Gender Differences in the Cost of Corrections in Group Work

Yuki Takahashi∗

Preliminary draft. Please do not cite or circulate.

Click here for the latest version

August 13, 2021

Abstract

Corrections among colleagues is an integral part of group work. Pointing out a colleague’s mistake has the potential to improve group performance. However, people may take corrections as personal criticism and dislike colleagues who corrected them. If people dislike female colleagues more, women face a higher hurdle in their career success, and groups cannot fully benefit from their female colleagues. This paper studies whether people dislike collaborating with someone who corrects them and more so when that person is a woman. I find that people are less willing to collaborate with a person who has corrected them even if the correction improves group performance. Nevertheless, people equally dislike corrections from women and men. These findings suggest that while women do not face a higher hurdle, correcting colleagues is costly and reduces group efficiency.

JEL codes: J16, M54, D91, C92

Keywords: correction, collaboration, group work, gender differences

∗PhD Candidate, Department of Economics, University of Bologna. Email: yuki.takahashi2@unibo.it. I am grateful to Maria Bigoni, Siri Isaksson, Bertil Tungodden, Laura Anderucci, and Natalia Montinari whose feedback was essential for this project. I am also grateful to participants of the experiment for their participation and cooperation. Sonia Bhalotra, Francesca Cassanelli, Alessandro Castagnetti, Mónica Costa-Dias, Valeria Ferraro, Lenka Fiala, Ria Granzier-Nakajima, Silvia Griselda, Annalisa Loviglio, Yoko Okuyama, Monika Pompeo, Olivind Scheyen, Vincenzo Scrutinio, Erik Ø. Sørensen, Ludovica Spinola, Florian Zimmermann, and PhD students at the NHH and the University of Bologna all provided many helpful comments. This paper also benefited from participants’ comments at the Annual Southern PhD Economics Conference, ESA Conference, FROGEE Workshop, Gender Gaps Conference, PhD-EVS, Stanford Institute for Theoretical Economics, Warwick Economics PhD Conference, WEAI Conference, Webinar in Gender and Family Economics, and seminars at Ca’ Foscari University, Catholic University of Brasilia, the NHH, Tilburg University, the University of Bologna, and the University of Copenhagen. Ceren Ay, Tommaso Batistoni, Philipp Chapkovski, Sebastian Fest, Christian König genannt Kersting, and oTree help & discussion group kindly answered my questions about oTree programming; in particular, my puzzle code was heavily based on Christian’s code. Michela Boldrini and Boon Han Koh conducted the quasi-laboratory experiments ahead of me and kindly answered my questions about their implementations. 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 (#262643).
1 Introduction

Receiving corrections from colleagues is an integral part of group work. Consider academic research. From the development of ideas to the writing up of the final draft, researchers discuss their research project with their colleagues, receive criticisms, and refine the ideas and the analysis. However, people may take the corrections personally.\textsuperscript{1} Imagine a researcher presents his paper for which he spent several years, and someone points out a possible flaw in his identification assumption or his experimental design. Since the validity of identification assumptions and experimental designs are debatable, he may take it as a personal criticism and be less willing to collaborate with the person who corrects him in the future. While not collaborating is a least aggressive way to express his discomfort, it could be detrimental for her career success because having collaborations is essential in academia, where people co-author the majority of papers (Jones 2021; Wuchty, Jones, and Uzzi 2007).

Women’s corrections may receive stronger negative reactions. Evidence suggests that people use double standard for women and men: men undervalue women when they criticize them (Sinclair and Kunda 2000) and that people punish women more harshly when they make mistakes (Sarsons 2019) and commit misconducts (Egan, Matvos, and Seru 2021). If so, women face higher hurdle in their career success. It is also detrimental for group efficiency as group members cannot fully benefit from female colleagues.

This paper studies whether people dislike collaborating with someone who corrects them and more so when that person is a woman. Answering this question using secondary data poses two challenges. First, group formation is not random and group corrections are endogenous. Second, different corrections are not necessarily comparable to each other.

Thus, I design a quasi-laboratory experiment, a hybrid of physical laboratory and online experiments, where group formation is randomized and define corrections such that researchers can track its quality mathematically. Specifically, participants are allocated to a group of eight and solve one joint task with each group member one by one. Each time participants finish the task, they state whether they would like to collaborate with the group member with whom they have just solved the task for the same task in the next stage, which is the main source of earnings. This gives a strong incentive for participants to select as good a collaborator as possible. The order of the group members with whom participants solve the task is randomized. As a joint task, I use Isaksson (2018)’s number-sliding puzzle which allows me to calculate an objective measure of each participant’s contribution to the joint task as well as to classify each move as good or bad.\textsuperscript{2} I define a correction as reversing a group member’s move, which is comparable across different participants and can be classified as either good or bad.

I find that participants correctly understand the notion of good and bad moves; that is, the higher your contribution is to solving the puzzle, the more likely it is that you will be asked to join a team. This is in line with what one would expect and validates my experimental design. I

\textsuperscript{1} An extreme example is a controversy around a paper that argues hydroxychloroquine is effective in treating COVID-19 patients; the paper’s authors ended up suing the scientist who pointed out the flaw in the paper’s analysis (Davey 2021).

\textsuperscript{2} Participants solve a 3x3 number-sliding puzzle in pairs by alternating their moves. A good move is defined as a move that reduces the number of moves away from the solution, and a bad move is defined as a move that increases the number of moves away from the solution.
also show that women and men contribute equally to the puzzle and neither women nor men underestimate women’s contribution, suggesting that any gender differences I would find is unlikely to be due to people’s belief about difference in women’s and men’s ability.

Nonetheless, after controlling for the contribution, people are significantly less willing to collaborate with a person who has corrected their moves even if the corrections move the puzzle closer to the solution. Yet, people respond to corrections by women as negative as by men.

As an exploratory analysis, I also find that while women are less willing to collaborate with a person who made either good or bad corrections, while men are less willing to collaborate only with a person who made good corrections. This is true especially high-ability men, implying that men’s behavior is irrational. Also, men are less able to take into account for partners’ contribution in collaborator selections. However, the number of observations are not large enough and I may be picking up some irregularities in the data; I would need to conduct another experiment to examine these points.

This paper primarily relates to studies on gender differences in the contribution of ideas in group work. Coffman (2014) finds that women are less likely to contribute their ideas to the group in a male task due to self-stereotyping and Gallus and Heikensten (2019) find that debiasing their self-stereotyping by giving an award for their high ability increases women’s contribution of their ideas: they put women’s idea further ahead without involving open correction of their group member. However, on some occasions, the contribution of ideas has to be made openly, for example in academic seminars and business meetings. In such cases, group members’ response plays an important role in the effectiveness of the intervention. Coffman, Flikkema, and Shurchkov (2021) find that group members are less likely to choose women’s answers as a group answer in male-typed questions. Guo and Recalde (2020) find that group members correct women’s ideas more often than men’s ideas. Dupas et al. (2021) find that female economists receive more patronizing and hostile questions during seminars. Isaksson (2018) finds that men are more likely to correct their group member’s bad moves in the same puzzle used in my experiment. My paper introduces correction in the contribution of ideas and examines its cost on women and on group efficiency.

