THE ECONOMIC JOURNAL

The Economic Journal, **130** (*April*), 716–739 DOI: 10.1093/ej/uez067
Advance Access Publication Date: 28 January 2020

I (DON'T) LIKE YOU! BUT WHO CARES? GENDER DIFFERENCES IN SAME-SEX AND MIXED-SEX TEAMS*

Leonie Gerhards and Michael Kosfeld

We study the effect of likeability on women's and men's team behaviour 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 likeability, we find that female teams lower their contribution to the public good in the event of low likeability, while male teams achieve high levels of co-operation irrespective of the level of mutual likeability. In mixed-sex teams, both women's and men's contributions depend on mutual likeability. Similar results are found in the minimum effort game. Our results offer a new perspective on gender differences in labour market outcomes: mutual dislikeability impedes team behaviour, except in all-male teams.

Women earn lower wages then men (Altonji and Blank, 1999; Bertrand *et al.*, 2010; Goldin, 2014; Blau and Kahn, 2017), are less likely to be promoted to top management positions (Cassidy and Kauhanen, 2016; Frederiksen and Kato, 2017) and are underrepresented in higher levels of the corporate hierarchy (Bertrand and Hallock, 2001; Kleven *et al.*, 2018). A growing empirical literature has identified important gender differences in economic behaviour and decision making, helping us to understand and explain at least part of these findings. Seminal papers document, inter alia, gender differences in overconfidence (Barber and Odean, 2001), competitiveness (Gneezy *et al.*, 2003; Niederle and Vesterlund, 2007), risk attitudes (Dohmen and Falk, 2011) and negotiation style (Babcock and Laschever, 2003).[1]

In this article, we identify a new behavioural finding that, to the best of our knowledge, has not been analysed or documented so far, offering a novel perspective on the determinants of gender differences in labour market outcomes: women significantly react to (mutual) likeability in social interactions. In contrast, men react to likeability only when they interact with women; if men interact with men, they don't care. In other words, likeability is not an influencing factor in all-male interactions, while in all others it is. As overall our results suggest that dislikeability hurts more than likeability helps, the observed gender difference in social interactions has important *negative* implications on gender differences in economic outcomes: women significantly suffer from the variation in likeability and achieve overall worse outcomes than men.

* Corresponding author: Leonie Gerhards, Department of Economics, Universität Hamburg, Von-Melle-Park 5, D-20146 Hamburg, Germany. Email: leonie.gerhards@uni-hamburg.de

This paper was received on 29 March 2018 and accepted on 26 October 2019. The Editor was Rachel Kranton.

We are grateful for valuable comments from the Editor as well as two anonymous referees. 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 at the European Workshop on Experimental and Behavioral Economics 2013, the Nordic Conference on Behavioral and Experimental Economics 2013 and the Workshop on Gender Heterogeneity in Political Economy 2018 for their helpful comments. We particularly thank Thomas Buser and participants at the Grüneburg Seminar in Frankfurt for their useful suggestions. We are grateful for 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] See also the 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.

Our results come from a novel laboratory experiment that allows us to control important aspects of the economic situation to a degree that is unfeasible in the field. We thereby secure a high degree of internal validity that enables us to disentangle, for example, likeability from attractiveness. Several studies have shown that attractive people on average achieve better economic outcomes (Hamermesh and Biddle, 1994; Mobius and Rosenblat, 2006; Ravina, 2008) and that attractiveness significantly influences decision making in economic contexts (Mulford *et al.*, 1998; Takahashi *et al.*, 2006; Wilson and Eckel, 2006; Andreoni and Petrie, 2008). It therefore seems crucial to control for attractiveness when identifying the effect of likeability on individual and social behaviour. Our results show, however, that attractiveness has only little effect in our context and leaves the effects of likeability on behaviour unaltered.

In our experiment, participants see photographs of the faces of other participants, whom they are asked to rate in terms of likeability. Subsequently, participants interact in two two-player games with the same partner: a one-shot co-operation game and a multi-period co-ordination game. Before participants play the two games, they are informed about the 'identity' of their partner by showing them a photograph of the participant with whom they are matched. We remind them of their likeability judgment about the other participant and also inform them of the other participant's judgment about them. We thereby carefully control the information that each participant holds with respect to the level of mutual likeability in his or her team (see Section 1 for discussion and further details). Participants do not receive any other information about each other. Furthermore, they take part in the experiment in two different labs (about 250 km apart from each other) such that encounters after the sessions are impossible.

Prior lab studies have shown that facial likeability enhances co-operation (Bell *et al.*, 2012; Mieth *et al.*, 2016).[2] However, these studies do not consider mutual likeability and also do not look at gender effects, which is what we are interested in. Casciaro and Sousa Lobo (2008) document that likeability plays a significant role in workplace interactions outside the lab. They find that employees are more likely to seek advice from competent co-workers only if they like them. In case of dislike, competence as a criterion for the choice of work partners is significantly reduced. Again, no gender effects are analysed in this study.

So why should we expect gender differences to play a role in our context in the first place? There are at least three reasons why we believe that gender matters. First, research in economics and social psychology has shown that women are generally more sensitive to details of the social interaction than men. See, e.g., Croson and Gneezy (2009) for a survey on gender differences in economic preferences arguing in the context of co-operation that '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] To the extent that (mutual) likeability is a relevant aspect of social interaction, we therefore expect that women react more strongly to variation in likeability. Whether this effect also depends on the sex of the interaction partner, however, is an open question.

Second, women form different networks and peer relationships than men. Evidence shows that women, on average, have fewer direct connections and their connections are more likely to be linked to each other. Thus, women's networks are smaller and more clustered, while men's

[2] In addition, Bell *et al.* (2012) show that likeable-looking cheaters are better remembered than likeable-looking co-operators, while Mieth *et al.* (2016) find that likeable-looking cheaters are punished more severely than unlikeable-looking cheaters. There also exists a larger literature on facial cues and co-operation in laboratory experiments. See, e.g., DeBruine (2002); Stirrat and Perrett (2010); Giang *et al.* (2012); Declerck *et al.* (2013).

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

networks are larger and looser. See, e.g., a recent study by Ductor *et al.* (2018) who show in co-authorship networks in economics and sociology that female researchers have fewer collaborators, collaborate more often with the same co-authors, and a higher fraction of their co-authors are co-authors of each other. Lindenlaub and Prummer (2018) provide similar evidence for co-authorship networks in computer science and also build a model that links network differences to differences in performance. Complementary and consistent findings come from phone data showing that women, on average, make fewer calls that last longer (Friebel and Seabright, 2011). As it seems plausible that likeability influences social networking behaviour, these results are consistent with the hypothesis that gender differences in network formation may at least partly be driven by women being more responsive to variation in likeability than men.[4]

Lastly, Benenson *et al.* (2009) provide direct evidence for our gender hypothesis. They show in a questionnaire study with college roommates that women have lower thresholds of tolerance for same-sex roommates than men. Tolerance is defined as the acceptance of stresses and strains within a personal relationship. Results 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. Benenson (2019) provides a recent discussion of these and related findings offering evolutionary arguments to explain the observed gender difference in peer relationships. See also Benenson and Markovits (2014).

In sum, while there exist various sources of evidence suggesting that likeability may matter more in social interactions for women than men and that the sex of the interaction partner may also play a role, no study has yet shown whether this is indeed true and—most importantly—whether these potential gender differences in behaviour also result in gender differences in economic outcomes. Our study is aimed at filling this gap. Being able to identify thoroughly the effect of likeability on individual and team behaviour, our results show that likeability has a significant effect for female participants but this does not hold in the same way for male participants. More precisely, when mutual likeability in same-sex teams is low, women reduce their co-operation significantly and co-ordinate on Pareto-inferior equilibria. Men, in contrast, achieve high levels of co-operation and co-ordination outcomes independent of mutual likeability in their team. In mixed-sex teams, on the other hand, female and male participants respond significantly to mutual likeability, i.e., both co-operate less and co-ordinate on less efficient Nash equilibria if likeability in their team is low. In sum, the results document a clear and consistent finding: variation in likeability matters for behaviour and economic outcomes in teams, except in all-male teams.

Our results offer a new perspective on gender differences in labour market outcomes by highlighting a behavioural asymmetry in the way male and female performance is influenced, creating significant advantages for men. In many organisational contexts, performance is, in large part, determined by successful co-operation and co-ordination with others, be it with individual colleagues or clients or in teams and departments (Simon, 1947; Gibbons and Roberts, 2013). The same holds for career and promotion opportunities. For one thing, promotions often depend on revealed or expected performance; secondly, they may result from direct co-operation with others as well. Our results suggest that whenever two men interact with each other, mutual likeability is not expected to matter for performance outcomes. As soon as one of them (or both) is a woman, however, the situation changes. Then, likeability considerations become relevant, turning low

[4] Further, results in Ductor *et al.* (2018) show that co-authorship networks in economics exhibit a strong degree of gender homophily and that this effect is larger for men. Intriguingly, this is exactly what our results would predict. Cf. Section 3.

likeability into a disruptive factor—in a sense an exogenous 'hurdle'—that impedes successful co-operation and reduces performance outcomes. Women *always* face this potential hurdle, men don't.

