I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10825

# I (Don't) Like You! But Who Cares? Gender Differences in Same Sex and Mixed Sex Teams

Leonie Gerhards
Michael Kosfeld

JUNE 2017

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10825

# I (Don't) Like You! But Who Cares? Gender Differences in Same Sex and Mixed Sex Teams

**Leonie Gerhards**
*University of Hamburg*

**Michael Kosfeld**
*Goethe University Frankfurt and IZA*

JUNE 2017



**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

# I (Don't) Like You! But Who Cares? Gender Differences in Same Sex and Mixed Sex Teams*

We study the effect of likability on female and male team behavior in a lab experiment. Extending a two-player public goods game and a minimum effort game by an additional pre-play stage that informs team members about their mutual likability we find that female teams lower their contribution to the public good in case of low likability, while male teams achieve high levels of cooperation irrespective of the level of mutual likability. In mixed sex teams, both females' and males' contributions depend on mutual likability. Similar results are found in the minimum effort game. Our results offer a new perspective on gender differences in labor market outcomes: mutual dislikability impedes team behavior, except in all-male teams.

**JEL Classification:** C90, J16
**Keywords:** gender differences, likability, experiment, team behavior

**Corresponding author:**
Michael Kosfeld
Chair of Organization and Management
Goethe University Frankfurt
Grüneburgplatz 1
60323 Frankfurt/Main
Germany
E-mail: kosfeld@econ.uni-frankfurt.de

* We would like to thank seminar participants at Goethe University Frankfurt, University of Lausanne, University of Zurich, University of Copenhagen, Kiel Institute of the World Economy, University of Birmingham and participants of the European Workshop on Experimental and Behavioral Economics 2013 and the Nordic Conference on Behavioral and Experimental Economics 2013 for their helpful comments. We particularly thank Thomas Buser and participants of the Grüneburg Seminar in Frankfurt for their useful suggestions. We are grateful for the help and support from Frank Drzensky, Yann Girard, Florian Hett, Alexander Nix, Neele Siemer and Bernard Richter in conducting the experimental sessions. Financial support from the Alfons und Gertrud-Kassel Stiftung is gratefully acknowledged.

# 1 Introduction

A growing empirical literature documents significant gender effects in economic preferences and decision-making, thereby providing important explanations for persistent differences in labor market outcomes between women and men.[1] Seminal papers have identified, inter alia, differences in overconfidence, competitiveness, risk attitudes and negotiation style (Barber and Odean, 2001; Gneezy et al., 2003; Niederle and Vesterlund, 2007; Dohmen and Falk, 2011; Babcock and Laschever, 2003).[2] In this paper, we document a new behavioral finding that to the best of our knowledge has not been identified or analyzed so far, and that offers a novel perspective on the determinants of gender differences in labor market outcomes.

The starting point for our study is the role of likability in social interactions. In many economic situations we deal and interact with others, who we either like or don't like. Suppose, for example, that you work on a small team project, together with some colleague who you are expected to cooperate with in the firm's as well as your joint interest. However, for some reason you simply don't like your colleague very much; and the same also holds for your colleague, i.e., he or she is not a big fan of you either. What does this imply for your willingness to cooperate with each other in the team? It seems reasonable to presume that the effect is negative, i.e., team members who don't like each other are less likely to cooperate than if they like each other. Our conjecture in this study is that there exists a gender difference in this effect: women may react more strongly to mutual (dis)likability than men.

There is evidence that women seem more responsive to details of the social interaction (Croson and Gneezy, 2009), but we are not aware of any study that considers gender differences in likability effects explicitly. To this extent, the approach in this paper is exploratory. The only paper that comes close is Benenson et al. (2009). In a questionnaire study with college roommates the authors analyze men's and women's thresholds of tolerance for genetically unrelated same sex individuals. Tolerance is defined as the acceptance of stresses and strains within a personal relationship. Benenson et al. show that women are less likely to accept a given level of stress with their roommate than men. This is not only reflected in their answers to the questionnaire, but also in a higher frequency of room switches and relocations within a given period of time.

The results in our study confirm these effects. Based on a novel lab experiment, which allows us to control the degree of mutual likability in a team, we find that mutual likability has indeed a stronger impact on behavior for female participants than for male participants. Moreover, the effect exclusively comes from same sex teams. In the experiment we focus on two different games, a one-shot cooperation game and a repeated coordination game. We choose these games because cooperation and coordination are essential features of teamwork

[1] With regard to gender differences in wage outcomes see, for instance, Altonji and Blank (1999), Bertrand et al. (2010) and Goldin (2014). Bertrand and Hallock (2001) and Kleven et al. (2015) provide evidence for the underrepresentation of women in higher levels of corporate hierarchies. In a similar vein, Cassidy and Kauhanen (2016) and Frederiksen and Kato (forthcoming) document that promotion probabilities within and across firms and the likelihood of being appointed to top management positions are significantly lower for women than for men.

[2] See also studies surveyed in Bertrand (2011), Croson and Gneezy (2009), and Eckel and Grossman (2008a). Buser et al. (2014) show that gender differences in competitiveness and risk taking are important drivers of career relevant choices.

across many organizational contexts. Our results reveal the same behavioral pattern in both games. In particular, in same sex teams women cooperate significantly less and also coordinate on Pareto-inferior equilibria if mutual likability in their team is low than if it is high. Men, on the other hand, achieve high levels of cooperation and coordination outcomes independent of mutual likability in their team.

Our results are in line with evidence suggesting that women are in general more responsive to details of the social interaction. As Croson and Gneezy (2009) write in their meta-study: "social preferences of women are more situationally specific than those of men; women are neither more nor less socially oriented, but their social preferences are more malleable" (p.448).[3] Interestingly, however, in case of likability this only holds for same sex teams, as our results show. In mixed sex teams, we find that both female and male participants react similarly and significantly to mutual likability in the two games. Both cooperate less and coordinate on less efficient Nash equilibria if likability in the mixed sex team is low. In sum, our experimental results document a consistent and robust finding: (dis)likability matters for team performance, except for all-male teams.

This finding has interesting implications for our understanding of gender differences in firms and organizations. If mutual likability is less important in all-male environments, dislikability becomes a hurdle – or, equivalently, likability becomes an asset – for women but not for men. Consider, for example, wage negotiations with a male superior (Babcock and Laschever, 2003). Since female employees may fear that their superior perceives them as less likable if they demand a high wage, they may be reluctant to ask for it, whereas male employees do not need to care. The important conclusion and policy recommendation is that firms should downplay (dis)likability considerations at the workplace. This may not always be possible completely, but our results suggest that a professional and "likability-neutral" work culture may enhance the career perspectives and the more general labor market outcomes of women significantly.

The remainder of the paper is organized as follows. Section 2 provides details of the design of our experiment. In Section 3 we discuss the results. Section 4 concludes.

## 2 Experimental design

In the experiment, participants were either matched with another participant from the same or from the opposite sex. We refer to the resulting two-player teams as "same sex" and "mixed sex" matchings, respectively. To achieve maximum prior anonymity between matching partners, we conducted the experiment simultaneously at two different locations in Germany that are approximately 250 kilometers apart (Frankfurt and Düsseldorf). All teams consisted of one participant from either location. Before describing the experimental games, we provide more information about the matching protocol.

[3] See also Stuhlmacher and Walters (1999) and Eckel and Grossman (2008b) for the more general influence of contextual factors on decision-making.

## 2.1 Matching protocol

In each session, always eight female and eight male subjects participated at both locations. Upon arrival at the experimental site, we took standardized portrait photographs of each participant. Participants were asked to express a neutral look and were not allowed to smile as this might have affected their impression on others (Centorrino et al., 2015). Immediately thereafter participants were presented with their own photographs and in case someone was dissatisfied with it, a second photograph was taken. Subsequently, participants drew a lottery ticket that assigned them to a computer in the lab. For the purpose of the experiment we created a virtual lab of 32 participants by connecting sixteen computers from either lab via the internet.

The experiment consisted of three parts. In each part, participants received information in form of written instructions that were distributed at the beginning of the respective part.[4] In the first part of the experiment, participants were randomly divided into four matching groups (of eight subjects each). Each matching group comprised four subjects from either location. While subjects received no information about the other group members from their own location, the first decision screen on the computer presented them with the photographs of the four group members from the other location. In the same sex matchings, these participants were all of the same sex as the subject him- or herself. In the mixed sex matchings, conversely, all of them were of the opposite sex. The instructions emphasized that the participants shown on the photographs simultaneously participated in the experiment in an experimental lab located in another German city. Apart from that, we did not provide any further details about the location or the displayed participants.