More generally, my paper contributes to the literature on gender differences in group work. Isaksson (2018) finds that women under-claim their contribution compared to men in group work despite their equal contribution. Haynes and Heilman (2013) find similar results. Sarsons et al. (2021) find that people attribute less credit to a female economist when she co-authors a paper with a male economist(s). Born, Ranehill, and Sandberg (2020) and Stoddard, Karpowitz, and Preece (2020) find that women are less willing to lead a male-majority group. Shan (2020) finds that female students are more likely to drop out from an introductory economics class when they are assigned to a male-majority study group. Babcock et al. (2017) find that women are more likely to volunteer and be asked to do non-promotable tasks. My paper promotes our understanding of gender differences in group work.

My paper also speaks to the literature on organizational efficiency. Bandiera, Barankay, and Rasul (2009) find that managers favor workers who have connections with the managers, which distorts allocation of talent and reduces firm efficiency as theory predicts (MacLeod

3. As the puzzle was originally used by Isaksson (2018).
2003; Prendergast and Topel 1996). Li (2020) finds that this managers’ favoritism not only
distorts the optimal allocation of talent but also reduces non-favored workers’ performance.
Cullen and Perez-Truglia (2021) find that managers favor workers with whom they presumably
take smoking break together. Fang and Huang (2017) find that institutional investors give
positively-biased evaluations for corporate analysts who graduated the same university especially
when the analysts are men. Kennedy and Pronin (2008) find that people tend to view others
who disagree with them as biased. Ronayne and Sgroi (2019) find that people often stick to their
decisions rather than accepting decisions suggested by people with higher ability. My finding
that people are less willing to collaborate with people who have corrected them can be another
source of organizational inefficiency.

2 Experiment

Introducing a quasi-laboratory format I run the experiment in a quasi-laboratory format
where we experimenters connect us to the participants via Zoom throughout the experiment (but
turn off participants’ camera and microphone except at the beginning of the experiment) and
conduct it as we usually do in a physical laboratory but participants participate remotely using
their computers. Appendix A discusses pros and cons of the quasi-laboratory format relative to
physical laboratory and standard online experiments.

Group task As the group task I use Isaksson (2018)’s puzzle, a sliding puzzle with 8 numbered
tiles, which should be placed in numerical order within a 3x3 frame (see panel A of figure 3
for an example). To achieve this goal, participants play in pairs, alternating their moves. This
puzzle has nice mathematical properties: I can define the puzzle difficulty and classify a given
move as either good or bad by the Breadth-First Search algorithm. From the number of good
and bad moves one makes, I can calculate individual contributions to the group task; I measure
it by net good moves, the number of good moves minus the number of bad moves an individual
makes in a given puzzle:

\[
\text{Player i’s contribution} \equiv \text{i’s # good moves} - \text{i’s # bad moves} \in \mathbb{Z}
\]  

I can also determine the quality of corrections of different participants objectively and
comparably.\footnote{The difficulty is defined as the number of moves away from the solution, a good move is defined as a move
that reduces the number of moves away from the solution, and a bad move is defined as a move that increases the
number of moves away from the solution.} Further, the puzzle-solving captures an essential characteristic of teamwork 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).

At each stage of the puzzle, there is only one best strategy which is to make a good move.\footnote{Conditional on that both players are trying to solve the puzzle; I show later that the results are robust to
exclusion of puzzles where either player might not be trying to solve the puzzle.} There can be more than one good and bad moves, but different good/bad moves are equal.
There is no path dependence either: the history of the puzzle moves does not matter.

4.
The experiment consists of three parts as summarized in figure 1 and described in detail below. At the beginning of each part, participants must answer a set of comprehension questions to make sure they understand the instructions.

2.1 Design and procedure

Registration

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

Pre-experiment

On the day and the time of the session they have registered for, participants enter the Zoom waiting room.7 They receive a link to the virtual room for the experiment and enter their first name, last name, and their email they have used in the registration. They also draw a virtual coin numbered from 1 to 40 without replacement.

Then I admit participants to the Zoom meeting room one by one and rename them by the first name they have just entered. This information is necessary to match up their earnings in this experiment and their payment information stored in the laboratory database, so participants have strong incentive to provide their true name and email address. 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 roll call, a way to reveal participants’ gender to other participants without making gender salient (Bordalo et al. 2019; Coffman, Flikkema, and Shurchkov 2021). Specifically, I take call each participant’s first name one by one and ask her or him to respond via microphone. This process ensures other participants that the called participant’s first name corresponds to her or his gender. If there are more participants than I would need for the session (I need 16 participants), I draw random numbers from 1 to 40 and ask those who drew the coins with the same number to leave.8 Those who leave the session receive the 2€ show-up fee. Figure 2 shows a Zoom screen participants would see during the roll call (the person whose camera is on is the experimenter; participants would see this screen throughout the experiment but the experimenter’s camera may be turned off).

I then read out the instructions about the rules of the experiment and take questions on Zoom. Once participants start the main part, they can communicate with the experimenter only via Zoom’s private chat.

---

6. I recruit a few more participants than I would need for a given session in case some participants would not show up to the session.
7. Zoom link is sent with an invitation email; I check that they have indeed registered for a given session before admitting them to the Zoom meeting room.
8. I draw with replacement a number from 1 to 40 using Google’s random number generator (which is displayed by searching with “random number generator”). If no participant has a coin with the drawn number, I draw next number until the number of participants is 16. I share my computer screen so that participants see the numbers are actually drawn randomly.
Part 1: Individual practice stage

Participants work on the puzzle individually with an incentive (0.2€ for each puzzle they solve). They can solve as many puzzles as possible with increasing difficulty (maximum 15 puzzles) in 4 minutes. This part familiarizes them with the puzzle and provides us with a measure of their ability given by the number of puzzles they solve. After the 4 minutes are over, they receive information on how many puzzles they have solved.

Part 2: Collaborator selection stage

Part 2 contains 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 allocated to a group of 8 based on their ability similarity as measured in part 1. This is done to reduce ability difference among participants and participants do not know this grouping criterion.

Second, participants are paired with another randomly chosen participant in the same group and 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 the paired participant’s move. Reversing the partner’s move is what I call correction in this paper. Each participant’s contribution in a given puzzle is measured by net good moves as defined in equation 1 above. Panel A of figure 3 shows a sample puzzle screen where a participant is paired with another participant called Giovanni and waiting for Giovanni to make his move.

