Dislike Being Corrected?
Corrections and Gender in Collaborative Work

Yuki Takahashi*

February 21, 2024

Abstract

Collaboration is integral in workplaces but also involves correcting one's colleagues, which can cause interpersonal friction, especially when women correct men. Using a pre-registered, quasi-laboratory experiment, I show that men are no less willing to collaborate with a woman than a man who corrected them regardless of the correction quality. However, the exploratory analysis shows that people are less willing to collaborate with others who have corrected them, even if the correction was high quality. Thus, although women correcting men does not cause larger interpersonal frictions, correcting colleagues is costly and can be detrimental to group efficiency and peer learning.

JEL codes: J16, M54, D91, C92

Keywords: Correction, collaboration, group work, gender, quasi-laboratory experiment

*Department of Economics, Tilburg University. Warandelaan 2, Koopmans Building K333, 5037AB Tilburg, The Netherlands. Email: y.takahashi@uvt.nl. This paper was previously circulated as "*Gender Differences in the Cost of Corrections in Group Work*." I am grateful to Maria Bigoni, Siri Isaksson, Bertil Tungodden, and Boon Han Koh, whose feedback was essential for this project, and to the participants of the experiment. This paper also benefited from comments at the CSQIEP Job Market Seminar, Discrimination and Diversity Workshop, Stanford Institute for Theoretical Economics conference, Warwick Economics PhD Conference, among others. The puzzle code used in the experiment is heavily based on Christian König genannt Kersting's code. Lorenzo Golinelli provided excellent technical and administrative assistance. This study was pre-registered with the OSF registry (https://osf.io/tgyc5) and approved by the IRB at the University of Bologna on November 3, 2020 (ref. no. 262643).

"As women, we might be tempted to say, 'Excuse me, you made a small error here.' DO NOT SAY THIS. As non-threatening women, we must avoid that instinct because it serves no one, least of all ourselves."

— Sarah Cooper *How to Be Successful without Hurting Men's Feelings*

1 Introduction

Collaboration is a core element of workplace environments, as most workplaces require group work (Jones 2021; Lazear and Shaw 2007; Wuchty, Jones, and Uzzi 2007). However, collaborative interactions often involve correcting one's colleagues. For example, one may have to correct presentation slides a colleague has prepared that contain errors, or one may have to point out an error in the identification assumption that the presenter is making to improve the paper. These corrections are essential for groups to function well, but they may also damage collaborative relationships if people take corrections personally. This potential psychological friction can be detrimental to group efficiency and peer learning. It is particularly a concern for women: because men dislike being led by women (Abel 2023; Born, Ranehill, and Sandberg 2022; Chakraborty and Serra 2023; Husain, Matsa, and Miller 2023), men may also dislike to be corrected by women. This men's dislike can result in women having difficulty sustaining collaborative relationships with their male colleagues and be detrimental to their careers. In this paper, I examine this question.

In the pre-registered analysis, I study whether men are less willing to collaborate with a woman who corrected them, whether such behavior leads to a suboptimal collaborator selection, and whether men's gender bias is the underlying mechanism. In the non-pre-registered exploratory analysis, I focus on a more general aspect of the interplay between correction and collaboration and study whether people are less willing to collaborate with someone who corrected them (regardless of the decision maker's or the corrector's gender), whether such behavior leads to a suboptimal collaborator selection, and whether it is due to misunderstanding of the correction quality. I define collaboration as working with others toward the same goal and correction as overriding what others have done. For the collaborative task, I use Isaksson (2018)'s number-sliding puzzle, which allows me to calculate an objective measure of each participant's contribution to the collaborative task and classify each move either as good (moving the puzzle closer to the solution) or bad (moving the puzzle further away from the solution).

I first confirm that the participants understand the notion of good and bad moves; the more a participant contributes to solving the puzzle, the more likely they are selected to be a collaborator. Also, the puzzles are not too difficult for the participants. Further, men and women contribute equally well to the puzzle and do not underestimate or overestimate women's contribution relative to men's.[1]

In the pre-registered analysis, I do not find evidence that men are less willing to collaborate with a woman who corrected them than with a man who corrected them after controlling for the individual contributions. I also do not find men's differential response to women's vs. men's corrections even

1. This is a composite of their belief about women's ability relative to men's and their preference for working with women relative to men.

when I split the corrections into good and bad corrections at the 5% significance level or when I split men into high and low gender bias measured by the sexism questions used in Karpowitz et al. (2023).

In the non-pre-registered exploratory analysis, I find that people, both men and women, are less willing to collaborate with someone who corrected them, even if the corrections are good, efficiency-improving ones. The effect size is about a 12-15 percentage points reduction, or about a 15%-19% reduction relative to the baseline mean, in one's willingness to collaborate, which requires making about 0.50-0.61 standard deviations of additional good moves to compensate. It is unlikely that the participants misunderstood good corrections as bad ones, as even higher-ability participants – those who should be better able to identify good and bad corrections – respond negatively to corrections. These results suggest that while an average man is no less willing to collaborate with a woman than a man who corrected them, at least in a gender-neutral task, correcting colleagues is costly regardless of gender and can be detrimental to group efficiency and peer learning in collaborative work.

This paper's contribution is threefold. First, it contributes to the literature on peer learning at the workplace by showing that aversion to being corrected can inhibit peer learning. The literature shows peer learning reduces wage inequality among workers (Jarosch, Oberfield, and Rossi-Hansberg 2021) and increases worker productivity (Sandvik et al. 2020). However, Chandrasekhar, Golub, and Yang (2019) show that the shame of revealing and signaling ignorance are potential barriers to peer learning. My paper reveals that an aversion to being corrected is an additional barrier to peer learning.

Second, my paper contributes to the literature on workplace climate and productivity by showing that aversion to being corrected can distort group efficiency. My findings complement Alan, Corekcioglu, and Sutter (2022), who find that a better workplace climate increases worker satisfaction and the degree of mutual reciprocation while reducing toxic competition and worker turnover. Aside from Alan et al., my findings also relate to the organizational economics literature: for example, Edmans (2011) finds that firms with high employee satisfaction exhibit higher stock prices. Although some studies find the same environment can affect women and men differently, for example, at economics seminars (Dupas et al. 2021), I do not find such evidence. It seems genderness of the environment matters whether women and men receive differential treatment, as Folke and Rickne (2022) find that women in male-dominant jobs receive more harassment, *and* men in female-dominant jobs receive more harassment.

Finally, this paper contributes to the literature on the differential treatment of women's opinions. My paper primarily relates to Coffman, Flikkema, and Shurchkov (2021), who find that group members are less likely to choose women's answers as a group answer in male-typed questions, and Guo and Recalde (2023), who find that group members correct women's ideas more often than men's in slightly male-typed tasks. As a correction is one form of expressing one's ideas and the task is gender-neutral, my findings seem to suggest that the task genderness matters in how people treat women's opinions differently to men's opinions. This is consistent with the social psychology literature that argues it is gender-incongruent behaviors that receive backlash (Eagly and Karau 2002; Heilman et al. 2004; Rudman et al. 2012). Relatedly, Klinowski (2023) finds that female scientists are less likely to publish comments that criticize or correct other scientists' publications than male scientists, which could be due to fears of a backlash.

The paper closest to mine is Isaksson (2018), who uses the same puzzle and shows that women claim their contribution less than men, especially in difficult puzzles, and that men correct their partners more often than women. I answer different research questions with a different experimental design using Isaksson's puzzle: whether receiving corrections reduces one's willingness to collaborate with that person and whether men react more negatively to women's corrections using a design adapted from Fisman et al. (2006, 2008)'s speed dating experiments but for preference for a working partner instead of a romantic partner.

The remainder of the paper proceeds as follows. In Section 2, I describe the experimental design, procedure, and implementation. Next, I briefly describe the data obtained from the experiment in Section 3. Then, I proceed to empirical analysis: I present the empirical strategy in Section 4, the pre-registered results in Section 5, and the non-pre-registered results in Section 6. I conclude the paper in Section 7.

2 Experiment

Introducing the quasi-laboratory format I run the experiment in a quasi-laboratory format where we experimenters and the participants are connected via Zoom throughout the experiment, but turn off participants' cameras and microphones except at the beginning of the experiment. Aside from that participants participate remotely using their computers, the experiment is conducted as it would be in a physical laboratory.

The collaborative task I use Isaksson (2018)'s puzzle as the collaborative task. It is a sliding puzzle with eight numbered tiles which need to be placed in numerical order within a 3x3 frame (see Figure 1 for an example). To achieve this goal, participants play in pairs, alternating their moves.[2] This puzzle has nice mathematical properties: I can objectively define the puzzle's difficulty and classify a given move as either good or bad via the Breadth-First Search algorithm.[3] Based on the number of good and bad moves a participant makes, I can measure individual contributions to the task as net good moves, the number of good moves minus the number of bad moves an individual makes in a given puzzle.

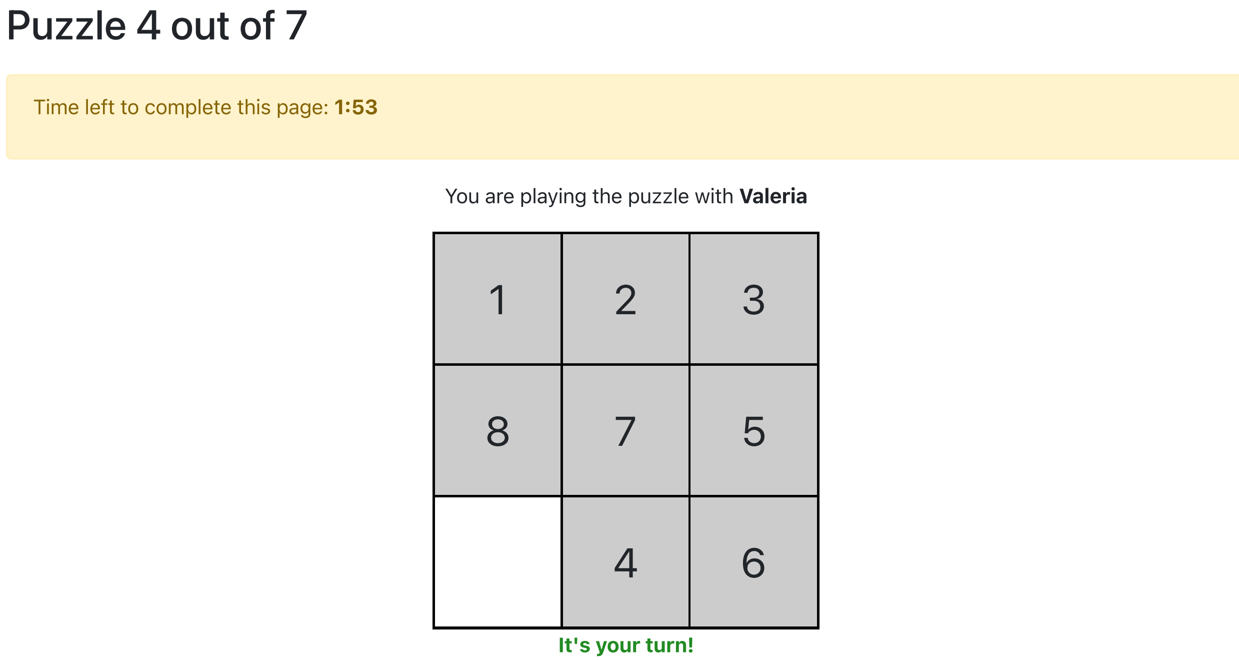
I define a correction as reversing the pair's move, which allows me to objectively compare the quality of corrections by different participants because corrections are also moves.[4] Further, puzzle-solving captures an essential characteristic of collaborative work in which two or more people work towards the same goal (Isaksson 2018), but the quality of each move and correction is only partially observable to participants (but fully observable to the experimenter). This partial observability allows participants

2. Each participant has to make a move during their turn; they cannot pass.

3. The difficulty is defined as the number of moves away from the solution; a good move is defined as a move that reduces the distance (in number of moves) to the solution, while a bad move is defined as a move that increases the distance to the solution.

4. Because some corrections happen early in the puzzle and the others later in the puzzle, what I capture in the analysis is the average effect of a correction.

Figure 1: Puzzle screen



Notes: This shows a sample puzzle screen where a participant is matched with another participant called Valeria in the 4th round of the puzzle and makes their move. All the texts are in Italian in the experiment.

to engage in motivated reasoning (Kunda 1990; Chance and Norton 2015; Gino, Norton, and Weber 2016), interpreting the quality of corrections in a self-serving manner.

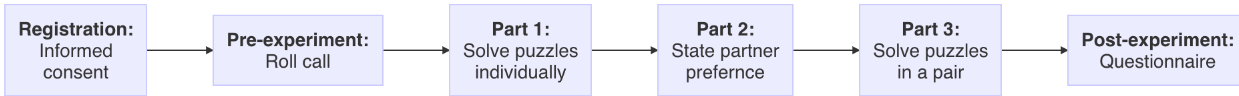
At each stage of the puzzle, there is only one good strategy, which is to make a good move, and one bad strategy which is to make a bad move.[5] There can be more than one good and bad move, but different good/bad moves are equal. There is no path dependence either: the history of the puzzle moves does not matter.

2.1 Design and procedure

The experiment consists of three parts. In part 1, participants work on the puzzle individually. The purposes of this part are to familiarize participants with the puzzle and to measure their puzzle solving ability. In part 2, participants first learn the rules of part 3, then state their collaborator preference for part 3 by solving one puzzle with each of the potential collaborators. In part 3, participants work on the puzzles with the collaborator selected based on their preference stated in part 2. At the beginning of each part, participants must answer a set of comprehension questions to ensure they understand the instructions. Figure 2 summarizes the flow of the experiment, which I explain in detail below.

5. This assumes that both players are trying to solve the puzzle; I show in Figure 6 that the results are robust to the exclusion of puzzles where either player might not be trying to solve the puzzle.

Figure 2: Flowchart of the experiment



Registration

Upon receiving an invitation email to the experiment, participants register for the session they want to participate in, and upload their ID documents as well as a signed consent form.[6]

Pre-experiment

Participants enter the Zoom waiting room on the day and at the time of the session they have registered for.[7] They receive a link to the virtual room for the experiment and enter their first name, last name, and the email address they used in the registration. This information is necessary to match up their earnings in this experiment to their payment information stored in the laboratory database, so participants have a strong incentive to provide their true name and email address. After that, they draw a virtual coin numbered from 1 to 40 without replacement.

