

Gürerk, Özgür; Irlenbusch, Bernd; Rockenbach, Bettina

Conference Paper

Endogenously Emerging Gender Diversity in an Experimental Team Work Setting

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Demography and Gender, No. D10-V3

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Gürerk, Özgür; Irlenbusch, Bernd; Rockenbach, Bettina (2017) : Endogenously Emerging Gender Diversity in an Experimental Team Work Setting, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Demography and Gender, No. D10-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/168067>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Endogenously Emerging Gender Diversity in an Experimental Team Work Setting

Özgür Gülerk, RWTH Aachen University

Bernd Irlenbusch, University of Cologne

Bettina Rockenbach, University of Cologne

This version: Jan 30, 2017

Abstract:

We study gender diversity and performance in *endogenously* formed teams in a repeated work setting. Participants can choose whether to perform a cooperation task only with members of the own gender or in a mixed-gender team. We find that independent of the team choice, in the initial periods, men contribute significantly more to the team projects than women. This induces men to prefer the successful men-only teams in the subsequent periods, resulting in significantly higher profits for men compared to women. Only over time, this endogenously emerged “gender profit gap” closes.

Keywords: gender differences, endogenously formed teams, cooperation, punishment

JEL classification: C92, J71, M54

1. Introduction

“Gender equality is a moral and a business imperative.” This succinct statement opening the blurb of Bohnet’s (2016) recent book summarizes a widely shared view, not only in the business world but also in most western societies. Yet, the appropriate means to achieve gender equality at the workplace are not at all trivial and have many facets. Bohnet’s book presents an impressive collection of different (institutional) design features, based on behavioral insights that help overcoming biases and stereotypes. In addition to these (in part soft) interventions to promote gender equality, many countries have embarked in more radical and schematic interventions by implementing quotas for female representation in management boards and committees. The implementation of quotas addresses the moral facet of gender equality, independent of its (direct) economic consequences. The economic consequences of gender diversity of work teams have been studied in the field, as well as in laboratory experimental contexts. Ahern and Dittmar (2012) use data from the natural field experiment of implementing gender quotas in Norway and report a decline in stock prices and operating profits after the implementation of the gender quota in boards, and attribute this to less CEO experience of the newly appointed female directors. In a field experiment, Hoogendoorn et al. (2013) report that teams with an equal gender mix perform better than male-dominated teams in terms of sales and profits. Azmat and Petrongolo (2014) present an overview over laboratory as well as field experiments on the economic consequences of gender diversity in work teams. While they report a number of relevant findings, they point to the important research issue that team compositions in the workplace generally evolve endogenously¹: *“One of the main problems with studying gender and groups is that groups are typically formed in an endogenous way. While experiments can go some way to solve this issue through random assignment into groups, they tend to create an artificial environment, in which it becomes difficult to distinguish group diversity and group dynamics.”* (p. 38).

In this paper, we address the issue of gender diversity and team performance in *endogenously* formed teams. In a laboratory experiment, male and female participants choose whether to perform a task in a team with members of the same (own) gender (*women-only* and *men-only*) or whether to perform the task in a mixed-gender team. Repeated choice with feedback on the gender composition and the performance of these teams allows studying the dynamics over

¹ Endogenously formed teams have rarely been studied with regard to gender diversity. In this respect, we are only aware of an attitude survey conducted by Chatman and O'Reilly (2004) who report that women express a greater likelihood of leaving homogeneous groups than do men in the clothing and retail industries.

time. The experimental task asks for individually costly contributions to a team project that benefits all team members equally, independent of the individual contributions. Our experimental setup allows addressing the question of preferences of women and men for same- or mixed-gender teams, in their initial choice as well as over time. Moreover, we are able to investigate whether and how the gender composition of a team influences women's and men's contributions and payoffs. This allows investigating whether endogenously selected same-gender teams perform differently compared to endogenously selected mixed-gender teams.

Remarkably, we find an endogenously occurring “gender profit gap”, primarily driven by gender differences in the initial behavior. In the first round, men exhibit a clear preference for the mixed-gender team, while women seem to be indifferent between the mixed-gender and the women-only team. Independent of the team choice, men contribute significantly more to the team project than women. These initial differences seem to induce men to predominantly choose the more successful men-only team in the further rounds, resulting in significantly higher profits for men compared to women. It takes until the second half of the thirty repetitions, until women “recover” from the disadvantage of the low initial contributions and the “gender profit gap” closes.

2. Experimental Design and Procedures

2.1 Experimental Design

Our experiment is based on a three-stage game with twelve players, six of them female and six male. Each group of twelve players plays the game repeatedly over 30 periods in a partner matching. These groups constitute the independent experimental observation. The course of the game is as follows:

Stage 1: Team Choice Stage

In the team choice stage, each of the six female and the six male players decides on the team membership. While women can join either the women-only team, abbreviated as the *W-team* or the mixed *WM-team*, men can join either the men-only *M-team* or the mixed *WM-team*. This means, in any given period, a maximum of 12 players could join the *WM-Team*, while the maximum possible number of players, each in the *W-team* and the *M-team* is six. After all players made their team choice, and before the contribution stage starts, each player learns how many individuals are in each team. Players also learn how many men and women joined the *WM-team*.

Stage 2: Contribution Stage

In the contribution stage, each player receives an endowment of 20 tokens. If there is only one player in a team, then this team cannot play a team game and the single player's endowment is added automatically to the private account. If a team has two or more members, then a team game is played within the own team, independently of the other teams. The team game simulates team production with conflicting individual and collective interests through a public goods game. Team members decide how many of their 20 tokens to invest into the team account. Each token invested creates a benefit of 1.6 for the entire team, which is then equally distributed among the team members. Thus, investing into the team account benefits the entire team (the productivity of each token invested into the team account is larger than 1). Yet, the individual return (marginal per capita return MPCR) on the invested token is lower than 1 ($\text{MPCR} = 1.6 \text{ divided by the number of team members}$). This means that for each player it is individually and myopically more profitable to save the token in the private account (with a private return of 1 token), rather than putting it into the team account (with a private return of less than 1 token). From the collective perspective of the team, however, contributions increase efficiency. The discrepancy in the benefits of investment between the team and the individual constitutes the conflict in individual and collective interests.