Once they finish the puzzle, participants state 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 7 members of their group once and only once. In each round, participants solve another puzzle in a pair, then state whether they would like to collaborate with the same participant in part 3. The sequence of puzzles is the same for all pairs in all sessions. The puzzle difficulty is kept the same across the seven rounds. The minimum number of moves to solve the puzzles is set to 8 based on the pilot.

The paired participant’s first name is displayed on the computer screen throughout the puzzle and when participants select their collaborator to subtly inform the paired participant’s gender. Panel B of figure 3 shows an example of the collaborator selection screen where a participant finished playing a puzzle with another participant called Giovanni and must state whether she or he would like to collaborate with Giovanni in part 3.

At the end of part 3, participants are paired according to the following algorithm:

1. For every participant, call it i, I count the number of matches; that is, the number of other participants in the group who were willing to be paired 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 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 two participants that 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.

**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 and this is told to both participants as in part 2. They can solve as many puzzles as possible with increasing difficulty (maximum 20 puzzles).

**Post-experiment**

Each participant answers a short questionnaire which consists of (i) the six hostile and benevolent sexism questions used in Stoddard, Karpowitz, and Preece (2020) with US college students and (ii) their basic demographic information and what they have thought about the experiment. The answer to the sexism questions is used to construct a gender bias measure (see Appendix B for the construction of the measure) and their demographic information is used to know participants’ characteristics as well as casually check whether they have anticipated that the experiment is about gender.  

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

**2.2 Implementation**

The experiment was programmed with oTree (Chen, Schonger, and Wickens 2016) and conducted in Italian during November-December 2020. I recruited 464 participants (244 female and 220 male) registered on the Bologna Laboratory for Experiments in Social Science’s ORSEE (Greiner 2015) 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). The first two conditions were to reduce noise coming from differences in socio-demographic backgrounds and race or/and ethnicity that may be inferred from participants’ first name or/and voice and the last condition was to reduce experimenter demand effects. The number of participants was determined by a power simulation in the pre-analysis plan to achieve 80% power. The experiment is pre-registered with the OSF.

---

10. None has anticipated that the puzzle is about gender.
11. The laboratory prohibits deception, so no participant has participated in an experiment with deception.
12. 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 exclusion of these 16 participants.
13. The pre-registration documents are available at the OSF registry: https://osf.io/tgyc5.
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 the maximum 25€ and the minimum 2€, all including the 2€ show-up fee.

3 Data

I use part 2 data in the analysis as part 2 is where we can observe collaborator selection decisions. I aggregate the move-level data at each puzzle so that we can associate behaviors in the puzzle to the collaborator selection decisions.

3.1 Participants’ characteristics

Table 1 describes participants’ characteristics. Male participants are slightly older than female participants by 1.4 years and more gender-biased. People from southern Italy are slightly overrepresented for both female and male participants. Female participants are more likely to major in humanities and male participants are more likely to major in natural sciences and engineering, a tendency observed in most OECD countries (see, for example, Carrell, Page, and West 2010). Most female and male participants are either bachelor or master students (97% of female and 94% of male).

3.2 Move-level summary

Figure 4 shows the average move quality along with 95% confidence intervals (panel A), the fraction of total moves in each move (panel B), and the probability of corrections in each move (panel C), separately for female only (gray), male only (white), and mixed gender pairs (blue). Panel A shows that for all kinds of pairs, the average move quality is around 0.8 (8 out of 10 are good moves) until the 8th move (the minimum number of moves to solve a puzzle). After the 8th move, move quality deteriorates and stays around 0.6 (6 out of 10 are good moves). Panel B shows that for all kinds of pairs, about 71% of the puzzles are solved within 8 moves ((0.0875-0.025)/0.0875 \approx 0.71), which is the minimum number of moves to solve the puzzle, then the other 30% takes more. Panel C shows that corrections happen across the moves, but are more likely to happen after the 8th move.

3.3 Puzzle-level summary

Table 2 describes own (panel A) and partner’s puzzle behaviors (panel B) and puzzle outcomes (panel C). Panel A shows that there is no gender differences in puzzle solving ability: both contribution in part 2 and the number of puzzles solved in part 1, the difference between female and male participants are slightly older than female participants by 1.4 years and more gender-biased. People from southern Italy are slightly overrepresented for both female and male participants. Female participants are more likely to major in humanities and male participants are more likely to major in natural sciences and engineering, a tendency observed in most OECD countries (see, for example, Carrell, Page, and West 2010). Most female and male participants are either bachelor or master students (97% of female and 94% of male).

---

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

15. Individual fixed effects in the analysis control for one’s major. However, I do not run heterogeneity analysis by major because major choice is endogenous to one’s gender.
and male participants are statistically insignificant at 5% and quantitatively insignificant. 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 difference I would find is unlikely to come from their ability difference. Panel A also shows that there are no gender differences in propensity to correct partners, suggesting any gender differences I would find are not coming from either gender corrects more than the other gender.

Figure 5 presents the distribution of contribution to further elaborate panel A of table 2 that women and men are equally good at puzzle solving. First, panel A shows that most participants contributed the same degree, in about 70% of the puzzles participants’ contribution is 4 (total good moves minus total bad moves) and women’s and men’s distributions almost overlap. Second, panel B shows that the same conclusions hold when we limit the sample to solved puzzles: nearly 80% of participants contributed the same degree and women’s and men’s distributions almost overlap.

Panel B shows that puzzle solving ability of partners paired with female and male participants is the same as well we propensity to make corrections (both of a mistake and of a right move), suggesting random pairing was successful and that any gender differences I would find is not coming from partners of either gender correct more often. Participants are corrected by their partner in 15-16% of the total puzzles, of which 10-11% are corrections of mistakes and 5% are corrections of a right move.

Panel C shows that participants state they want to collaborate with the partner 71-72% of the time. Participants spend on average 43-44 seconds for each puzzle (the maximum time a pair can spend is 120 seconds) and take 11 moves (remember the minimum number of moves to solve the puzzle is 8). 85-86% of the puzzles are solved and participants and the partner correct each other’s move consecutively in 4% of the puzzles. There is no gender difference in any of these outcomes, suggesting any gender differences cannot be attributed to imbalance in these outcomes.

16. This definition of contribution is what Isaksson (2018) defines as “performance.” In the pre-analysis plan, I defined i’s contribution as “performance” of i divided by sum of “performance” of i and j and truncated values outside [0,1]. However, in my data, there is truncation in more than 10% of the puzzle and the original contribution measure may not appropriately reflect participants’ actual contribution. Thus, I instead use this “performance” measure in my analysis; since I add individual fixed effects, whether a measure is relative or absolute does not matter. Nonetheless, the same results hold when I use original contribution measure.

17. The correlation coefficient between contribution and number of puzzles solved in part 1 is 0.1059 and the p-value is below 0.00000005 (with standard errors clustered at individual level).