As participants arrive and are verified, I admit them to the Zoom meeting room one by one and rename them with their first name. If there is more than one participant with the same first name, I add a number after their first name (e.g., Giovanni2).

After admitting all the participants to the Zoom meeting room, I do a roll call to reveal participants' gender to other participants without making their gender salient (Bordalo et al. 2019; Coffman, Flikkema, and Shurchkov 2021). Specifically, I take attendance by calling each participant's first name one by one and asking them to respond verbally via their microphone. This process ensures other participants that the called participant's first name corresponds to their gender. If there are more participants than I need for the session (i.e., more than 16 participants), I draw random numbers from 1 to 40 and ask those who have the coins with the drawn number to leave.[8] Those who leave the session receive a 2€ show-up fee. Figure 3 shows the Zoom screen the participants would see during the roll call: the experimenter's camera is turned on while the participants' cameras are turned off; participants see this screen throughout the experiment.[9]

I then read out the instructions explaining the rules of the experiment and take questions on Zoom via microphone. Once participants start the main part of the experiment, they can only communicate

6. I recruited a few more participants than needed for each session in case some participants did not show up to the session.

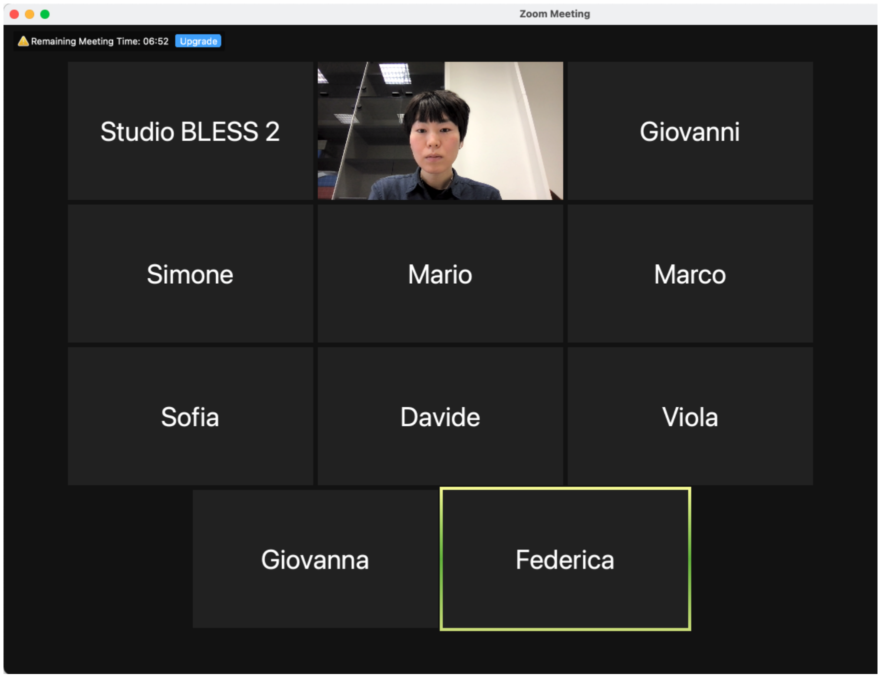
7. The Zoom link is sent with an invitation email; I checked that each participant in the waiting room had indeed registered for that session before admitting them to the main room.

8. I draw with replacement a number from 1 to 40 using Google's random number generator (https://www.google.com/search?q=random+number). If no participant has a coin with the drawn number, I draw the next number until the number of participants is 16. I share my computer screen during this process so that participants see the numbers are actually drawn randomly.

9. The experimenter's camera may occasionally be turned off during the experiment.

with the experimenter via Zoom's private chat.

Figure 3: Zoom screen



Notes: This figure shows the Zoom screen participants would see during the roll call. The the experimenter's camera is turned on while the participants' cameras are turned off; participants see this screen throughout the experiment.

Part 1: Individual practice stage

At the beginning of part 1, I explain to participants in depth how to solve the puzzle efficiently (in minimum moves) and provide comprehension questions about the solution strategies; see the instructions in Appendix C.[10] Participants then work on the puzzle individually with an incentive (0.2€ for each puzzle they solve). They can solve as many puzzles as possible in 4 minutes (maximum 15 puzzles) with increasing difficulty. After the 4 minutes are over, they receive information on how many puzzles they have solved. This part familiarizes them with the puzzle and gives me a measure of their ability based on how many puzzles they solve.

Part 2: Collaborator selection stage

Part 2 consists of seven rounds, and participants learn the rules of part 3 before starting part 2. This part is based on Fisman et al. (2006, 2008)'s speed dating experiments and proceeds as follows. First, participants are divided into groups of eight, with participants of similar ability measured in part 1 placed in the same group. This is to reduce ability differences among participants to make corrections and gender more salient; participants are not told about this grouping criterion.

10. However, I do not tell participants that they can correct others to reduce experimenter demand effects.

Second, participants are paired with another randomly chosen participant in the same group, and they solve one puzzle together by alternating their moves. The participant who makes the first move is drawn at random, and both participants know this first-mover selection criterion. If they cannot solve the puzzle within 2 minutes, they finish the puzzle without solving it. Participants are allowed to reverse – that is, correct – their partner's move.[11] Each participant's contribution to a given puzzle is measured by net good moves. Figure 1 shows a sample puzzle screen where one participant is paired with another participant called Valeria and is making their move. Each partner's first name is displayed on the computer screen throughout the puzzle, and when participants select their collaborator, thus subtly informing on the partner's gender.

Once they finish the puzzle, participants state in private whether they would like to collaborate with the same participant in part 3 (yes/no). At the end of the first round, new pairs are formed with a perfect stranger matching procedure, so that every participant is paired with each of the other seven members of their group once and only once and rounds 2 to 7 proceed exactly in the same way as round 1. The sequence of puzzles is the same for all pairs in all sessions. The puzzle difficulty is kept the same across all seven rounds. I set the minimum number of moves to solve the puzzles to be 8, based on a pilot, so that the puzzles are neither too easy nor too difficult to solve.

At the end of part 2, participants are paired according to the following algorithm, adapted from Fisman et al. (2006, 2008):

1. For every participant i, I count the number of matches; that is, the number of other participants in the group who were willing to collaborate with i and with whom i is willing to collaborate in part 3.
2. I randomly choose one participant.
3. If the chosen participant has only one match, I pair them up and let them work together in part 3.
4. If the chosen participant has more than one match, I randomly choose one of the matches.
5. I exclude participants who have been paired and repeat (1)-(3) until no feasible match is left.
6. If some participants are still left unpaired, I pair them up randomly.

At the beginning of part 2, I explain in depth this pairing algorithm along with comprehension questions so that the collaborator preference statement is incentivized.

The assumption for this matching algorithm to be incentive compatible is that payoff is the primary concern for participants. While this may not be a valid assumption in real life because people may also care about their ability relative to their collaborator, for example, these factors do not play important roles in my experiment. Other individual-specific factors are controlled for by exploiting within-subject variation. Abilities of the past and the future partners' abilities are random because of the random pairing of participants within the same group.

11. Solving the puzzle itself is not incentivized, so participants who do not want to collaborate with a given partner or fear receiving a bad response may not reverse that partner's move, even if they think the move is wrong. However, since I am interested in the effect of correction on collaborator selection, participants' *intentions* to correct that do not end up as an actual correction do not confound the analysis.

Part 3: Group work stage

The paired participants work together on the puzzles by alternating their moves for 12 minutes and earn 1€ for each puzzle solved. Which participant makes the first move is randomized at each puzzle, which is told to both participants as in part 2. They can solve as many puzzles as possible (maximum 20), with increasing difficulty.

Post-experiment

Each participant answers a short questionnaire, which consists of (i) the six hostile and benevolent sexism questions used by Karpowitz et al. (2023) with US college students and (ii) their basic demographic information and their impressions about the experiment. Answers to the first set of questions is used to construct the participants' gender bias measure and the answer to the second set of questions is used to know participants' characteristics and their impressions to casually check whether they have anticipated that the experiment was about gender.[12]

After participants answer all the questions, I tell them their earnings privately and let them leave the virtual room and close Zoom. They later receive their earnings via PayPal.

2.2 Implementation and participant characteristics

The experiment was programmed with oTree (Chen, Schonger, and Wickens 2016) and conducted in Italian during November-December 2020. I recruited 464 participants (220 male and 244 female), all registered at the Bologna Laboratory for Experiments in Social Science's ORSEE (Greiner 2015). I restrict the participant pool to those who (i) were students, (ii) were born in Italy, and (iii) had not participated in gender-related experiments before (as far as I could check).[13] The first two conditions were to reduce noise coming from differences in socio-demographic backgrounds and race and/or ethnicity that may be inferred from participants' first names and/or voices; the last condition was to reduce experimenter demand effects.[14] The number of participants was determined by a power simulation in the pre-analysis plan to achieve 80% power for the first hypothesis (H1 in Section 4); I used two-sided t-test and set the type I error probability to 5%.[15] The experiment was pre-registered with the OSF.[16]

I ran 29 sessions with 16 participants each. The average duration of a session was 70 minutes. The average total payment per participant was 11.55€ with a maximum of 25€ and a minimum of 2€, including the 2€ show-up fee.

Online Appendix Table B1 describes the participants' characteristics. It shows that male participants are slightly older (1.41 years) and more gender-biased (0.12 points) than female participants. In

12. I did not find any evidence that the participants anticipated that it was an experiment about gender.

13. The laboratory prohibits deception, so no participant participated in an experiment with deception.

14. Despite only recruiting people born in Italy, 1 male participant answered in the post-questionnaire that he was born abroad. I included this participant in the analysis anyway but the results are robust to excluding him.

15. This number includes 16 participants from a pilot session run before the pre-registration, where the experimental instructions were slightly different. The results are robust to the exclusion of these 16 participants.

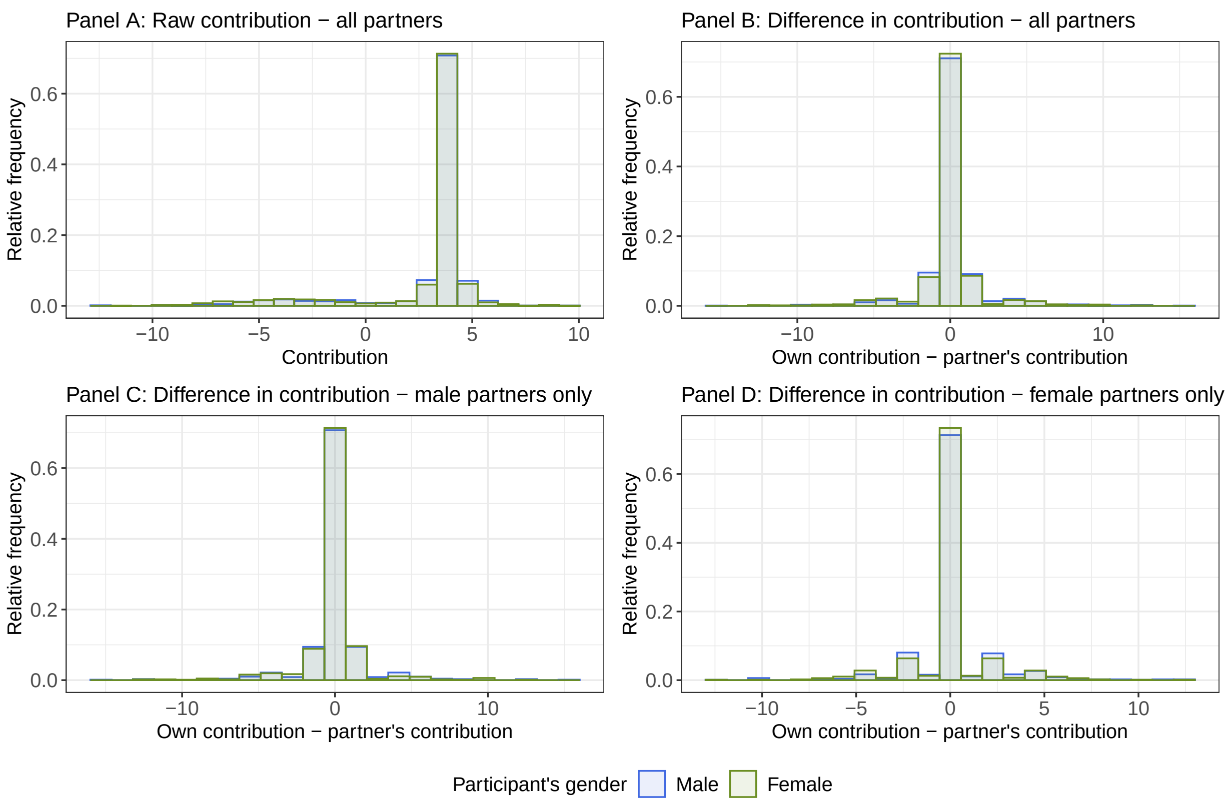
16. You can find the OSF registry at the following URL: https://osf.io/tgyc5.

addition, male participants are more likely to major in natural sciences and engineering and less likely to major in humanities, a tendency observed in most OECD countries (see, for example, Carrell, Page, and West 2010).[17] Also, most participants are either bachelor's or master's students, and only a few are PhD students. I confirmed no economics PhD students participated in the experiment.

3 Data

I use part 2 data in the analysis, as that is where we can observe collaborator selection decisions. I aggregate the move-level data at each puzzle level so that I can associate behaviors in the puzzle to the collaborator selection decisions.[18] Online Appendix A describes the data in detail; I briefly discuss its key aspects below.