The instructions on the first decision screen asked the participants to rate, based on the photographs, how likable (in German: "sympathisch") they found each of the four displayed individuals. Ratings had to be entered separately for each photograph on an 11-point Likert scale ranging from "very unlikable" to "very likable".[5] Based on these ratings, the computer generated an ordinal ranking that was shown to the participant on decision screen 2. Rank 1 was assigned to the photograph that the participant granted the highest likability rating, rank 2 to the photograph that he or she granted the second highest likability rating, and so on. In case of a tie, multiple photographs were assigned the same rank. Then, the second decision screen asked participants to enter their final unambiguous likability ranking from 1 to 4, to rule out ties in likability ranks.

We used these likability ranks to inform subjects about their mutual likability in later stages of the experiment. The main advantage compared to absolute ratings is that relative ranks are directly interpretable and less susceptible to individual taste differences in the assessment of likability. We will, however, include subjects' absolute ratings in our analyses, as well.

To generate sufficient variation with regard to mutual (dis)likability within teams, we matched participants based on the likability rankings they provided us with. More precisely,

[4] A full set of the instructions is available at `https://www.dropbox.com/s/aed3p9rd2ds38vw/Gender_Cooperation_Coordination%5BInstructions%5D.pdf?dl=0`.

[5] See sample screen shots in Appendix B.

our matching algorithm, first, calculated the 16 sums of mutually assigned likability ranks for all possible combinations of the four plus four participants within a matching group from both locations. For example, if two participants had assigned each other rank 1, their sum of ranks was equal to 2, if both had assigned each other rank 3, their sum of ranks was equal to 6, and so forth. Across sessions, we implemented two different matching protocols. In half of the sessions, subjects with the highest sum of ranks were matched first, those with the second highest sum thereafter, and so on. In the other half of the sessions, the matching protocol worked the other way round: Subjects with the lowest sum of ranks were matched first and the remaining subjects thereafter. This procedure allowed us to make maximal use of the given ranks to obtain sufficient variation with regard to mutual likability within teams. Note that we did not provide participants with details about our matching procedure.[6]

## 2.2 Cooperation game

At the beginning of the second part of the experiment, participants received the instructions to a one-shot, linear public goods game. Notably, at this point, they were not informed yet, with whom they were going to be matched in this experimental game.

In our public goods game, two players can each contribute any integer value out of an initial endowment of 6 Euros to a joint project. Contributions to the project are multiplied by 1.5 and equally distributed between the two players. Formally, the payoff for each player is given by

$$\Pi_i^{Coop} = 6 - c_i + 0.75\ (c_1 + c_2),$$

where $c_i$ represents player $i$'s individual contribution for $i = 1, 2$. Assuming that both players maximize their individual payoff, the game has a unique Nash equilibrium, in which every player contributes zero. However, players maximize their joint payoff by contributing 6 each, yielding a payoff of 9 Euros to every player.

After participants had read the instructions and correctly answered a series of control questions, they were informed about their matching partner. For this, we displayed the matching partner's photograph on the computer screen (see sample screen shots in Appendix B). Moreover, a short text below the photograph informed both players of their mutual likability rankings. Precisely, it read: "In the following part of the experiment you will be matched to this person. You elected this person on rank 1 [2, 3, or 4] in your likability ranking, meaning you evaluated this person as most likable [rather likable, rather unlikable, or most unlikable] compared to the other persons. This person elected you on rank 1 [2, 3, or 4] in his/her likability ranking, meaning he/she evaluated you as most likable [rather likable, rather unlikable, or most unlikable] compared to other persons that were displayed to him/her."[7]

[6] Importantly, there is neither a statistically significant difference in the distributions of mutually assigned likability ranks between male and female teams nor between same sex and mixed sex teams (Fisher exact test, $p \geq 0.446$).

[7] It is possible that participants might update their likability evaluation of the other player after they receive information about how they are evaluated by the other person. The likability measure we use in our analysis does not capture any such effects. Since updating after being positively (negatively) surprised would presumably result in an even lower (higher) likability rank for the matching partner, we consider our results a lower bound of (dis)likability effects.

Subsequently, both players made their decision in the public goods game. Moreover, we elicited their beliefs concerning the expected contribution of the other player. Participants could earn an additional 2 Euros, if their estimation was in the range of +/− 1 Euros of the other player's actual contribution. We informed participants about the outcome of the public goods game and the correctness of their belief only at the end of the experiment.

### 2.3 Coordination game

In the third part of the experiment, participants played a classic coordination game, in which they interacted once again with the same person as in the public goods game. The payoff structure of the coordination game is based on the minimum-effort game of van Huyck et al. (1990), applied to a two-player set-up.

In this game, two players simultaneously choose an integer between 1 and 7. A player's payoff (in Euros) is given by

$$\Pi_i^{Coord} = 6 - g_i + 2\, min(g_1, g_2),$$

where $g_i$ represents player $i$'s chosen number for $i = 1, 2$. Table 1 shows the payoff table for the possible combinations of one's own and the other player's chosen number:

Table 1: Payoffs in the coordination game

| Own number | Other player's chosen number | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 7; 7 | 7; 6 | 7; 5 | 7; 4 | 7; 3 | 7; 2 | 7; 1 |
| 2 | 6; 7 | 8; 8 | 8; 7 | 8; 6 | 8; 5 | 8; 4 | 8; 3 |
| 3 | 5; 7 | 7; 8 | 9; 9 | 9; 8 | 9; 7 | 9; 6 | 9; 5 |
| 4 | 4; 7 | 6; 8 | 8; 9 | 10; 10 | 10; 9 | 10; 8 | 10; 7 |
| 5 | 3; 7 | 5; 8 | 7; 9 | 9; 10 | 11; 11 | 11; 10 | 11; 9 |
| 6 | 2; 7 | 4; 8 | 6; 9 | 8; 10 | 10; 11 | 12; 12 | 12; 11 |
| 7 | 1; 7 | 3; 8 | 5; 9 | 7; 10 | 9; 11 | 11; 12 | 13; 13 |

Assuming that both players aim at maximizing their individual payoff, the coordination game has seven strict Nash equilibria, in each of which both players choose the same number. The Nash equilibria are Pareto ranked, with the equilibrium in which both players choose 7 yielding the highest payoff of 13 Euros to both players. However, this equilibrium is not without risks, because if the other player deviates, a player suffers a payoff loss of up to 12 Euros (in case the other player chooses 1). In contrast, the equilibrium $(1, 1)$, in which both players choose their maximin strategy, provides a secure payoff of 7 Euros.

To allow for learning and investigate possible dynamic effects, participants played the coordination game over ten rounds. At the end of the experiment one round was randomly selected for payment. To make sure that the participants understood the rules of the game, they again answered a series of control questions before taking decisions in the game. Moreover, the computer screen displayed once more the photograph of the matching partner together with the information about mutual likability rankings (same information screen as in part 2) before the game started.

All sessions were conducted simultaneously at the FLEX computer lab of Goethe University Frankfurt and the DICE computer lab of the University of Düsseldorf in the summer of 2012. Subjects were students from different disciplines at the two universities and were recruited via ORSEE (Greiner, 2004). In total, 224 subjects participated in the experiment with 128 observations in the same sex matchings and 96 in the mixed sex matchings. All decisions were made on a computer using z-Tree (Fischbacher, 2007). A session lasted approximately 75 minutes, in which subjects earned on average 21 Euros.

### 2.4 Additional independent ratings

To assess the robustness of our likability effects, in particular with regard to potential correlates – like, e.g., attractiveness – we obtained additional, independent ratings for each photograph. In total, 137 students (68 female, 69 male) from a third German University (University of Mainz) participated in these rating sessions. In a post-participation survey, none of these raters indicated to having participated in the original experiment, nor did anyone state to know any of the photographed individuals. Each independent rater was presented with a set of photographs from one session of the original experiment comprising 16 male and 16 female individuals. The order in which photographs were presented was randomized across raters. On average, we obtained 20 independent ratings per photograph (approx. 10 entered by female raters, 10 by male raters). These ratings covered six different variables: likability, attractiveness, cooperativeness, dominance, kindness, and trustworthiness. Just as likability in the original experiment, each variable was evaluated on an 11-point Likert scale. Raters were paid a flat fee of 10 Euros for participation. The rating sessions lasted on average about 50 minutes.

## 3 Results

We present our results as follows. We first show that there are no significant gender differences in the original likability ratings of female and male participants in the main experiment. Next we turn to the effect of (dis)likability in the cooperation and coordination game. In both games we first run separate non-parametric tests for same sex and mixed sex teams. In a second step, we then analyze both matching types in a joint regression analysis.

### 3.1 Likability ratings

Do men and women rate the likability of others differently? Recall that every participant rated four photographs – either from the same sex (same sex matchings) or from the opposite sex (mixed sex matchings) at the beginning of the experiment. Kolmogorow-Smirnov tests confirm that neither the distribution of within-person means of likability ratings, nor the distribution of within-person standard deviations of these ratings differ significantly across gender or matching type ($p \geq 0.238$).[8] We thus conclude that male and female participants assess the likability of others based on photographs similarly.

[8] Unless stated otherwise, all $p$-values are based on two-sided tests.