18. Of the 3180 puzzle, there are 495 puzzles where at least one correction occurred, of which 325 puzzles experienced only good corrections and 110 only bad corrections. The remaining 60 puzzles experienced both good and bad corrections. In order for good and bad corrections to capture only good and bad correction effect, I classify these 60 puzzles to good corrections if there were more good corrections than bad corrections (19 puzzles) and to bad corrections otherwise (41 puzzles). This classification is a bit arbitrary, but the results are robust to excluding these 60 puzzles, which I show in section 8.

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

20. Note that time spent to solve a puzzle is endogenous to correction and not a good control. For example, if one corrects a mistake, then it takes fewer time to solve the puzzle. If one corrects a right move, on the other hand, then it takes more time to solve the puzzle.
3.4 Balance across rounds

Remember that each participant plays the puzzle for seven rounds and variables unaffected by treatment (interactions within a randomly-formed pair) must be balanced. Figure 6 plots average partner gender balance (fraction of female partners, panel A) and puzzle outcomes (panels B-H) across seven rounds along with their 95% confidence intervals, separately for female (blue) and male participants (green).

First, panel A shows that partner gender is roughly balanced across rounds, except in the first round where female participants are less likely to face female partners and male participants more likely to female participants. Second, panels B-H show that most outcome variables are unbalanced across rounds both for female and male participants; specifically, whether a participant is selected as a collaborator and a puzzle is solved are lower in rounds 6 and 7. Also, while the number of corrections, time a pair spends on the puzzle, and total moves – all of which are likely to affect collaborator selection – are higher in rounds 6 and 7. It is unclear why there are these imbalances across rounds because all puzzles are the same difficulty: it could be that participants got tired in later rounds, puzzles in rounds 6 and 7 are perceived more difficult, etc.

However, they are all outcomes of a particular pair so they are just correlations. I show in section 8 that the results are robust to exclusion of rounds 6 and 7.

4 Theoretical framework

I provide a simple theoretical framework to provide a benchmark for rational agent’s behaviors.

I consider a participant i who maximizes her or his expected utility by selecting their collaborator j from a set of i’s potential collaborators \( J \equiv \{1, 2, 3, 4, 5, 6, 7\} \). i’s utility depends on her or his payoff and emotion. The utility is increasing in the payoff and the payoff is increasing in i’s belief about j’s ability. Thus, if i would select with whom to play in part 3, she or he would face the following problem:

\[
\max_{j \in J} E_{\mu_j}[u_i(\pi(\mu_j(\tilde{a}_j, c_j, f_j)), \kappa_i(c_j, f_j))|\theta_i, \omega_i], \quad \partial u_i/\partial \pi > 0, \quad \partial \pi/\partial \mu_j > 0
\]

where each term is defined as follows:

- \( \mu_j \): i’s belief about j’s ability
- \( \tilde{a}_j \): j’s ability perceived by i
- \( c_j \): j’s correction (=1 if j corrected i, =0 not corrected)
- \( f_j \): j’s gender (=1 if female, =0 if male)
- \( \theta_i \): i’s belief about her or his ability relative to other participants (>0 if high, =0 if same, <0 if low)
- \( \omega_i \): j’s belief about women’s ability relative to men (>0 if high, =0 if same, <0 if low)

I assume:

- \( \mu_j \) is increasing in j’s ability perceived by i: \( \partial \mu_j/\partial \tilde{a}_j > 0 \)
- i’s utility is decreasing in her or his emotion: \( \partial u_i/\partial \kappa_i < 0 \)
- emotion is irrelevant if i is fully rational: \( u_i(\pi, \kappa_i) \propto u_i(\pi) \)
If i can fully observe j’s move quality and i is fully rational, then j’s correction, $c_j$, and gender, $f_j$, do not convey any information about j’s ability and is irrelevant for i’s decision making. However, since i can only partially observe j’s move quality, j’s correction and gender convey information about j’s ability even if i is fully rational.\textsuperscript{21}

4.1 When i is fully rational

First, keeping j’s ability perceived by i ($\tilde{a}_j$) fixed, the information j’s correction conveys depends on $\theta_i$. If i believes she or he is good at the puzzle, she or he would consider a correction as a signal of low ability because i believes her or his move is correct. On the other hand, if i believes her or his ability is low, then she or he would consider a correction as a signal of high ability. If i believes his ability is the same as j’s, then a correction would not convey any information. Thus,

- $\partial \mu_j / \partial c_j < 0$ if $\theta_i > 0$,
- $\partial \mu_j / \partial c_j = 0$ if $\theta_i = 0$, and
- $\partial \mu_j / \partial c_j > 0$ if $\theta_i < 0$.

Similarly, if i believes women is better at the puzzle, she or he would consider a correction from a woman as a signal of high ability relative to men’s correction. On the other hand, if i believes women is worse at the puzzle, then she or he would consider a correction from a woman as a signal of low ability relative to men’s correction. If i believes women and men are equally good at the puzzle, then a correction from a woman or man is irrelevant. Thus,

- $\partial^2 \mu_j / \partial c_j \partial f_j > 0 \forall \theta_i > 0$,
- $\partial^2 \mu_j / \partial c_j \partial f_j > 0 \forall \theta_i = 0$, and
- $\partial^2 \mu_j / \partial c_j \partial f_j < 0 \forall \theta_i < 0$.

4.2 When i is not fully rational

When i is not fully rational, i’s emotion, $\kappa_i$, enters in her or his maximization problem. Specifically, I assume that j’s correction induces i’s negative feeling towards j. Also, I assume corrections by women induce i’s stronger negative feeling towards j. Thus,

- $\partial \kappa_i / \partial c_j < 0$ and
- $\partial^2 \kappa_i / \partial c_j \partial f_j < 0$.

5 Response to being corrected

In this section, I document evidence that people – both women and men – understand the notion of good and bad moves. However, they are less willing to work with a person who corrected their move after controlling for that person’s contribution to the puzzle, even if that person makes good corrections.

\textsuperscript{21} I nonparametrically control for j’s gender, but I also examine the effect of interaction term between j’s correction and j’s gender.
5.1 Response to being corrected: Estimating equation

I run the following OLS regression.

\[
\text{Select}_{ij} = \beta_1 \text{Corrected}_{ij} + \beta_2 \text{Female}_j + \delta \text{Contribution}_j + \mu_i + \epsilon_{ij}
\]  

(3)

where each variable is defined as follows:

- \( \text{Select}_{ij} \in \{0, 1\} \): an indicator variable equals 1 if \( i \) selects \( j \) as their collaborator, 0 otherwise.
- \( \text{Corrected}_{ij} \in \{0, 1\} \): an indicator variable equals 1 if \( i \) is corrected by \( j \), 0 otherwise.
- \( \text{Female}_j \in \{0, 1\} \): an indicator variable equals 1 if \( j \) is female, 0 otherwise.
- \( \text{Contribution}_j \in \mathbb{Z} \): \( j \)'s contribution to a puzzle played with \( i \).
- \( \epsilon_{ij} \): omitted factors that affect \( i \)'s likelihood to select \( j \) as their collaborator.

and \( \mu_i \equiv \sum_{k=1}^{N} \mu_k 1[i = k] \) is individual fixed effects, where \( N \) is the total number of participants in the sample and 1 is the indicator variable. Standard errors are clustered at the individual level.\(^{22}\)

The key identification assumption is that \( \text{Contribution}_j \) fully captures \( j \)'s ability perceived by \( i \) (not true ability).\(^{23}\) This assumption is reasonable if we think participants’ willingness to collaborate is increasing in the partner’s contribution to the puzzle, which is consistent with that participants can partially observe their partners’ ability and their expected utility is increasing in their payoff.