Figure 4: Distribution of contributions



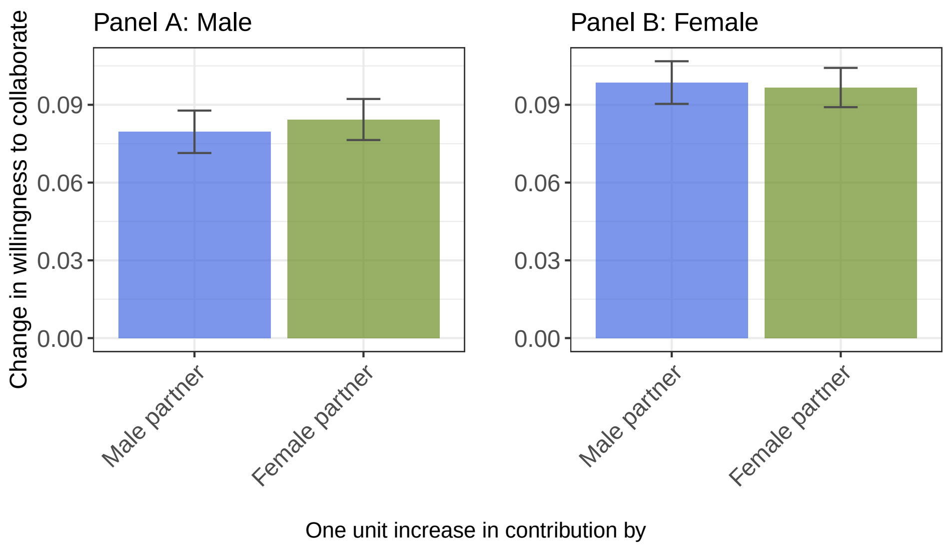
Notes: This figure shows the distribution of individual contributions by gender for raw contribution (panel A), difference in contribution between own and partner's (panel B), same as Panel B but with male partners only (Panel C), with female partners only (Panel D). Contribution is defined as one's net good moves in a given puzzle (the number of good moves minus the number of bad moves).

17. Individual fixed effects in the analysis control for participants' major.

18. Online Appendix Figure B1 summarizes the move-level data and shows no statistically significant differences in move quality by one's own gender or the gender of one's partner, no systematic differences in the probability that correction is happening by one's own gender or the gender of one's partner, and one's own gender or the gender of one's partner does not affect how fast participants solve the puzzle.

Panel A of Figure 4 presents the distribution of contribution by participants' gender and shows most participants contribute to the same degree and that men and women are equally good at puzzle-solving: in about 70% of the puzzles, each participant's contribution is 4 (total good moves minus total bad moves), and the men's and women's distributions almost overlap. Also, Panel B presents the distribution of difference in contribution between own and partner's, and shows that the pattern in Panel A does not come from different puzzles: both participants contributed the same degree in a given puzzle in about 70% of the cases, and it is true for both male and female participants. This is true regardless of the partner's gender (Panel C for male partners and Panel D for female partners).

Figure 5: Change in willingness to collaborate by one unit increase in partner's contribution



Notes: This figure shows men's (Panel A) and women's (Panel B) change in willingness to collaborate by one unit increase in male (blue) and female (green) partner's contribution, along with the 95% confidence intervals. The confidence intervals are calculated with standard errors are clustered at the individual level.

In addition, men or women do not underestimate or overestimate women's contribution. Figure 5 shows men's (Panel A) and women's (Panel B) change in willingness to collaborate by one unit increase in male (blue) and female (green) partner's contribution, along with the 95% confidence intervals. The figure shows men and women are more willing to collaborate with someone who contributes more, regardless of that person's gender. The increase is slightly smaller for male participants, possibly due to the overconfidence in their ability (and hence underestimation of the partner's ability). Figure 5 also suggests that participants understand the notion of good and bad moves; they do not respond to partner's contributions otherwise.[19]

Finally, key variables are balanced across rounds even ex-post as shown in Online Appendix Figure B3. A few variables have some imbalance, especially in rounds 6 and 7, where participants are less

19. As inferred from the positive contribution, the fact that most puzzles were solved, and that participants are more willing to collaborate with someone who contribute more, the puzzles are not too difficult to the participants; see Online Appendix Figure B2 that plots the participants' perceived puzzle difficulty from the post-questionnaire.

willing to collaborate, experience more corrections, and are less likely to solve the puzzle. I will show in Figure 6 that the results are robust to the exclusion of these rounds.[20]

4 Empirical strategy

I test the following pre-registered hypotheses:

H1 Men are less willing to collaborate with a woman than with a man who corrects them

H2 The behavior conjectured in H1 leads to a suboptimal collaborator selection

H3 Men's gender bias is the underlying mechanism of the behavior conjectured in H1

I also test the following non-pre-registered exploratory hypotheses to better understand a more general aspect of the interplay between correction and collaboration and provide policy implications of the pre-registered hypotheses.

H1S People are less willing to collaborate with someone who corrects them

H2S The behavior conjectured in H1S leads to a suboptimal collaborator selection

H3S The behavior conjectured in H1S is not due to misunderstanding of the correction quality

4.1 Empirical strategy for pre-registered hypotheses

Test for H1

To test H1, I estimate the following model with OLS using male decision maker sample only.

$$Select_{ij} = \beta_1 Corrected_{ij} \times Female_j + \beta_2 Corrected_{ij} + \beta_3 Female_j + \delta Contribution_j + \mu_i + \epsilon_{ij} \tag{1}$$

where each variable is defined as follows:

- $Select_{ij} \in \{0,1\}$: an indicator variable equals 1 if i selects j as their collaborator, 0 otherwise.
- $Corrected_{ij} \in \mathbb{N}$: the number of times j corrected i in a given puzzle (including 0).
- $Female_j \in \{0,1\}$: an indicator variable equals 1 if j is female, 0 otherwise.
- $Contribution_j \in \mathbb{Z}$: j's contribution to a puzzle played with i.
- ϵ_{ij}: omitted factors that affect i's likelihood to select j as their collaborator.

and $\mu_i \equiv \sum_{k=1}^{N} \mu^k \mathbb{1}[i = k]$ are the individual fixed effects, where N is the total number of male participants in the sample and $\mathbb{1}$ is the indicator variable. Standard errors are clustered at the individual level.[21]

I exploit the random pairing of participants conditional on individual fixed effects for causal identification. It is conditional because the grouping of eight people in part 2 is based on each participant's performance in part 1. Specifically, I control for everything a participant can observe

20. Another imbalance is the average partner gender balance (fraction of female partners) between male and female participants in round 1. However, it is balanced in rounds 2-7.

21. This is because the treatment unit is i. Although the same participant appears twice (once as i and once as j), j is passive in collaborator selection. However, the results are robust when we cluster standard errors at the pair level instead; see Online Appendix Table B2. The same is true for non-pre-registered results; see Online Appendix Table B5.

about their partner during the puzzle: the partner's gender, whether the partner makes a correction, and the partner's puzzle-solving ability the participant perceives. Conditional on them, correction occurs due to a specific initial puzzle configuration with the same objective difficulty across rounds but some participants may find it more difficult than others. I show in Figure 6 that time to solve the puzzle is not mediating the results.

The coefficient of interest is β_1, which shows whether male participants are less willing to collaborate with a woman than with a man who corrected them.

Test for H2

To test H2, I estimate the following model with OLS using male decision maker sample only.

$$\begin{aligned} Select_{ij} =& \beta_1 Corrected_{ij} \times Female_j + \beta_2 Corrected_{ij} + \beta_3 Female_j + \beta_4 CorrectedGood_{ij} \times Female_j \\ &+ \beta_5 CorrectedGood_{ij} + \delta Contribution_j + \mu_i + \epsilon_{ij} \end{aligned} \tag{2}$$

where each variable is defined as follows:

- $CorrectedGood_{ij} \in \mathbb{N}$: the number of times j corrected i that moved the puzzle closer to the solution in a given puzzle (including 0).

Other variables are defined in the same way as equation 1.

The coefficient of interest is β_4, which shows whether male participants are much less willing to collaborate with a woman than with a man who corrected their mistake, compared to another woman than another man who corrected their right move.

Test for H3

To test H3, I estimate the following model with OLS using male decision maker sample only.

$$\begin{aligned} Select_{ij} =& \beta_1 Corrected_{ij} + \beta_2 Female_j + \beta_3 Corrected_{ij} \times Female_j + \beta_4 Corrected_{ij} \times HigherBias_i \\ &+ \beta_5 Female_j \times HigherBias_i + \beta_6 Corrected_{ij} \times Female_j \times HigherBias_i \\ &+ \delta Contribution_j + \mu_i + \epsilon_{ij} \end{aligned} \tag{3}$$

where each variable is defined as follows:

- $HigherBias_i \in \{0, 1\}$: an indicator variable equals 1 if i's gender bias score from the six hostile and benevolent sexism questions is above median among male participants, 0 otherwise.

Other variables are as defined in the same way as equation 1.

The coefficient of interest is β_6, which shows compared to male participants with lower gender bias, whether male participants with higher gender bias are less willing to collaborate with a woman than with a man who corrected them.

Changes to the pre-registered specifications

I make one change to the pre-registered specifications: I modify the definition of contribution from the one in the pre-analysis plan to prevent truncation: the pre-specified contribution measure is truncated from below and above in more than 10% of the puzzle. Nonetheless, the results are quantitatively the same when I use the original contribution measure, reported in Online Appendix Table B3.

Significance level cutoff

I use 5% (two-sided) as the cutoff of statistical significance as I did in the power simulation in the pre-analysis plan.

4.2 Empirical strategy for non-pre-registered exploratory hypotheses

Test for H1S

To test H1S, I estimate the following model with OLS using the whole decision maker sample (both male and female).

$$Select_{ij} = \beta_1 Corrected_{ij} + \beta_2 Female_j + \delta Contribution_j + \mu_i + \epsilon_{ij} \tag{4}$$

where variables are as defined in the same way as equation 1.

The coefficient of interest is β_1, which shows whether participants are less willing to collaborate with someone who corrected them than others who did not correct them.

Test for H2S

To test H2S, I estimate the following model with OLS using the whole decision maker sample (both male and female).

$$Select_{ij} = \beta_1 Corrected_{ij} + \beta_2 CorrectedGood_{ij} + \beta_3 Female_j + \delta Contribution_j + \mu_i + \epsilon_{ij} \tag{5}$$

where variables are as defined in the same way as equations 1 and 2.

The coefficient of interest is β_2, which shows whether participants are less willing to collaborate with someone who corrected their mistakes than others who corrected their right moves.

Test for H3S

To test H3S, I estimate the following model with OLS using the whole decision maker sample (both male and female).

$$\begin{aligned} Select_{ij} =& \beta_1 Corrected_{ij} + \beta_2 CorrectedGood_{ij} + \beta_3 Female_j + \beta_4 Corrected_{ij} \times HigherAbility_i \\ &+ \beta_5 CorrectedGood_{ij} \times HigherAbility_i + \delta Contribution_j + \mu_i + \epsilon_{ij} \end{aligned} \tag{6}$$

where each variable is defined as follows:

- $HigherAbility_i \in \{0, 1\}$: an indicator variable equals 1 if i solved an above-median number of puzzles in part 1 in a session they participated in, 0 otherwise.

Other variables are as defined in equations equations 1 and 2.

The coefficient of interest is β_5, which shows compared to lower-ability participants, whether higher-ability participants are less willing to collaborate with someone who corrected their mistakes than others who corrected their right moves.

To interpret the results from this test as evidence in support of H3S requires some caution, because higher-ability participants could be more over-confident (they receive feedback about their absolute puzzle-solving ability in the part 1 practice stage). Thus those higher-ability participants are more able than lower-ability participants to distinguish good and bad moves, but also are more likely to believe that their partners' corrections are bad even if they are good. Thus, the estimate of β_5 will likely capture the both opposing effects. Yet, it will provide still suggestive evidence of H3S.

5 Pre-registered results

5.1 Results for H1

Table 1 presents the regression results of equations 1 (columns 1-2), 2 (columns 3-4), and 3 (columns 5-6). Columns 1, 3, and 5 omit the partner's contribution.

Looking at column 1, the coefficient estimate on the correction is negative and statistically significant.[22] On the other hand, the coefficient estimate on the interaction between the correction and the female partner dummy is statistically insignificant. The coefficient estimate on the partner's contribution in column 2 is positive and quantitatively and statistically significant, suggesting that men understand the notion of good and bad moves, as we saw in Figure 5.[23] Also, the coefficient estimate on the correction halved after controlling for the partner's contribution, and the coefficient estimate on the interaction between the correction and the female partner dummy is mostly unchanged. Thus, I do not find evidence in support of H1 and men are no less willing to collaborate with a woman than a man who corrected them.

5.2 Results for H2

Looking at column 3, the coefficient estimate on the correction, which captures the effect of bad corrections on one's willingness to collaborate, is statistically significant and very negative – more negative than the coefficient estimate on the correction in column 1 that captures both good and bad corrections. Also, the coefficient estimate on the good correction is positive and statistically significant. These further suggest that men understand the notion of good and bad moves.[24] The coefficient estimate on the interaction between the correction and the female partner dummy is statistically

22. This is true for female participants as shown in Online Appendix Table B4.
23. This is true for women as shown in Online Appendix Table B4.
24. This is true for women as shown in Online Appendix Table B4.

Table 1: Men's response to women's corrections

Outcome:	Willing to collaborate (yes=1, no=0)					
Sample:	Male decision makers					
	(1)	(2)	(3)	(4)	(5)	(6)
Correction	-0.152***	-0.070*	-0.267***	0.001	-0.154***	-0.086*
	(0.028)	(0.030)	(0.052)	(0.046)	(0.041)	(0.040)
Good correction			0.182**	-0.104+		
			(0.064)	(0.062)		
Female partner	0.013	0.023	0.018	0.020	0.024	0.018
	(0.026)	(0.022)	(0.026)	(0.022)	(0.035)	(0.029)
Partner's contribution		0.078***		0.084***		0.079***
		(0.003)		(0.003)		(0.003)
Correction x Female partner	-0.027	-0.021	-0.115	0.053	-0.011	-0.010
	(0.045)	(0.035)	(0.082)	(0.071)	(0.054)	(0.046)
Good correction x Female partner			0.091	-0.082		
			(0.104)	(0.085)		
Correction x Higher bias					0.001	0.029
					(0.056)	(0.058)
Female partner x Higher bias					-0.029	0.006
					(0.053)	(0.045)
Correction x Female partner x Higher bias					-0.026	-0.017
					(0.092)	(0.068)
Individual FE	✓	✓	✓	✓	✓	✓
H0: Correction x Female partner +Good correction x Female partner=0 (p-value)			0.656	0.490		
H0: Correction x Female partner x Higher bias +Correction x Female partner=0 (p-value)					0.619	0.587
Baseline mean	0.747	0.747	0.747	0.747	0.749	0.749
Baseline SD	0.435	0.435	0.435	0.435	0.434	0.434
Adj. R-squared	0.061	0.304	0.075	0.308	0.060	0.303
Observations	1510	1510	1510	1510	1503	1503
Individuals	220	220	220	220	219	219

Notes: This table presents the regression results of equations 1 (columns 1-2), 2 (columns 3-4), and 3 (columns 5-6). Baseline mean and standard deviation are participants' willingness to collaborate with male partners who do not make any corrections. Standard errors in parentheses are clustered at the individual level. Significance levels: + 10%, * 5%, ** 1%, and *** 0.1%.

insignificant, neither does the coefficient estimate on the interaction between the good correction and the female partner dummy. Thus, the evidence does not support H2 and men's behavior does not lead to a suboptimal collaborator selection.