Firms and organisations can mitigate this problem by implementing work cultures that highlight performance but do not see likeability as a prerequisite for successful co-operation among employees. Also, the mere recognition of likeability as an important factor may already create awareness of an otherwise hidden dimension of gender discrimination and inequality—similar to other well-documented differences, e.g., in risk taking, confidence or negotiation behaviour (cf. references above). Together, this may influence, as our results suggest, the career perspectives and the more general economic outcomes of women in firms and organisations positively.

The remainder of the article is organised as follows. In the next section, we present and discuss details of our experimental design. Section 2 contains the results. Finally, Section 3 concludes.

1. Experimental Design

We are interested in the effect of likeability on male and female behaviour conditional on the sex of the interaction partner. Therefore, participants in our experiment are randomly matched with another participant either from the same sex 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 km apart (Frankfurt/Main and Düsseldorf). All teams consisted of one participant from each location. Before describing the experimental games, we provide more information about the matching protocol.

1.1. *Matching Protocol*

In each session, eight female and eight male participants always participated at both locations. Upon arrival at the experimental site, we took standardised 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 our experiment we created a virtual lab of 32 participants by connecting 16 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. A full set of the instructions is provided in the Online Appendix. In the first part of the experiment, participants were randomly divided into four matching groups of eight participants each. Each matching group comprised four participants from either location. While participants 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 same-sex matchings, these participants were all of the same sex as the participant him- or herself. In mixed-sex matchings, all were of the opposite sex. The instructions explained 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 participants.

On the decision screen, participants were asked to rate, based on the photograph they saw, how likeable (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 unlikeable' to 'very likeable'.[5] Based on these ratings, the computer generated an ordinal ranking that was shown to the participant on the subsequent screen. In particular, rank 1 was assigned to the photograph that the participant granted the highest likeability rating, rank 2 to the photograph that he or she granted the second-highest likeability rating, and so forth. In the event of a tie, multiple photographs were assigned the same rank. Then, participants were asked to enter and confirm their final likeability ranking from 1 to 4 where ties were ruled out.

We used the ordinal *rankings* to inform participants about their mutual likeability in later stages of the experiment (see below). The main advantage compared with absolute *ratings* is that ordinal rankings are more easily interpretable and less susceptible to taste differences in the individual assessment of likeability. However, in our robustness checks we will include participants' absolute ratings as well.

To generate sufficient variation with regard to mutual (dis)likeability within teams, we matched participants based on the likeability rankings they provided us with. More precisely, our matching algorithm, first, calculated the 16 sums of mutually assigned likeability 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, participants 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: participants with the lowest sum of ranks were matched first and the remaining participants thereafter. This procedure allowed us to make maximal use of the given rankings to obtain sufficient variation with regard to mutual likeability within teams. Note that we did not provide participants with details about our matching procedure.[6]

1.2. *Co-operation Game and Information about the Matching Partner*

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

In the 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. Thus, the pay-off function for player i is given by

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

where c_i represents player i's individual contribution with $i = 1, 2$. Under the assumption that both players maximise their own individual pay-off, the game has a unique Nash equilibrium, in which every player contributes zero. However, players maximise their joint pay-off by contributing 6 euros each, yielding a pay-off of 9 euros to every player.

[5] See sample screen shots in the Online Appendix.

[6] Importantly, we neither find a statistically significant difference in the distributions of mutually assigned likeability rankings between male and female nor same-sex and mixed-sex teams (Fisher exact tests, $p = 0.446$ and $p = 0.557$, respectively).

After participants had read the instructions to the game 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 the Online Appendix). Moreover, a short text below the photograph informed both players of their mutual likeability ranks. 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 likeability ranking, meaning you evaluated this person as most likeable [rather likeable, rather unlikeable, or most unlikeable] compared to the other persons. This person elected you on rank 1 [2, 3, or 4] in his/her likeability ranking, meaning he/she evaluated you as most likeable [rather likeable, rather unlikeable, or most unlikeable] compared to other persons that were displayed to him/her.'

We chose this procedure to ensure that all participants in the experiment have exactly the same level of information regarding mutual likeability in their team. This allows us to identify the effect of *actual likeability* on behaviour and separate it from the effect of *prior beliefs* about likeability, which are likely to differ between individuals and may also depend on gender. From a methodological point of view, our design is thus not based on an attempt to make our experiment as similar as possible to social interactions outside the lab. Rather, it is chosen to maximise the internal validity of our results, i.e., identify effects thoroughly by controlling all variables of interest, which we see as a key advantage of running lab experiments compared with gathering data in the field. Nevertheless, we are convinced that the situation participants face in our experiment captures important elements of relevant situations outside the experiment. First, it does happen also outside the lab that we sometimes receive explicit information about how much we are liked by somebody else. Second, even if we do not know explicitly how much we are liked by others, in case we like (or dislike) another person, it is likely that we will behave in a manner that makes the other person actually like (or dislike) us as well, i.e., likeability tends to be reciprocal. Indeed, Curtis and Miller (1986) show that such beliefs can turn into a self-fulfilling prophecy: people, who are told that another person (dis)likes them, actually behave in a way that makes them be (dis)liked. In our experiment, we control these different channels and make sure that teams are, with respect to knowledge of likeability, in a sense on a level playing field.[7]

After participants were informed about their matching partner, everybody made his or her decision in the public goods game. Moreover, we elicited participants' beliefs about 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.

1.3. *Co-ordination Game*

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

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

Table 1. *Pay-offs in the Co-ordination Game.*

	Other player's chosen number						
Own 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

In this game, two players simultaneously choose an integer between 1 and 7. A player's pay-off (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 with $i = 1, 2$. Table 1 shows the pay-off table for all possible combinations of one's own and the other player's chosen number.

Assuming that both players maximise their own individual pay-off, the co-ordination 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 seven yielding the highest pay-off of 13 euros to both players. However, this equilibrium is not without risks, because if the other player deviates, a player suffers a pay-off 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 pay-off of 7 euros.

To allow for learning and investigate potential dynamic effects, participants played the co-ordination game over ten rounds. At the end of the experiment one round was randomly selected for payment. Participants again answered a series of control questions before taking decisions in the co-ordination game. Moreover, on the computer screen we displayed once more the photograph of the matching partner together with the information about mutual likeability ranks (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. Participants were students from different disciplines at the two universities and were recruited via the Online Recruitment System for Economic Experiments (ORSEE) (Greiner, 2004). In total, 224 participants 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 participants earned on average 21 euros.

1.4. *Additional Independent Ratings*

To assess the robustness of our likeability effects, in particular with regard to potential correlates—such as 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 having participated in the original experiment, nor did anyone state that they knew 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 randomised across raters. On average, we obtained 20 independent ratings per photograph (approx. ten entered by female raters, ten by male raters). These ratings covered six different variables: likeability, attractiveness, co-operativeness, dominance, kindness and trustworthiness. Just as for likeability in the original experiment, each variable was evaluated on an 11-point Likert scale. Raters were paid a flat fee of 10 euros for their participation. The rating sessions lasted on average about 50 minutes.

2. Results

We present and discuss our results in the following way. We first show that there are no significant gender differences in the original likeability ratings of female and male participants in the main experiment. Next we turn to the effect of (dis)likeability in the co-operation and in the co-ordination game. In both games, we run separate non-parametric tests for same-sex and mixed-sex teams and analyse both matching protocols in a joint regression analysis.

2.1. *Likeability Ratings*

Do men and women rate the likeability 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 likeability ratings, nor the distribution of within-person standard deviations of these ratings differ significantly across gender or matching protocol ($p \geq 0.301$).[8] We thus conclude that male and female participants assess other people's likeability based on a photograph similarly.

Further support comes from a comparison of the original likeability ratings in the experiment and the additional items we elicited in the post-experimental survey with independent raters. Of all elicited variables, the original likeability rating is the strongest correlated with the likeability rating as assessed by the independent raters (see Table A1 in the Appendix). This is true when considering the entire sample and when examining the sub-samples of male and female participants separately. This suggests that independent and original raters closely agree on what it means to evaluate a particular face in terms of likeability, irrespective of the sex of the rated person.[9] We further find that likeability is positively correlated with other positive characteristics like kindness, co-operativeness, trustworthiness and attractiveness. This is true for male and female participants. In order to separate the effect of likeability from attractiveness, we will run regressions both with and without controlling for the latter variable. We control for attractiveness in a 'gender-adjusted' way, i.e., whenever we consider the behaviour of male participants, we include the average attractiveness rating of their matching partner assessed by independent male raters as a control variable. We proceed in a similar way with female participants. As we will discuss in detail below, our results do not depend on the inclusion of this control variable.