5.2 Response to being corrected: Results

Table 3 presents the regression results of equation 3. Column 3 shows that when we do not control for partner’s contribution, the coefficient estimate on bad correction is negative and very large: the point estimate is 0.580 (p-value < 0.01). That is, participants are 58% less willing to collaborate with partners who made a bad correction, a correction which made the puzzle far away from the solution. This is evidence that my experimental design is valid: participants correctly understand the notion of good and bad moves and that participants are more willing to collaborate with partners who contributed more. This is true both for women (column 5, the point estimate -0.634 and the p-value < 0.01) and men (column 7, the point estimate -0.522 and the p-value < 0.01).

Looking at columns 2, 4, 6, and 7, the coefficient estimate on the partner’s contribution is positive and quantitatively and statistically highly significant: in column 2, the point estimate is 0.083 (p-value < 0.01). This suggests that participants are 8.3% more willing to collaborate with partners who make one more good move.

\(^{22}\) 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.

\(^{23}\) By random pairing of participants, the paired participant’s gender is exogenous to participant’s unobservables. However, correction is not exogenous for two reasons: (i) correction can be correlated with the paired participant’s ability and paired participant’s ability can affect participant’s collaborator selection; (ii) There is an effect similar to the reflection effect: participant’s puzzle behavior affects the paired participant’s behavior and vice versa; for example, a participant’s meanness can increase the paired participant’s correction and can also affect her of his collaborator selection. The identification assumption concerns the former point. To address the latter point, I add individual fixed effects.
The coefficient estimate on being corrected in column 2 is negative and quantitatively and statistically highly significant with the point estimate -0.198 (p-value < 0.01). This suggests that people are 19.8% less willing to collaborate with those who corrected their move, which corresponds to an increase in contribution by 0.85 standard deviation.\(^{24}\)

When we split the correction into good and bad corrections, column 4, people rightly respond negatively to bad corrections: the point estimate -0.135 (p-value < 0.01). This suggests that people are 13.5% less willing to work with a person who made a good correction. Columns 6 and 8 show this effect is mainly driven by women’s response (to be discussed more in section 7). However, people also respond negatively to good corrections: the point estimate is -0.223 (p-value < 0.01). This suggests that people are 22.3% less willing to work with a person who corrected their mistakes. The effect is present for both women (column 6) and men (column 8). Thus, people are less willing to collaborate with a person who corrected their moves, even if the corrections move the puzzle closer to the solution.

6 Do women’s corrections receive stronger negative reactions?

In this section, I document that neither men nor women underestimate women’s contribution and that women’s corrections do not receive stronger negative reactions by either women or men.

6.1 Do women’s corrections receive stronger negative reactions? Estimating equation

I run the following OLS regression.

\[
Select_{ij} = \beta_1 \text{Corrected}_{ij} + \beta_2 \text{Female}_j + \beta_3 \text{Corrected}_{ij} \times \text{Female}_j + \delta_1 \text{Contribution}_j + \delta_2 \text{Contribution}_j \times \text{Female}_j + \mu_i + \epsilon_{ij}
\]  

(4)

Where each variable is defined as in equation 3.

6.2 Do women’s corrections receive stronger negative reactions? Results

Table 4 presents the regression results of equation 4. Columns 1 and 3 show that the coefficient estimate on the interaction between partner’s contribution and female partner is almost 0 and statistically insignificant even at 10%. This suggests that neither women nor men underestimate women’s contribution when selecting a partner.

Looking at columns 1 and 2, the coefficient estimate on the interaction between female partner and correction, female partner and good correction, and female partner and bad correction are all positive although statistically insignificant. This suggest that if anything, women respond slightly less negatively to women’s correction.

Looking at columns 3 and 4, the coefficient estimate on the interaction between female partner and correction, and female partner and good correction are negative although statistically insignificant. The coefficient estimate on the interaction between female partner and bad

\(^{24}\) The standard deviation is taken from panel B of table 2 and is simple arithmetic average of partners faced by women and men: (2.73+2.87)/2=2.8.
correction is positive although statistically insignificant. This suggest that men do not respond more negatively to women’s correction. Corroborating this, columns 5 and 6 show that there is no statistically significant difference between women’s and men’s response to women’s corrections. Thus, women’s corrections do not receive stronger negative reactions by either women or men.

7 Why do women and men respond differently to corrections?

In table 3, we saw that men do not respond negatively to bad corrections while women correctly do so. In this section, as an exploratory analysis, I document evidence that men’s negative response is irrational and that men are less able to take into account for partners’ contribution in collaborator selections.

A caveat is that the number of good and bad corrections are not large enough and I may be picking up some irregularities in the data.

7.1 Why do women and men respond differently to corrections? Estimating equation

I run the following OLS regression.

\[
\text{Select}_{ij} = \beta_1 \text{Corrected}_{ij} + \beta_2 \text{Female}_j + \beta_3 \text{Corrected}_{ij} \times \text{HighAbility}_i \\
+ \delta_1 \text{Contribution}_j + \delta_2 \text{Contribution}_j \times \text{HighAbility}_i + \mu_i + \epsilon_{ij}
\] (5)

where each variable is defined as follows:

- \( \text{HighAbility}_i \in \{0, 1\} \): an indicator variable equals 1 if i solved the above-median number of puzzles in part 1 in a session she or he has participated, 0 otherwise.

Other variables are as defined in equations 3.

The idea of this regression is the following. Negative response to corrections is not necessarily irrational; because the quality of corrections is not fully observable, it could be that participants have interpreted the correction as a signal of a partner’s low ability. This is likely because they must have believed their move was good, otherwise they have not taken that move. However, people who solved a larger number of puzzles in part 1 (what I define as high-ability in equation 5) should be able to observe move quality better than people who solved a fewer number of puzzles in part 1 (what I define as low-ability in equation 5; it is the baseline category so not shown in the equation). Thus, if the negative response to corrections is rational, people who solved a larger number of puzzles in part 1 should respond less negatively to good corrections and more negatively to bad corrections.