We decided to hold the productivity (return) of 1.6 constant for each contributed token, independent of the number of members in a team. This design choice eliminates the possible motivation for joining a larger team just because of pure efficiency reasons. Therefore, the MPCR in our game varies with the number of players in a team. For a player, who is in a team with only one other player, we chose the MPCR to be equal to 0.8 ($1.6/2$), while it linearly decreases when the team grows. In a team with 12 members, it amounts to 0.13 ($1.6/12$).²

Stage 3: Punishment Stage

In the punishment stage, each player receives an additional amount of 20 tokens. If there is only one player in a team, there is no action to be taken for this player. If there are two players or more, each player learns the individual contributions of all other members of the own team and their genders. The individual data of the fellow members is presented to the players in a table. The entries in the table are anonymous (without any subject ids) and placed randomly, meaning

² Previous studies that we have conducted share the basic design with the current study, in particular with respect to varying the MPCR (see e.g., Gülerk et al. 2006, 2014). In these studies, we observe large teams emerging, which means that MPCRs endogenously emerge to be small.

that it is not possible to track a team member's contributions over periods. Each player can allocate the 20 tokens at her discretion to punish others in the team or save the tokens in the private account. We apply a commonly used experimental punishment technology with 1:3 effectiveness. This means, each allocated punishment token costs the punishing player one token, and it reduces the payoff of the punished player by three tokens. Losses are theoretically possible. To reduce the risk of bankruptcy, we provide each player with an additional endowment of 400 tokens at the start of the experiment. Each player can allocate punishment tokens to one or several other team members. The sum of allocated tokens in one period could not exceed 20 tokens.

2.2 Theoretical Considerations

Nash equilibrium with myopic and selfish preferences

If we consider the stage game assuming that each player is a money maximizer with myopic, selfish preferences, then no player has an incentive to punish any other player since punishment is costly. By backward induction, this implies that no player contributes to the public good in the first stage. Since the structure is the same in all teams, a player is indifferent between choosing the team with only own-gender or the mixed-gender team. The payoff of a player then equals to 40 tokens. Since the number of periods is finite, and the players know this, by backward induction the predicted behavior of the last period unravels to all periods.³

Socially optimal behavior

If one assumes that there are more than one player in a team, payoffs would be maximized when all players contribute fully to the public good, and no player punishes any other player. In this case, the payoff of each player amounts to 52 tokens, regardless whether the number of the members of a team is two or more.

2.3 Hypotheses on Gender Differences

The rapidly growing literature on gender-related characteristics in economic environments reveals many cases with inconclusive gender differences (see Croson and Gneezy (2009) for an excellent overview, and Balliet et al. (2011) for a large meta-study on cooperation and gender

³ From previous experimental studies we know that even in finitely repeated games players show a behavior that is more in line with predictions from an infinitely repeated game. Thus, the threats of no cooperation in the future, punishment, or leaving the team might lead to more cooperative behaviour. There are, however, no obvious differences for our three teams with respect to these patterns of behaviour.

in exogenously formed groups). One of the robust findings seems to be that women are more responsive to the experimental conditions, in particular when they know the gender of the counterparts. Additionally, there is evidence that both genders prefer to interact with women. In a 2-person winner-takes-it-all competition, for example, Geraldles (2016) shows that 65% of men and 80% of women choose to be paired with a woman. The preference to be paired with a woman does not depend on the deciding subject's "confidence-level" on the task, for neither gender. These insights lead to hypothesis 1.

Hypothesis 1:

- a. *Women tend to prefer the W-team to the WM-team.*
- b. *Men tend to prefer the WM-team to the M-team.*

Dufwenberg and Muren (2006) study the influence of gender composition when a group of three dictators allocated money between them and a fourth person. They find that groups are more generous and equalitarian when the dictator group has a majority of women. In a business game, Apesteguia et al. (2012) study how the gender composition of teams affects their economic performance. While they find that teams formed by three women are significantly outperformed by any other gender combination, they attribute it to women teams being less aggressive and investing more in social sustainability initiatives. Chatman and O'Reilly (2004) report that women are more cooperative in all-female work groups. Together with the finding that women seem to be more responsive to the experimental conditions, especially to knowing the gender of their counterpart (Croson and Gneezy, 2009), we formulate hypothesis 2a.

Charness and Rustichini (2011) report that in a 2-person prisoners' dilemma game with an audience, men are more cooperative when (passive) women are watching than when (passive) men are watching. These finding nurture our hypothesis 2b.

Hypothesis 2:

- a. *Women contribute more in the W-team than in the WM-team.*
- b. *Men contribute more in the WM-team than in the M-team.*

Nowell and Tinkler (1994) report that exogenously formed, all-female groups are more cooperative than either exogenously formed, all-male or exogenously formed, mixed-gender groups. There is, however, also evidence pointing in the opposite direction (e.g., Brown-Kruse and Hummels 1993) or reporting no difference at all (Mason et al., 1991). Due to this

inconclusive findings, we abstain from formulating a hypothesis on which team (women-only, men-only, or mixed) is more cooperative and treat this as an explorative question.

2.4 Experimental Procedures

The experiments were programmed with z-Tree (Fischbacher, 2007). The experimental sessions were conducted at the experimental laboratory of the University of Erfurt (Germany) in December 2011 and June 2012. Each experimental session comprised two independent observational groups with 12 subjects each. In each group, six males and six females participated. Subjects were invited with the recruitment software ORSEE (Greiner, 2015) and were randomly matched to groups. Most participants were undergraduate students of social sciences. None of them participated in a similar experiment before. In total, 168 subjects participated in 14 independent experimental groups.

As the participants arrived in the laboratory, they were seated in separate cabins and received a copy of the experimental instructions (see Appendix B). To assure common information on the experimental setup, the instructor read aloud the instructions. The same instructor was present in each session. After instructions were read, the participants were allowed to ask clarifying questions on the rules of the game. On average, an experimental session took 2 hours. Participants were paid privately. The conversion rate from tokens to cash was publicly announced in the instructions. Average pay was 21 Euro.

3. Results

In this section, we present our experimental results. We do this in two steps. First, we look at behavior in the first period, where players' team choices cannot be influenced by other players' decisions and players have no information about what others contribute, when they decide about their own contribution. Second, we look at behavior and the dynamics in later periods.

3.1 First Period Behavior

Team Choices

In the first period, women appear to be rather indifferent between their two team options. In total, 56.0% of the women choose the *WM-team*, compared to 44.0% who choose the *W-team*. A clear majority of men prefer the mixed-gender team in the first period: 70.2% of the men choose the *WM-team*, compared to 29.8% who choose the men-only *M-team* ($p < 0.001$,

Whitney-Mann U test)⁴. While the second observation is in line with our hypothesis 1, the first is not.

Result “Initial Team Choice”: *In the first period, about half of the women choose the mixed-gender team while a clear majority of men chooses this team.*

Contributions, Punishment and Pay

Figure 1 shows average contributions, pay (profit), and sent and received punishments in the first period, for both genders and all teams. Over all teams, in period 1, women contribute on average 8.1 which is significantly less than men’s contributions of 11.4 ($p = 0.026$, Wilcoxon matched-pairs test)⁵. While women predominantly contribute 5 and 10, men predominantly contribute 15 or 20. The same is true if one focuses on the *WM-team*, where men contribute on average 10.9 while women contribute on average 7.6 which is significantly less ($p = 0.027$, WMPT, see Figure 1, panel a). Contributions of women do not significantly differ between the *W-team* and the *WM-team* ($p = 0.508$, WMPT). The same is true for men in the *M-team* and the *WM-team* ($p = 0.860$, WMPT). Thus, we do not find support for hypothesis 2 that contribution behavior of women and men differ between teams.