[8] Note that all p-values that we report in this article are based on two-sided tests.

[9] We find no significant gender differences in any of the correlations ($p \geq 0.147$). We control for rater-specific effects by standardising all ratings at the rater level before averaging across raters, similar to the approach suggested by Darai and Grätz (2013).

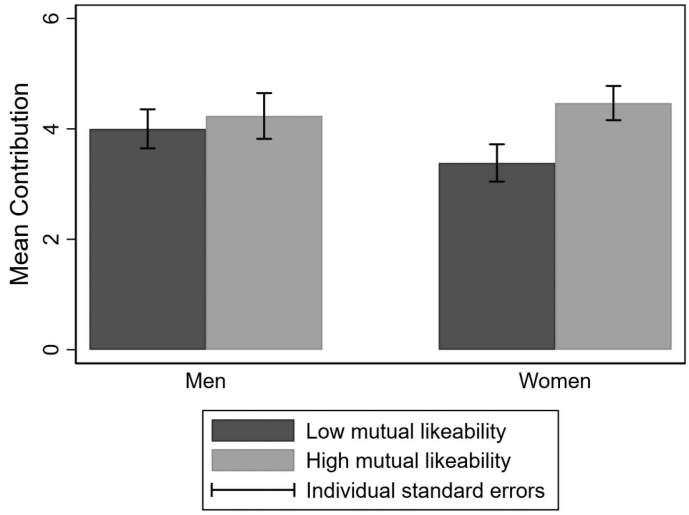


Fig. 1. *Average Contributions to the Public Good (Same-sex Teams).*

We now come to behaviour in the co-operation game (Subsection 2.2) and in the co-ordination game (Subsection 2.3).

2.2. *Behaviour in the Co-operation Game*

We observe no general difference in men's and women's co-operative behaviour 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, $p = 0.285$). This finding is consistent with results reported in Balliet *et al.* (2011), who show that there exist no significant gender differences in general co-operation.

2.2.1. *Same-sex teams*

When we consider same-sex teams, we see that men contribute on average 4.11 euros (s.d. 2.15). Women's average contribution amounts to 3.89 euros (s.d. 1.91). While the difference still remains insignificant (Mann-Whitney test, $p = 0.288$), the tendency for it to become larger is in line with Balliet *et al.* (2011), who document that co-operation in male–male interactions is significantly larger than co-operation in female–female interactions. Intriguingly, the gender difference becomes also more pronounced in our setting, once we condition on low and high mutual likeability.

Figure 1 shows the mean contribution of men and women in low and high mutual likeability same-sex teams. To create these subgroups, we divide the two gender samples according to a median split based on the sum of likeability ranks within each team. Recall that the smallest sum of likeability ranks is 2, which is the case if both players rank each other on rank 1. The largest sum is 8, which results if both rank each other on rank 4. The median value of the sum of ranks in our data is equal to 5 in both samples. With this, we obtain 30 and 34 observations of high and

Table 2. *Determinants of Behaviour in the Co-operation Game.*

	Men		Women	
	Same-sex	Mixed-sex	Same-sex	Mixed-sex
High mutual likeability	0.220	1.648	1.027	1.098
	(0.693)	(0.009)	(0.037)	(0.036)
Mutual likeability	0.171	0.439	0.269	0.319
	(0.250)	(0.014)	(0.039)	(0.020)
Likeability from	−0.209	0.476	0.426	0.448
	(0.420)	(0.124)	(0.065)	(0.172)
Likeability for	0.550	0.397	0.117	0.216
	(0.021)	(0.231)	(0.593)	(0.438)

Notes: The table displays main effects of selected variables, *p*-values from corresponding *F*-tests are given in parentheses. The dummy 'High mutual likeability' assumes value 1 if the pair's sum of mutual likeability ranks is smaller than the median value of all pairs' sum of ranks (=5). 'Mutual likeability' is defined as 10 − 'sum of likeability ranks'. 'Likeability for' and 'Likeability from' are defined as 5 − 'likeability rank assigned to', and 5 − 'likeability rank received from the matched partner', respectively. Information taken from the regression analysis presented in columns (1), (3) and (5) in Table A2 in the Appendix.

low mutual likeability teams, respectively; these observations comprise all teams, in which the sum of ranks is either strictly below or above or equal to the median value.

As Figure 1 shows, men in low as well as high mutual likeability teams contribute similar amounts to the public good (4.00 vs. 4.23 euros; Mann-Whitney test, $p = 0.462$). Even if they do not like each other, men do not seem to care. The difference in women's contributions, however, is significant. If mutual likeability in all-female teams is low, women contribute on average about 30% less (3.38 vs. 4.47 euros; Mann-Whitney test, $p = 0.020$).

We test these differences in a regression analysis. Regressions come with the advantage of enabling us to control for the matching partner's attractiveness as well as other visible characteristics, such as the wearing of glasses or jewellery, hair colour and (non-European) ethnicity that might influence both likeability ratings and contributions. Table A2 in the Appendix provides the complete regressions both with and without the inclusion of these controls. Table 2 summarises the main effects including controls.

The first two rows in Table 2 corroborate our findings from above. We consider mutual likeability either as a dummy variable (*high mutual likeability*), similar to the non-parametric tests above, or as a non-binary variable (*mutual likeability*), which is defined as 10 − 'sum of likeability ranks' in each team.[10] As can be seen, mutual likeability has a positive association with co-operation in both men's and women's same-sex matchings. However, the effect is both large in magnitude and statistically significant only in female teams. In detail, row 1 reveals that out of the endowment of 6 euros, men in same-sex teams contribute on average only 0.22 euros more if they like each other than if they don't and the effect is insignificant. In contrast, for women high mutual likeability translates into an additional 1.03 euros and the effect is significant on the 5% level. When considering mutual likeability as a non-dummy variable (row 2), we find similar results. A unit increase in mutual likeability translates into an insignificant increase of 0.17 euros for men, but a significant increase of 0.27 euros for women. While the regression results confirm that mutual likeability in same-sex teams has a significant effect for women but not for men,

[10] Recall that the maximum sum of ranks is 8 if both matching partners rank each other on rank 4, and the minimum sum of ranks is 2 if the matching partners rank each other on rank 1. The full regressions for the first two rows of Table 2 are presented in columns (1) and (3) of Table A2.

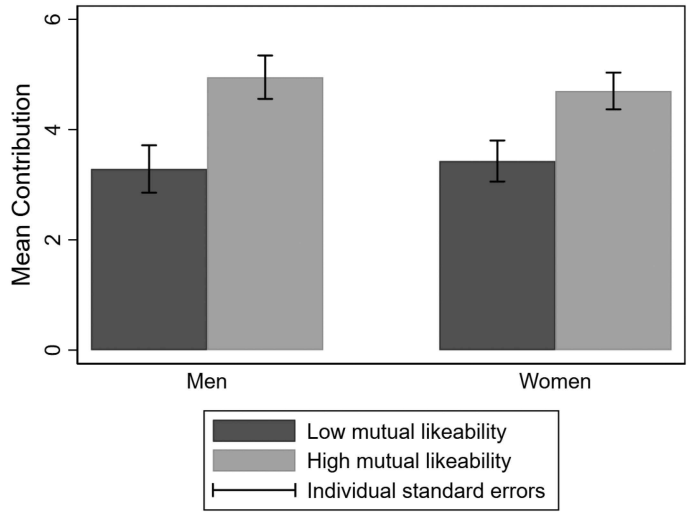


Fig. 2. *Average Contributions to the Public Good (Mixed-sex Teams).*

the gender difference in the co-operation game turns out not to be statistically significant (F-test results: $p = 0.274$ for the effects reported in row 1 and $p = 0.619$ for row 2).

2.2.2. *Mixed-sex teams*

Next, we consider co-operation in mixed-sex teams. Results show that the likeability effect for women remains; but, now, men care as well.

First, considering the entire sample of mixed-sex teams we, again, find no general gender difference with regard to co-operation. On average, men contribute 3.98 euros (s.d. 2.22) to the public good, women contribute 3.96 euros (s.d. 1.88, Mann-Whitney test, $p = 0.684$). To assess the impact of mutual likeability, we again perform a median split on the sum of likeability ranks.[11] From the total number of 48 mixed-sex teams, 20 teams are in the high and 28 are in the low mutual likeability group. Figure 2 shows the corresponding contribution decisions of men and women for these sub samples.