7.2 Why do women and men respond differently to corrections? Results

In column 1 of Table 5, the coefficient estimate on the interaction between correction and male is positive and marginally statistically significant at 10%, suggesting that men respond less negatively than women to being corrected. However, this does not mean men are more efficient in selecting their collaborator: in column 2, while the coefficient estimate on the interaction
between good correction and male is statistically insignificant, the coefficient estimate on the interaction between bad and male is positive and statistically significant at 5% level: men are 16.9% more willing to collaborate with those who made bad corrections than women. In addition, while not very robust, the coefficient estimate on the interaction between partner’s contribution and male is negative and statistically significant at 5%-10% (columns 1 and 2), suggesting that men are less able to take into account for partner’s contribution when selecting a collaborator.

Column 3 of Table 5 presents the regression results of equation ?? for women. The coefficient estimate on the interaction between good correction and high ability and the coefficient estimate on the interaction between bad correction and high ability are both negative but statistically insignificant. Thus, we cannot say whether women’s negative response to corrections are rational or irrational.

Column 4 of Table 5 presents the regression results of equation ?? for men. The coefficient estimate on the interaction between good correction and high ability is negative and statistically significant with the point estimate -0.183 and the p-value < 0.05. However, the coefficient estimate on the interaction between bad correction and high ability is negative but not statistically significant. Thus, at least for men, the negative response to corrections are irrational.

8 External validity and robustness

In this section, I argue that the findings so far – that people are less willing to collaborate with those who have corrected them even if it is a good correction – are likely to be lower bound. I also show that the findings are robust to alternative explanations.

8.1 External validity

While the laboratory setting is different from the real-world workplace, my findings are likely to be a lower bound because of the three reasons. First, being corrected is not observed by others in my experiment: those who have been corrected do not face any reputational cost, unlike in the real-world workplace. Second, the emotional stake is much smaller: it is a puzzle not very relevant for their work or study; it is not something people have been devoting much of their time to, such as university exams, academic research, and corporate investment projects. Third, participants are equal in my experiment; in a real-world, on the other hand, there are sometimes senior-junior relationships and corrections from junior people may induce stronger negative reactions.

But there are two caveats. First is that every participants are stranger to each other in my experiments while people know each other in the real-world workplace. Thus, it is possible that repeated interactions would mitigate people’s negative response to corrections (but they may magnify the negative response due to rivalry, failure to build a good rapport, etc.). Second is that most participants are bachelor or master students who are supposed to have weaker gender bias. Women’s corrections may receive stronger negative reactions if participants are older.
8.2 Robustness

Excluding unsolved puzzles  Whether participants can solve a puzzle is an outcome of a particular pairing that is random. However, “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). If a participant is not trying to solve the puzzle, then a pair is unlikely to solve the puzzle and good and bad corrections may not be meaningful.

To address this concern, I re-estimate equations 3, 4, and 5 with solved puzzles only. Green dots and lines in figures 7, 8, and 9 show that the results we have seen so far (plotted with blue dots and lines in these figures) are robust to limiting the sample to solved puzzles only. However, results that high-ability men react more negatively to good corrections lose statistical significance, may be due to drop in the sample size.

Excluding rounds 6 and 7  We see in figure 6 that participants are less willing to collaborate with the paired participants in rounds 6 and 7. Also, there are more corrections in rounds 6 and 7 than in other rounds. Although they are both outcomes of particular pairs, one may wonder whether rounds 6 and 7 are driving the results.

To address this concern, I re-estimate equations equations 3, 4, and 5 with rounds 1-5 only. Red dots and lines in figures 7, 8, and 9 show that the results we have seen so far (plotted with blue dots and lines in these figures) are robust to excluding rounds 6 and 7.

Excluding puzzles where both good and bad corrections occurred  As discussed in the footnote of section 3.3, there are 495 puzzles in which at least one correction occurred, of which 325 puzzles experienced good corrections only, 110 puzzles bad corrections only, and 60 puzzles experienced both good and bad corrections. In puzzles that experienced both good and bad corrections, I considered that the puzzles experienced good corrections if there were more good corrections than bad corrections, and experienced bad corrections otherwise. However, some people may think classification is a bit arbitrary.

To address this concern, I re-estimate equations equations 3, 4, and 5 in which only good or bad corrections occurred. Purple dots and lines in figures 7, 8, and 9 show that the results we have seen so far (plotted with blue dots and lines in these figures) are robust to excluding puzzles where both good and bad corrections occurred.

9 Conclusions

This paper studies whether people dislike working together with someone who corrects them and more so when that person is a woman. I design a quasi-laboratory experiment where participants are paired with seven other participants, solve one number-sliding puzzle together, and express a preference on which of them to be paired with in the final, payoff-relevant, part of the experiment. I find that participants understand the notion of good and bad moves and more willing to work with people who contributed more to the puzzle, validating my experimental design. However, once I control for the paired participants’ contribution to the puzzle, participants are significantly less willing to collaborate with a person who has corrected their moves even if the corrections
move the puzzle closer to the solution. However, I do not find people respond to corrections by women more negatively than corrections by men.

This paper have three contributions. First, I introduce correction in the contribution of ideas and examines its cost on women and on group efficiency. Second, I promote our understanding of gender differences in group work. Third, I show the fact that people are less willing to collaborate with people who have corrected them can be another source of organizational inefficiency.
References


Figures

Figure 1: Timeline of the experiment

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

Figure 2: Zoom screen

Notes: This figure shows a Zoom screen participants would see during the roll call. The experimenter’s camera is on during the roll call. Participants would see this screen throughout the experiment but the experimenter’s camera may be turned off.
Figure 3: Puzzle and collaborator selection screen

(a) Puzzle screen

Il puzzle 4 su 7

Tempo rimasto per completare questa pagina: 1:54

Stai risolvendo il puzzle con Giovanni

1 2 3
8 7 5
4 6

Aspetta il tuo partner!

(b) Collaborator selection screen

Il puzzle 4 su 7

Hai risolto il puzzle con Giovanni. Sei disposto a lavorare con Giovanni nella parte 3?