Result “Initial Contributions”: *In the first period, men contribute more than women do. Contributions of women do not differ between the two teams. The same is true for men.*

As can be seen from panels c) and d) in Figure 1, average punishment points sent and received per subject tend to be lower in the same-gender teams, compared to the mixed gender teams. Indeed, women in the *WM-team* receive significantly more punishment points compared to women in the *W-team* ($p = 0.047$, WMPT) and men in the *WM-team* send weakly significantly more punishment points than men do in the *M-team* ($p = 0.051$, WMPT).

Result “Initial Punishment”: *In the first period women are punished more in the mixed-gender team than in the women only team.*

⁴ All tests in the result section report two-tailed significances. Except the very first test on the first period team choice, all following within and across team comparisons are tested with Wilcoxon matched-pairs test, because of the interdependency of the decisions, within and across teams. To save space, we abbreviate the Wilcoxon matched-pairs test with WMPT.

⁵ This finding is in contrast to Ortmann and Tichy (1999), who report that women cooperate significantly more than men do in the first round of a repeated, binary choice prisoner’s dilemma game in exogenously formed two-player teams. This difference disappears by the last period.

In the first period women in the *W-team* earn significantly less than men in the *M-team* (women earn 38.6 and men earn 43.9; $p = 0.028$, WMPT, see Figure 1, panel b).⁶ Overall in the first period women earn 35.4 on average while men earn 38.8 ($p = 0.096$, WMPT). Women in the *WM-team* earn significantly less than women in the *W-team* ($p = 0.023$, WMPT). The main reason is that the relatively low contributing women in the *WM-team* are punished more by the high-contributing men compared to the punishment women receive by other women in the *W-team*. Also men in the *WM-team* earn significantly less than men in the *M-team* ($p = 0.002$, WMPT). The main reason here is that fellow women in the *WM-team* tend to contribute less than fellow men in the *M-team*. A second reason is that men in the *WM-team* also tend to contribute more to the second order public good of costly punishment than men in the *M-team*.

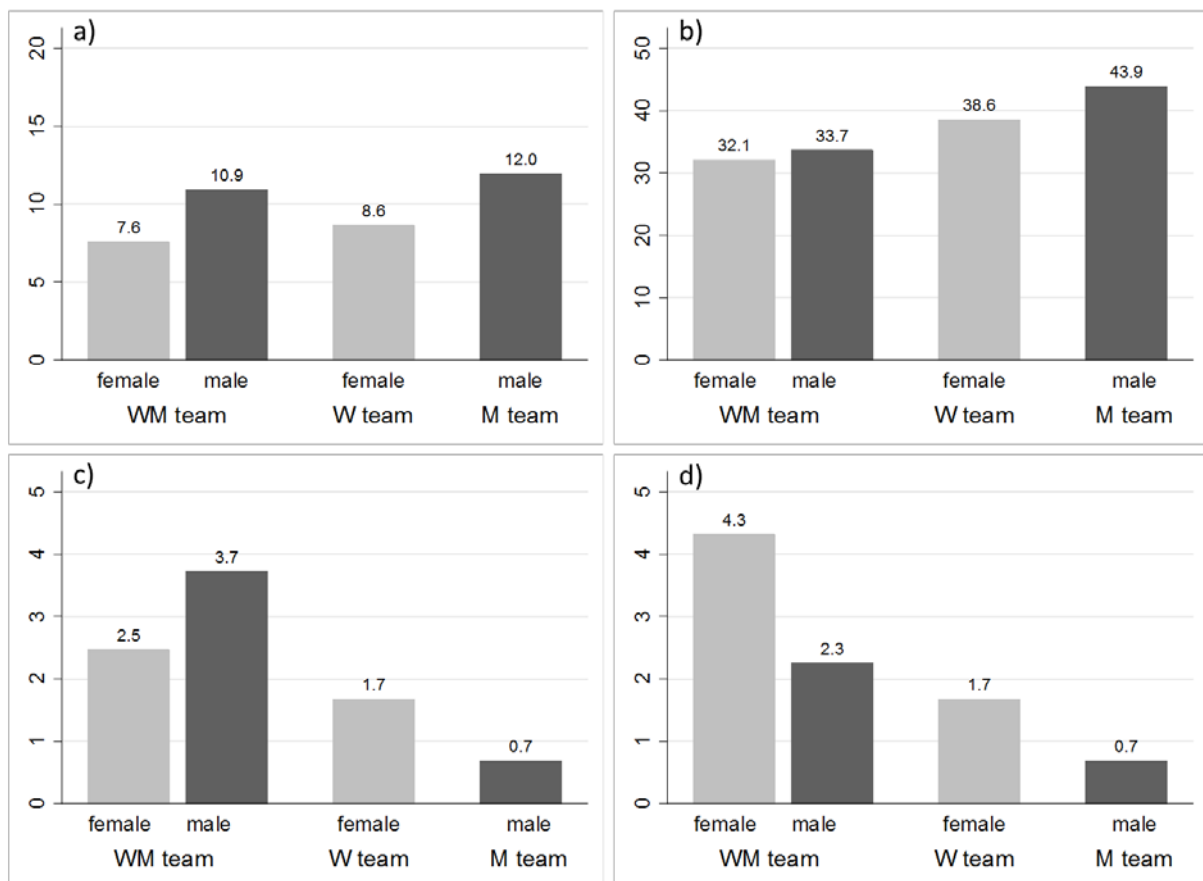


Figure 1: Average a) contributions, b) pay, c) sent punishment points, and d) received punishment points in the first period for each of the teams, respectively.

Result “Initial Pay”: *In the first period, women earn more in the women-only team than in the mixed-gender team. The analogous result is also true for men. Women earn less in their women-*

⁶ This is in line with Rapoport and Chammah (1965), who report that male-male interactions are more cooperative than female-female interactions.

only team than men do in their men-only team. The same tends to be the case when averaging over all teams.