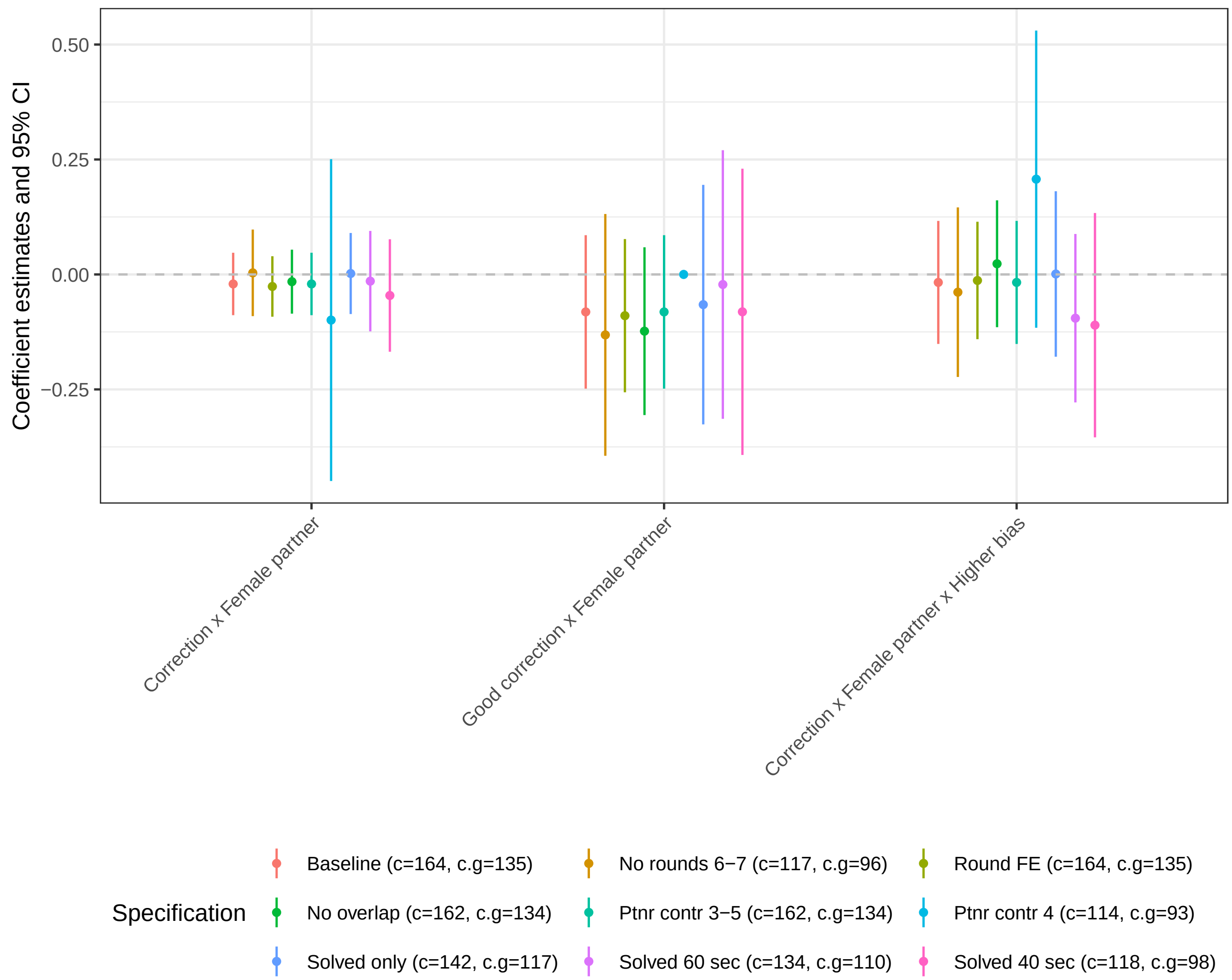
5.3 Results for H3

Looking at column 5, the coefficient estimate on the correction is again statistically significantly negative and about the same magnitude as column 1. However, the coefficient estimate on the interaction between the correction, the female partner dummy, and the higher bias dummy is statistically insignificant. While the coefficient estimate on the correction halves after controlling for the partner's contribution, shown in column 6, the coefficient estimate on the interaction between the correction, the female partner dummy, and the higher bias dummy remains statistically insignificant.

Thus, there is no evidence that men with higher gender bias respond more negatively to women's corrections.

5.4 Robustness of the main results

Figure 6: Men's response to women's corrections: Robustness



Notes: This figure re-estimates and plots the coefficient estimates (dots) and 95% confidence intervals (lines) of columns 2, 4, and 6 of Table 1 with various different specifications, with the specification "Baseline" being the same as specifications of those tables for comparison. The number of corrections (c) and good corrections (c.g) in each sample are indicated in the parenthesis next to the specification.

Figure 6 re-estimates and plots the coefficient estimates (dots) and 95% confidence intervals (lines) of columns 2, 4, and 6 of Table 1 with various different specifications to address concerns raised in Sections 3 and 4. "Baseline" specification is what is presented in the Table 1 for a reference. "No rounds 6-7" excludes rounds 6 and 7 to address across round imbalance in Figure B3. "Round FE" adds round fixed effects. "No overlap" excludes 60 puzzles where both good and bad corrections occurred. "Ptnr contr 3-5" restricts the sample to puzzles where the partner's contribution is between

3 to 5 and "Ptnr contr 4" further restricts the sample to puzzles where the partner's contribution is 4 – about 70% of the sample – to show the robustness to that the partner's contribution is linear in the main specification. "Solved only" restricts the sample to solved puzzles only to address the concern that "a good move is only preferable if you are playing with a partner who is also trying to solve the puzzle" (Isaksson 2018, p. 25). "Solved 60 sec" restricts the sample to puzzles that are solved within 60 seconds (1/2 of the total time), which is a subset of solved puzzles only by definition, and "Solved 40 sec" further restricts the sample to puzzles that are solved within 40 seconds (1/3 of the total time) to show that the results are not mediated through the time to solve the puzzle. None of them present qualitatively different results.

6 Non-pre-registered exploratory results

The results from the pre-registered analysis show no evidence that men are less willing to collaborate with a woman relative to a man who corrected them. The natural question we ask next is whether it is costly to correct one's colleagues and how much, even though women and men's costs are not different. In this section, I test this non-pre-registered exploratory question by examining a more general aspect of the interplay between correction and collaboration.

6.1 Results for H1S

Table 2 presents the regression results of equations 4 (columns 1-2), 5 (columns 3-4), and 6 (columns 5-6). Columns 1, 3, and 5 omit the partner's contribution.

Looking at column 1, the coefficient estimate on the correction is negative and statistically significant. Even after we take out the contribution part of the correction, in column 2, the coefficient estimate remains negative and statistically significant: making a correction reduces a chance to be selected as a collaborator by 12 percentage points or about 15% relative to the baseline mean. To offset this negative effect, one needs to make 1.4 units ($\approx -.12/.084$) or 0.50 standard deviations (1.4/(2.87+2.73)*2) of additional contributions.[25] Thus, the evidence is consistent with H1S and people are less willing to collaborate with someone who corrects them.

6.2 Results for H2S

Looking at column 3, the coefficient estimate on the correction is negative and statistically significant while the coefficient estimate on the good correction is positive and statistically significant, as we should expect. However, after taking out the contribution part of the good correction, in column 4, the coefficient estimate on the good correction becomes negative and statistically significant: making a good correction reduces a chance to be selected as a collaborate by 15 percentage points ($\approx$-0.147=-0.044-0.103) or about 19% relative to the baseline mean, and one needs to make additional 1.7 units ($\approx$-0.147/.087) or about 0.61 standard deviations of additional contributions to offset it.

25. 2.87 is standard deviation of contributions of partners paired with male participants and 2.73 is that of partners paired with female participants; see Online Appendix Table A1.

Table 2: People's response to corrections (non-pre-registered)

Outcome:	Willing to collaborate (yes=1, no=0)					
Sample:	All decision makers					
	(1)	(2)	(3)	(4)	(5)	(6)
Correction	-0.205***	-0.120***	-0.362***	-0.044	-0.397***	-0.040
	(0.015)	(0.013)	(0.037)	(0.027)	(0.049)	(0.041)
Good correction			0.220***	-0.103**	0.275***	-0.089+
			(0.046)	(0.032)	(0.059)	(0.047)
Female partner	-0.002	0.010	-0.002	0.011	-0.002	0.010
	(0.017)	(0.014)	(0.017)	(0.014)	(0.017)	(0.014)
Partner's contribution		0.084***		0.087***		0.087***
		(0.002)		(0.002)		(0.002)
Correction x Higher ability					0.077	-0.008
					(0.070)	(0.049)
Good correction x Higher ability					-0.129	-0.037
					(0.092)	(0.062)
Individual FE	✓	✓	✓	✓	✓	✓
H0: Correction +Good correction=0 (p-value)			0.000	0.000		
H0: Correction x Higher ability +Good correction x Higher ability=0 (p-value)					0.196	0.787
Baseline mean	0.780	0.780	0.780	0.780	0.780	0.780
Baseline SD	0.414	0.414	0.414	0.414	0.414	0.414
Adj. R-squared	0.073	0.337	0.086	0.340	0.086	0.340
Observations	3180	3180	3180	3180	3180	3180
Individuals	464	464	464	464	464	464

Notes: This table presents the regression results of equations 4 (columns 1-2), 5 (columns 3-4), and 6 (columns 5-6). Baseline mean and standard deviation are participants' willingness to collaborate with male partners who do not make any corrections. Standard errors in parentheses are clustered at the individual level. Significance levels: + 10%, * 5%, ** 1%, and *** 0.1%.

This evidence is consistent with H2S and the behavior conjectured in H1S leads to a suboptimal collaborator selection.

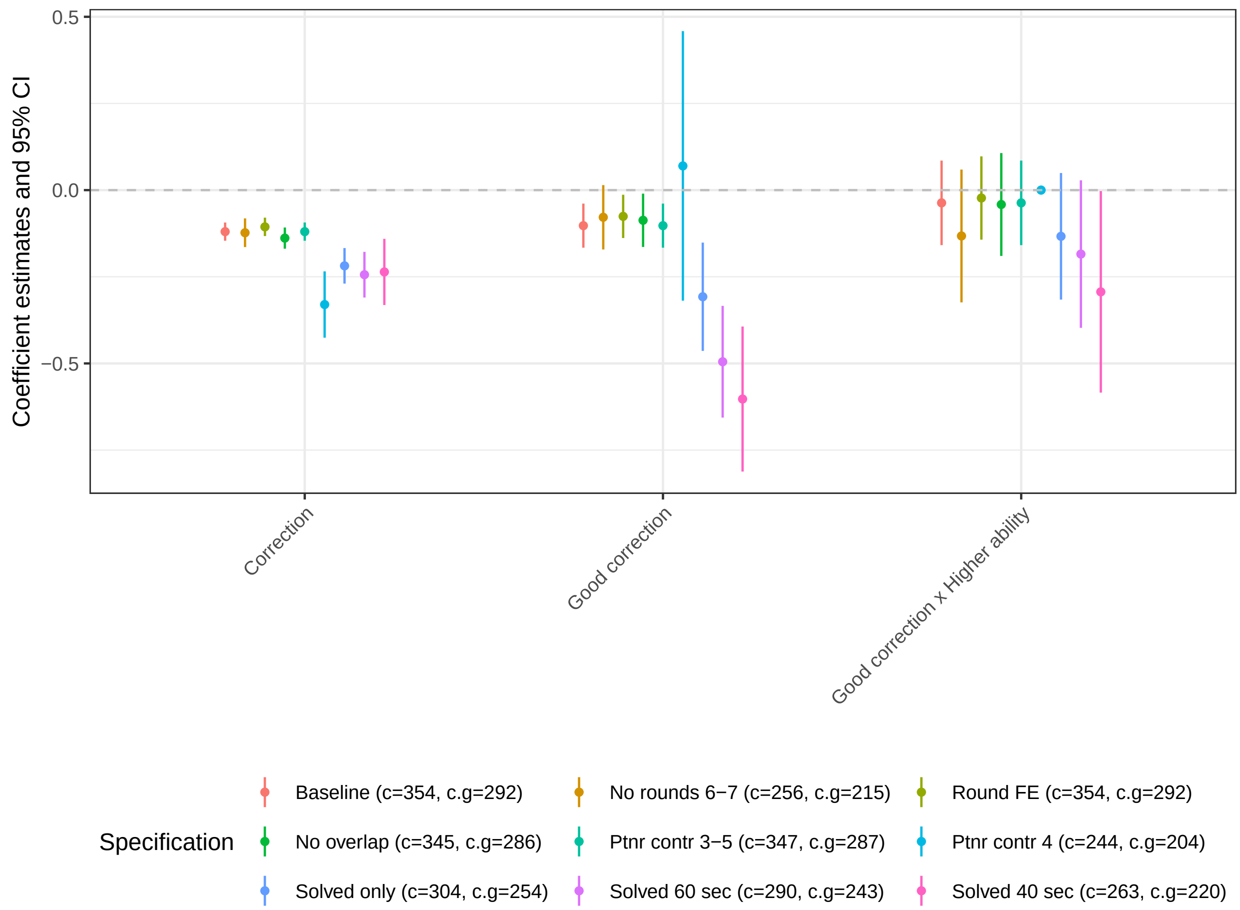
6.3 Results for H3S

Looking at column 5, the coefficient estimates on the interaction between the correction and the higher-ability dummy and the interaction between the good correction and the higher-ability dummy are both statistically insignificant. Even after controlling for the partner's contribution, they are both statistically insignificant. Thus, even the higher-ability people are less willing to collaborate with someone who corrected them, which is suggestive of H3S that the behavior conjectured in H1S is not due to misunderstanding of the correction quality.