☐ Sì
☐ No

Successivo

Notes: Panel A shows a sample puzzle screen where a participant is matched with another participant called Giovanni at the 4th round puzzle and waiting for Giovanni to make his move. Panel B shows a sample collaborator selection screen where a participant finished solving the 4th round puzzle with Giovanni and deciding whether she or he would like to collaborate with Giovanni in part 3.
Figure 4: Move quality, fraction of total moves, and probability of corrections

Panel A: Average move quality and 95% confidence intervals (1 if good, 0 if bad)

Panel B: Fraction of total moves

Panel C: Probability of corrections

Notes: The average move quality along with 95% confidence intervals (panel A), the fraction of total moves in each move (panel B), and the probability of corrections in each move (panel C), separately for female only (gray), male only (white), and mixed gender pairs (blue). The confidence interval of panel A is 95% confidence intervals of $\beta$s from the following OLS regression: $\text{MoveQuality}_{ijt} = \beta_1 + \sum_{k=2}^{58} \beta_k \mathbb{1}[t_{ij} = k] + \epsilon_{ijt}$, where $t_{ij}$ is the pair i-j’s move round and $\mathbb{1}$ is an indicator variable. $\text{MoveQuality}_{ijt}$ takes a value of 1 if a move of a pair i-j in tth move is good and 0 if bad. I add an estimate of $\beta_1$ to estimates of $\beta_2$-$\beta_{58}$ to make the figure easier to look at. Standard errors are clustered at the pair level.
Figure 5: Distribution of contribution and gender differences

Notes: This figure shows the distribution of individual contribution in all puzzles (panel A) and in solved puzzles (panel B). They show that most participants contributed the same degree and that there is no gender difference in contribution. Equation 1 provides definition of contribution.
Figure 6: Balance across rounds

Panel A. Partner gender balance (frac. female)

Panel B: Willing to collaborate (yes=1, no=0)

Panel C: Correction

Panel D: Correction of a mistake

Panel E: Correction of a right move

Panel F: Time spent (sec.)

Panel G: Total moves

Panel H: Puzzle solved

Participant’s gender

Notes: This figure shows point estimates and 95% confidence intervals of $\beta$s from the following OLS regression with gender balance (female dummy) and different puzzle outcomes separately for female (blue) and male participants (green): $y_{ij} = \beta_1 + \sum_{k=2}^{7} \beta_k [t_{ij} = k] + \epsilon_{ij}$, where $t_{ij} \in \{1, 2, 3, 4, 5, 6, 7\}$ is the puzzle round in which i and j are playing, $I$ is an indicator variable, and $y_{ij}$ is outcome variable indicated in each panel. I add the estimate of $\beta_1$ to estimates of $\beta_2$-$\beta_7$ to make the figure easier to look at. Standard errors are clustered at the individual level.
Figure 7: Response to being corrected: Robustness

Panel A. Female

Panel B. Male

Notes:
Figure 8: Do women’s corrections receive stronger negative reactions? Robustness

Figure 9: Why do women and men respond differently to corrections? Robustness

Notes:
### Table 1: Participants’ characteristics

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

**Notes:** This table describes participants’ characteristics. Gender bias is measured with the 6 hostile and benevolent sexism questions and constructed as in Appendix B. P-values of the difference between female and male participants are calculated with heteroskedasticity-robust standard errors.
Table 2: Own and partners’ puzzle behaviors and puzzle outcomes

<table>
<thead>
<tr>
<th>Panel A: Own behaviors</th>
<th>Female (N=1708)</th>
<th>Male (N=1540)</th>
<th>Difference (Female – Male)</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contribution</td>
<td>2.98 2.93</td>
<td>3.14 2.64</td>
<td>-0.16 0.10</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># puzzles solved in pt. 1</td>
<td>8.36 2.41</td>
<td>8.80 2.34</td>
<td>-0.44 0.22</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correction</td>
<td>0.15 0.36</td>
<td>0.16 0.36</td>
<td>0.00 0.01</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correction of a mistake</td>
<td>0.10 0.31</td>
<td>0.11 0.32</td>
<td>-0.01 0.01</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correction of a right move</td>
<td>0.05 0.22</td>
<td>0.05 0.21</td>
<td>0.00 0.01</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Partner’s behaviors

<table>
<thead>
<tr>
<th></th>
<th>Female (N=1708)</th>
<th>Male (N=1540)</th>
<th>Difference (Female – Male)</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contribution</td>
<td>3.04 2.73</td>
<td>3.07 2.87</td>
<td>-0.03 0.10</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># puzzles solved in pt. 1</td>
<td>8.58 2.35</td>
<td>8.57 2.43</td>
<td>0.01 0.16</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correction</td>
<td>0.16 0.37</td>
<td>0.15 0.36</td>
<td>0.01 0.01</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correction of a mistake</td>
<td>0.11 0.31</td>
<td>0.10 0.31</td>
<td>0.01 0.01</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correction of a right move</td>
<td>0.05 0.21</td>
<td>0.05 0.21</td>
<td>0.00 0.01</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel C: Puzzle outcomes

<table>
<thead>
<tr>
<th></th>
<th>Female (N=1708)</th>
<th>Male (N=1540)</th>
<th>Difference (Female – Male)</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SE</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Willing to collaborate (yes=1, no=0)</td>
<td>0.72 0.45</td>
<td>0.71 0.45</td>
<td>0.01 0.02</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time spent (sec.)</td>
<td>43.74 36.15</td>
<td>42.99 35.76</td>
<td>0.74 1.28</td>
<td>0.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total moves</td>
<td>11.18 7.46</td>
<td>11.21 7.70</td>
<td>-0.03 0.28</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puzzle solved</td>
<td>0.85 0.36</td>
<td>0.86 0.35</td>
<td>-0.01 0.01</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consecutive correction</td>
<td>0.04 0.20</td>
<td>0.04 0.21</td>
<td>0.00 0.01</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 female and male participants are calculated with standard errors clustered at the individual level. Equation 1 provides definition of contribution.
Table 3: Response to being corrected

<table>
<thead>
<tr>
<th>Outcome: Willing to collaborate (yes=1, no=0)</th>
<th>Sample: All</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Correction</td>
<td>-0.367***</td>
<td>-0.198***</td>
<td></td>
</tr>
<tr>
<td>Good correction</td>
<td>-0.267***</td>
<td>-0.223***</td>
<td>-0.304***</td>
</tr>
<tr>
<td>Bad correction</td>
<td>-0.580***</td>
<td>-0.135***</td>
<td>-0.634***</td>
</tr>
<tr>
<td>Female partner</td>
<td>-0.006</td>
<td>0.008</td>
<td>-0.003</td>
</tr>
<tr>
<td>Partner’s contribution</td>
<td>0.083***</td>
<td>0.085***</td>
<td>0.090***</td>
</tr>
<tr>
<td>Individual FE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Notes: This table presents the regression results of equation 3 and shows that people – both women and men – understand the notion of good and bad moves. However, they are less willing to work with a person who corrected their move after controlling for that person’s contribution to the puzzle, even if that person makes good corrections. Standard errors in parentheses are clustered at the individual level. Significance levels: * 10%, ** 5%, and *** 1%.
Table 4: Do women’s corrections receive stronger negative reactions?