3.2 Behavior over Time

Team Choice Dynamics

The bars in Figure 2 show the team choices of both genders. Panel a) provides the average number of women (grey bars) and men (black bars) in the *WM-team* over periods. The grey bars in panel b) show the corresponding average number of women in the *W-team* and the black bars in panel c) show the corresponding average number of men in the *M-team*. As discussed above, and as can be seen from all three panels, many players choose the *WM-team* in period 1, but most of them are men. In periods two and three, both genders tend to move to the same-gender teams, respectively. Thus, in the third period, about 60% of each gender choose the exclusive-gender teams, respectively. This basically does not change for men in the remaining 27 periods. Women, however, more and more move to the mixed-gender team. In period 10, about 60% of the women choose the *WM-team*. This percentage tends to be stable for the next 15 periods after which more women start to move to the *W-team*, so that in period 30 about 50% of the women stay in the *W-team*, and 50% of the women stay in the *WM-team*, respectively. Overall, a woman switches on average 3.7 times from one of the two teams to the other while a man does so only 2.8 times ($p = 0.068$, WMPT). This observation is in line with the impression from Figure 2 of a relatively stable high number of men in the *M-team*.

As one might expect, being punished increases the likelihood of switching to the other team. In a probit regression with a binary dependent variable *change_team*, denoting a switch from one team to the other, the binary (independent) variable *being_punished* turns out to be highly significant (see regression (a) in Table 1). Women and men tend not to behave differently in this respect, as suggested by the non-significant interaction variable *being_punished_men*. The regression (a) also confirms the impression mentioned above that men switch less often than women do, as indicated by the negatively significant indicator variable *men*. The men's likelihood of switching to the other team tends to increase, if the average contribution in the last period was higher in the other team compared to the own one (cf. the variable *contr_higher_in_other_team_men*).

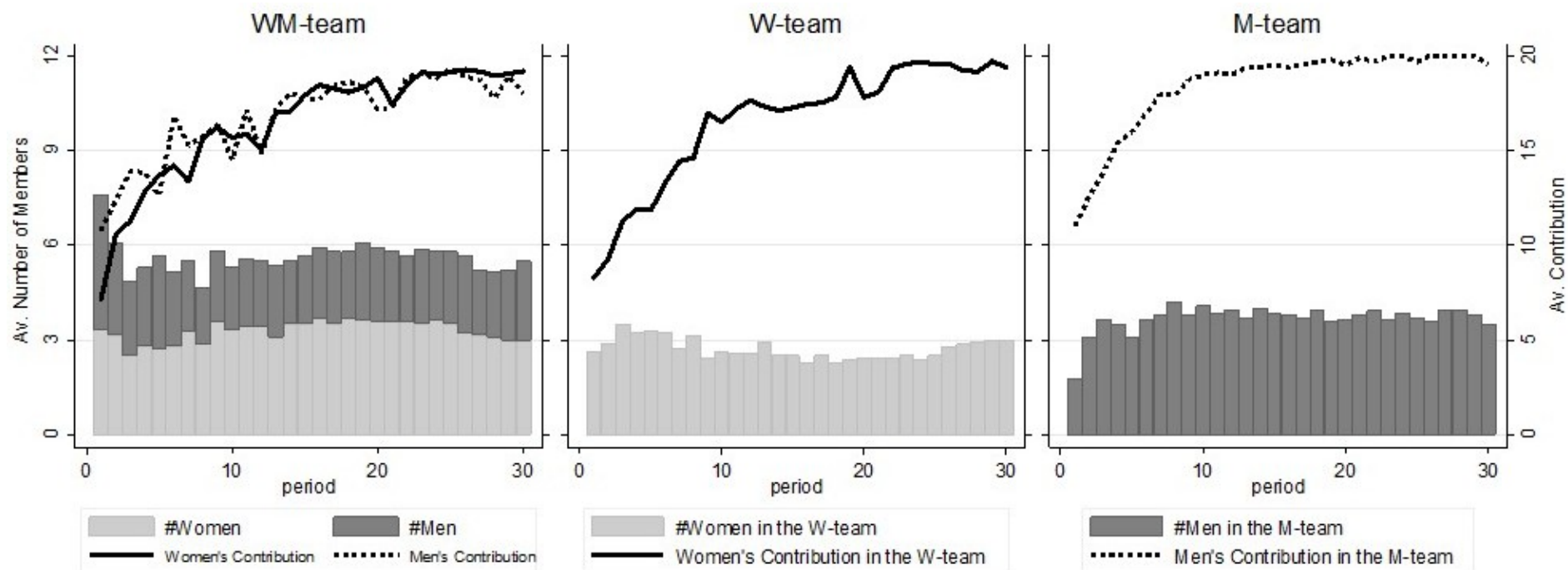


Figure 2: Average numbers (bars) and average contributions (lines) of women and men in each period. Bars in panel a) provide the average number of women (grey bars) and men (black bars) in the WM-team over periods. The grey bars in panel b) show the corresponding average number of women in the W-team and the black bars in panel c) show the corresponding average number of men in the M-team.

Table 1: Regression results on (a) team change behavior, (b) conditional cooperation, (c) sent punishment, and (d) reaction to punishment

	(a) Probit Regression DepVar = change_team	(b) Tobit Regression DepVar = contribution	(c) Tobit Regression DepVar = sent_punishment	(d) Tobit Regression DepVar = deltacontr
being_punished	1.030 (0.100)***			
contr_higher_in_other_team	0.210 (0.135)			
Men	-0.264 (0.124)**	1.102 (2.411)	0.074 (0.294)	-0.243 (0.347)
being_punished_men	-0.102 (0.085)			
contr_higher_in_other_team_men	0.341 (0.204)*			
others_contr_prev_period_own_team		1.272 (0.108)***		
others_contr_prev_period_own_team_men		-0.026 (0.137)		
delta_to_other_contr			0.121 (0.026)***	
delta_to_other_contr_men			0.069 (0.073)	
received_punishment_prev_period				0.321 (0.054)***
received_punishment_prev_period_men				0.104 (0.146)
Constant	-1.501 (0.136)***	-0.592 (1.125)	0.994 (0.099)***	0.576 (0.189)***
N	4763	4749	1333	931
Std. Err. adjusted for 14 cluster groups	Yes	yes	Yes	Yes

Panels employ the following models: (a) Random effects probit model (*change_team* ‘yes’ = 1 or ‘no’ = 0), (b) Tobit model with *contribution* as the dependent variable, (c) Tobit model with *sent_punishment* as the dependent variable, and (d) Tobit model with difference of contributions in period *t* and contribution in period *t* – 1. All models are clustered by independent observation groups (including six women and six men each, respectively).

If we focus on the *WM-team*, women are significantly more likely to leave when being punished than when they are not punished (37.2% compared to 17.0%, $p=0.007$, WMPT). Men also leave the *WM-team* when being punished (32.6 %), but not significantly more often than when they are not punished (27.8%, $p = 0.340$, WMPT). As suggested by the regression result above, one

reason for this observation might be that men predominantly leave the *WM-team* because of the higher contributions in the *M-team*. Thus, team choice behavior over time does not support our hypothesis 1.