Further support comes from a comparison of the original likability ratings in the experiment and the additional variables that we elicited in the post-experimental survey with independent raters. Of all elicited variables, the original likability rating is the strongest correlated with the likability rating as assessed by the independent raters. This is true when considering the entire sample and when examining the sub-samples of male and female subjects separately. This suggests that independent and original raters closely agree on what it means to evaluate a particular face with respect to likability – irrespective of the gender of the rated person.[9] We further find that likability is positively correlated with other positive characteristics like kindness, cooperativeness, trustworthiness, and attractiveness. This is true for male and female subjects. Of these, the correlations between the original likability rating and attractiveness are always the lowest. In order to differentiate between the effects of likability and attractiveness, we will run robustness checks in which we control for the latter in a "gender adjusted" way. That is, when we consider the behavior of male (female) subjects, we include the average attractiveness rating of their matching partner, assessed by male (female) independent raters as a control variable in the regressions. As we will discuss in detail below, our results regarding the effects of (dis)likability do not depend on the inclusion of this control variable.

## 3.2 Behavior in the cooperation game

In line with previous research (Balliet et al., 2011) we observe no general difference in male and female cooperative behavior in the public goods game. Overall, men contribute on average 4.05 Euros (s.d. 2.17) to the public good, women contribute 3.92 Euros (s.d. 1.89). The distributions of these contributions are not significantly different from each other (Mann-Whitney test, $p = 0.285$).

### 3.2.1 Same sex teams

When restricting our analysis to same sex teams, we again do not observe a significant overall gender difference. The average contribution of men when being matched with a male partner amounts to 4.11 Euros (s.d. 2.15), the contribution of women in same sex pairs is only slightly and insignificantly smaller: 3.89 Euros on average (s.d. 1.91, Mann-Whitney test, $p = 0.288$). The picture, however, changes once we include the effect of likability. Figure 1 shows the mean contribution of men and women in low and high mutual likability same sex teams (the displayed standard errors are calculated at the individual level). To create these subgroups, we divide the two gender samples according to a median split based on the sum of likability ranks within each team. Recall that the smallest sum of likability ranks is 2, which is the case if both players rank each other on rank 1. The largest sum is 8, which happens if both rank each other on rank 4. The median value of the sum of ranks in both samples is equal to 5. We obtain 30 (34) observations of high (low) mutual likability teams in each gender sample, comprising all teams in which the sum of ranks is below (above or equal to) the median value.

[9] See the corresponding correlations in Table A.1 in Appendix A. There are no significant gender differences in any of the correlations ($p \geq 0.147$). We control for rater-specific effects by standardizing all ratings at the

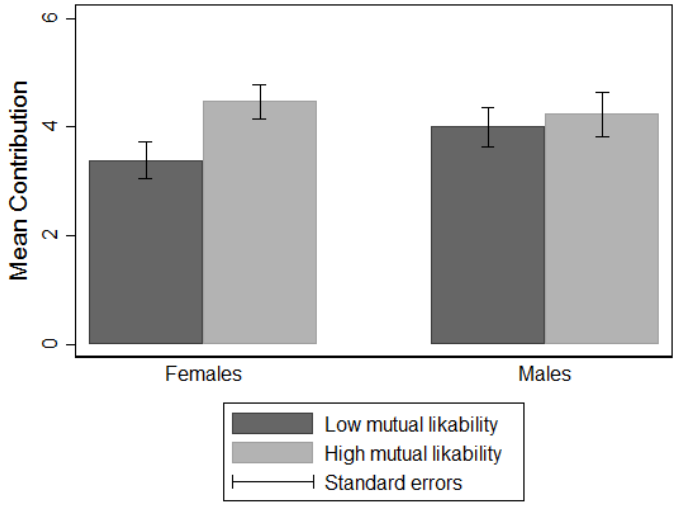



Figure 1: Average contribution in the public goods game (same sex teams)

Table 2: Average contribution and expectation in the public goods game (same sex teams)

| | Contributions | Expectations | Difference | Precision |
|---|---|---|---|---|
| Men | | | | |
| Low mutual likability pairs | 4.00 (2.06) | 3.92 (1.52) | -0.09 (2.64) | 2.03 (1.66) |
| High mutual likability pairs | 4.23 (2.27) | 4.17 (1.49) | -0.07 (2.78) | 2.13 (1.74) |
| Women | | | | |
| Low mutual likability pairs | 3.38 (1.97) | 3.53 (1.46) | 0.15 (2.28) | 1.74 (1.46) |
| High mutual likability pairs | 4.47 (1.70) | 4.47 (0.78) | 0 (2.05) | 1.60 (1.25) |

The table displays mean values and standard deviations that are given in parentheses. Difference is defined as the difference between the expectation about the matching partner's contribution and his/her actual contribution. Precision is defined as the absolute difference between the expectation about the matching partner's contribution and his/her actual contribution.

As Figure 1 (cf. also Table 2) shows, men in low as well as high mutual likability teams contribute similar amounts to the public good (Mann-Whitney test, $p = 0.462$). Even if they do not like each other, men do not seem to care. The effect of likability on women's contributions, on the other hand, is sizable. If mutual likability is low, they contribute about 24% less (Mann-Whitney test, $p = 0.011$).

### 3.2.2 Mixed sex teams

How about cooperation in mixed sex teams? Do gender differences remain, or do men or women (or both) change their behavior? Results show that the likability effect for women remains; but now, men care as well.

First, considering the entire sample of mixed sex teams we find, as before, no general gender difference with regard to cooperation. On average, men contribute 3.98 Euros (s.d. 2.22) to

rater level before averaging across raters, similar to the approach suggested by Darai and Grätz (2013).

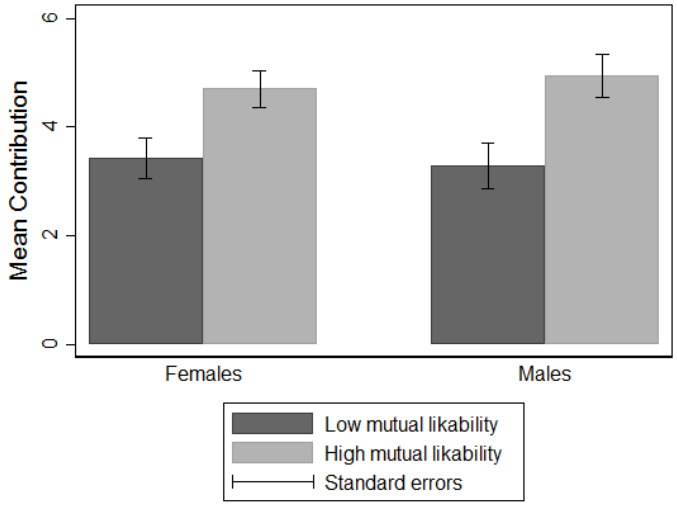



Figure 2: Average contribution in the public goods game (mixed sex teams)

the public good, women contribute 3.96 Euros (s.d. 1.88, Mann-Whitney test, $p = 0.684$). To assess the impact of mutual likability, we again perform a median split on the sum of likability ranks.[10] From the total number of 48 mixed sex teams, 20 teams are in the high, 28 are in the low mutual likability group. Figure 2 and Table 3 display the corresponding contribution decisions of men and women for these subsamples (the displayed standard errors and standard deviations are calculated at the individual level).

Table 3: Average contribution and expectation in the public goods game (mixed sex teams)

| | Contributions | Expectations | Difference | Precision |
|---|---|---|---|---|
| Men | | | | |
| Low mutual likability pairs | 3.29 (2.27) | 3.32 (1.58) | -0.11 (2.57) | 2.04 (1.53) |
| High mutual likability pairs | 4.95 (1.76) | 4.60 (0.68) | -0.10 (1.74) | 1.40 (0.99) |
| Women | | | | |
| Low mutual likability pairs | 3.43 (1.97) | 3.32 (1.79) | 0.34 (2.63) | 2.25 (1.29) |
| High mutual likability pairs | 4.70 (1.49) | 4.35 (1.27) | -0.60 (2.35) | 1.90 (1.45) |

The table displays mean values and standard deviations that are given in parentheses. Difference is defined as the average of the differences between the expectation about the matching partner's contribution and his/her actual contribution. Precision is defined as the absolute difference between the expectation about the matching partner's contribution and his/her actual contribution.

Just like in same sex teams we find a strong likability effect for women in mixed sex teams. Table 3 shows that female participants contribute on average 4.70 Euros to the public good in high mutual likability teams, while they contribute more than 27% less in low mutual likability teams (Mann-Whitney test, $p = 0.022$). However, in contrast to same sex teams, we observe a significant and even slightly stronger (dis)likability effect for men in mixed sex teams as well.

[10]The median sum of ranks is equal to 5. Low and high mutual likability team are formed analogously as in the same sex matching sample.

In low mutual likability teams they contribute almost 34% less than in high mutual likability teams (Mann-Whitney test, $p = 0.005$).