Just as in same-sex teams, we find a strong positive likeability effect for women in mixed-sex teams. Female participants contribute on average 4.70 euros to the public good in high mutual likeability teams, while they contribute about 37% less in low mutual likeability teams (3.43 (s.d. 1.97) vs. 4.70 (s.d. 1.49); Mann-Whitney test, $p = 0.022$). However, in contrast to same-sex teams, we now observe a significant likeability effect for men in mixed-sex teams as well. If mutual likeability is low, men's contribution is on average about 50% lower than if mutual likeability is high (3.29 (s.d. 2.27) vs. 4.95 (s.d. 1.76); Mann-Whitney test, $p = 0.005$).

The reported main effects from the first two rows of Table 2 confirm our findings from non-parametric tests.[12] For men in mixed-sex matchings, the effect of high mutual likeability can be quantified as an additional 1.65 euros on average. For women, high mutual likeability translates into an additional 1.10 euros. Similarly, a unit increase in mutual likeability when measured

[11] The median sum of ranks is equal to 5. Low and high mutual likeability teams are formed analogously to the same-sex matching sample.

[12] See full regressions in columns (1) and (3) of Table A2 in the Appendix.

as a non-dummy variable translates into an increase in contributions of 0.44 euros for men and 0.32 euros for women in mixed-sex matchings. All main effects within gender groups are significant on at least the 5% level. The differences between gender are insignificant (F-test result: $p = 0.494$ for the effects in row 1 and $p = 0.593$ for row 2).

2.2.3. *Further results*

The regression analysis provides us with further interesting results. First, we disentangle the effect of mutual likeability by separating the likeability rank a participant *assigns to* the matched partner from the likeability rank he or she *receives from* the other person. Here, we can also look at participants' absolute ratings instead of relative ranks. Second, we consider the role of attractiveness as well as other facial characteristics. Finally, we look at participants' beliefs.

While the main focus of our research is on *mutual* likeability as reflected in a team's sum of likeability ranks, we can use the regressions to take a closer look at the two components that this measure comprises: a participant's likeability rank received from the matched partner and the likeability rank assigned to the matched partner. To facilitate the interpretation of coefficients, we define the variable *likeability for* (the matched partner) as 5 − 'likeability rank assigned to the matched partner' and, analogously, *likeability from* (the matched partner) as 5 − 'likeability rank received from the matched partner'. As before, we summarise the main effects in Table 2, including controls, and report full regressions with and without controls in Table A2 in the Appendix.

Comparing men's and women's behaviour in same-sex matchings, the results in Table 2 show that women react significantly to *likeability from* (the matched partner). On average, a unit increase in this variable translates into an increase in contributions of 0.43 euros for women. At the same time, men show no significant reaction to the likeability rank they receive from the matched partner in same-sex teams. Instead, their contribution significantly depends on *likeability for* (the matched partner). A unit increase in this variable is associated with a significant increase of 0.55 euros in contributions on average. For women, no such association is found. F-tests reveal that the gender difference in these main effects is either marginally significant or insignificant (*likeability from*: $p = 0.065$; *likeability for*: $p = 0.182$).

In mixed-sex matchings, we find that neither *likeability from* nor *likeability for* (the matched partner) is significantly associated with behaviour in the public goods game (there are also no significant gender differences; F-test results: $p \geq 0.648$). Interestingly however, Table 2 shows that men's reaction to receiving a high likeability rank *from* a female matching partner is comparably more positive than receiving the same likeability rank from a male matching partner. On average, a unit increase in this variable translates into additional 0.48 euros in mixed-sex teams versus −0.21 euros in same-sex teams (F-test result: $p = 0.094$). All results hold if we replace a participants' likeability rank by his or her likeability rating as columns (7) and (8) in Table A2 in the Appendix show.

With respect to attractiveness of the matched partner as assessed by independent raters, we find that attractiveness matters only very little; it has no direct effect on contributions ($p \geq 0.218$) and does not alter the estimated effects of likeability. In general, all our results hold similarly irrespective of whether we include additional facial controls or not (see columns (1) to (5) in Table A2).

Lastly, we consider participants' beliefs about the matched partner's co-operativeness. Table 3 provides summary statistics for participants' expectations about their matching partners' contributions to the public good as well as the difference between these expectations and

Table 3. *Average Contributions and Expectations in the Public Goods Game.*

	Men		Women	
	Same-sex	Mixed-sex	Same-sex	Mixed-sex
Low mutual likeability pairs				
Contribution	4.00 (2.06)	3.29 (2.27)	3.38 (1.97)	3.43 (1.97)
Expectation	3.91 (1.52)	3.32 (1.59)	3.53 (1.46)	3.32 (1.79)
Difference	−0.09 (2.64)	−0.11 (2.57)	0.15 (2.28)	0.34 (2.63)
Absolute difference	2.03 (1.66)	2.04 (1.53)	1.74 (1.46)	2.25 (1.29)
High mutual likeability pairs				
Contribution	4.23 (2.27)	4.95 (1.76)	4.47 (1.70)	4.70 (1.49)
Expectation	4.17 (1.49)	4.60 (0.68)	4.47 (0.78)	4.35 (1.27)
Difference	−0.07 (2.78)	−0.10 (1.74)	0.00 (2.05)	−0.60 (2.35)
Absolute difference	2.13 (1.74)	1.40 (0.99)	1.60 (1.25)	1.90 (1.45)

Notes: The table displays mean values and standard deviations (calculated at the individual level) that are given in parentheses. (Absolute) differences are defined as the (absolute) differences between the expectation about the matching partner's contribution and his/her actual contribution.

actually observed contributions. On average, the difference between participants' beliefs and their matching partner's actual contributions is found to be relatively small. This, however, hides the fact that some participants overestimate the matched partner's contribution considerably while others underestimate it. If we consider the absolute difference between beliefs and actual contributions, the data show that participants' beliefs are on average about 2 euros off, at the aggregate. We do not find any significant effects of gender, type of matching (same sex vs. mixed sex) or mutual likeability considerations on participants' ability to predict their matching partner's contributions.[13]

In Table A3 and Table A4 in the Appendix we analyse how far beliefs about the matching partner's co-operativeness vary with mutual likeability. Table A3 summarises the main effects; Table A4 parallels the regressions from Table A2, the only difference being that, instead of contributions, we now use participants' beliefs about their matching partner's contribution as a dependent variable. The results mirror closely our results on contributions from above. This indicates that participants' beliefs depend in a similar way on mutual likeability, suggesting that low contributions in the event of low mutual likeability are driven not by a desire for retaliation, but rather by an expression of low expectations regarding the contribution of the other player.

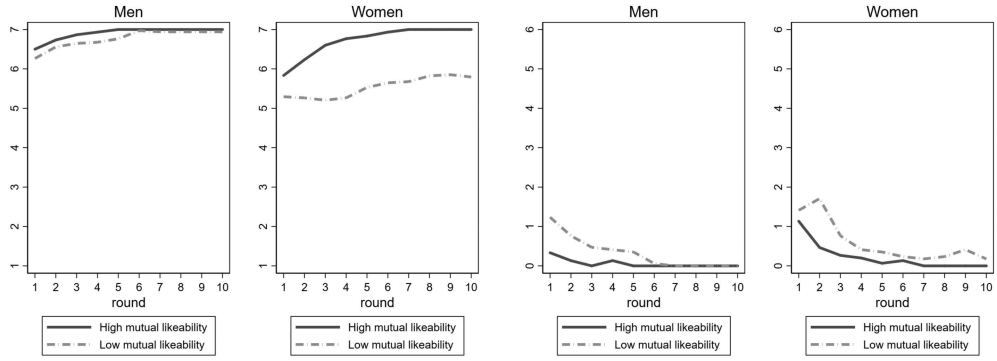
In sum, our results in the co-operation game show that women's co-operation is negatively affected by mutual dislikeability in both same-sex and mixed-sex teams. To the contrary, men seem to care about mutual likeability only in mixed-sex teams.

Next, we analyse behaviour in the co-ordination game, to see whether our results carry over to a different strategic setting.

2.3. *Behaviour in the Co-ordination Game*

Recall that in each of the ten rounds of the co-ordination game team members choose numbers between one and seven. The game has seven strict Nash equilibria in each of which both players choose the same number. These equilibria are Pareto ranked with pay-offs ranging from 7 euros in equilibrium (1, 1) to 13 euros in equilibrium (7, 7).

[13] Ranksum test results: $p \geq 0.143$.



(a) *Average Numbers Over Rounds* (b) *Difference in Numbers Over Rounds*

Fig. 3. *Behaviour in the Co-ordination Game (Same-sex Teams).*

We again start by considering same-sex teams first and then analyse behaviour in mixed-sex teams.