Result “Team Choice over Time”: *Over all periods, women tend to relatively often switch between teams and exhibit a slight preference for the mixed-gender team. Men switch less often and tend to prefer the men-only team.*

Contributions

As in period 1, also averaged over all periods and teams, women contribute significantly less than men do (women contribute 16.9 on average and men contribute 18.1 on average; $p = 0.022$, WMPT)⁷. The difference in contributions between genders is mainly due to contribution differences in the same-gender teams. Averaged over all periods, women contribute 17.0 in *W-teams*, while men’s average contributions in *M-teams* (18.5) are significantly higher ($p = 0.036$, WMPT). The difference in contributions between women and men is particularly pronounced during the first half of the periods 1-15 (on average, women contribute 14.7, and men 16.8; $p = 0.013$). In the second half, the difference in contributions is not significant anymore (women contribute 19.0 and men contribute 19.4; $p = 0.572$, WMPT). The development of contribution differences is shown in more detail in Figure 3, panel a).

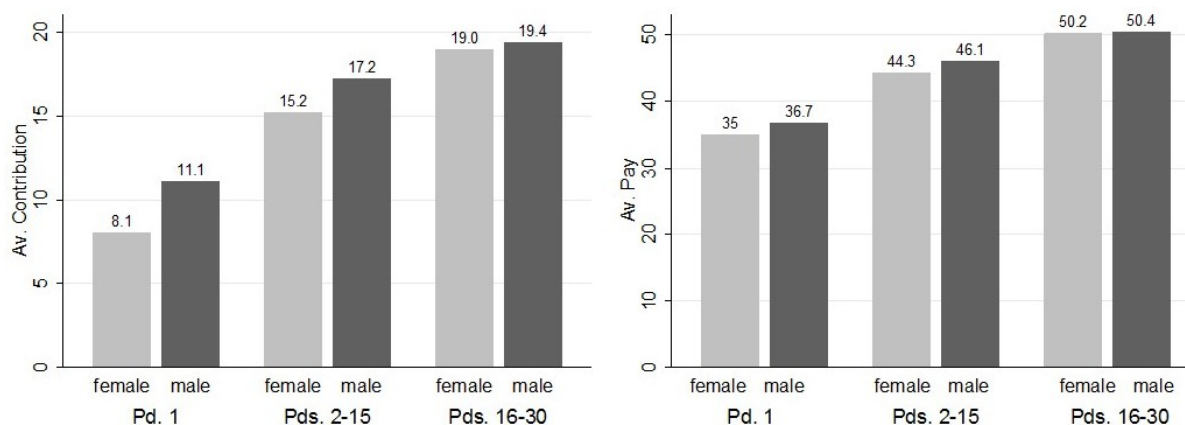


Figure 3: Average contributions (panel a) and pay (panel b) of women (grey bars) and men (black bars) over periods. The first pair of bars shows averages in the first period. The second

⁷ For a visual comparison of average a) contributions, b) pay, c) sent punishment points, and d) received punishment points over all periods for each of the teams, respectively, see Figure A.1.

pair of bars shows averages over periods 2-15 and the third pair of bars shows averages over periods 16-30, respectively.

Thus, contributions of both genders tend to converge towards each other over periods. It is also true that on average, women and men exhibit a similar extent of conditional cooperation. The regression b) in Table 1 explaining contributions in period t shows that they are highly correlated with the average respective contributions of others in the same team in period $t - 1$ (cf. the variable *others_contr_prev_period_own_team*). This suggests that participants are conditionally cooperative. In fact, both genders seem to be conditionally cooperative to a similar extent since the interaction variable *others_contr_prev_period_own_team_men* turns out not to be significant. Again also over time we do not find differences in contributions between teams, neither for women nor for men, which is not in line with our hypothesis 2.

Result “Contributions over Time”: *Over all periods, men contribute more than women do, but the difference in contributions disappears over periods. Women and men are quite similar in their extent of behaving conditionally cooperative.*

Punishment

Overall, women on average send 0.7 punishment points while men send 0.8 points. Women receive 0.8 punishment points, and men receive 0.6 punishment points. Both differences are not significant ($p = 0.397$, and $p = 0.158$, respectively, WMPT). Women send and receive only slightly more punishment points (on average 0.6) in *W-teams*, compared to men do in *M-teams* (on average 0.5). There is, however, a difference in *WM-teams*, where men send on average significantly more punishment points than women (women send 0.7 on average and men send 1.2 on average; $p = 0.004$, WMPT).

The Tobit regression model shown in panel c in Table 1 explains the number of sent punishment points as dependent variable, with the variable *delta_to_other_contr*, as the difference between the contribution of the punisher and the contribution of the punished subject in the current period. The corresponding coefficient turns out to be highly significant indicating a positive correlation between the difference in contributions and exerted punishment. In this respect, there seems to be no difference between women and men as the interaction variable *delta_to_other_contr_men* is not significant.⁸

⁸ Analogous Tobit regressions restricted to the punisher being a woman or a man, respectively, show that neither women nor men discriminate between whether the target is a woman or a man (see Table A1 in the appendix).

Received punishment points in *WM-teams* are higher for women (on average 1.0) compared to men (on average 0.8), however, the difference being only weakly significant ($p = 0.084$, WMPT). Panel d in Table 1 reports the results of a Tobit regression model explaining the change in contributions from period $t - 1$ to period t by the amount of punishment received in period $t - 1$, *received_punishment_prev_period*. Being punished seems to lead to an increase in contributions. This effect seems not to be different between genders as the non-significant interaction effect *received_punishment_prev_period_men* indicates.

Result “Punishment over Time”: *Over all periods and teams, sent punishment and received punishment tend to be similar for women and men. Also the reaction to punishment in terms of contribution change seems not to be different between women and men. Only in the mixed-gender teams, men tend to exert more punishment than women do.*

Pay

As in the first period, women’s average pay over all 30 periods is significantly lower than that of men (women earn 46.9 and men earn 47.9; $p = 0.026$, WMPT). This is primarily driven by the (weakly significantly) lower pay of women in the *W-teams* compared to men’s pay in the *M-teams* (on average, women earn 47.5 and men earn 49.1; $p = 0.084$, WMPT). Women in the *WM-teams* and in the *W-teams* earn non-significantly different amounts ($p = 0.925$, WMPT). The same is true when one compares pay for men in the *WM-teams* and *M-teams* ($p = 0.245$, WMPT). The difference in pay between genders comes into being in periods 1-15. Average pay of women in these periods over all teams is significantly lower than average pay of men (women earn 43.6 and men earn 45.4; $p = 0.019$, WMPT). This is not the case in the second half of the periods. There is no significant difference of pay averaged over periods 16-30 (women earn 50.2 and men earn 50.4; $p = 0.594$, WMPT).