Table 2 and Table 3 provide summary statistics for subjects' expectations about their matching partners' contributions to the public good as well as for the difference between these expectations and actually observed contributions. Considering the last two columns in these tables, we find that the average difference between subjects' expectations and their matching partner's actual contributions is relatively small. This, however, hides the fact that some participants overestimate the other player's contribution while others underestimate it. If instead we take the absolute difference between expectations and actual contributions (what we call "precision"), we see that the average is considerably larger. In general, we do not find any significant effects of gender, type of matching (same sex or mixed sex) or mutual likability considerations on subjects' ability to predict their matching partner's contributions.[11]

### 3.2.3 Regression analysis

We corroborate and extend our above findings in a regression analysis. Regressions come with the advantage of allowing us to control for the matching partner's attractiveness and other visible characteristics such as wearing of glasses or jewelry, hair color, as well as (non-European) ethnicity that might influence both likability ratings and contributions. Moreover, while our main focus is on *mutual* likability that is reflected in the matching pairs' sum of ranks, regressions allow us to disentangle the effects of likability ranks *for* and the likability ranks received *from* the matching partner while controlling for the respective other. In order to compare men's and women's behavior across same sex and mixed sex matchings, we consider three-way interaction models. Table A.2 in the Appendix displays the complete regression results, Table 4 summarizes the main effects for men and women in same sex and mixed sex matchings.

Table 4: Determinants of subjects' contributions

| | Men | | Women | |
|---|---|---|---|---|
| | Same sex | Mixed sex | Same sex | Mixed sex |
| Mutual likability | −0.177 | −0.446 | −0.279 | −0.344 |
| (Sum or ranks) | (0.235) | (0.009) | (0.025) | (0.011) |
| Rank from m. p. | 0.217 | −0.487 | −0.429 | −0.461 |
| | (0.396) | (0.097) | (0.056) | (0.163) |
| Rank for m. p. | −0.571 | −0.397 | −0.1291 | −0.244 |
| | (0.017) | (0.205) | (0.521) | (0.356) |

The table displays main effects of selected variables, $p$-values from corresponding F-tests are given in parentheses. Information taken from the regression analysis presented in columns (1) and (3) in Table A.2 in the Appendix.

The first row of Table 4 confirms our previous findings from non-parametric tests. While mutual likability (i.e., the sum of likability ranks in each team) has a negative effect on cooperation in all types of matchings, its effect is larger and only statistically significant in

[11] Test results not shown, but available upon request.

female same sex teams and in mixed sex matchings (both for males and females). Men's contributions in same sex matchings, on the other hand, are not significantly affected by mutual likabililty.

To investigate deeper what factors drive the average mutual likability effect, we separate the sum of ranks into its two components: the likability rank received from the matched person and the likability rank assigned to the matched person. Note that a higher rank number indicates lower likability here. The respective full regression analysis is presented in column (3) of Table A.2 in the Appendix. The resulting main effects are presented in the lower two rows of Table 4. Comparing men's and women's behavior in same sex matchings, we find that women react significantly more negatively to the likability rank they receive from the matched person than men. For this, compare the corresponding entries in Table 4 and see the significantly negative coefficient of *Rank from m. p.* $\times$ *Female* in Table A.2. Men, on the other hand, contribute significantly less if they assigned a low likability rank to their male matching partner (see the significantly negative main effect of *Rank for m. p.* in Table 4). Men do, however, not significantly react to the likability rank they receive from their male partner. Their reaction to receiving a low likability rank from a female matching partner is comparably stronger (see the significantly negative coefficient of *Rank from m. p.* $\times$ *Mixed sex* in Table A.2). In fact, the corresponding main effect of *Rank from m. p.* for mixed sex matchings in Table 4 is significantly negative for men, but not for women.

Our results do not change if we include additional controls for facial characteristics (compare models (1) and (2) as well as models (3) and (4) of Table A.2 in the Appendix). In particular, if we include the matching partner's attractiveness as assessed by our independent ratings as an additional variable, it neither has a significant effect on contributions ($p \geq 0.608$) nor does it alter the effect of likability. Also, controlling for the absolute likability *rating* instead of the likability rank that is assigned to the matching partner does not change our results (see the robustness checks in model (5) and (6) of Table A.2).

In Table A.3 and Table A.4 in the Appendix we address the question in how far expectations about the matching partner's cooperativeness shape subjects' own contributions. Table A.3 summarizes the main effects; Table A.4 parallels the regressions from Table A.2. The only difference is that, instead of contributions, we now use subjects' expectations about their matching partner's contribution as dependent variable. The results mirror closely our findings from above. This indicates that subjects' expectations depend significantly on mutual likability, suggesting that low contributions in case of low mutual likability are not driven by a desire for retaliation, but rather manifestations of a similarly low expectation regarding the contribution of the other player.

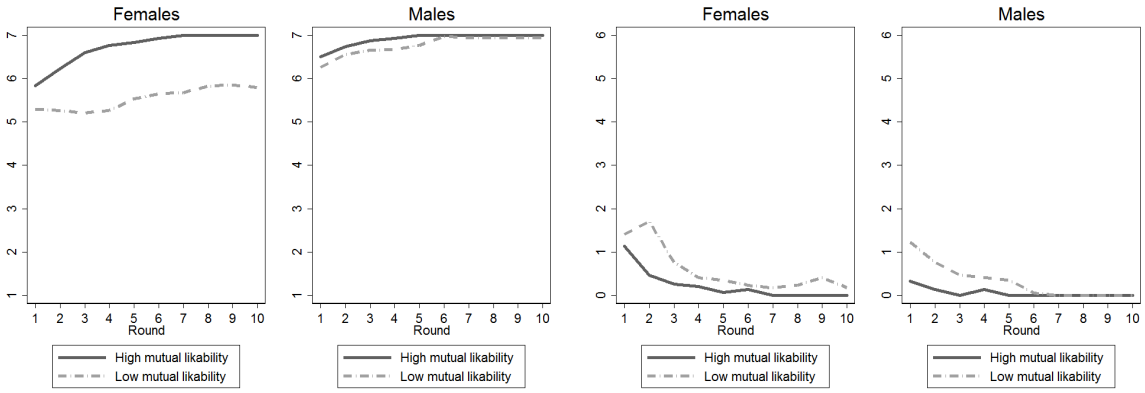
In sum, our results in the cooperation game show that female cooperation is negatively affected by mutual dislikability in both same sex and mixed sex teams. To the contrary, men care about mutual likability only in mixed sex teams. In particular, both men and women react negatively to low likability ranks received from a female matching partner. We next analyze behavior in the coordination game, to see whether these results generalize to a different strategic setting.

## 3.3 Behavior in the coordination game

### 3.3.1 Same sex teams

Recall that in each of the 10 rounds of the coordination game team members choose numbers between 1 and 7. The game has seven strict Nash equilibria in which both players choose the same number. These equilibria are Pareto ranked with payoffs ranging from 7 Euros in equilibrium $(1, 1)$ to 13 Euros in equilibrium $(7, 7)$.

Figure 3 illustrates men's and women's behavior in the coordination game in same sex teams. Panel (a) displays for each round the average of the individually chosen numbers; panel (b) shows the average within-team difference between those numbers. As before, we divide the two gender samples based on a median split of the sum of likability ranks to compare behavior in low and high mutual likability teams (cf. Section 3.2). The broken lines show the average figures in low mutual likability teams, the solid lines analogously display averages in high mutual likability teams.




(a) Chosen numbers over rounds (b) Difference in chosen numbers over rounds

Figure 3: Behavior in the coordination game (same sex teams)

Consider panel (b) first. The graphs reveal that both male and female teams coordinate very quickly in case of high mutual likability. The within-team difference is basically zero from round three (five) onward in male (female) teams. In low mutual likability teams, coordination takes a bit more time, but also here participants coordinate remarkably well – in particular in male teams. While female teams do not always reach full coordination in case of low mutual likability, they do get close: from round 6 onward, at most one out of the 17 female low mutual likability teams chooses numbers that differ by more than 1.