2.3.1. *Same-sex teams*

Figure 3 illustrates men's and women's behaviour in the co-ordination game in same-sex teams. Panel (a) displays for each round the average number chosen across all teams; panel (b) shows the average within-team difference of numbers chosen by the two members in a team. As before, we divide the two gender samples based on a median split of the sum of likeability ranks to compare behaviour in low and high mutual likeability teams (cf. Subsection 2.2). The broken lines show low mutual likeability teams; the solid lines show high mutual likeability teams.

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

However, when looking at panel (a) we see that female and male teams behave very differently with respect to *what equilibrium* they co-ordinate on. The data reveal two key findings: First, women choose significantly lower numbers in low mutual likeability teams than in high mutual likeability teams. While average numbers are above 6 and very quickly converge to 7 if mutual likeability in all-female teams is high, the average numbers basically stay around 5.5 over the ten rounds if mutual likeability is low (Mann-Whitney tests comparing women in high vs. low mutual likeability teams: $p \leq 0.040$ from round 4 onward).[14] Second and to the contrary, male participants choose high numbers close to 7 right from the start. More importantly, they do so irrespective of the level mutual likeability (Mann-Whitney tests comparing men's high vs. low mutual likeability teams: $p \geq 0.199$ in all rounds except round 5).[15] When looking at gender

[14] In round 1 the difference is not significant ($p = 0.328$), in round 2 the difference is marginally significant ($p = 0.100$) and in round 3 the difference is significant at the 10% level ($p = 0.059$).

[15] In round 5: $p = 0.093$.

Table 4. *Determinants of Behaviour in the Co-ordination Game.*

	Men		Women	
	Same-sex	Mixed-sex	Same-sex	Mixed-sex
High mutual likeability	0.189	0.848	1.101	0.125
	(0.109)	(0.037)	(0.014)	(0.712)
Mutual likeability	0.080	0.300	0.279	0.084
	(0.022)	(0.023)	(0.019)	(0.367)
Likeability from	0.065	0.324	0.320	0.120
	(0.106)	(0.127)	(0.008)	(0.426)
Likeability for	0.096	0.269	0.237	0.053
	(0.052)	(0.145)	(0.076)	(0.794)

Notes: The table displays main effects of selected variables, p-values from corresponding F-tests are given in parentheses. The dummy 'High mutual likeability' assumes value 1 if the pair's sum of mutual likeability ranks is smaller than the median value of all pairs' sum of ranks (=5). 'Mutual likeability' is defined as 10 − 'sum of likeability ranks'. 'Likeability for' and 'Likeability from' are defined as 5 − 'likeability rank assigned to', and 5 − 'likeability rank received from the matched partner', respectively. Information taken from the regression analysis presented in columns (1), (3) and (5) in Table A5 in the Appendix.

differences conditional on the level of mutual likeability, we find that with the exception of rounds 1 and 2 there are no significant gender differences in high mutual likeability teams.[16] However, in low mutual likeability teams differences are significant in all rounds (Mann-Whitney tests, $p \leq 0.066$).

We again test and extend our results in a regression analysis that allows us to additionally control for the matching partner's attractiveness as well as further facial characteristics. Table A5 in the Appendix contains the full results with and without controls, Table 4 summarise the main effects with controls. All variables are as defined in Subsection 2.2.

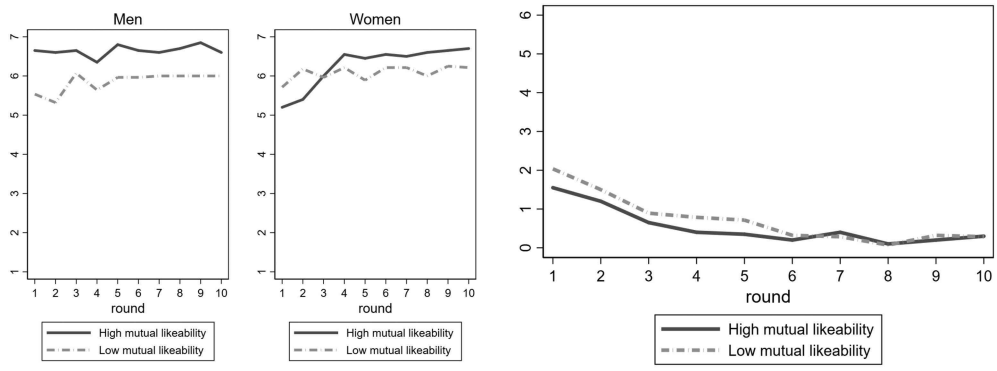
As the first two rows in Table 4 show, co-ordination in same-sex teams is significantly affected by mutual likeability for women, while for men this does not hold or only to a much smaller degree. On average, men who are matched to the same sex in high mutual likeability teams choose only slightly larger numbers compared with low mutual likeability teams (main effect size equal to 0.189: see first row in Table 4). For women, in contrast, this effect is more than five times as large (main effect size equal to 1.101). Similarly, a unit increase in mutual likeability translates into a mean increase of only 0.080 in the number chosen for men but to an increase of 0.279 for women. Unlike our results in the co-operation game, these gender differences are now also (marginally) significant (F-test results: $p = 0.044$ for the effects reported in row 1 and $p = 0.099$ for row 2).

Thus, also in the co-ordination game we see that female teams react to mutual likeability, while male teams do so, if at all, only to a much smaller degree. While there is no gender effect on participants' ability to co-ordinate per se in same-sex teams, we observe a gender effect with regard to equilibrium outcomes. Literally *all* male teams successfully co-ordinate on the Pareto efficient equilibrium (7, 7) eventually, irrespective of the level of mutual (dis)likeability. This, however, does not hold for female teams.

2.3.2. *Mixed-sex teams*

Figure 4 depicts the average chosen numbers (panel (a)) and within-team differences (panel (b)) of women and men in low (broken lines) and high (solid lines) mutual likeability teams when matched with the opposite sex.

[16] Mann-Whitney tests, $p = 0.072$ in round 1, $p = 0.061$ in round 2 and $p \geq 0.150$ in the remaining rounds.



(a) *Chosen Numbers Over Rounds* (b) *Difference in Chosen Numbers Over Rounds*

Fig. 4. *Behaviour in the Co-ordination Game (Mixed-sex Teams).*

Again, consider panel (b) first. As evident, although co-ordination takes a bit more time and differences between numbers are initially slightly larger compared to same-sex teams, almost all mixed-sex teams co-ordinate eventually on equilibrium play, i.e., choose the same number, irrespective of the level of mutual likeability. Mann-Whitney tests reveal for none of the ten rounds a significant difference between the high and low mutual likeability groups ($p \geq 0.353$).

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

The main effects displayed in the first two rows of Table 4 confirm that in mixed-sex teams, mutual likeability is positively associated with the number chosen for both women and men.[17] Intriguingly and similar to our results on co-operation in mixed-sex teams, this effect is larger for men (F-test results comparing gender differences in the main effects: $p = 0.008$ for the results in row 1 and $p = 0.010$ for row 2).

2.3.3. *Further results*

Again, we find that the partner's attractiveness has no significant effect on chosen numbers in the co-ordination game ($p \geq 0.201$ in models (1) and (3)). Also, the main effects summarised in the first two rows of Table 4 remain unchanged if we drop the controls for facial characteristics in models (2) and (4) of Table A5.

When we break down a team's mutual likeability into its two components, the lower two rows of Table 4 reveal that only in same-sex teams women react more strongly to likeability

[17] Cf. columns (1) and (3) in Table A5 for the full results.

received from their matching partner than men (F-test result: $p = 0.038$).[18] All remaining F-test results on gender differences between the main effects extracted from model (5) are insignificant at conventional levels. Similar results are obtained, if we drop the control variables for facial characteristics or take participants' likeability ratings instead of likeability ranks (see models (6) to (8) in Table A5).

Taken as a whole, in the co-ordination game we find a very similar behavioural pattern as in the co-operation game: female teams' ability to co-ordinate on the Pareto-efficient equilibrium is comparably more affected by concerns for mutual likeability than male teams'. In mixed teams, on the other hand, both men and women react to mutual likeability leading to sizeable efficiency losses.

3. Conclusion

The results in our study provide new evidence for gender differences in economic and social interactions. In both a one-shot co-operation and a repeated co-ordination game, we show that mutual likeability significantly shapes players' behaviour. If mutual likeability is low (i.e., players like each other relatively little), players co-operate less and co-ordinate on less efficient equilibria than if mutual likeability is high (i.e., players like each other relatively more). Importantly, this result holds only if one of the team members is a woman. If men are among themselves, they do not seem to care much about mutual likeability, i.e., they both co-operate and co-ordinate on high pay-off equilibria independent of the level of mutual likeability in their team. Our likeability ratings are based on participants' photographs. We therefore control for attractiveness and other facial characteristics in our analysis, and find that it leaves our results unaffected.