Result “Pay over Time”: *Over all periods, women earn less than men. This is particularly true when comparing the same-gender teams. The difference in pay, however, disappears over periods.*

4. Summary and Conclusion

We study gender diversity and team performance in *endogenously* formed teams in a repeated team game. In each repetition, male and female participants can freely choose whether to perform a task in a team solely with members of the same (own) gender (*women-only* and *men-only*), or whether to perform the task in a mixed-gender team. We find an endogenously

occurring “gender profit gap”, driven by gender differences in the initial behavior. Independent of the team choice, men contribute significantly more to the team project than women⁹. These initial differences seem to induce men to predominantly choose the more successful men only team in the further periods, resulting in significantly higher profits for men compared to women. It takes until the second half of the thirty repetitions, until women “recover” from the disadvantage of the low initial contributions and the “gender profit gap” closes.

We provide evidence that these results cannot be explained by a lower cooperativeness of women per-se: men and women do not exhibit different degrees of conditional cooperation in response to the groups’ past contribution level, and they do also not exhibit differences in punishment behavior. In the first period, however, when a group’s cooperativeness is not known, cooperation is a matter of risk and trust. In their review paper, Croson and Gneezy (2009) discuss gender differences in trust games and conclude that “*women trust less than or the same as men in these settings. But women’s trust levels are more context-sensitive than those of men*” (p. 460), in line with Gilligan (1982). Simpson and Van Vugt (2009) describe two motivations for non-cooperation: greed and fear. Greed corresponds to the temptation to free-ride on others’ cooperation, while fear refers to the prospect that one’s cooperation may be exploited. They argue that from an evolutionary point, these motivations affect women and men differently. From an evolutionary perspective, women avoid taking too many risks to avoid being exploited in social interactions, while men have evolved a “high-risk-high-stakes” strategy. As a consequence, they argue that women’s non-cooperation is mainly due to the fear of being exploited by others, while men’s non-cooperation is motivated by greed. Our observation that men’s main motivation for switching the team seems to be a higher profit in the other team nicely fits into this picture. Even more importantly, this evolutionary perspective also fits to the observed low cooperativeness of women in the first period (out of fear of being exploited) and the lacking difference in conditional cooperation compared between women and men in later rounds, when experience allows to better calibrate the risk of being exploited.

Our results point to the eminent importance to identify contexts, that encourage women to be initially as cooperative as men, since the low initial contributions have detrimental effects for women’s profits which only slowly improve. Our data suggest that creating a “reserved space for women”, i.e., creating a context ensuring that women interact with other women only, is not

⁹ The question whether women or men contribute more in a public goods game provided inconclusive answers. While some studies find – as we do – men to be more cooperative than women (e.g., Brown-Kruse and Hummels 1993) other find the opposite (e.g., Charness and Villeval 2009) or report no differences (e.g., Bolton and Katok 1995, Eckel and Grossman 2008). See Croson and Gneezy (2009) for an extensive discussion on this issue.

sufficient. It seems that more rigorous measures have to be taken to initially foster women's trust and remove the initial gender contribution difference. In this respect, our results point to a new direction for future avenues in gender research.

References

- Ahern, K. R., Dittmar, A. K., 2012. The changing of the boards: The impact on firm valuation of mandated female board representation, *Quarterly Journal of Economics*, 127(1): 137-197.
- Apesteguia, J., Azmat, G., Iriberry, N., 2012. The impact of gender composition on team performance and decision making: Evidence from the field, *Management Science*, 58(1), 78-93.
- Azmat, G., Petrongolo, B., 2014. Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics*, 30: 32-40.
- Balliet, D., Li, N.P., Macfarlan, S.J., Van Vugt, S.J., 2011. Sex differences in cooperation: A meta-analytic review of social dilemmas, *Psychological Bulletin*, 137(6), 881–909.
- Berge, L. I. O., Juniwaty, K. S., Sekei, L. H., 2016. Gender composition and group dynamics: Evidence from a laboratory experiment with microfinance clients, *Journal of Economic Behavior & Organization*, 131: 1-20.
- Bohnet, I. 2016. What works: Gender equality by design, Harvard University Press.
- Bolton, G., Katok, E., 1995. An experimental test for gender differences in beneficent behavior, *Economics Letters*, 48(3-4), 287-292.
- Brown-Kruse, J., Hummels, D., 1993. Gender effects in laboratory public goods contribution: Do individuals put their money where their mouth is? *Journal of Economic Behavior & Organization*, 22(3), 255-267.
- Chatman, J.A., O' Reilly, C.A., 2004. Asymmetric reactions to work group sex diversity among men and women. *Academy of Management Journal*, 47(2), 193-208.
- Charness, G., Villeval, M-C., 2009. Cooperation and competition in intergenerational experiments in the field and the laboratory, *American Economic Review*, 99(3), 956-978.
- Charness, G., Rustichini, A., 2011. Gender differences in cooperation with group membership, *Games and Economic Behavior*, 72, 77-85.
- Croson, R., Gneezy, U., 2009. Gender differences in preferences, *Journal of Economic Literature*, 47(2): 448-474.
- Dufwenberg, M., Muren, A., 2006. Generosity, anonymity, gender", *Journal of Economic Behavior & Organization*, 61(1), 42-49.

- Eckel, C.C., Grossman, P.J., 2008. Men, women and risk aversion: Experimental evidence, *Handbook of Experimental Economics Results*, 1, 1061-1073.
- Fischbacher, U. 2007. z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics*, 10(2): 171-178.
- Geraldes, D. 2016. Women dislike competing against men, discussion paper.
- Gilligan, C. 1982. In a different voice, Harvard University Press.
- Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE, *Journal of the Economic Science Association*, 1(1): 114-125.
- Gürrer, Ö., Irlenbusch, B., Rockenbach, B. 2006. The competitive advantage of sanctioning institutions, *Science*, 312, 5770, 108-111.
- Gürrer, Ö., Irlenbusch, B., Rockenbach, B. 2014. On cooperation in open communities, *Journal of Public Economics*, 120: 220-230.
- Hoogendoorn, S., Oosterbeek, H., Van Praag, M., 2013. The impact of gender diversity on the performance of business teams: Evidence from a field experiment, *Management Science*, 59(7), 1514-1528.
- Mason, C.F., Phillips, O.R., Redington, D.B. 1991, The role of gender in a non-cooperative game, *Journal of Economic Behavior & Organization*, 15(2), 215-235.
- Nowell, C., Tinkler, S., 1994. The influence of gender on the provision of a public good, *Journal of Economic Behavior & Organization*, 25(1), 25-36.
- Ortmann, A., Tichy, L.K., 1999. Gender differences in the laboratory: evidence from prisoner's dilemma games, *Journal of Economic Behavior & Organization*, 39(3), 327-339.
- Rapoport, A., Chammah, A. M., 1965. Sex differences in factors contributing to the level of cooperation in the prisoner's dilemma game. *Journal of Personality and Social Psychology*, 2, 831-838.
- Simpson, B., Van Vugt, M., 2009. Sex differences in cooperation: Integrating the evolutionary and social psychological perspectives. *Advances in Group Processes*, 26, 81-103.