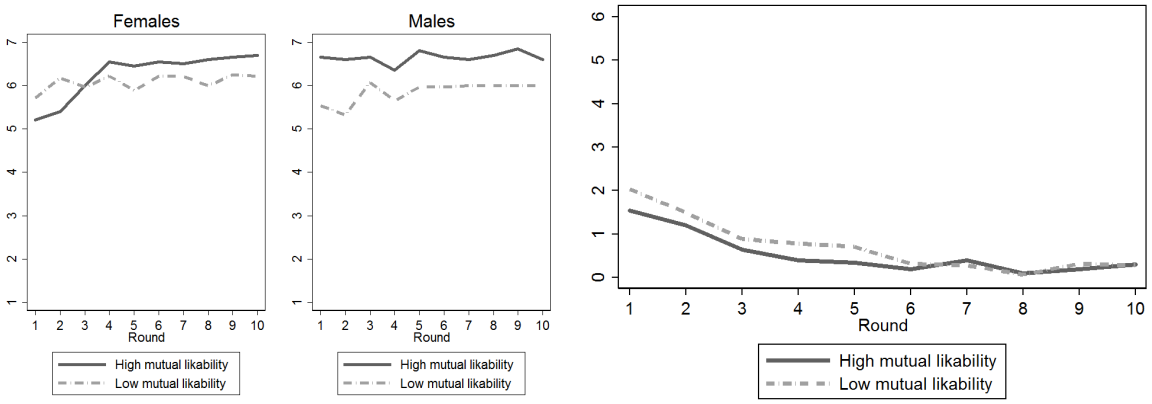
Yet, when looking at panel (a) we see that female and male teams behave very differently with respect to *what* equilibrium they coordinate on. The figure reveals two key findings: First, women choose significantly lower numbers in low mutual likability teams than in high mutual likability teams. While average numbers are above 6 and very quickly converge to 7 if mutual likability is high in female teams, the average numbers basically stay around 5.5 over the ten rounds if mutual likability is low (Mann-Whitney test comparing female high vs. low

mutual likability teams: $p \leq 0.005$ from round 3 onward).[12] Second and in contrary, male participants choose high numbers close to 7 right from the start. More importantly, they do so irrespective of mutual (dis)likability (Mann-Whitney test comparing male high vs. low mutual likability teams: $p \geq 0.181$ in all rounds except round 5).[13] When comparing gender differences conditional on mutual likability, we find that with the exception of rounds 1, 2 and 5 there are no significant differences in chosen numbers between female and male high mutual likability teams.[14] However, in low mutual likability teams these differences are highly significantly (Mann-Whitney test, $p \leq 0.016$ in all rounds).

Thus, also in the coordination game we see that female teams react to mutual likability, while male teams do not, or at least to a much smaller degree. While there is no gender effect on participants' ability to coordinate per se in same sex teams, we find a strong gender effect of likability with regard to equilibrium outcomes. Literally *all* male teams successfully coordinate on the Pareto efficient equilibrium $(7, 7)$ eventually, irrespective of the level of mutual (dis)likability. This does, however, not hold for female teams.

### 3.3.2 Mixed sex matchings

Next, we present the results for coordination in mixed sex teams. Figure 4 depicts the average chosen numbers (panel (a)) and within-team differences (panel (b)) of female and male participants in low (broken lines) and high (solid lines) mutual likability teams.




(a) Chosen numbers over rounds (b) Difference in chosen numbers over rounds

Figure 4: Behavior in the coordination game (mixed sex teams)

Again, consider panel (b) first. As evident, although coordination takes a bit more time and differences between numbers are initially slightly larger compared to same sex teams, almost all mixed sex teams coordinate eventually on equilibrium, i.e., choose the same number, irrespective of the level of mutual likability. Mann-Whitney tests reveal for none of the ten

[12] In round 1 the difference is not significant ($p = 0.236$), in round 2 the difference is marginally significant ($p = 0.0995$).

[13] In round 5: $p = 0.054$.

[14] Mann-Whitney tests, $p = 0.059$ in round 1, $p = 0.024$ in round 2, $p = 0.040$ in round 5 and $p \geq 0.133$ in the remaining rounds.

rounds a significant difference between the high and low mutual likability groups ($p \geq 0.353$).

Panel (a), however, shows that equilibrium outcomes crucially depend on (dis)likability. While high mutual likability teams reach equilibria $(6, 6)$ or $(7, 7)$ relatively quickly, numbers in low mutual likability teams stay between 5 and 6 in most of the rounds. Interestingly, in high mutual likability teams women choose significantly lower numbers than men in the first two rounds (Mann-Whitney test, $p = 0.006$ in round 1, $p = 0.053$ in round 2) but then reach similarly high levels from round 3 onward ($p \geq 0.213$). Presumably, the men's unresponsiveness in the first rounds facilitates coordination on a high-payoff equilibrium later on. This is not the case in low mutual likability teams. Here the chosen numbers of both male and female team members stay at comparably lower levels over all rounds.

It seems remarkable that mixed teams find it more difficult to coordinate on the Pareto efficient Nash equilibrium, given the easiness with which same sex teams are able to solve the problem (cf. Section 3.3.1) – in particular, in case of high mutual likability. In sum, our results from the coordination game clearly show that very much like in the cooperation game, women and men are equally affected by mutual likability concerns in mixed sex teams.

### 3.3.3 Regression analysis

We again corroborate and extend our non-parametric findings in a regression analysis. Model (1) in Table A.5 in the Appendix together with the corresponding F-test results displayed in Table 5 show that the chosen numbers of both men and women in their respective same sex matchings are significantly affected by mutual likability ($p = 0.071$ for men, $p = 0.024$ for women). However, while the effect is small for men, it is large and also significantly stronger for women (see the coefficient of the significantly negative interaction term *Sum of ranks* $\times$ *Female* in Table A.5).

Again, the above results remain unchanged if we include further controls for facial characteristics in model (2) of Table A.5. Also, the partner's attractiveness has no significant effect on chosen numbers in the coordination game ($p \geq 0.201$ in all regressions in Table A.5). When we separate team members' sum of likability ranks, model (3) of Table A.5 reveals that in same sex matchings women react significantly more negatively than men to the likability rank they receive from the other player (consider also the corresponding main effect and F-test result in Table 5, $p = 0.008$). The effect remains robust to the inclusion of control variables for facial characteristics in model (4) and also if we take subjects' full likability ratings into account instead of their likability ranks (see model (5) and (6) in Table A.5).

Taken as a whole, in the coordination game we find a similar behavioral pattern as in the cooperation game: Female pairs' ability to coordinate on the Pareto efficient equilibrium is significantly more affected by mutual likability considerations than male pairs'. In mixed sex matchings, on the other hand, also men react to the mutual likability information leading to sizable efficiency losses.

Table 5: Determinants of subjects' chosen numbers

| | Men | | Women | |
|---|---|---|---|---|
| | Same sex | Mixed sex | Same sex | Mixed sex |
| Mutual likability | −0.058 | −0.273 | −0.275 | −0.083 |
| (Sum of ranks) | (0.071) | (0.033) | (0.024) | (0.388) |
| Rank from m. p. | −0.047 | −0.326 | −0.341 | −0.125 |
| | (0.108) | (0.132) | (0.008) | (0.366) |
| Rank for m. p. | −0.068 | −0.212 | −0.209 | −0.039 |
| | (0.112) | (0.247) | (0.087) | (0.852) |

The table displays main effects of selected variables, $p$-values from corresponding F-tests are given in parentheses. Information taken from the regression analysis presented in columns (1) and (3) in Table A.5 in the Appendix.

# 4 Conclusion

Our lab experiment provides new evidence for gender differences in economic and social interactions. In a two-player one-shot cooperation game and a repeated coordination game, we document that mutual likability significantly shapes players' behavior. If mutual likability is low (i.e., players like each other relatively little), players cooperate less and coordinate on less efficient equilibria than in case mutual likability is high (i.e., players like each other relatively more). Importantly, this result holds only for teams that contain at least one female member. It does, conversely, not hold for all-male teams. In the latter type of teams, men do not seem to care much about likability. They both cooperate and coordinate well independent of the degree of mutual likability in their team. Since our likability ratings are based on a photograph of the other player, we include robustness checks in which we control for attractiveness and further facial characteristics. This leaves our main results unaffected.

We believe that these findings offer a new perspective on gender differences in labor market outcomes. Basically, what our results say is that for women, likability is an asset – or equivalently, dislikability is a hurdle – in *every* of their interactions. For men, on the other hand, likability matters only if they are to interact with the opposite sex. This finding is all the more important, as our results show that the overall effect of (dis)likability is negative, i.e., dislikability looms larger than likability.

Our results document a clear disadvantage for women in terms of average payoffs in the experiment. Aggregated over both experimental games, female participants earn 4.36 percent less than males. In same-sex teams, average earnings are even 7.75 percent lower.[15] Note that these average earnings are based on a balanced matching between male and female participants in a controlled lab experiment in which further influences such as discrimination in wage-setting can be ruled out. Given that in many labor market contexts – and in particular in better-paid jobs and top management positions – interaction with (and between) male colleagues represent the present standard, our results most likely underestimate the impact of likability effects on gender differences in labor market outcomes in the field.

[15]Both differences are statistically significant in Mann-Whitney ranksum tests on the five and one percent level, respectively.

A number of important questions emerge for future research. First, it would be interesting to see how behavior is affected if players can adjust their likability rating after their first interaction in the experimental game. It is conceivable that the effects in our experimental setting, which so far depend only on the first visual impression of the other player and not on his or her actual behavior in the game will become stronger as mutual likability and behavior may reinforce each other. Second, different modes of interaction, e.g. via the phone, Skype or face to face, are worth being explored. Finally, it seems important to investigate what mechanisms can be used to mitigate the negative effects of dislikability or, alternatively, improve the positive effects of likability. Most likely gender effects exist here as well. These additional experiments seem the logical next step to further understand the role of likability in social interaction and its implication for gender differences in labor market outcomes.

