195	195	195	195	195	195	195

Notes: This table shows that dictators were matched recipients of different gender and IQ in a balanced way even ex-post. The estimates are obtained by running OLS regression of various recipient characteristics and round and proximity dummies on covariates of interest with dictator IQ fixed effects. The F statistic shows the joint significance of all covariates. The standard errors are clustered at the dictator level with Pustejovsky and Tipton (2018)'s small cluster bias adjustment are reported below each coefficient estimate. R-squared is net of dictator IQ fixed effects. Unit of observation: dictator-recipient pair. Significance levels: * 10%, ** 5%, and *** 1%.

Table C4: Summary statistics: Dictator data by gender

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

Notes: This table shows summary statistics separately for female and male dictators: the dictators' and the paired recipients' characteristics, how well dictators knew the paired recipients, and dictators' allocation. Recipients whose name is non-Italian sounding and whom the dictator declared they knew them "very well" at least one are not included. Standard deviation of demeaned allocation is calculated as sample standard deviation of $\widetilde{Allocate}_{ik} = Allocate_{ik} - \overline{Allocate}_k$, where $Allocate_{ik}$ is allocation by dictator i whose IQ is k and $\overline{Allocate}_k = \sum_{i \in k} Allocate_{ik}$ is average allocation of dictators with IQ k. P-values for difference in means are calculated with the two-sample t-test with HC2 heteroskedasticity-robust standard errors (MacKinnon and White 1985) with Bell and McCaffrey (2002)'s small sample bias adjustment.

Table C5: Summary statistics: Evaluator data vs. dictator data

	Evaluator		Dictator		Difference
	Mean	SD	Mean	SD	p-value
Panel A: Evaluator / Dictator					
IQ level	6.84	1.14	6.69	1.23	0.21
IQ rank	3.40	1.74	3.58	1.67	0.30
Age	23.34	2.78	23.47	2.72	0.63
From Emilia-Romagna	0.20	0.40	0.18	0.39	0.76
Humanities	0.34	0.48	0.46	0.50	0.02
Social sciences	0.27	0.44	0.19	0.40	0.08
STEM	0.39	0.49	0.35	0.48	0.42
Post bachelor	0.49	0.50	0.46	0.50	0.48
Overconfidence	0.49	0.75	0.43	0.76	0.42
Time on feedback (sec.)	93.26	83.96	107.60	95.60	0.12
Observations	193		195		
Panel B: Paired recipients					
IQ level	6.84	1.16	6.84	1.16	1.00
IQ rank	3.42	1.74	3.42	1.74	0.98
IQ higher	0.50	0.50	0.53	0.50	0.46
Age	23.35	2.80	23.35	2.77	0.99
Female	0.47	0.50	0.47	0.50	0.99
From Emilia-Romagna	0.19	0.40	0.20	0.40	0.87
Observations	368		558		
Panel C: Proximity					
Did not know at all	0.98	0.14	0.96	0.19	0.08
Knew but not well	0.02	0.14	0.03	0.17	0.34
Saw before	0.00	0.00	0.01	0.09	0.03
Observations	368		558		
Panel D: Belief on the recipient's IQ					
Belief on IQ level	6.34	1.89			
Belief on IQ level (demeaned)		1.87			
Observations	368				

Notes: This table shows summary statistics for the evaluators and dictators: the evaluators'/dictators' and the paired recipients' characteristics, how well evaluators/dictators knew the paired recipients, and evaluators' belief. Recipients whose name is non-Italian sounding and whom the dictator declared they knew them "very well" at least one are not included. P-values for difference in means are calculated with the two-sample t-test with HC2 heteroskedasticity-robust standard errors (MacKinnon and White 1985) with Bell and McCaffrey (2002)'s small sample bias adjustment.

Table C6: The role of the recipients' gender and IQ in the dictators' allocation: Further robustness checks

Outcome:	Dictator's allocation (fraction of endowment)		
Sample:	Excluding IQ rank 1 and 6	Excluding proximity 3	Excluding proximity 2 and 3
	(1)	(2)	(3)
IQHigher	0.006	0.011	0.005
	(0.036)	(0.033)	(0.034)
	[-0.065, 0.077]	[-0.056, 0.077]	[-0.062, 0.073]
Female	0.019	0.006	0.008
	(0.029)	(0.026)	(0.027)
	[-0.040, 0.077]	[-0.045, 0.058]	[-0.046, 0.062]
IQHigherxFemale	0.001	0.029	0.029
	(0.044)	(0.037)	(0.038)
	[-0.087, 0.088]	[-0.043, 0.102]	[-0.047, 0.104]
Female+IQHigherxFemale	0.019	0.036	0.037
	(0.034)	(0.026)	(0.026)
	[-0.047, 0.086]	[-0.016, 0.087]	[-0.015, 0.088]
Outcome Mean	0.412	0.402	0.404
Outcome SD	0.234	0.239	0.239
R-squared	0.102	0.083	0.079
Observations	386	553	537
Clusters	135	195	194

Notes: This table shows OLS estimates of the role of recipient's gender and IQ in dictators' giving for various subsamples. The outcome variable is giving in the dictator game as a fraction of endowment. All specifications include dictator IQ fixed effects, round fixed effects, proximity fixed effects, dictator controls, and recipient controls, except column 3 where proximity fixed effects are excluded. Joint statistical significance of coefficient estimate on Female+IQHigherxFemale is calculated using the t-test. The standard error (in parenthesis) and the 95% confidence interval (in bracket) are reported below each coefficient estimate. Standard errors are clustered at dictator level with Pustejovsky and Tipton (2018)'s small cluster bias adjustment. R-squared is net of the number of questions corrects fixed effects. Unit of observation: dictator's allocation. Significance levels: * 10%, ** 5%, and *** 1%.