Appendix A

Table A1: Regression results on sent punishment for (a) men and (b) women

	(a)	(b)
DepVar =	Tobit Regression	Tobit Regression
Sentpunishment	Senders = Men	Senders = Women
delta_to_other_contr	0.251 (0.117)**	0.121 (0.027)***
receiver_men	0.064 (0.543)	-0.104 (0.147)
delta_to_other_contr_receiver_men	-0.087 (0.099)	0.002 (0.020)
Constant	0.982 (0.571)*	1.014 (0.111)***
N	627	706
Std. Err. adjusted for 14 cluster groups	yes	yes

Panels employ the following models: Tobit model with sent punishment as the dependant variable for (a) men as senders and (b) women as senders; all models are clustered by independent observation groups (6 women and 6 men, respectively)

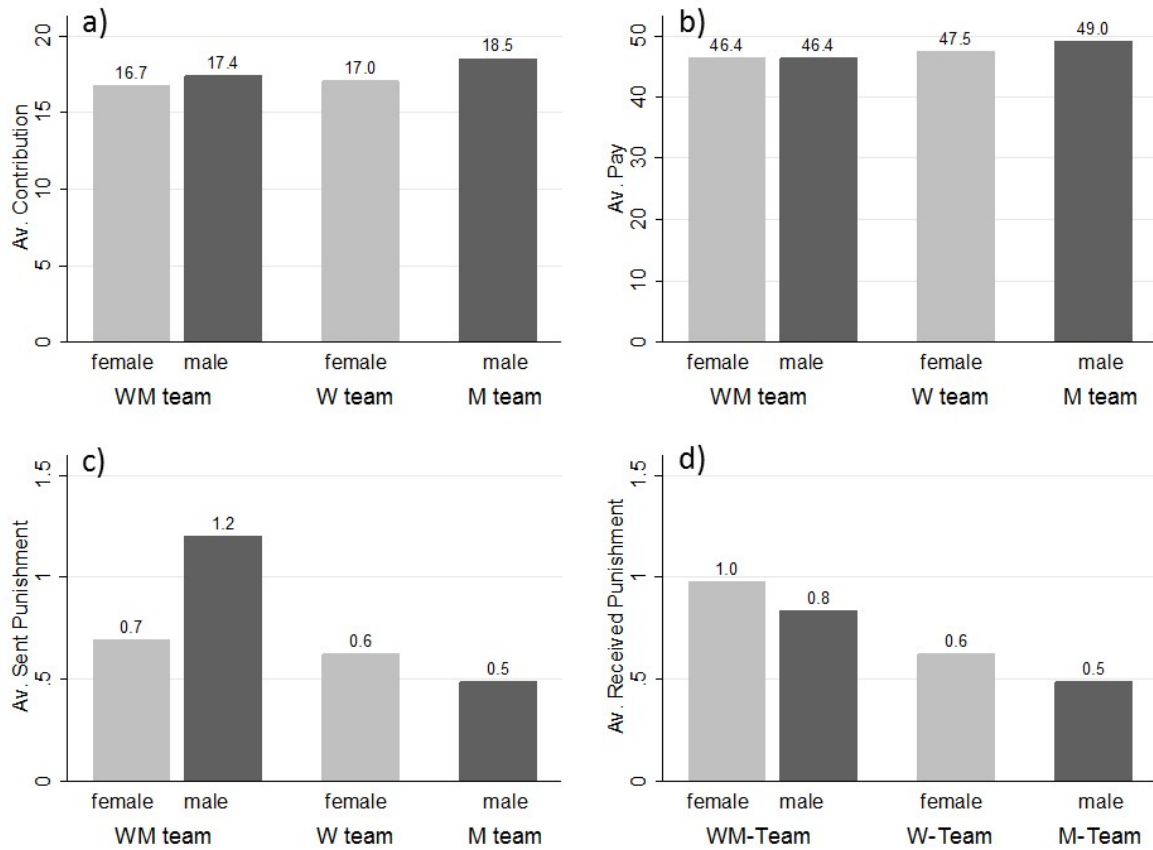


Figure A1: Average (a) contributions, (b) pay, (c) sent punishment points, and (d) received punishment points over all periods for each of the teams, respectively

Appendix B

Instructions to the Experiment (Original instructions were provided in German. They are available from the authors on request.)

General Information: At the beginning of the experiment, you will be randomly assigned to one of **two subpopulations, each consisting of 12 participants**. During the whole experiment, you will interact only with the members of your subpopulation. To each of the subpopulations, **6 men** and **6 women** will be assigned.

Course of Action: The experiment consists of **30 rounds**. Each round consists of two stages. In Stage 1, the group choice and the decision regarding the contribution to the project take place. In Stage 2, participants may influence the earnings of the other group members.

Stage 1 (i) The Group Choice: In Stage 1, each participant decides which group she wants to join. There are three different groups:

- **G-Group:** Women as well as men can enter this group.
- **F-Group:** Only women can enter this group.
- **M-Group:** Only men can enter this group.

Each participant can choose between two of these three groups. Men can choose between the M-Group and the G-Group. Women choose between the F-Group and the G-Group.

After the group choice has been completed, you will learn how many members each group have. For the G-Group, you will also learn how many women and men are in that group.

The screenshot shows a web-based experiment interface. At the top, a header bar contains the text 'Runde' followed by '1 von 2'. Below this, the main area is titled '1. Stufe: Die Wahl der Gruppen'. The central question is 'Welcher Gruppe möchten Sie beitreten?'. Below the question, there are two radio button options: 'Gruppe M (Männer)' and 'Gruppe G (Gemischt)'. To the left of these options is the label 'Gruppe'. At the bottom right of the choice area is a red 'OK' button.

(ii) Contributing to the Project: In stage 1 of each round, each group member is endowed with 20 tokens. You have to decide how many of the 20 tokens you are going to contribute to the project. The remaining tokens will be kept by you.

Your Earnings from the project:

$$= 1.6 \times \text{Sum of the Contributions of all Group Members} / \text{Number of Group Members}$$

For each group member, the earnings from the project are calculated according to this formula. Please note: Each group member receives the same earnings from the project, i.e., each group member benefits from **all** contributions to the project.