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Willing to collaborate (yes=1, no=0)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Correction</td>
<td>-0.260***</td>
</tr>
<tr>
<td>Good correction</td>
<td>-0.263***</td>
</tr>
<tr>
<td>Bad correction</td>
<td>-0.255***</td>
</tr>
<tr>
<td>Female partner</td>
<td>-0.001</td>
</tr>
<tr>
<td>Partner’s contribution</td>
<td>0.090***</td>
</tr>
<tr>
<td>Correction x Female partner</td>
<td>0.047</td>
</tr>
<tr>
<td>Good correction x Female partner</td>
<td>0.035</td>
</tr>
<tr>
<td>Bad correction x Female partner</td>
<td>0.077</td>
</tr>
<tr>
<td>Partner’s contribution x Female partner</td>
<td>-0.001</td>
</tr>
<tr>
<td>Correction x Male</td>
<td>0.135**</td>
</tr>
<tr>
<td>Good correction x Male</td>
<td>0.127*</td>
</tr>
<tr>
<td>Bad correction x Male</td>
<td>0.151</td>
</tr>
<tr>
<td>Partner’s contribution x Male</td>
<td>-0.013</td>
</tr>
<tr>
<td>Correction x Female partner x Male</td>
<td>-0.097</td>
</tr>
<tr>
<td>Good correction x Female partner x Male</td>
<td></td>
</tr>
<tr>
<td>Bad correction x Female partner x Male</td>
<td></td>
</tr>
<tr>
<td>Partner’s contribution x Female partner x Male</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individual FE</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline mean</td>
<td>0.780</td>
<td>0.780</td>
<td>0.778</td>
<td>0.780</td>
</tr>
<tr>
<td>Baseline SD</td>
<td>0.414</td>
<td>0.414</td>
<td>0.416</td>
<td>0.416</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.365</td>
<td>0.364</td>
<td>0.305</td>
<td>0.309</td>
</tr>
<tr>
<td>Observations</td>
<td>1670</td>
<td>1670</td>
<td>1510</td>
<td>1510</td>
</tr>
<tr>
<td>Clusters</td>
<td>244</td>
<td>244</td>
<td>220</td>
<td>220</td>
</tr>
</tbody>
</table>

Notes: This table presents the regression results of equation 4 and shows that neither men nor women underestimate women’s contribution and that women’s corrections do not receive stronger negative reactions by either women or men. Standard errors in parentheses are clustered at the individual level. Significance levels: * 10%, ** 5%, and *** 1%.
Table 5: Why do women and men respond differently to corrections?

<table>
<thead>
<tr>
<th>Sample:</th>
<th>Willing to collaborate (yes=1, no=0)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correction</td>
<td>-0.237***</td>
<td>(0.030)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good correction</td>
<td>-0.245***</td>
<td>(0.034)</td>
<td>-0.226***</td>
<td>(0.044)</td>
</tr>
<tr>
<td>Bad correction</td>
<td>-0.215***</td>
<td>(0.051)</td>
<td>-0.182**</td>
<td>(0.070)</td>
</tr>
<tr>
<td>Female partner</td>
<td>0.002</td>
<td>(0.018)</td>
<td>0.008</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Partner’s contribution</td>
<td>0.090***</td>
<td>(0.004)</td>
<td>0.090***</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Correction x Male</td>
<td>0.085*</td>
<td>(0.044)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good correction x Male</td>
<td>0.052</td>
<td>(0.050)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad correction x Male</td>
<td>0.169**</td>
<td>(0.073)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partner’s contribution x Male</td>
<td>-0.012**</td>
<td>(0.005)</td>
<td>-0.010*</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Good correction x High ability</td>
<td>-0.040</td>
<td>(0.068)</td>
<td>-0.183**</td>
<td>(0.076)</td>
</tr>
<tr>
<td>Bad correction x High ability</td>
<td>-0.075</td>
<td>(0.102)</td>
<td>-0.077</td>
<td>(0.110)</td>
</tr>
<tr>
<td>Partner’s contribution x High ability</td>
<td>-0.003</td>
<td>(0.007)</td>
<td>-0.002</td>
<td>(0.008)</td>
</tr>
</tbody>
</table>

Notes: This table presents the regression results of equation 5 and shows that men’s negative response is irrational and that men are less able to take into account for partners’ contribution in collaborator selections. However, the number of good and bad corrections are not large enough and I may be picking up some irregularities in the data. Standard errors in parentheses are clustered at the individual level. Significance levels: * 10%, ** 5%, and *** 1%.
Appendix A  Pros and cons of the quasi-laboratory format

On top of logistical convenience and complying with the COVID pre-caution measures, the quasi-laboratory format has an additional benefit over physical laboratory experiments in that participants cannot see each other when they enter the laboratory which adds an additional layer of anonymity among participants. A drawback is that participants can potentially be distracted while participating.

However, unlike standard online experiments such as on MTurk and Prolific where participants’ identity is fully anonymous by the platforms’ rule, we have participants’ personal information and participants know it as we recruit them from our standard laboratory subject pool. Also, they are connected to us via Zoom throughout the experiment. These mostly prevent participants’ attrition that can be endogenous to their decisions or treatments and the main problem of online interactive experiments (Arechar, Gächter, and Molleman 2018) and experiments where treatments affect the probability of attrition, e.g., experiments with intertemporal decision making. In my experiment, we experienced no participant attrition. A drawback is that we could not collect a large number of observations.

Another benefit of quasi-laboratory experiments over standard online experiments is that we can screen participants based on their participation status in previous experiments. This allows us to collect cleaner data; in particular, this allows us to screen out participants who have participated in experiments with deception, which is another problem of online experiments (Arechar, Gächter, and Molleman 2018).

There are already a few other studies that use a quasi-laboratory format, for example, Goeschl, Oestreich, and Soldà (2021).

Appendix B  Construction of the gender bias measure

I construct the gender bias measure following Stoddard, Karpowitz, and Preece (2020) who use the measure to measure sexism of US undergraduate students.

As discussed in section 2.1, I ask participants to answer the following six hostile and benevolent sexism questions Stoddard, Karpowitz, and Preece (2020) have chosen from Glick and Fiske (1996)’s full-length sexism questionnaire.

Instructions: 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.

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

Answer choices to each question: Strongly agree, Agree a little, Neither agree nor disagree, Disagree a little, Strongly disagree

I assign a value of 4 to “Strongly agree,” 3 to “Agree a little,” 2 to “Neither agree nor disagree,” 1 to “Disagree a little,” and 0 to “Strongly disagree.” Then I sum up the values for each participant and divide the sum by 24 which is the highest value one can receive. Thus, the measure takes a value from 0 to 1, and the higher the measure, the more gender-biased the person is. In the experiment, I use a certified Italian translation from Manganelli Rattazzi, Volpato, and Canova (2008) and Rollero, Glick, and Tartaglia (2014).
Appendix C  Experimental instructions
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]
General instructions

Overview: 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.

Your payment: 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.

Confidentiality: 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).

General rules: 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.

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

**Example 1:**

First, consider the following puzzle.

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.
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.

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.
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.

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

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.
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.

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

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.
This time, let’s try moving 7 down.

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

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

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

- \( \sqrt{2} \)
- 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
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
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]

---

**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.