Our results hint at the existence of a *likeability factor* that offers a novel perspective on gender differences in labour market outcomes. Basically, what the likeability factor says is that for women, likeability is an asset (or equivalently, dislikeability is a hurdle) in *every* of their interactions. For men, on the other hand, likeability matters only if they interact with the opposite sex. This difference is all the more important, as our results suggest that the overall effect of variation in likeability is negative, i.e., dislikeability hurts more than likeability helps. The likeability factor thus creates a behavioural asymmetry, leading to considerable advantages in terms of average performance and economic outcomes for men.

These advantages are well visible in our experiment, which can be seen as a lab model of human behaviour in team interactions. Aggregated over all rounds of both games and all teams, women earn on average 4.36% less than men in our experiment. In same-sex teams, the gender pay gap is even larger and amounts to 7.75% lower earnings for women on average.[19] Note that these earnings differences are based on a balanced matching between male and female participants in a controlled lab environment in which further influencing factors such as discrimination in wage setting are ruled out. Given that in many organisational and labour market contexts such factors also play a role and—in particular in better-paid jobs and top management positions—interactions with (and between) male colleagues still represent the standard, our results most

[18] For the corresponding full regression see model (5) of Table A5.
[19] These differences are statistically significant on the 5% and 1% level, respectively (Mann-Whitney tests).

likely underestimate the impact of likeability effects on gender differences in labour market outcomes in the field.

What can society and firms do about it? First, research shows that the mere awareness of a behavioural asymmetry or bias can sometimes have surprisingly positive effects leading to a significant reduction or even disappearance of the bias (Pope *et al.*, 2018). In this sense, the recognition of the likeability factor by firms, workers and in particular human resources managers seems important and may already have some effect. More generally, our results suggest that firms are advised to implement and promote what we would call 'likeability neutral' work cultures, i.e., environments that highlight performance and professional behaviour among employees but do not see likeability as a prerequisite for successful co-operation. To the extent that it is feasible and desirable, our results also have implications for how to optimally form teams and how to make rules regarding requests for re-assignments. In sum, we believe that the likeability factor can provide a useful new perspective on how modern workplaces shape male and female interactions and how they can be further developed to provide equal opportunities for women and men.

From a research perspective, our results contribute to the literature by showing that likeability is one of the important context-dependent variables to which women seem more sensitive than men, which can potentially explain gender differences in a wider range of studies (Croson and Gneezy, 2009). In this way, the results also shed light on some of these differences from a new angle. For example, our results suggest that the higher co-operation rates observed in male–male interactions compared with female–female interactions (Balliet *et al.*, 2011) may be driven by gender differences in same-sex, low mutual likeability teams. Further, they offer a possible mechanism for recent findings in co-authorship networks documenting a stronger degree of gender based homophily (i.e., a preference to form links with others of the same gender) for men than women (Ductor *et al.*, 2018).

Still, a number of important questions remain for future research. First, it would be interesting to see whether and how behaviour is affected differently if players can adjust their likeability rating after their first interaction with the other player. It is conceivable that the effects in our setting, which so far depend only on the first visual impression and not on the other player's actual behaviour, will become stronger as mutual likeability and behaviour may reinforce each other. Second, different modes of interaction, e.g., via the phone or face to face, are worth being explored. Finally, it seems important to investigate more explicitly what instruments can be used to mitigate the negative effects of dislikeability or, alternatively, improve the positive effects of likeability. Most likely gender differences exist in the effectiveness of these instruments as well. We regard these studies as the logical next step to further explore and better understand the role of likeability in social interaction and its implication for gender differences in labour market outcomes.

Appendix A. Additional Tables

Table A1. *Correlation Matrix of Standardised Facial Appearance Variables Assessed by Participants in the Original Experiment and Independent Raters.*

	Original likeability rating			
	All 224 participants	112 female participants	112 male participants	Test for equality of men's and women's correlation coefficients
Likeability	0.629***	0.663***	0.618***	$p = 0.570$
Attractiveness	0.400***	0.395***	0.491***	$p = 0.374$
Co-operativeness	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***	$p = 0.303$

Notes: $^{**}p < 0.05$, $^{***}p < 0.01$. The original likeability rating is calculated as the mean of the four standardised ratings that were given during the original experiment. The remaining variables are calculated as the mean standardised ratings given by the on average 20 independent raters per photograph.

Table A2. *Determinants of Behaviour in the Co-operation Game.*

	Dep. var.: own contribution							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High mutual likeability	0.220	0.233						
	(0.556)	(0.558)						
High mutual likeability ×	1.428*	1.431*						
Mixed sex	(0.827)	(0.807)						
High mutual likeability ×	0.807	0.851						
Female	(0.734)	(0.738)						
High mutual likeability ×	−1.358	−1.244						
Female × Mixed sex	(1.074)	(1.068)						
Mutual likeability			0.171	0.177				
			(0.148)	(0.148)				
Mutual likeability ×			0.268	0.269				
Mixed sex			(0.231)	(0.223)				
Mutual likeability ×			0.098	0.102				
Female			(0.196)	(0.193)				
Mutual likeability ×			−0.217	−0.203				
Female × Mixed sex			(0.296)	(0.290)				
Likeability from					−0.209	−0.217	−0.220	−0.221
					(0.258)	(0.255)	(0.255)	(0.256)
Likeability from ×					0.685*	0.704*	0.723*	0.751*
Mixed sex					(0.406)	(0.387)	(0.418)	(0.404)
Likeability from ×					0.635*	0.646*	0.610*	0.625*
Female					(0.341)	(0.338)	(0.346)	(0.339)
Likeability from ×					−0.663	−0.672	−0.722	−0.754
Female × Mixed sex					(0.535)	(0.525)	(0.537)	(0.528)

Table A2. *Continued*

	Dep. var.: own contribution							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Likeability for					0.550**	0.571**		
					(0.235)	(0.237)		
Likeability for ×					−0.153	−0.174		
Mixed sex					(0.413)	(0.391)		
Likeability for ×					−0.433	−0.442		
Female					(0.322)	(0.310)		
Likeability for ×					0.253	0.289		
Female × Mixed sex					(0.521)	(0.481)		
Rating for							0.897***	0.897***
							(0.290)	(0.296)
Rating for ×							−0.567	−0.597
Mixed sex							(0.554)	(0.529)
Rating for ×							−0.802*	−0.792*
Female							(0.410)	(0.400)
Rating for ×							1.020	1.062*
Female × Mixed sex							(0.673)	(0.638)
Female	−0.676	−0.618	−0.729	−0.731	−0.730	−0.731	−1.717*	−1.755*
	(0.535)	(0.529)	(1.089)	(1.092)	(1.102)	(1.103)	(0.981)	(0.963)
Mixed sex	−0.836	−0.714	−1.432	−1.384	−1.399	−1.365	−1.939	−1.951*
	(0.612)	(0.550)	(1.222)	(1.188)	(1.232)	(1.195)	(1.177)	(1.157)
Female × Mixed sex	1.053	0.761	1.357	1.203	1.257	1.139	2.023	2.016
	(0.877)	(0.772)	(1.658)	(1.615)	(1.672)	(1.631)	(1.591)	(1.532)
Constant	3.909***	4.000***	3.098***	3.191***	3.093***	3.191***	4.615***	4.650***
	(0.757)	(0.342)	(0.957)	(0.785)	(0.973)	(0.793)	(0.896)	(0.710)
Controls	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*
Adjusted R^2	0.037	0.049	0.047	0.066	0.050	0.070	0.060	0.078
Number of clusters	112	112	112	112	112	112	112	112
Observations	224	224	224	224	224	224	221	221

Notes: OLS regressions with standard errors that are clustered at the matching pair level and given in parentheses: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dummy 'High mutual likeability' assumes value 1 if the pair's sum of mutual likeability ranks is smaller than the median value of all pairs' sum of ranks (=5). 'Mutual likeability' is defined as the difference 10 − sum of ranks. 'Likeability for' and 'Likeability from' are defined as the differences 5 − likeability rank assigned to, and 5 − likeability rank received from the matched partner, respectively. Likeability ratings are standardised at the rater level. Controls include attractiveness of the matched partner as well as dummies indicating whether the matched partner wears glasses or jewellery, or has dark hair or a European appearance.