6.4 Robustness of the non-pre-registered exploratory results

Figure 6 re-estimates and plots the coefficient estimates (dots) and 95% confidence intervals (lines) of columns 2, 4, and 6 of Table 2 with various different specifications to address concerns raised in Sections 3 and 4, just as in Figure 6. None of them present qualitatively different results; while the

Figure 7: People's response to corrections (non-pre-registered): Robustness



Notes: This figure re-estimates and plots the coefficient estimates (dots) and 95% confidence intervals (lines) of columns 2, 4, and 6 of Table 2 with various different specifications, with the specification "Baseline" being the same as specifications of those tables for comparison. The number of corrections (c) and good corrections (c.g) in each sample are indicated in the parenthesis next to the specification.

coefficient estimate on the good correction becomes positive when we restrict the sample to puzzles where the partner's contribution is 4, the confidence interval is too wide. Also, the results above hold even when I split the sample into men and women; see Online Appendix Table B6 for men and Online Appendix Table B7 for women.

Taken together, the results from the pre-registered and non-pre-registered analysis show that while women's corrections do not receive stringer backlash, correcting colleagues is very costly. As discussed in the introduction, it poses a barrier to group efficiency and peer learning in collaborative work.

7 Discussion

This paper shows that men are no less willing to collaborate with a woman than with a man who corrected them, regardless of the correction quality with a pre-registered analysis. However, a non-pre-registered exploratory analysis shows both men and women are less willing to collaborate with someone who corrected them, regardless of the correction quality and the gender of the corrector. This evidence demonstrates that correcting colleagues is costly, reducing group efficiency and inhibiting peer learning.

Of course, the effects we observe in this experiment may be larger or smaller in a different setting. On the one hand, the factors that are not present in my experiment that can amplify the negative effect of corrections in the workplace are (i) reputation cost (Bénabou and Tirole 2006), (ii) emotional stake in the task, (iii) seniority, and (iv) genderness of the task. First, the emotional cost would be larger if being corrected is observed by others. Second, we are usually corrected for the mistakes not in the puzzles but in our work, in which we spend a long time; we often take corrections on our papers personally at academic seminars or in referee reports. Thus, we should expect a stronger negative effect for corrections in the workplace. Third, everyone is equal in my experiment, but there are usually junior-senior relationships in the workplace, and we should expect corrections by juniors to seniors to have stronger effects. In fact, in academic seminars, we often are more careful correcting senior colleagues than correcting equal or junior colleagues. Fourth, although the experimental task I use is gender-neutral, the tasks in the workplace are often framed as male-typed or female-typed. For example, leadership positions such as politicians and corporate executives, for example, are often considered male-typed (Rudman et al. 2012), while positions related to care work are often considered female-typed (Delfino 2022), and studies document discrimination against women and men in gender-incongruent jobs (Folke and Rickne 2022). Thus, it is possible that in a different setting, the negative effect of corrections is larger for women in leadership positions or men in care work sectors.

On the other hand, the factors that are not present in this experiment that can mitigate the negative effect of corrections are (i) rapport and (ii) ambiguity of the corrections. First, participants are strangers to each other in my experiment, while people usually know their colleagues in the workplace. Thus, the negative effect of corrections can be weaker when the corrections come from someone in a good relationship, although the effect can be stronger if the relationship with that person is not good. Second, whether one corrects others is very obvious in my experiment – it is a binary variable – but it is not necessarily very clear in the workplace. In fact, we strive to be constructive when speaking up in academic seminars.[26] These factors seem to suggest important policy implications of my experiment, namely, the importance of cooperative environments to improve group efficiency and promote peer learning in the workplace. My paper provides a benchmark to study these issues in a controlled (quasi-)laboratory experiment.

26. See, for example, "Guidance for a Constructive Culture of Exchange in MIT Economics Seminars" available on the AEA website: https://docs.google.com/document/d/1Pf9fYEtDLg-IbooxHraO8tPL5I9GW5vxC-TW_CmQBcE.

References

Abel, Martin. 2023. "Do Workers Discriminate against Female Bosses?" *Journal of Human Resources.*

Alan, Sule, Gozde Corekcioglu, and Matthias Sutter. 2022. "Improving Workplace Climate in Large Corporations: A Clustered Randomized Intervention." *The Quarterly Journal of Economics.*

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–1678.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2019. "Beliefs about Gender." *American Economic Review* 109 (3): 739–773.

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.

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.

Chakraborty, Priyanka, and Danila Serra. 2023. *Gender and Leadership in Organizations: The Threat of Backlash.* Working Paper.

Chance, Zoë, and Michael I. Norton. 2015. "The What and Why of Self-Deception." *Current Opinion in Psychology,* Morality and Ethics, 6:104–107.

Chandrasekhar, Arun G., Benjamin Golub, and He Yang. 2019. *Signaling, Shame, and Silence in Social Learning.* Working Paper 3261632.

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, Clio Bryant Flikkema, and Olga Shurchkov. 2021. "Gender Stereotypes in Deliberation and Team Decisions." *Games and Economic Behavior* 129:329–349.

Delfino, Alexia. 2022. *Breaking Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs.* Working Paper.

Dupas, Pascaline, Alicia Sasser Modestino, Muriel Niederle, Justin Wolfers, and Seminar Dynamics Collective. 2021. *Gender and the Dynamics of Economics Seminars.* Working Paper.

Eagly, Alice H., and Steven J. Karau. 2002. "Role Congruity Theory of Prejudice toward Female Leaders." *Psychological Review* 109 (3): 573–598.

Edmans, Alex. 2011. "Does the Stock Market Fully Value Intangibles? Employee Satisfaction and Equity Prices." *Journal of Financial Economics* 101 (3): 621–640.

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.

———. 2008. "Racial Preferences in Dating." *The Review of Economic Studies* 75 (1): 117–132.

Folke, Olle, and Johanna Rickne. 2022. "Sexual Harassment and Gender Inequality in the Labor Market." *The Quarterly Journal of Economics* 137 (4): 2163–2212.

Gino, Francesca, Michael I. Norton, and Roberto A. Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically." *Journal of Economic Perspectives* 30 (3): 189–212.

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

Guo, Joyce, and María P. Recalde. 2023. "Overriding in Teams: The Role of Beliefs, Social Image, and Gender." *Management Science* 69 (4): 2239–2262.

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.

Husain, Aliza N., David A. Matsa, and Amalia R. Miller. 2023. "Do Male Workers Prefer Male Leaders?: An Analysis of Principals' Effects on Teacher Retention." *Journal of Human Resources* 58 (5): 1480–1522.

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

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg. 2021. "Learning From Coworkers." *Econometrica* 89 (2): 647–676.

Jones, Benjamin F. 2021. "The Rise of Research Teams: Benefits and Costs in Economics." *Journal of Economic Perspectives* 35 (2): 191–216.

Karpowitz, Christopher F., Stephen D. O'Connell, Jessica Preece, and Olga Stoddard. 2023. "Strength in Numbers? Gender Composition, Leadership, and Women's Influence in Teams." *Journal of Political Economy.*

Klinowski, David. 2023. "Voicing Disagreement in Science: Missing Women." *The Review of Economics and Statistics,* 1–40.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108 (3): 480–498.

Lazear, Edward P., and Kathryn L. Shaw. 2007. "Personnel Economics: The Economist's View of Human Resources." *Journal of Economic Perspectives* 21 (4): 91–114.

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.

Sandvik, Jason J, Richard E. Saouma, Nathan T. Seegert, and Christopher T. Stanton. 2020. "Workplace Knowledge Flows." *The Quarterly Journal of Economics* 135 (3): 1635–1680.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. "The Increasing Dominance of Teams in Production of Knowledge." *Science* 316 (5827): 1036–1039.

Online Appendix

A Data description

Table A1: Own and partner's puzzle behaviors and puzzle outcomes

	Male decision makers (N=1540)		Female decision makers (N=1708)		Difference (Male – Female)		
	Mean	SD	Mean	SD	Mean	SE	P-value
Panel A: Own behaviors							
Contribution	3.14	2.64	2.98	2.93	0.16	0.10	0.11
# puzzles solved in part 1	8.80	2.34	8.36	2.41	0.44	0.22	0.05
Correction	0.23	0.65	0.23	0.63	0.00	0.03	0.93
Good correction	0.17	0.52	0.17	0.53	0.00	0.02	0.98
Bad correction	0.06	0.32	0.06	0.27	0.00	0.01	0.88
Correction (0/1 dummy)	0.16	0.36	0.15	0.36	0.00	0.01	0.85
Good correction (0/1 dummy)	0.12	0.33	0.12	0.33	0.00	0.01	0.90
Bad correction (0/1 dummy)	0.05	0.22	0.06	0.23	0.00	0.01	0.70
(Fraction of female partners)	0.54	0.50	0.51	0.50	0.03	0.02	0.03
Panel B: Partner's behaviors							
Contribution	3.07	2.87	3.04	2.73	0.03	0.10	0.77
# puzzles solved in part 1	8.57	2.43	8.58	2.35	-0.01	0.16	0.93
Correction	0.22	0.63	0.24	0.66	-0.02	0.03	0.48
Good correction	0.16	0.52	0.17	0.53	-0.02	0.02	0.47
Bad correction	0.06	0.30	0.06	0.29	0.00	0.01	0.85
Correction (0/1 dummy)	0.15	0.36	0.16	0.37	-0.01	0.01	0.51
Good correction (0/1 dummy)	0.12	0.32	0.13	0.33	-0.01	0.01	0.44
Bad correction (0/1 dummy)	0.05	0.22	0.06	0.23	-0.01	0.01	0.44
Panel C: Puzzle outcomes							
Willing to collaborate (yes=1, no=0)	0.71	0.45	0.72	0.45	-0.01	0.02	0.49
Willing to collaborate (residualized)	0.00	0.42	0.00	0.42	0.00	0.00	0.46
Time spent (second)	42.99	35.76	43.74	36.15	-0.74	1.28	0.56
Total moves	11.21	7.70	11.18	7.46	0.03	0.28	0.92
Puzzle solved	0.86	0.35	0.85	0.36	0.01	0.01	0.43
Consecutive correction	0.04	0.21	0.04	0.20	0.00	0.01	0.81

Notes: This table describes own (panel A) and partner's puzzle behaviors (panel B) and puzzle outcomes (panel C). P-values of the difference between male and female participants are calculated with standard errors clustered at the individual level. Contribution is defined as one's net good moves in a given puzzle (the number of good moves minus the number of bad moves).

This section describes the data in detail. Table A1 describes own (panel A) and partner's puzzle behaviors (panel B) and puzzle outcomes (panel C). Panel A shows no gender differences in puzzle-solving ability: for both the contributions in part 2 and the number of puzzles solved in part 1, the difference between male and female participants is statistically insignificant at 5% and quantitatively insignificant.[27,28] This is consistent with Isaksson (2018), who also finds no gender difference in contribution or number of puzzles solved alone using the same puzzle, suggesting that any gender differences I find are unlikely to come from the ability differences between male and female participants. Panel A also shows that there are no gender differences in propensity to correct partners, unlike Isaksson (2018), who finds that men correct their partners more often than women, although their result is from move-level data, and Klinowski (2023), who finds men are more likely to point out (and penalize) others' mistakes. In addition, there can be more than one correction in one puzzle, which is mostly due to good corrections, but not many puzzles experience more than one correction.[29] Finally, the last row of Panel A shows that male participants are slightly more likely to have female partners, although only by three percentage points.

Panel B shows that the puzzle-solving ability, as well as the propensity to correct partners' moves (both mistakes and right moves), are the same for partners paired with male and female participants, suggesting that the random pairing was successful. Participants are corrected by their partners in 15-16% of the total puzzles, of which 12-13% are good corrections, and 5-6% are bad corrections; there are no gender differences in the propensity to be corrected.[30]

Panel C shows that participants state they want to collaborate with the paired partner 71-72% of the time, and that there is enough within-subject variation.[31] Participants spend on average 43-44 seconds for each puzzle (the maximum time a pair can spend is 120 seconds) and make 11 moves. 85-86% of the puzzles are solved, and participants correct their partner's moves consecutively in 4% of the puzzles.[32] There is no gender difference in any of these outcomes, suggesting any gender differences cannot be attributed to the imbalance in these outcomes.

B Additional figures and tables

27. The number of puzzles solved in part 1 is marginally significant but quantitatively insignificant.

28. The correlation coefficient between contribution and number of puzzles solved in part 1 is 0.1059 and the p-value is below 0.001 (with standard errors clustered at the individual level).

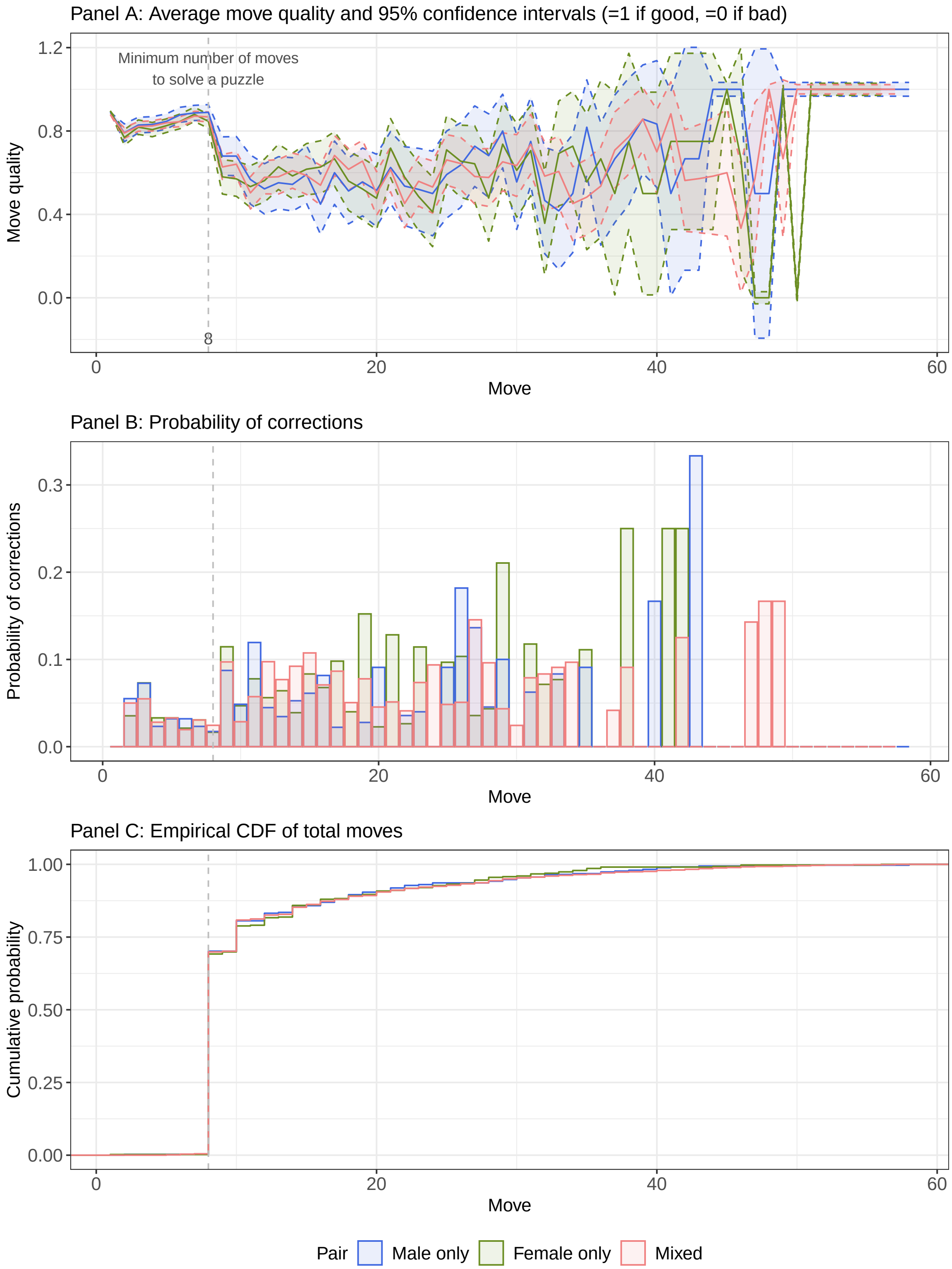
29. Of the 495 puzzles where at least one correction occurred, 72% of them experienced only one correction. Note that of the 495 puzzles, there are 60 puzzles experienced both good and bad corrections (325 puzzles experienced good corrections only and 110 puzzles bad corrections only). The results are robust to excluding these 60 puzzles as shown in Figure 6.