## References


Altonji, J. G., Blank, R. M., 1999. Race and gender in the labor market. Handbook of Labor Economics 3, 3143–3259.

Babcock, L., Laschever, S., 2003. Women don't ask: Negotiation and the gender divide. Princeton University Press.

Balliet, D., Li, N. P., Macfarlan, S. J., van Vugt, M., 2011. Sex differences in cooperation: A meta-analytic review of social dilemmas. Psychological Bulletin 137 (6), 881–909.

Barber, B. M., Odean, T., 2001. Boys will be boys: Gender, overconfidence, and common stock investment. The Quarterly Journal of Economic 116 (1), 261–292.

Benenson, J. F., Markovits, H., Fitzgerald, C., Geoffroy, D., Flemming, J., Kahlenberg, S. M., Wrangham, R. W., 2009. Males' greater tolerance of same-sex peers. Psychological Science 20 (2), 184–190.

Bertrand, M., 2011. New perspectives on gender. Handbook of Labor Economics 4, 1543–1590.

Bertrand, M., Goldin, C., Katz, L. F., 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. American Economic Journal: Applied Economics 2 (3), 228–255.

Bertrand, M., Hallock, K. F., 2001. The gender gap in top corporate jobs. ILR Review 55 (1), 3–21.

Buser, T., Niederle, M., Oosterbeek, H., 2014. Gender, competitiveness, and career choices. The Quarterly Journal of Economics 1409, 1447.

Cassidy, Hugh, D. J., Kauhanen, A., 2016. Promotion signaling, gender, and turnover: New theory and evidence. Journal of Economic Behavior & Organization 126, 140–166.

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., Seabright, P., 2015. Honest signaling in trust interactions: smiles rated as genuine induce trust and signal higher earning opportunities. Evolution and Human Behavior 36 (1), 8–16.

Croson, R., Gneezy, U., 2009. Gender differences in preferences. Journal of Economic Literature 47 (2), 448–474.

Darai, D., Grätz, S., 2013. Attraction and cooperative behavior. Working Paper, University of Zürich.

Dohmen, T., Falk, A., 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. The American Economic Review 101 (2), 556–590.

Eckel, C. C., Grossman, P. J., 2008a. Differences in the economic decisions of men and women: experimental evidence. Handbook of Experimental Economics Results, 509–519.

Eckel, C. C., Grossman, P. J., 2008b. Men, women and risk aversion: Experimental evidence. Handbook of Experimental Economics Results, 1061–1073.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10 (2), 171–178.

Frederiksen, A., Kato, T., forthcoming. Human Capital and Career Success: Evidence from Linked Employer-Employee Data. Economic Journal.

Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: Gender differences. The Quarterly Journal of Economics 118 (3), 1049–1074.

Goldin, C., 2014. A grand gender convergence: Its last chapter. The American Economic Review 104 (4), 1091–1119.

Greiner, B., 2004. The online recruitment system ORSEE - A guide for the organization of experiments in economics. Papers on Strategic Interaction 2003-10, Max Planck Institute of Economics, Strategic Interaction Group.

Kleven, H. J., Landais, C., Sogaard, J. E., 2015. Children and gender inequality: Evidence from denmark. Unpublished manuscript, London School Econ.

Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? do men compete too much? The Quarterly Journal of Economics 122 (3), 1067–1101.

Stuhlmacher, A. F., Walters, A. E., 1999. Gender differences in negotiation outcome: A meta-analysis. Personnel Psychology 52 (3), 653–677.

# Appendix

# A Additional tables

Table A.1: Correlation matrix of standardized facial appearance variables assessed by subjects in the original experiment and independent raters

| | Original likability rating | | | Test for equality of males' and females' correlation coefficients |
|---|---|---|---|---|
| | All 224 subjects | 112 female subjects | 112 male subjects | |
| Likability | $0.629^{***}$ | $0.663^{***}$ | $0.618^{***}$ | $p = 0.570$ |
| Attractiveness | $0.400^{***}$ | $0.395^{***}$ | $0.491^{***}$ | $p = 0.374$ |
| Cooperativeness | $0.559^{***}$ | $0.568^{***}$ | $0.574^{***}$ | $p = 0.949$ |
| Dominance | $-0.165^{**}$ | $-0.255^{***}$ | -0.064 | $p = 0.147$ |
| Kindness | $0.604^{***}$ | $0.653^{***}$ | $0.576^{***}$ | $p = 0.361$ |
| Trustworthiness | $0.530^{***}$ | $0.604^{***}$ | $0.507^{***}$ | $p = 0.303$ |

** p<0.05 *** p<0.01. The original likability rating is calculated as the mean of the 4 standardized ratings that were given during the original experiment. The remaining variables are calculated as the mean standardized ratings given by the on average 20 independent raters per photograph.

Table A.2: Determinants of subjects' contributions

| | Dep. var.: Own contribution | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Sum of ranks | −0.177 | −0.171 | | | | |
| | (0.148) | (0.148) | | | | |
| Sum of ranks × Mixed sex | −0.269 | −0.268 | | | | |
| | (0.223) | (0.231) | | | | |
| Sum of ranks × Female | −0.102 | −0.098 | | | | |
| | (0.193) | (0.196) | | | | |
| Sum of ranks × Female × Mixed sex | 0.203 | 0.217 | | | | |
| | (0.290) | (0.296) | | | | |
| Rank from m.p. | | | 0.217 | 0.209 | 0.221 | 0.220 |
| | | | (0.255) | (0.258) | (0.256) | (0.255) |
| Rank from m. p. × Mixed sex | | | −0.704* | −0.685* | −0.751* | −0.723* |
| | | | (0.387) | (0.406) | (0.404) | (0.418) |
| Rank from m. p. × Female | | | −0.646* | −0.635* | −0.625* | −0.610* |
| | | | (0.338) | (0.341) | (0.339) | (0.346) |
| Rank from m. p. × Female × Mixed sex | | | 0.672 | 0.663 | 0.754 | 0.722 |
| | | | (0.525) | (0.535) | (0.528) | (0.537) |
| Rank for m.p. | | | −0.571** | −0.550** | | |
| | | | (0.237) | (0.235) | | |
| Rank for m. p. × Mixed sex | | | 0.174 | 0.153 | | |
| | | | (0.391) | (0.413) | | |
| Rank for m. p. × Female | | | 0.442 | 0.433 | | |
| | | | (0.310) | (0.322) | | |
| Rank for m. p. × Female × Mixed sex | | | −0.289 | −0.253 | | |
| | | | (0.481) | (0.521) | | |
| Rating for m.p. | | | | | 0.897*** | 0.897*** |
| | | | | | (0.296) | (0.290) |
| Rating for m. p. × Mixed sex | | | | | −0.597 | −0.567 |
| | | | | | (0.529) | (0.554) |
| Rating for m. p. × Female | | | | | −0.792* | −0.802* |
| | | | | | (0.400) | (0.410) |
| Rating for m. p. × Female × Mixed sex | | | | | 1.062* | 1.020 |
| | | | | | (0.638) | (0.673) |
| Female | 0.290 | 0.249 | 0.290 | 0.277 | 1.369 | 1.333 |
| | (0.967) | (1.031) | (0.976) | (1.041) | (0.890) | (0.939) |
| Mixed sex | 1.301 | 1.245 | 1.285 | 1.259 | 1.805* | 1.675 |
| | (1.179) | (1.277) | (1.196) | (1.294) | (1.026) | (1.143) |
| Female × Mixed sex | −0.829 | −0.816 | −0.775 | −0.794 | −1.752 | −1.586 |
| | (1.474) | (1.619) | (1.516) | (1.653) | (1.320) | (1.466) |
| Constant | 4.962*** | 4.809*** | 4.962*** | 4.800*** | 3.545*** | 3.515*** |
| | (0.792) | (1.086) | (0.800) | (1.109) | (0.682) | (0.997) |
| Controls | *No* | *Yes* | *No* | *Yes* | *No* | *Yes* |
| Adjusted $R^2$ | 0.066 | 0.047 | 0.070 | 0.050 | 0.078 | 0.060 |
| Observations | 224 | 224 | 224 | 224 | 221 | 221 |

OLS regressions with standard errors that are clustered at the matching pair level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Likability ratings are standardized at the rater level. Controls include attractiveness of the matched player as well as dummies indicating whether the matched player wears glasses or jewelry, has dark hair, or a European appearance.