Table A3. *Determinants of Participants' Expected Contributions.*

	Men		Women	
	Same-sex	Mixed-sex	Same-sex	Mixed-sex
High mutual likeability	0.197	1.229	0.924	0.988
	(0.571)	(0.001)	(0.002)	(0.032)
Mutual likeability	0.114	0.330	0.258	0.295
	(0.228)	(0.001)	(0.001)	(0.011)
Likeability from	−0.066	0.363	0.392	0.450
	(0.667)	(0.049)	(0.009)	(0.138)
Likeability for	0.293	0.288	0.123	0.165
	(0.064)	(0.141)	(0.446)	(0.511)

Notes: The table displays main effects of selected variables, *p*-values from corresponding *F*-tests are given in parentheses. The dummy 'High mutual likeability' assumes value 1 if the pair's sum of mutual likeability ranks is smaller than the median value of all pairs' sum of ranks (=5). 'Mutual likeability' is defined as the difference 10 − sum of ranks. 'Likeability for' and 'Likeability from' are defined as the differences 5 − likeability rank assigned to, and 5 − likeability rank received from the matched partner, respectively. Information taken from the regression analysis presented in columns (1), (3) and (5) in Table A4.

Table A4. *Determinants of Participants' Expected Contributions.*

	Dep. var.: expected contribution							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High mutual likeability	0.197	0.255						
	(0.347)	(0.358)						
High mutual likeability ×	1.032**	1.024**						
Mixed sex	(0.487)	(0.491)						
High mutual likeability ×	0.727	0.682						
Female	(0.451)	(0.441)						
High mutual likeability ×	−0.967	−0.932						
Female × Mixed sex	(0.711)	(0.711)						
Mutual likeability			0.114	0.125				
			(0.094)	(0.099)				
Mutual likeability ×			0.216	0.216				
Mixed sex			(0.135)	(0.136)				
Mutual likeability ×			0.144	0.138				
Female			(0.121)	(0.117)				
Mutual likeability ×			−0.179	−0.192				
Female × Mixed sex			(0.192)	(0.189)				
Likeability from					−0.066	−0.058	−0.026	−0.008
					(0.153)	(0.153)	(0.155)	(0.159)
Likeability from ×					0.429*	0.429*	0.386	0.387
Mixed sex					(0.242)	(0.236)	(0.254)	(0.253)
Likeability from ×					0.458**	0.433**	0.417*	0.389*
Female					(0.212)	(0.215)	(0.225)	(0.229)
Likeability from ×					−0.371	−0.389	−0.373	−0.402
Female × Mixed sex					(0.417)	(0.410)	(0.410)	(0.407)
Likeability for					0.293*	0.307*		
					(0.157)	(0.162)		
Likeability for ×					−0.005	−0.001		
Mixed sex					(0.257)	(0.258)		
Likeability for ×					−0.170	−0.157		
Female					(0.222)	(0.223)		
Likeability for ×					0.047	0.029		
Female × Mixed sex					(0.394)	(0.387)		
Rating for							0.301	0.280
							(0.200)	(0.203)
Rating for ×							0.017	0.026
Mixed sex							(0.322)	(0.330)
Rating for ×							−0.152	−0.090
Female							(0.300)	(0.306)
Rating for ×							0.251	0.213
Female × Mixed sex							(0.494)	(0.502)
Female	−0.351	−0.382	−0.686	−0.762	−0.683	−0.762	−1.011	−1.058
	(0.348)	(0.347)	(0.704)	(0.704)	(0.711)	(0.710)	(0.654)	(0.658)
Mixed sex	−0.576	−0.590	−1.108	−1.193	−1.084	−1.180	−1.000	−1.078
	(0.436)	(0.402)	(0.833)	(0.822)	(0.844)	(0.834)	(0.764)	(0.747)
Female × Mixed sex	0.367	0.382	0.679	0.919	0.584	0.857	0.694	0.943
	(0.633)	(0.558)	(1.173)	(1.120)	(1.196)	(1.142)	(1.236)	(1.165)
Constant	4.288***	3.912***	3.758***	3.385***	3.793***	3.385***	4.472***	4.034***
	(0.541)	(0.266)	(0.658)	(0.583)	(0.655)	(0.588)	(0.563)	(0.472)
Controls	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*
Adjusted R^2	0.074	0.075	0.092	0.101	0.089	0.096	0.088	0.093
Number of clusters	112	112	112	112	112	112	112	112
Observations	224	224	224	224	224	224	221	221

Notes: OLS regressions with standard errors that are clustered at the matching pair level and given in parentheses: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dummy 'High mutual likeability' assumes value 1 if the pair's sum of mutual likeability ranks is smaller than the median value of all pairs' sum of ranks (=5). 'Mutual likeability' is defined as the difference 10 − sum of ranks. 'Likeability for' and 'Likeability from' are defined as the differences 5 − likeability rank assigned to, and 5 − likeability rank received from the matched partner, respectively. Likeability ratings are standardised at the rater level. Controls include attractiveness of the matched partner as well as dummies indicating whether the matched partner wears glasses or jewellery, or has dark hair or a European appearance.

Table A5. *Determinants of Behaviour in the Co-ordination Game.*

	Dep. var.: chosen number							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High mutual likeability	0.189	0.139						
	(0.117)	(0.108)						
High mutual likeability ×	0.659	0.656						
Mixed sex	(0.416)	(0.412)						
High mutual likeability ×	0.912**	1.046**						
Female	(0.448)	(0.470)						
High mutual likeability ×	−1.634***	−1.667***						
Female × Mixed sex	(0.533)	(0.521)						
Mutual likeability			0.080**	0.058*				
			(0.034)	(0.032)				
Mutual likeability ×			0.219	0.216				
Mixed sex			(0.134)	(0.131)				
Mutual likeability ×			0.199*	0.218*				
Female			(0.120)	(0.124)				
Mutual likeability ×			−0.415***	−0.407***				
Female × Mixed sex			(0.151)	(0.144)				
Likeability from					0.065	0.047	0.077*	0.069*
					(0.040)	(0.029)	(0.046)	(0.040)
Likeability from ×					0.259	0.279	0.244	0.270
Mixed sex					(0.213)	(0.217)	(0.226)	(0.234)
Likeability from ×					0.255**	0.294**	0.147	0.185
Female					(0.122)	(0.129)	(0.124)	(0.124)
Likeability from ×					−0.459	−0.485	−0.296	−0.326
Female × Mixed sex					(0.322)	(0.329)	(0.291)	(0.294)
Likeability for					0.096*	0.068		
					(0.049)	(0.043)		
Likeability for ×					0.174	0.143		
Mixed sex					(0.189)	(0.187)		
Likeability for ×					0.141	0.141		
Female					(0.136)	(0.128)		
Likeability for ×					−0.358	−0.313		
Female × Mixed sex					(0.346)	(0.341)		
Rating for							0.065	0.020
							(0.061)	(0.039)
Rating for ×							0.192	0.173
Mixed sex							(0.243)	(0.243)
Rating for ×							0.298	0.303
Female							(0.218)	(0.205)
Rating for ×							−0.666	−0.622
Female × Mixed sex							(0.460)	(0.441)
Female	−1.376***	−1.229***	−1.912**	−1.850**	−1.911**	−1.850**	−1.250**	−1.170**
	(0.463)	(0.456)	(0.830)	(0.862)	(0.831)	(0.863)	(0.481)	(0.511)
Mixed sex	−1.149***	−0.915**	−1.876**	−1.682**	−1.863**	−1.658**	−1.410*	−1.274*
	(0.421)	(0.368)	(0.855)	(0.822)	(0.841)	(0.807)	(0.736)	(0.727)
Female × Mixed sex	1.891***	1.465***	3.101***	2.753***	3.071***	2.709***	1.774**	1.471*
	(0.567)	(0.489)	(0.997)	(0.962)	(1.002)	(0.965)	(0.856)	(0.844)
Constant	6.080***	6.300***	5.724***	6.065***	5.743***	6.065***	5.896***	6.176***
	(0.348)	(0.144)	(0.363)	(0.238)	(0.373)	(0.238)	(0.348)	(0.188)
Controls	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*
Round dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Adjusted R^2	0.127	0.113	0.143	0.125	0.142	0.124	0.130	0.115
Number of clusters	112	112	112	112	112	112	112	112
Observations	2240	2240	2240	2240	2240	2240	2210	2210

Notes: OLS regressions with standard errors that are clustered at the matching pair level and given in parentheses: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dummy 'High mutual likeability' assumes value 1 if the pair's sum of mutual likeability ranks is smaller than the median value of all pairs' sum of ranks (=5). 'Mutual likeability' is defined as the difference 10 − sum of ranks. 'Likeability for' and 'Likeability from' are defined as the differences 5 − likeability rank assigned to, and 5 − likeability rank received from the matched partner, respectively. Likeability ratings are standardised at the rater level. Controls include attractiveness of the matched partner as well as dummies indicating whether the matched partner wears glasses or jewellery, or has dark hair or a European appearance.