1 von 2	Verbleibende Zeit [sec]: 37
1. Stufe: Beitrag zum Projekt	Information
Bitte wählen Sie Ihren Beitrag zum Gemeinschaftsprojekt <input type="text"/>	Sie befinden sich in der Gruppe M (Männer) . Die Gruppe M hat 2 Mitglieder. Die Gruppe G hat 1 Mitglieder; davon 1 Männer und 0 Frauen. Die Gruppe F hat 3 Mitglieder. Jedes Gruppenmitglied hat eine Ausstattung von 20 .
OK	

Calculation of your payoff in stage 1: Your payoff in stage 1 consists of two components:

- **tokens you have kept** = endowment – your contribution to the project
- **earnings from the project** = $1.6 \times \text{sum of the contributions of all group members} / \text{number of group members}$

	Your Contribution to the Project	The Sum of Others' Contributions in Your Group	Number of Team Members in Your Group	Your Earnings from the Project	Your Earnings from Stage 1
Example 1	7	45	5	$45 \times 1,6/5 = 14,4$	$20 - 7 + 14,4 = 27,4$
Example 2	1	33	7	$33 \times 1,6/7 = 7,5$	$20 - 1 + 7,5 = 26,5$
Example 3	16	67	4	$67 \times 1,6/4 = 26,8$	$20 - 16 + 26,8 = 30,8$

Stage 2

Assignment of Tokens: In stage 2, it will be displayed how much each group member has contributed to the project. In the G-Group, you will also learn the gender of the respective member.

By assigning tokens, you can reduce the payoff of a group member or keep it unchanged.

In each round, each participant receives additional 20 tokens in stage 2. You have to decide how many of the 20 tokens you are going to assign to other group members. The remaining tokens are kept by you. You can check the costs of your token assignment by pressing the button *Calculation of Tokens*.

- Each **negative token** you assign to a group member **reduces her payoff by 3 tokens**.
- If you assign **0 tokens** to a group member, her **payoff won't change**.

Please note: Before each round, a display order will randomly be determined. Thus, it is not possible to identify any group member by her position on the displayed list throughout different rounds.

Runde

2 von 2

Verbleibende Zeit [sec]: 54

2. Stufe: Vergabe von Punkten

Information

Ergebnisse der anderen Gruppenmitglieder aus der 1. Stufe

Sie befinden sich in der Gruppe **G (Gemischt)**.

Die Gruppe **G** hat insgesamt **3** Mitglieder.

In der Spalte "**Geschlecht**" steht "**1**" für einen Mann und "**2**" für eine Frau.

Ihr Einkommen aus der 1. Stufe

Ihr Beitrag zum Gemeinschaftsprojekt	4
Summe der Beiträge aller Gruppenmitglieder zum Gemeinschaftsprojekt	8
Ihr Einkommen aus der 1. Stufe	20.3

In der 2. Stufe haben Sie **20** zusätzliche Punkte, die Sie entweder behalten oder zur Vergabe von Punkten verwenden können.

Drücken Sie den Knopf "Punkteberechnung", um ihre Punktevergabe zu überprüfen.

Die Summe Ihrer Punktevergabe -

Behaltene Punkte -

Punkteberechnung

OK

Geschlecht	Beitrag	Eink.1.Stufe	Negativpunkte
1	3	21.3	<input type="text"/>
2	1	23.3	<input type="text"/>

Calculation of your payoff in stage 2: Your payoff in stage 2 consists of two components:

- **tokens you have kept** = 20 – sum of the tokens that you have assigned to the other group members
- **less the threefold number of negative tokens** you obtained from other group members

Thus, **your payoff in Stage 2** amounts to:

20 – sum of the tokens that you assigned to other group members

– 3x (the number of tokens you obtained from other group members)

Calculation of your round payoff: Your round payoff is composed of

Your payoff from Stage 1	$20 - \text{your contribution to the project} + 1.6 \times \text{sum of the contributions of all group members} / \text{number of group members}$
+ Your payoff from Stage 2	$20 - \text{sum of the tokens you have assigned to other group members} - 3 \times (\text{number of tokens you obtained from other group members})$
= Your round payoff	

Special case: a single group member: If it happens that you are the only member in your group, you will receive 20 tokens in Stage 1 and 20 tokens in Stage 2, i.e., your round payoff will amount to 40. You do not have to take any action either on Stage 1 or on Stage 2.

Information at the end of the round: At the end of the round, a detailed overview will inform you of the results obtained in all groups. For each group member, you will learn the following about that person: His/her contribution to the project, payoff from Stage 1, assigned tokens (if applicable), received tokens (if applicable), payoff from Stage 2, round payoff, and for the G-Group you will learn the gender of the respective group member.

Rundenergebnisse der Mitglieder der Gruppe M				
	Beitrag	1	4	
	Eink.1.Stufe	23.0	20.0	
	Verg.Pkte	-1	-1	
	Erh.Neg.pkte	-1	-1	
	Rundeneink.	39.0	36.0	
Rundenergebnisse der Mitglieder der Gruppe G (Anmerkung zur Zeile Geschlecht: 1=Mann, 2=Frau)				
	Geschlecht	1		
	Beitrag	0		
	Eink.1.Stufe	20.0		
	Verg.Pkte	0		
	Erh.Neg.pkte	0		
	Rundeneink.	40.0		
Ihr Rundenergebnis (F)		Rundenergebnisse der anderen Mitglieder der Gruppe F		
Beitrag	5	3	2	
Eink. 1. Stufe	20.3	22.3	23.3	
Verg.Pkte	-3	-3	-3	
Erh.Neg.pkte	-3	-2	-4	
Rundeneink.	28.3	33.3	28.3	

History: Starting from the second round, at the beginning of a new round, you will be informed about the overview of the average results (as above) of all previous rounds.

Runde
2 von 2

Verbleibende Zeit [sec]: 0

1. Stufe: Die Wahl der Gruppen

Welcher Gruppe möchten Sie beitreten?

Gruppe

☐ Gruppe M (Männer)
☐ Gruppe G (Gemischt)

OK

Die in der Geschichte angegebenen Werte sind Durchschnittswerte aus den Gruppen.

Runde	M: Größe	M: Beitrag	M: Erh. Neg.pkte	M: Einkommen	G: Größe	G: Beitrag	G: Erh. Neg.pkte	G: Einkommen	F: Größe	F: Beitrag	F: Erh. Neg.pkte	F: Einkommen
1	2	2.5	1.0	37.5	1	0.0	0.0	40.0	3	3.3	3.0	30.0

Total Payoff: The total payoff from the experiment is composed of the starting capital of 400 tokens plus the sum of round payoffs from all 30 rounds. At the end of the experiment, your total payoff will be converted into Euro with an exchange rate of 1,15 €per 100 tokens.

Please notice: No communication is allowed during the whole experiment. If you have a question, please raise your hand out of the cabin. All decisions will be made anonymously, i.e., no other participant will be informed about the identity of anyone who has made a certain decision. Payment will be anonymous, too, i.e., no participant will find out what the payoff of another participant was.

We wish you every success!