30. The percentage of good corrections and bad corrections does not sum up to the percentage of all corrections because there are puzzles where both good and bad corrections occurred. The results are robust to exclusion of these overlapping puzzles, as shown in Figure 6.

31. The latter point is from standard deviation of "Willingness to collaborate (residualized)." The residualized value is calculated by regressing willingness to collaborate on individual fixed effects and take the residuals, and shows the variation I can exploit after adding individual fixed effects in the analysis.

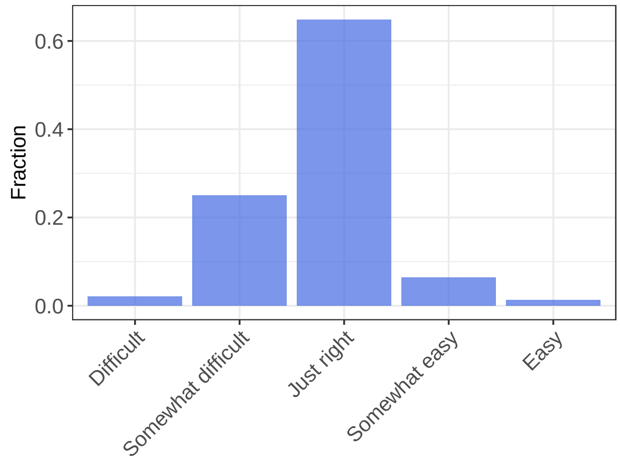
32. Indeed, in puzzles where consecutive correction happens, the probability of selecting the paired partner as a collaborator drops from 78.0% to 26.8%.

Figure B1: Move quality, probability of corrections, and empirical CDF of total moves



Notes: The average move quality along with 95% confidence intervals (panel A), the probability of corrections in each move (panel B), and the empirical CDF of total moves (panel C) separately for males only (blue), females only (green), and mixed gender pairs (red). The confidence interval of panel A is 95% confidence intervals of βs from the following OLS regression: $MoveQuality_{ijt} = \beta_1 + \sum_{k=2}^{58} \beta_k \mathbb{1}[t_{ij} = k] + \epsilon_{ijt}$, where t_{ij} is the pair i-j's move round and $\mathbb{1}$ is an indicator variable. $MoveQuality_{ijt}$ takes a value of 1 if a move of a pair i-j on the tth move is good and 0 if bad. I add an estimate of β_1 to estimates of β_2-β_{58} to make the figure easier to look at. Standard errors are clustered at the pair level.

Figure B2: Perceived puzzle difficulty



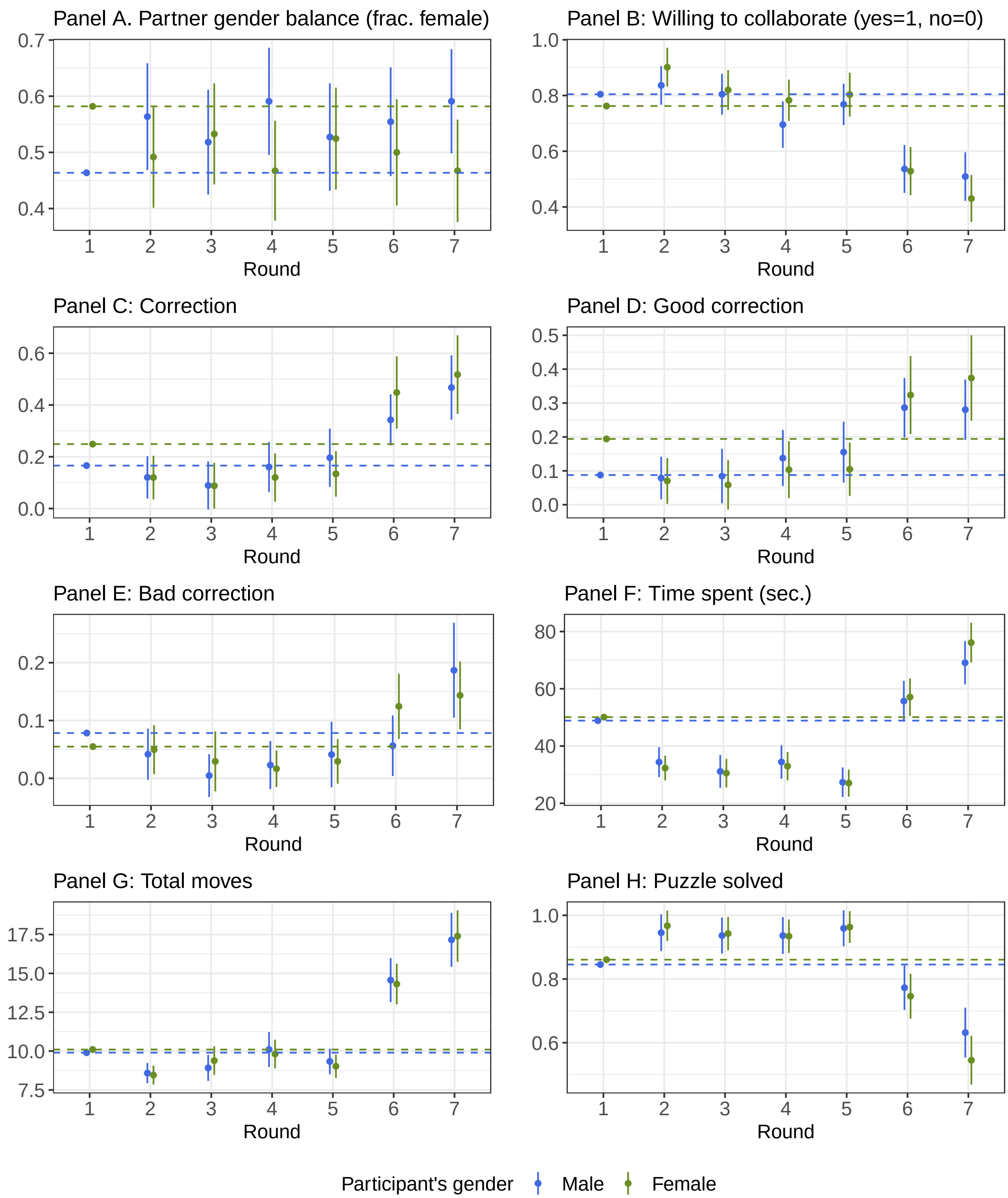
Notes: This figure shows the participants' perceived puzzle difficulty from the post-questionnaire.

Table B1: Participants' characteristics

	Male (N=220)			Female (N=244)			Difference (Male – Female)	
	Mean	SD	Median	Mean	SD	Median	Mean	P-value
Age	25.87	4.33	25	24.45	3.13	24	1.41	0.00
Gender bias (0-1)	0.29	0.19	0.29	0.17	0.16	0.12	0.12	0.00
Region of origin (within Italy)								
North	0.36			0.32			0.04	0.37
Center	0.24			0.23			0.01	0.77
South	0.40			0.45			-0.06	0.23
Major:								
Humanities	0.22			0.45			-0.23	0.00
Social sciences	0.27			0.24			0.03	0.52
Natural sciences	0.20			0.12			0.08	0.02
Engineering	0.23			0.05			0.17	0.00
Medicine	0.08			0.13			-0.05	0.08
Program:								
Bachelor	0.26			0.34			-0.08	0.06
Master	0.68			0.63			0.05	0.26
Doctor	0.06			0.03			0.03	0.11

Notes: This table describes participants' characteristics. P-values of the difference between male and female participants are calculated with heteroskedasticity-robust standard errors.

Figure B3: Balance across rounds



Notes: This figure shows point estimates and 95% confidence intervals of βs from the following OLS regression with gender balance (female dummy) and different puzzle outcomes separately for male (blue) and female participants (green): $y_{ij} = \beta_1 + \sum_{k=2}^{7} \beta_k \mathbb{1}[t_{ij} = k] + \epsilon_{ij}$, where $t_{ij} \in \{1, 2, 3, 4, 5, 6, 7\}$ is the puzzle round in which i and j are playing, $\mathbb{1}$ is an indicator variable, and y_{ij} is the dependent variable indicated in each panel. I add the estimate of β_1 to estimates of β_2-β_7 to make the figure easier to look at. Standard errors are clustered at the individual level.

Table B2: Men's response to women's corrections with the standard errors clustered at the pair level

Outcome:	Willing to collaborate (yes=1, no=0)					
Sample:	Male decision makers					
	(1)	(2)	(3)	(4)	(5)	(6)
Correction	-0.152***	-0.070*	-0.267***	0.001	-0.154***	-0.086*
	(0.027)	(0.030)	(0.053)	(0.048)	(0.040)	(0.034)
Good correction			0.182**	-0.104+		
			(0.060)	(0.055)		
Female partner	0.013	0.023	0.018	0.020	0.024	0.018
	(0.025)	(0.021)	(0.025)	(0.021)	(0.032)	(0.028)
Partner's contribution		0.078***		0.084***		0.079***
		(0.003)		(0.003)		(0.003)
Correction x Female partner	-0.027	-0.021	-0.115	0.053	-0.011	-0.010
	(0.041)	(0.036)	(0.078)	(0.067)	(0.053)	(0.045)
Good correction x Female partner			0.091	-0.082		
			(0.090)	(0.076)		
Correction x Higher bias					0.001	0.029
					(0.050)	(0.048)
Female partner x Higher bias					-0.029	0.006
					(0.048)	(0.041)
Correction x Female partner x Higher bias					-0.026	-0.017
					(0.080)	(0.063)
Individual FE	✓	✓	✓	✓	✓	✓
H0: Correction x Female partner +Good correction x Female partner=0 (p-value)			0.594	0.468		
H0: Correction x Female partner x Higher bias +Correction x Female partner=0 (p-value)					0.550	0.580
Baseline mean	0.747	0.747	0.747	0.747	0.749	0.749
Baseline SD	0.435	0.435	0.435	0.435	0.434	0.434
Adj. R-squared	0.061	0.304	0.075	0.308	0.060	0.303
Observations	1510	1510	1510	1510	1503	1503
Individuals	1165	1165	1165	1165	1162	1162

Notes: This table reports the same estimation results as Table 1 but with the standard errors clustered at the pair level instead of at the individual level. Baseline mean and standard deviation are participants' willingness to collaborate with male partners who do not make any corrections. Significance levels: + 10%, * 5%, ** 1%, and *** 0.1%.

Table B3: Men's response to women's corrections using the pre-registered contribution measure

Outcome:	Willing to collaborate (yes=1, no=0)					
Sample:	Male decision makers					
	(1)	(2)	(3)	(4)	(5)	(6)
Correction	-0.152***	-0.122**	-0.267***	-0.031	-0.154***	-0.139*
	(0.028)	(0.040)	(0.052)	(0.046)	(0.041)	(0.055)
Good correction			0.182**	-0.140*		
			(0.064)	(0.061)		
Female partner	0.013	0.016	0.018	0.013	0.024	0.019
	(0.026)	(0.022)	(0.026)	(0.022)	(0.035)	(0.030)
Partner's contribution		1.222***		1.304***		1.224***
		(0.068)		(0.077)		(0.069)
Correction x Female partner	-0.027	-0.003	-0.115	0.035	-0.011	-0.005
	(0.045)	(0.044)	(0.082)	(0.069)	(0.054)	(0.063)
Good correction x Female partner			0.091	-0.029		
			(0.104)	(0.079)		
Correction x Higher bias					0.001	0.028
					(0.056)	(0.079)
Female partner x Higher bias					-0.029	-0.012
					(0.053)	(0.045)
Correction x Female partner x Higher bias					-0.026	0.012
					(0.092)	(0.086)
Individual FE	✓	✓	✓	✓	✓	✓
H0: Correction x Female partner +Good correction x Female partner=0 (p-value)			0.656	0.914		
H0: Correction x Female partner x Higher bias +Correction x Female partner=0 (p-value)					0.619	0.905
Baseline mean	0.747	0.747	0.747	0.747	0.749	0.749
Baseline SD	0.435	0.435	0.435	0.435	0.434	0.434
Adj. R-squared	0.061	0.299	0.075	0.304	0.060	0.298
Observations	1510	1510	1510	1510	1503	1503
Individuals	220	220	220	220	219	219

Notes: This table reports the same estimation results as Table 1 but with the pre-registered contribution measure specified in the pre-analysis plan, and shows that the results are robust to using the pre-registered measure. Baseline mean and standard deviation are participants' willingness to collaborate with male partners who do not make any corrections. Standard errors in parentheses are clustered at the individual level. Significance levels: + 10%, * 5%, ** 1%, and *** 0.1%.