Universität Hamburg
Goethe University Frankfurt

Additional Supporting Information may be found in the online version of this article:

Online Appendix
Replication Package

References

Altonji, J.G. and Blank, R.M. (1999). 'Race and gender in the labor market', *Handbook of Labor Economics*, vol. 3, pp. 3143–259.
Andreoni, J. and Petrie, R. (2008). 'Beauty, gender and stereotypes: evidence from laboratory experiments', *Journal of Economic Psychology*, vol. 29, pp. 73–93.
Babcock, L. and Laschever, S. (2009). 'Women don't ask: negotiation and the gender divide', Princeton University Press.
Balliet, D., Li, N.P., Macfarlan, S.J. and van Vugt, M. (2011). 'Sex differences in cooperation: a meta-analytic review of social dilemmas', *Psychological Bulletin*, vol. 137, pp. 881–909.
Barber, B.M. and Odean, T. (2001). 'Boys will be boys: gender, overconfidence, and common stock investment', *Quarterly Journal of Economic*s, vol. 116, pp. 261–92.
Bell, R., Buchner, A., Kroneisen, M. and Giang, T. (2012). 'On the flexibility of social source memory: a test of the emotional incongruity hypothesis', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 38, pp. 1512–29.
Benenson, J.F. (2014). *Warriors and Worriers: The Survival of the Sexes*, Oxford University Press.
Benenson, J.F. (2019). 'Sex differences in human peer relationships: a primate's-eye view', *Current Directions in Psychological Science*, vol. 28, pp. 1–7.
Benenson, J.F., Markovits, H., Fitzgerald, C., Geoffroy, D., Flemming, J., Kahlenberg, S.M. and Wrangham, R.W. (2009). 'Males' greater tolerance of same-sex peers', *Psychological Science*, vol. 20, pp. 184–90.
Bertrand, M. (2011). 'New perspectives on gender', in *Handbook of Labor Economics*, vol. 4, pp. 1543–90, Elsevier.
Bertrand, M., Goldin, C. and Katz, L.F. (2010). 'Dynamics of the gender gap for young professionals in the financial and corporate sectors', *American Economic Journal: Applied Economics*, vol. 2, pp. 228–55.
Bertrand, M. and Hallock, K.F. (2001). 'The gender gap in top corporate jobs', *ILR Review*, vol. 55, pp. 3–21.
Blau, F.D. and Kahn, L.M. (2017). 'The gender wage gap: extent, trends, and explanations', *Journal of Economic Literature*, vol. 55, pp. 789–865.
Buser, T., Niederle, M. and Oosterbeek, H. (2014). 'Gender, competitiveness, and career choices', *Quarterly Journal of Economics*', vol. 1409, p. 1447.
Casciaro, T. and Sousa Lobo, M. (2008). 'When competence is irrelevant: the role of interpersonal affect in task-related ties', *Administrative Science Quarterly*, vol. 53, pp. 655–84.
Cassidy, H., DeVaro, J. and Kauhanen, A. (2016). 'Promotion signaling, gender, and turnover: new theory and evidence', *Journal of Economic Behavior & Organization*, vol. 126, pp. 140–66.
Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M. and Seabright, P. (2015). 'Honest signaling in trust interactions: smiles rated as genuine induce trust and signal higher earning opportunities', *Evolution and Human Behavior*, vol. 36, pp. 8–16.
Croson, R. and Gneezy, U. (2009). 'Gender differences in preferences', *Journal of Economic Literature*, vol. 47, pp. 448–74.
Curtis, R.C. and Miller, K. (1986). 'Believing another likes or dislikes you: behaviors making the beliefs come true', *Journal of Personality and Social Psychology*, vol. 51, p. 284.
Darai, D. and Grätz, S. (2013). 'Attraction and cooperative behavior', Working Paper, University of Zurich.
DeBruine, L.M. (2002). 'Facial resemblance enhances trust', *Proceedings of the Royal Society B: Biological Sciences*, vol. 269, pp. 1307–12.
Declerck, C.H., Boone, C. and Emonds, G. (2013). 'When do people cooperate? The neuroeconomics of prosocial decision making', *Brain and Cognition*, vol. 81, pp. 95–117.
Dohmen, T. and Falk, A. (2011). 'Performance pay and multidimensional sorting: productivity, preferences, and gender', *American Economic Review*, vol. 101, pp. 556–90.
Ductor, L., Goyal, S. and Prummer, A. (2018). 'Gender and collaboration', Working Paper, Queen Mary University of London.
Eckel, C.C. and Grossman, P.J. (2008a). 'Differences in the economic decisions of men and women: experimental evidence', in *Handbook of Experimental Economics Results*, vol. 1, pp. 509–19.
Eckel, C.C. and Grossman, P.J. (2008b). 'Men, women and risk aversion: experimental evidence', in *Handbook of Experimental Economics Results*, pp. 1061–73.

Fischbacher, U. (2007). 'z-Tree: Zurich toolbox for ready-made economic experiments', *Experimental Economics*, vol. 10, pp. 171–78.
Frederiksen, A. and Kato, T. (2017). 'Human capital and career success: evidence from linked employer-employee data', Economic Journal, vol. 128, pp. 1952–82.
Friebel, G. and Seabright, P. (2011). 'Do women have longer conversations? Telephone evidence of gendered communication strategies', *Journal of Economic Psychology*, vol. 32, pp. 348–56.
Giang, T., Bell, R. and Buchner, A. (2012). 'Does facial resemblance enhance cooperation?', *PLoS ONE*, vol. 7, pp. e47809–11.
Gibbons, R. and Roberts, J. (2013). *The Handbook of Organizational Economics*, Princeton University Press.
Gneezy, U., Niederle, M. and Rustichini, A. (2003). 'Performance in competitive environments: gender differences', *Quarterly Journal of Economics*, vol. 118, pp. 1049–74.
Goldin, C. (2014). 'A grand gender convergence: its last chapter', *American Economic Review*, vol. 104, pp. 1091–119.
Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE', *Journal of the Economic Science Association*, vol. 1, pp. 114–25.
Hamermesh, D.S. and Biddle, J.E. (1994). 'Beauty and the labor market', *American Economic Review*, vol. 84, pp. 1174–94.
van Huyck, J.B., Battalio, R.C. and Beil, R.O. (1990). 'Tacit coordination games, strategic uncertainty, and coordination failure', *American Economic Review*, vol. 80, pp. 234–48.
Kleven, H.J., Landais, C. and Sogaard, J.E. (2019). 'Children and gender inequality: evidence from Denmark', *American Economic Journal: Applied Economics*, vol. 11, pp. 181–209.
Lindenlaub, I. and Prummer, A. (2019). 'Network structure and performance', Working Paper, Queen Mary University of London.
Mieth, L., Bell, R. and Buchner, A. (2016). 'Facial likability and smiling enhance cooperation, but have no direct effect on moralistic punishment', *Experimental Psychology*, vol. 63, pp. 263–77.
Mobius, M.M. and Rosenblat, T.S. (2006). 'Why beauty matters', *American Economic Review*, vol. 96, pp. 222–35.
Mulford, M., Orbell, J., Shatto, C. and Stockard, J. (1998). 'Physical attractiveness, opportunity, and success in everyday exchange', *American Journal of Sociology*, vol. 103, pp. 1565–92.
Niederle, M. and Vesterlund, L. (2007). 'Do women shy away from competition? Do men compete too much?', *Quarterly Journal of Economics*, vol. 122, pp. 1067–101.
Pope, D.G., Price, J. and Wolfers, J. (2018). 'Awareness reduces racial bias', *Management Science*, vol. 64, pp. 4988–95.
Ravina, E. (2019). 'Love & loans—the effect of beauty and personal characteristics in credit markets', Working Paper, Kellogg School of Management.
Simon, H.A. (1947). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*, Macmillan.
Stirrat, M. and Perrett, D.I. (2010). 'Valid facial cues to cooperation and trust: male facial width and trustworthiness', *Psychological Science*, vol. 21, pp. 349–54.
Stuhlmacher, A.F. and Walters, A.E. (1999). 'Gender differences in negotiation outcome: a meta-analysis', *Personnel Psychology*, vol. 52, pp. 653–77.
Takahashi, C., Yamagishi, T., Tanida, S., Kiyonari, T. and Kanazawa, S. (2006). 'Attractiveness and cooperation in social exchange', *Evolutionary Psychology*, vol. 4, pp. 315–29.
Wilson, R.K. and Eckel, C.C. (2006). 'Judging a book by its cover: beauty and expectations in the trust game', *Political Research Quarterly*, vol. 59, pp. 189–202.