Table A.3: Determinants of subjects' expected contributions

| | Men | | Women | |
|---|---|---|---|---|
| | Same sex | Mixed sex | Same sex | Mixed sex |
| Sum of ranks | −0.125 | −0.341 | −0.2625 | −0.287 |
| | (0.213) | (0.000) | (0.000) | (0.011) |
| Rank from m. p. | 0.058 | −0.371 | −0.375 | −0.415 |
| | (0.705) | (0.041) | (0.015) | (0.165) |
| Rank for m. p. | −0.307 | −0.306 | −0.150 | −0.178 |
| | (0.060) | (0.131) | (0.329) | (0.461) |

The table displays main effects of selected variables, $p$-values from corresponding F-tests are given in parentheses. Information taken from the regression analysis presented in columns (1) and (3) in Table A.4 in the Appendix.

Table A.4: Determinants of subjects' expected contributions

| | Dep. var.: Expected contribution | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Sum of ranks | −0.125 | −0.114 | | | | |
| | (0.099) | (0.094) | | | | |
| Sum of ranks × Mixed sex | −0.216 | −0.216 | | | | |
| | (0.136) | (0.135) | | | | |
| Sum of ranks × Female | −0.138 | −0.144 | | | | |
| | (0.117) | (0.121) | | | | |
| Sum of ranks × Female × Mixed sex | 0.192 | 0.179 | | | | |
| | (0.189) | (0.192) | | | | |
| Rank from m.p. | | | 0.058 | 0.066 | 0.008 | 0.026 |
| | | | (0.153) | (0.153) | (0.159) | (0.155) |
| Rank from m. p. × Mixed sex | | | −0.429* | −0.429* | −0.387 | −0.386 |
| | | | (0.236) | (0.242) | (0.253) | (0.254) |
| Rank from m. p. × Female | | | −0.433** | −0.458** | −0.389* | −0.417* |
| | | | (0.215) | (0.212) | (0.229) | (0.225) |
| Rank from m. p. × Female × Mixed sex | | | 0.389 | 0.371 | 0.402 | 0.373 |
| | | | (0.410) | (0.417) | (0.407) | (0.410) |
| Rank for m.p. | | | −0.307* | −0.293* | | |
| | | | (0.162) | (0.157) | | |
| Rank for m. p. × Mixed sex | | | 0.001 | 0.005 | | |
| | | | (0.258) | (0.257) | | |
| Rank for m. p. × Female | | | 0.157 | 0.170 | | |
| | | | (0.223) | (0.222) | | |
| Rank for m. p. × Female × Mixed sex | | | −0.029 | −0.047 | | |
| | | | (0.387) | (0.394) | | |
| Rating for m.p. | | | | | 0.280 | 0.301 |
| | | | | | (0.203) | (0.200) |
| Rating for m. p. × Mixed sex | | | | | 0.026 | 0.017 |
| | | | | | (0.330) | (0.322) |
| Rating for m. p. × Female | | | | | −0.090 | −0.152 |
| | | | | | (0.306) | (0.300) |
| Rating for m. p. × Female × Mixed sex | | | | | 0.213 | 0.251 |
| | | | | | (0.502) | (0.494) |
| Female | 0.618 | 0.754 | 0.618 | 0.755 | 0.885 | 1.074* |
| | (0.533) | (0.602) | (0.538) | (0.611) | (0.575) | (0.582) |
| Mixed sex | 0.971 | 1.054* | 0.960 | 1.039 | 0.857 | 0.930 |
| | (0.618) | (0.630) | (0.626) | (0.639) | (0.616) | (0.642) |
| Female × Mixed sex | −0.997 | −1.107 | −0.944 | −1.033 | −1.065 | −1.172 |
| | (0.893) | (0.942) | (0.919) | (0.966) | (0.991) | (1.009) |
| Constant | 4.631*** | 4.895*** | 4.631*** | 4.927*** | 3.996*** | 4.342*** |
| | (0.470) | (0.704) | (0.474) | (0.702) | (0.407) | (0.646) |
| Controls | *No* | *Yes* | *No* | *Yes* | *No* | *Yes* |
| Adjusted $R^2$ | 0.101 | 0.092 | 0.096 | 0.089 | 0.093 | 0.088 |
| Observations | 224 | 224 | 224 | 224 | 221 | 221 |

OLS regressions with standard errors that are clustered at the matching pair level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Likability ratings are standardized at the rater level. Controls include attractiveness of the matched player as well as dummies indicating whether the matched player wears glasses or jewelry, has dark hair, or a European appearance.

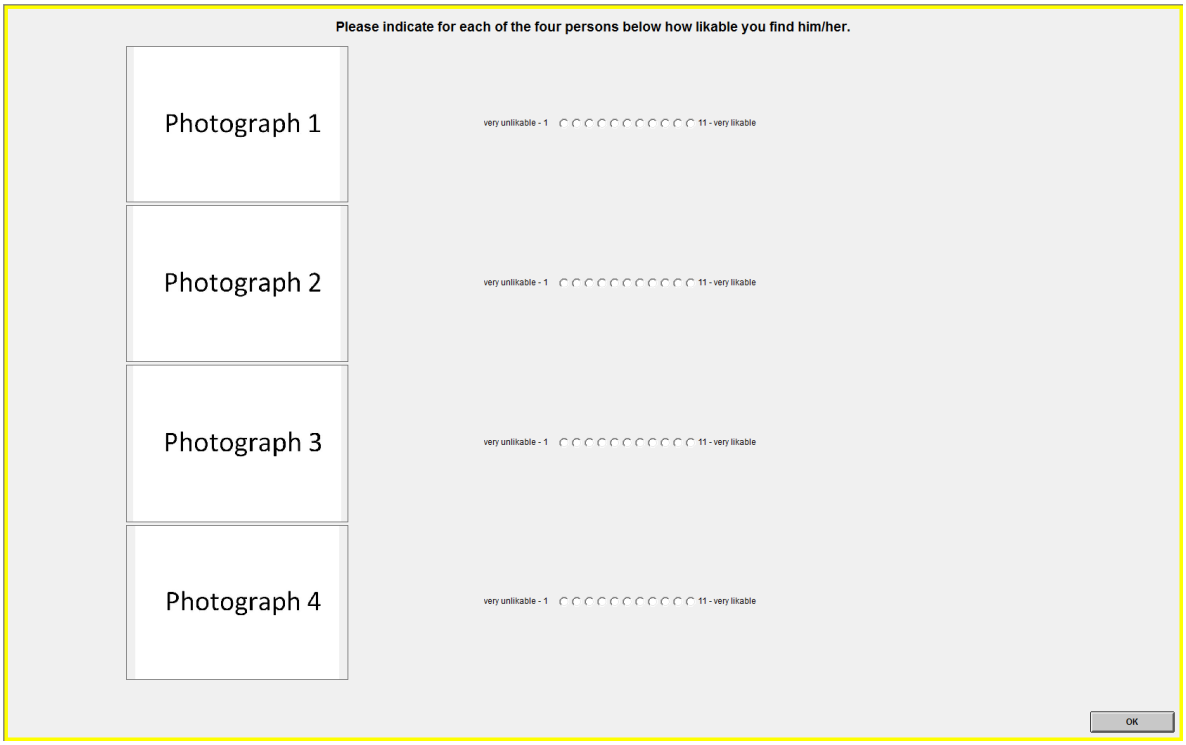
Table A.5: Determinants of subjects' chosen numbers