Table B4: Women's response to women's corrections

Outcome:	Willing to collaborate (yes=1, no=0)					
Sample:	Female decision makers					
	(1)	(2)	(3)	(4)	(5)	(6)
Correction	-0.238***	-0.148***	-0.341***	-0.133**	-0.228***	-0.151***
	(0.026)	(0.020)	(0.084)	(0.042)	(0.035)	(0.027)
Good correction			0.145	-0.020		
			(0.105)	(0.053)		
Female partner	-0.013	0.003	-0.011	0.003	-0.007	0.010
	(0.022)	(0.019)	(0.022)	(0.019)	(0.029)	(0.024)
Partner's contribution		0.089***		0.091***		0.090***
		(0.003)		(0.003)		(0.003)
Correction x Female partner	0.008	-0.001	-0.148	0.046	-0.015	0.018
	(0.033)	(0.027)	(0.100)	(0.072)	(0.045)	(0.046)
Good correction x Female partner			0.198	-0.062		
			(0.125)	(0.090)		
Correction x Higher bias					-0.022	0.006
					(0.052)	(0.038)
Female partner x Higher bias					-0.013	-0.017
					(0.045)	(0.038)
Correction x Female partner x Higher bias					0.043	-0.032
					(0.067)	(0.058)
Individual FE	✓	✓	✓	✓	✓	✓
H0: Correction x Female partner +Good correction x Female partner=0 (p-value)			0.279	0.652		
H0: Correction x Female partner x Higher bias +Correction x Female partner=0 (p-value)					0.567	0.680
Baseline mean	0.809	0.809	0.809	0.809	0.809	0.809
Baseline SD	0.393	0.393	0.393	0.393	0.393	0.393
Adj. R-squared	0.086	0.375	0.101	0.375	0.084	0.374
Observations	1670	1670	1670	1670	1670	1670
Individuals	244	244	244	244	244	244

Notes: This table reports the same estimation results as Table 1 but with female participants instead of with male participants. Baseline mean and standard deviation are participants' willingness to collaborate with male partners who do not make any corrections. Standard errors in parentheses are clustered at the individual level. Significance levels: + 10%, * 5%, ** 1%, and *** 0.1%.

Table B5: People's response to corrections with the standard errors clustered at the pair level (non-pre-registered)

Outcome:	Willing to collaborate (yes=1, no=0)					
Sample:	All decision makers					
	(1)	(2)	(3)	(4)	(5)	(6)
Correction	-0.205***	-0.120***	-0.362***	-0.044+	-0.397***	-0.040
	(0.013)	(0.012)	(0.034)	(0.026)	(0.046)	(0.039)
Good correction			0.220***	-0.103***	0.275***	-0.089*
			(0.040)	(0.030)	(0.053)	(0.044)
Female partner	-0.002	0.010	-0.002	0.011	-0.002	0.010
	(0.015)	(0.013)	(0.015)	(0.013)	(0.015)	(0.013)
Partner's contribution		0.084***		0.087***		0.087***
		(0.002)		(0.002)		(0.002)
Correction x Higher ability					0.077	-0.008
					(0.066)	(0.047)
Good correction x Higher ability					-0.129	-0.037
					(0.079)	(0.058)
Individual FE	✓	✓	✓	✓	✓	✓
H0: Correction +Good correction=0 (p-value)			0.000	0.000		
H0: Correction x Higher ability +Good correction x Higher ability=0 (p-value)					0.145	0.776
Baseline mean	0.780	0.780	0.780	0.780	0.780	0.780
Baseline SD	0.414	0.414	0.414	0.414	0.414	0.414
Adj. R-squared	0.073	0.337	0.086	0.340	0.086	0.340
Observations	3180	3180	3180	3180	3180	3180
Individuals	464	464	464	464	464	464

Notes: This table reports the same estimation results as Table 2 but with the standard errors clustered at the pair level instead of at the individual level. Baseline mean and standard deviation are participants' willingness to collaborate with partners who do not make any corrections. Significance levels: + 10%, * 5%, ** 1%, and *** 0.1%.

Table B6: Men's response to corrections (non-pre-registered)

Outcome:	Willing to collaborate (yes=1, no=0)					
Sample:	Male decision makers					
	(1)	(2)	(3)	(4)	(5)	(6)
Correction	-0.167***	-0.081***	-0.320***	0.025	-0.338***	0.052
	(0.023)	(0.021)	(0.044)	(0.038)	(0.064)	(0.051)
Good correction			0.220***	-0.145**	0.259***	-0.140*
			(0.054)	(0.048)	(0.078)	(0.064)
Female partner	0.007	0.019	0.008	0.019	0.007	0.017
	(0.026)	(0.021)	(0.026)	(0.021)	(0.026)	(0.021)
Partner's contribution		0.079***		0.083***		0.084***
		(0.003)		(0.003)		(0.003)
Correction x Higher ability					0.039	-0.070
					(0.083)	(0.062)
Good correction x Higher ability					-0.096	-0.027
					(0.106)	(0.080)
Individual FE	✓	✓	✓	✓	✓	✓
H0: Correction +Good correction=0 (p-value)			0.000	0.000		
H0: Correction x Higher ability +Good correction x Higher ability=0 (p-value)					0.458	0.753
Baseline mean	0.764	0.764	0.764	0.764	0.764	0.764
Baseline SD	0.425	0.425	0.425	0.425	0.425	0.425
Adj. R-squared	0.062	0.304	0.075	0.309	0.075	0.311
Observations	1510	1510	1510	1510	1510	1510
Individuals	220	220	220	220	220	220

Notes: This table reports the same estimation results as Table 2 but with male participants only. Baseline mean and standard deviation are participants' willingness to collaborate with partners who do not make any corrections. Standard errors in parentheses are clustered at the individual level. Significance levels: + 10%, * 5%, ** 1%, and *** 0.1%.

Table B7: Women's response to corrections (non-pre-registered)

Outcome:	Willing to collaborate (yes=1, no=0)					
Sample:	Female decision makers					
	(1)	(2)	(3)	(4)	(5)	(6)
Correction	-0.235***	-0.149***	-0.401***	-0.115***	-0.471***	-0.158***
	(0.018)	(0.016)	(0.062)	(0.034)	(0.063)	(0.045)
Good correction			0.229**	-0.045	0.314***	-0.003
			(0.077)	(0.042)	(0.080)	(0.053)
Female partner	-0.012	0.002	-0.011	0.003	-0.013	0.001
	(0.022)	(0.018)	(0.021)	(0.018)	(0.021)	(0.018)
Partner's contribution		0.089***		0.090***		0.090***
		(0.003)		(0.003)		(0.003)
Correction x Higher ability					0.144	0.085
					(0.107)	(0.060)
Good correction x Higher ability					-0.183	-0.084
					(0.142)	(0.083)
Individual FE	✓	✓	✓	✓	✓	✓
H0: Correction +Good correction=0 (p-value)			0.000	0.000		
H0: Correction x Higher ability +Good correction x Higher ability=0 (p-value)					0.183	0.225
Baseline mean	0.795	0.795	0.795	0.795	0.795	0.795
Baseline SD	0.404	0.404	0.404	0.404	0.404	0.404
Adj. R-squared	0.087	0.375	0.100	0.375	0.101	0.375
Observations	1670	1670	1670	1670	1670	1670
Individuals	244	244	244	244	244	244

Notes: This table reports the same estimation results as Table 2 but with female participants only. Baseline mean and standard deviation are participants' willingness to collaborate with partners who do not make any corrections. Standard errors in parentheses are clustered at the individual level. Significance levels: + 10%, * 5%, ** 1%, and *** 0.1%.

C Experimental instructions: English translation

App: pt0

Page: Reg

Registration

Please fill out the following information in order for us to pay you after the session. Please make sure that they correspond to the information you registered on ORSEE.

N.B. Please capitalize only the first letter of your first name and last name.

Good examples: Marco Rossi; Maria Bianchi; Anna Maria Gallo

Bad examples: MARCO ROSSI; maria bianchi; Anna maria Gallo

- First name: [Textbox]
- Last name: [Textbox]
- Email address registered on ORSEE: [Textbox]

[Check if there are any same first names. If so, add an integer (starting from 2) at the end of the first name]

Page: Draw

Draw a coin

Please draw a virtual coin by clicking the button below.

[Draw]

[Assign random number ranging from 1 to 40]

Page: Wait

Your coin

You drew the following coin.

Please wait until the session starts.

Page: Excess

Please click an appropriate button

[I was chosen to participate] [I was chosen to leave]

Page: Intro

General instructions

<u>Overview</u>: This study will consist of **3 parts** and a follow-up survey and is expected to take **1 hour**. At the beginning of each part, you will receive specific instructions, followed by a set of understanding questions. You must answer these understanding questions correctly to proceed.

<u>Your payment</u>: For completing this study, you are guaranteed **2€** for your participation, but can earn up to **25€** depending on how good you are at the tasks. The tasks involve solving sliding puzzles, like the one shown below.

puzzle_2_0.png

<u>Confidentiality</u>: Other people participating in this study can see your first name. Aside from your first name, other participants will not see any information about you. **At the conclusion of the study, all identifying information will be removed and the data will be kept confidential**. If there is more than one participant with the same first name, we add a number at the end of your first name (e.g. Marco2).

<u>General rules</u>: During the study, please turn off your camera and microphone, and do not communicate with anyone other than us. Also, please do not reload the page or close your browser because it may make your puzzle unsolvable. If you have any questions or face any problems, please send us a private chat on Zoom.

<u>App: pt1</u>

Page: Intro

Instructions for part 1 out of 3

In this part, you will solve the puzzle alone to familiarize yourself with it. You can solve as many puzzles as possible (but a maximum of 15 puzzles) in **4 minutes**. You will earn **0.2€ for each puzzle** you solve.

Your goal is to move the tiles and order them as follows:

puzzle_goal.png

Before you start, please go through the three examples below to understand how to solve the puzzle.

Example 1:

First, consider the following puzzle.

puzzle_1.png

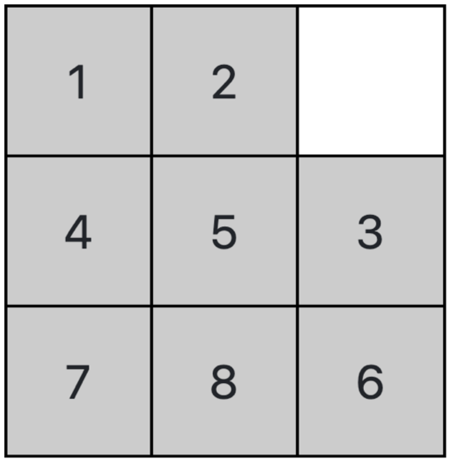
You can only move the tiles next to an empty cell and the tile you choose is moved to the empty cell. So, in this puzzle, there are 3 moves you can make: move 3 down, move 5 right, and move 6 up.

Among the 3 moves, moving 6 up is the only correct move: by moving 6 up, you can solve the puzzle. The other moves do not solve the puzzle.

When you click a tile next to an empty cell, the tile will be moved to the empty cell. So, in this case, you should click 6 to move it up.

Example 2:

Next, consider the following puzzle.



puzzle_2_0.png

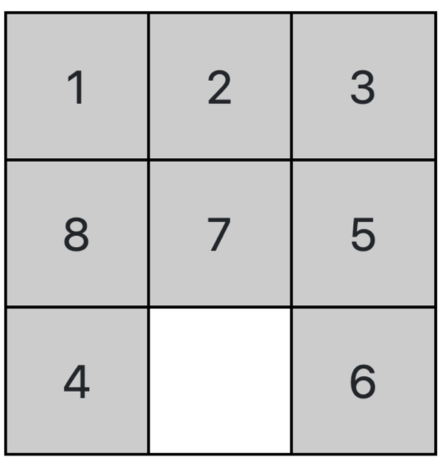
First, there are 2 moves you can make: move 2 right and move 3 up. Which moves should you make?

Observe that the only tiles that are not in the correct order are 3 and 6. So, you should move 3 up.

After moving 3 up, the puzzle will look like the one in example 1. Then you should move 6 up and the puzzle will be solved.

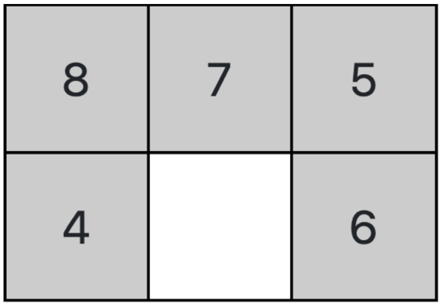
Example 3:

Finally, consider the following puzzle.



puzzle_3_0.png

This puzzle is a bit complicated but observe that the top row is already in the correct order. So, let's keep the top row as is, and think about the remaining part. **When the top row is in the correct order, you should always keep it as is**. So, think of this puzzle as the following simpler puzzle.



puzzle_3_0_2x3.png

You could solve the puzzle by trial and error. However, **after making the top row in the correct order, you should next make the left column in the correct order** to solve the puzzle faster. There are two moves you can make: move 4 right and move 7 down. Which is the faster way to make the left column in the correct order?

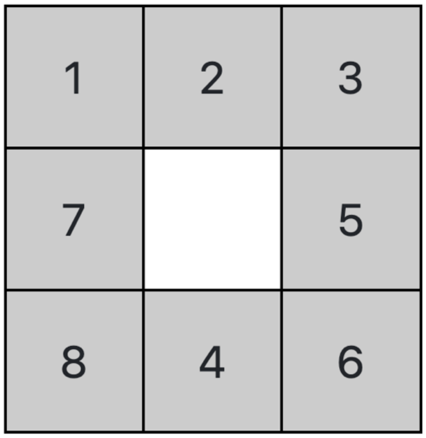
Let's try moving 4 right.

puzzle_3_1_bad_0.png

Now the only tile you can move is 8. So, let's move it down.

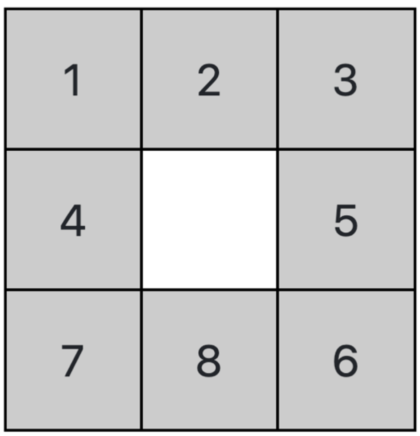
puzzle_3_1_bad_1.png

Now, if you ignore the top row which is already in the correct order, the only tile you can move is 7. So, let's move it to the left.



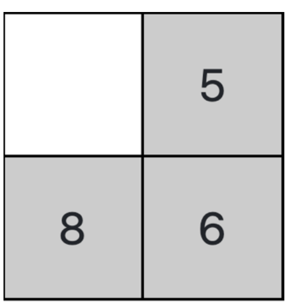
puzzle_3_1_bad_2.png

Then move 4 up, move 8 right, and move 7 down. Then you have made the left column in the correct order. You have moved tiles seven times until now.



puzzle_3_1_bad_3.png

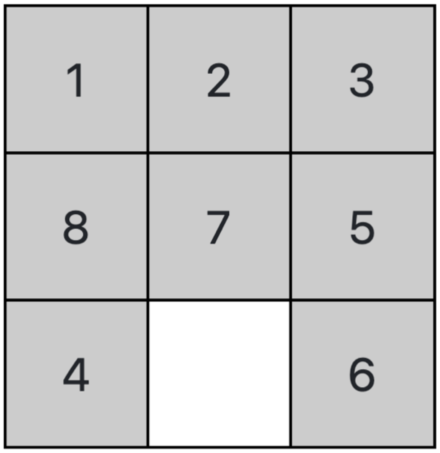
Now let's also keep the left column as is.



puzzle_3_1_bad_3_2x2.png

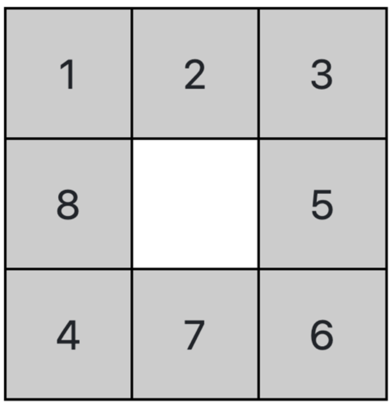
Then you can solve the puzzle by moving 5 left and then 6 up. With this method, **you have moved tiles nine times in total**.

Let's go back to the initial puzzle.



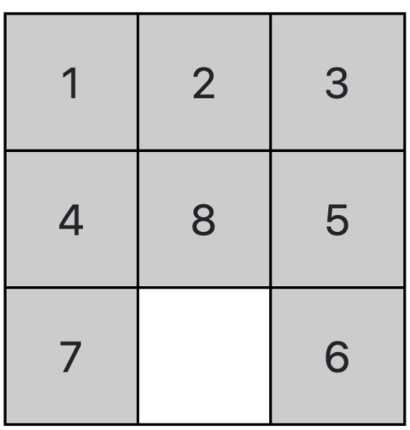
puzzle_3_0.png

This time, let's try moving 7 down.



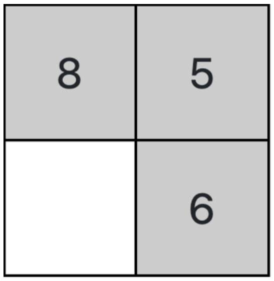
puzzle_3_1_good.png

Then move 8 right, 4 up, and 7 left. Now you have made the left column in the correct order only with four moves.



puzzle_3_4_good.png

Let's keep the left column as is (as well as the top row).



puzzle_3_4_good_2x2.png

Now it's easy to solve the puzzle: move 8 down, 5 left, and 6 up. With this method, **you have only moved tiles seven times in total**.

Because there is a time limit, it's better to solve the puzzle with the minimum number of moves. **We call a move a good move if it makes a puzzle closer to the solution, and a bad move if it makes a puzzle far from the solution. There are no neutral moves: all moves are either good or bad.**

In summary: when you solve the puzzle, first make the top row in the correct order, then make the left column in the correct order. Always try to make the number of moves as small as possible.

Understanding questions:

Before you proceed, please answer the following understanding questions. After you answer, please click Next.

1. Which of the following statements is true?

- ✓In this part, I will work on the puzzles individually for 4 minutes and earn 0.2€ for each puzzle I solve.
- In this part, I will work on the puzzles in pairs for 4 minutes and earn 0.2€ for each puzzle we solve.
- In this part, I will work on the puzzles individually for 4 minutes, but I will not earn anything.

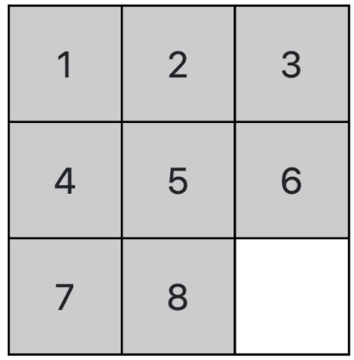
2. Which of the following puzzles is in the correct order?

- A
- ✓B

A

puzzle_2_0.png

B



puzzle_goal.png

3. What is the strategy you should use to solve the puzzle as fast as possible?

- First, make the left column in the correct order, then the bottom row. Always minimize the number of moves I make.
- First, make the top row in the correct order, then the right column. Always minimize the number of moves I make.
- ✓First, make the top row in the correct order, then the left column. Always minimize the number of moves I make.

4. Look at the following puzzle. Which is the good move?

- Move 4 down.
- ✓Move 7 left.

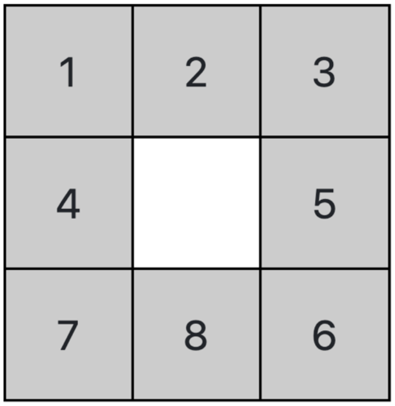
puzzle_3_3_good.png

5. Consider the puzzle in question 4. What is the minimum number of moves to solve the puzzle?

- 2
- 3
- ✓4

6. Look at the following puzzle. Which is the good move?

- ✓Move 5 left.
- Move 8 up.



puzzle_3_5_good.png

7. Consider the puzzle in question 6. What is the minimum number of moves to solve the puzzle?

- ✓2
- 3
- 4

Page: Ready

Be ready

[5 seconds time count]

Please be ready for the individual round.

Page: Game

Individual round

[4 minutes time count]

[max. 15 puzzles with increasing difficulty]

Page: Proceed

The individual round is over

The individual round is over. You have solved **xx puzzles**.

Please click Next to proceed.

App: pt2

Page: Intro

Instructions for part 2 out of 3

In this part, you will **choose your partner for part 3**, the next part.

Although you will not earn anything in this part, it is important to choose the best partner possible: in part 3, you will work on the puzzles for 12 minutes in a pair by moving the tiles in turn, and both you and your partner will earn 1€ for each puzzle you two solve. There is a maximum of 20 puzzles you and your partner can solve (so the maximum earning is 20€).

You will **meet 7 other people** participating in this session one by one and solve 1 puzzle together by moving tiles in turn as you would do in part 3. One of you will be randomly chosen to make the first move at the beginning of each puzzle. You will have a **2-minute limit** for each puzzle.

After solving the puzzle, you will **choose whether you want to work with this person in part 3 too**. This person or other people in this session will not see your choice. **You can choose as many people as you want**.

After you meet all the 7 people and state your choices, we will check all the choices you and the 7 other people have made, and decide each person's partner for part 3 as follows:

1. We randomly choose 1 person out of you and the other 7 people. Call this person Giovanni.
2. We then check if Giovanni has a "match": among people Giovanni has chosen, we check whether these people also have chosen Giovanni. If there is such a person, we make Giovanni and this person as partners for part 3.
3. If Giovanni has more than one match, we randomly choose one of the matches and make them as partners for part 3.
4. If Giovanni has not chosen anyone, the people Giovanni has chosen have not chosen Giovanni, or those people already have their partner, we put Giovanni on a waiting list and repeat points 1-3 above.
5. After we choose all people, we randomly match people on the waiting list as partners for part 3.

So, **even if you choose a particular person, you may not be able to work with that person in part 3**. So, choose everyone whom you want to work with in part 3.

Understanding questions:

Before you proceed, please answer the following understanding questions. After you answer, please click Next.

1. Which of the following statements is true?

- ✓In this part, I will choose my partner for part 3.
- In this part, I will work on the puzzles for 12 minutes in a pair by moving the tiles in turn.

2. How many people can you choose whom you want to work with in part 3?

- 1 person.
- 2 people.
- ✓As many people as you want.

3. Why is it important to choose the best partner for part 3?

- ✓ because how many puzzles I can solve in part 3 depends on my partner's moves.
- because my partner will solve puzzles for me.

4. Suppose you have chosen Giovanni and Valeria. However, while Valeria has chosen you, Giovanni has not. If we have randomly chosen you first, who will be your partner for part 3?

- Giovanni
- ✓Valeria
- Someone on the waiting list
- Randomly chosen from Giovanni and Valeria

5. Suppose you have chosen Giovanni and Valeria. However, unlike question 4, while Giovanni has chosen you, Valeria has not. If we have randomly chosen you first, who will be your partner for part 3?

- ✓Giovanni
- Valeria
- Someone on the waiting list
- Randomly chosen from Giovanni and Valeria

6. Suppose you have chosen Giovanni and Valeria. Also, both Giovanni and Valeria have chosen you. If we have randomly chosen you first, who will be your partner for part 3?

- Giovanni
- Valeria
- Someone on the waiting list
- ✓Randomly chosen from Giovanni and Valeria

7. Suppose you have chosen Giovanni and Valeria. Also, both Giovanni and Valeria have chosen you. However, we already matched Valeria with Giovanni before we choose you. Who will be your partner for part 3?

- Giovanni
- Valeria
- ✓Someone on the waiting list
- Randomly chosen from Giovanni and Valeria

8. Suppose you have not chosen anyone. Also, both Giovanni and Valeria have chosen you. If we have randomly chosen you first, who will be your partner for part 3?

- Giovanni
- Valeria

- ✔Someone on the waiting list
- Randomly chosen from Giovanni and Valeria

9. Suppose you have chosen Giovanni and Valeria. However, neither Giovanni nor Valeria has chosen you. If we have randomly chosen you first, who will be your partner for part 3?

- Giovanni
- Valeria
- ✔Someone on the waiting list
- Randomly chosen from Giovanni and Valeria

Page: Puzzle

Puzzle 1/2/3/4/5/6/7 out of 7

You are playing the puzzle with **[this person's ID]**

[2 minutes time count]

Page: Pref

Puzzle 1/2/3/4/5/6/7 out of 7

You have played the puzzle with **[this person's ID]**. Do you want to work with [this person's ID] in part 3?

[Yes, No]

App: pt3

Page: Partner

Your partner for part 3

Based on your and the 7 other people's choices, **[the partner's ID]** became your partner for part 3.

Page: Intro

Instructions for part 3 out of 3

In this part, you will work on the puzzles with your partner for **12 minutes** by moving the tiles in turn, and both you and your partner will earn **1€ for each puzzle** you two solve. There is a maximum of 20 puzzles you and your partner can solve (so the maximum earning is 20€). As in part 2, one of you will be randomly chosen to make the first move at the beginning of each puzzle.

Understanding questions:

Before you proceed, please answer the following understanding questions. After you answer, please click Next.

1. Which of the following statements is true?

- ✓In this part, you and your partner will both earn 1€ for each puzzle you two solve, which means you will earn 1€ for each puzzle you two solve.
- In this part, you and your partner will earn 1€ for each puzzle you two solve, which means you will earn 0.5€ for each puzzle you two solve.

2. You and your partner…

- ✓will work on the puzzles for 12 minutes by moving the tiles in turn. Which of you will make the first move is randomly determined at the beginning of each puzzle.
- will work on the puzzles for 12 minutes. Which of you will make the first move is randomly determined at the beginning of this part and fixed afterward.

Page: Ready

Be ready

[5 seconds time count]

Please be ready for the group round.

Page: Game

Puzzle 1/2/3/…/20

Your partner: **[the partner's ID]**

[12 minutes time count]

[max. 20 puzzles with increasing difficulty]

Page: Proceed

The group round is over

The group round is over. You have solved **xx puzzles**.

Please click Next to proceed.

App: pt4

Page: Intro

A follow-up survey

As the last task, we will ask you a series of questions in which there are no right or wrong answers. We are only interested in your personal opinions. We are interested in what characteristics are associated with people's behaviors in this study. **The answers you provide will in no way affect your earnings in this study and are kept confidential.**

Please click Next to start the survey.

Page: SurveyASI

Survey page 1 out of 2

Below is a series of statements concerning men and women and their relationships in contemporary society. Please indicate the degree to which you agree or disagree with each statement.

- Women are too easily offended.
- Many women are actually seeking special favors, such as hiring policies that favor them over men, under the guise of asking for "equality."
- Men should be willing to sacrifice their own wellbeing in order to provide financially for the women in their lives.
- Many women have a quality of purity that few men possess.
- No matter how accomplished he is, a man is not truly complete as a person unless he has the love of a woman.
- Women exaggerate problems they have at work.

[Choices: Strongly agree, Agree a little, Neither agree nor disagree, Disagree a little, Strongly disagree]

Page: SurveyDem

Survey page 2 out of 2

Please tell us about yourself and your opinion about this study.

- Your age: [Integer]
- Gender: [Male, Female]
- Region of origin: [Northwest, Northeast, Center, South, Islands, Abroad]
- Field of study: [Humanities, Law, Social Sciences, Natural Sciences/Mathematics, Medicine, Engineering]
- Degree program: [Bachelor, Master/Post-bachelor, Bachelor-master combined (1st, 2nd, or 3rd year), Bachelor-master combined (4th year or beyond), Doctor]
- What do you think this study was about? [Textbox]
- Was there anything unclear or confusing about this study? [Textbox]
- Were the puzzles difficult? [Difficult, Somewhat difficult, Just right, Somewhat easy, Easy]
- Do you have any other comments? (optional) [Textbox]

Page: ThankYou

Thank you for your participation

Thank you for your participation. You have completed the study.

Your earnings:

- **2€** for your participation.
- **xx.x€** for the puzzles you solved in part 1.
- **xx€** for the puzzles you and your partner solved in part 3.

Thus, you have earned **xx.x€** in this study. We will pay you your earnings via PayPal within 2 weeks. If you haven't received your earnings after 2 weeks, please contact us.

Optional: If you would like to know the results of this study, we are more than happy to send you the working paper via email once we finish this study.

[No, I do not want to receive the working paper] [Yes, I want to receive the working paper]

App: pt99

Page: ThankYou

Thank you for showing up

Thank you for showing up in this study. You will receive the show up fee of **2€** via PayPal within 2 weeks. If you haven't received your earnings after 2 weeks, please contact us.