| | Dep. var.: Chosen number | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Sum of ranks | −0.058* | −0.080** | | | | |
| | (0.032) | (0.034) | | | | |
| Sum of ranks × Mixed sex | −0.216 | −0.219 | | | | |
| | (0.131) | (0.134) | | | | |
| Sum of ranks × Female | −0.218* | −0.199* | | | | |
| | (0.124) | (0.120) | | | | |
| Sum of ranks × Female × Mixed sex | 0.407*** | 0.415*** | | | | |
| | (0.144) | (0.151) | | | | |
| Rank from m.p. | | | −0.047 | −0.065 | −0.069* | −0.077* |
| | | | (0.029) | (0.040) | (0.040) | (0.046) |
| Rank from m. p. × Mixed sex | | | −0.279 | −0.259 | −0.270 | −0.244 |
| | | | (0.217) | (0.213) | (0.234) | (0.226) |
| Rank from m. p. × Female | | | −0.294** | −0.255** | −0.185 | −0.147 |
| | | | (0.129) | (0.122) | (0.124) | (0.124) |
| Rank from m. p. × Female × Mixed sex | | | 0.485 | 0.459 | 0.326 | 0.296 |
| | | | (0.329) | (0.322) | (0.294) | (0.291) |
| Rank for m.p. | | | −0.068 | −0.096* | | |
| | | | (0.043) | (0.049) | | |
| Rank for m. p. × Mixed sex | | | −0.143 | −0.174 | | |
| | | | (0.187) | (0.189) | | |
| Rank for m. p. × Female | | | −0.141 | −0.141 | | |
| | | | (0.128) | (0.136) | | |
| Rank for m. p. × Female × Mixed sex | | | 0.313 | 0.358 | | |
| | | | (0.341) | (0.346) | | |
| Rating for m.p. | | | | | 0.020 | 0.065 |
| | | | | | (0.039) | (0.061) |
| Rating for m. p. × Mixed sex | | | | | 0.173 | 0.192 |
| | | | | | (0.243) | (0.243) |
| Rating for m. p. × Female | | | | | 0.303 | 0.298 |
| | | | | | (0.205) | (0.218) |
| Rating for m. p. × Female × Mixed sex | | | | | −0.622 | −0.666 |
| | | | | | (0.441) | (0.460) |
| Female | 0.325 | 0.079 | 0.325 | 0.074 | −0.247 | −0.512 |
| | (0.405) | (0.444) | (0.405) | (0.446) | (0.230) | (0.310) |
| Mixed sex | 0.475 | 0.319 | 0.455 | 0.303 | 0.076 | −0.191 |
| | (0.520) | (0.551) | (0.512) | (0.546) | (0.496) | (0.485) |
| Female × Mixed sex | −1.321** | −1.049 | −1.284** | −1.015 | −0.157 | 0.294 |
| | (0.524) | (0.649) | (0.517) | (0.652) | (0.713) | (0.748) |
| Constant | 6.642*** | 6.527*** | 6.642*** | 6.545*** | 6.521*** | 6.281*** |
| | (0.155) | (0.414) | (0.155) | (0.423) | (0.125) | (0.385) |
| Controls | *No* | *Yes* | *No* | *Yes* | *No* | *Yes* |
| Round dummies | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* | *Yes* |
| Adjusted $R^2$ | 0.125 | 0.143 | 0.124 | 0.142 | 0.115 | 0.130 |
| Number of clusters | 112 | 112 | 112 | 112 | 112 | 112 |
| Observations | 2240 | 2240 | 2240 | 2240 | 2210 | 2210 |

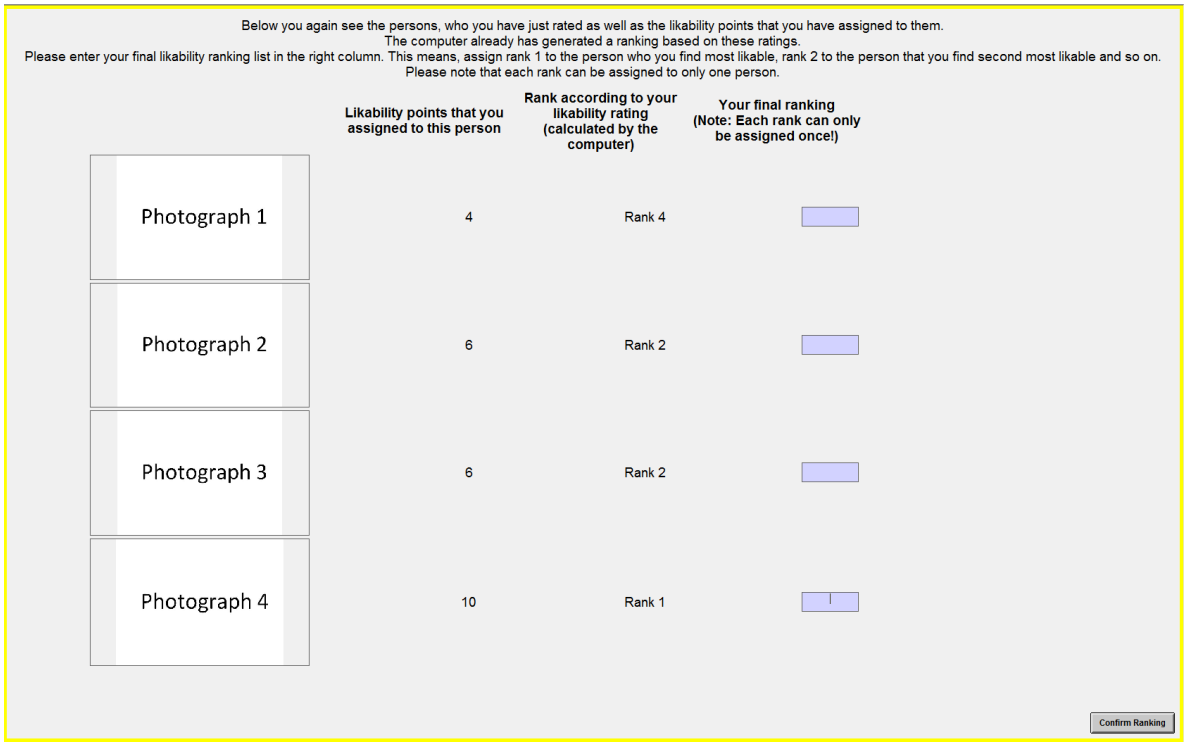
Pooled OLS regressions with standard errors that are clustered at the matching pair level and given in parentheses: * p<0.10, ** p<0.05, *** p<0.01. Controls include attractiveness of the matched player as well as dummies indicating whether the matched player wears glasses or jewelry, has dark hair, or a European appearance.

# B Translated zTree screenshots

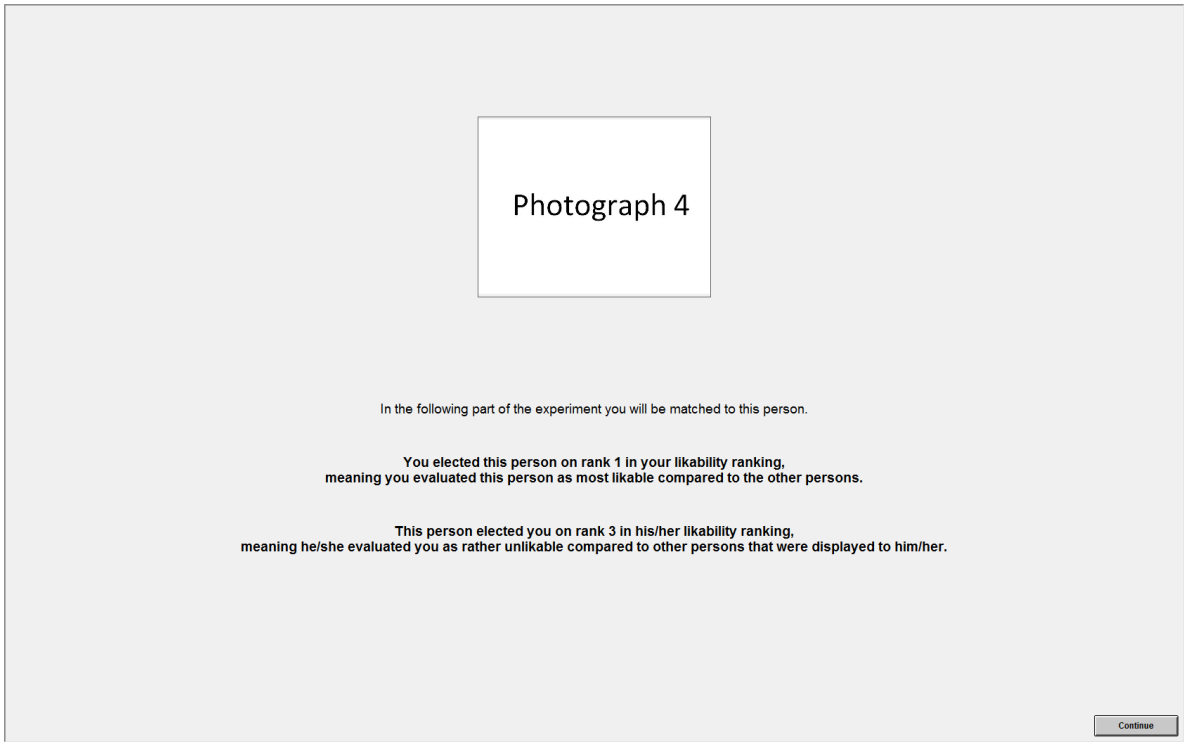
## Screen 1



## Screen 2


Below you again see the persons, who you have just rated as well as the likability points that you have assigned to them.
The computer already has generated a ranking based on these ratings.
Please enter your final likability ranking list in the right column. This means, assign rank 1 to the person who you find most likable, rank 2 to the person that you find second most likable and so on.
Please note that each rank can be assigned to only one person.
Likability points that you assigned to this person
Rank according to your likability rating (calculated by the computer)
Your final ranking (Note: Each rank can only be assigned once!)
Photograph 1
4
Rank 4
Photograph 2
6
Rank 2
Photograph 3
6
Rank 2
Photograph 4
10
Rank 1
Confirm Ranking

## Screen 3


Photograph 4
In the following part of the experiment you will be matched to this person.
You elected this person on rank 1 in your likability ranking,
meaning you evaluated this person as most likable compared to the other persons.
This person elected you on rank 3 in his/her likability ranking,
meaning he/she evaluated you as rather unlikable compared to other persons that were displayed to him/her.
Continue