Gender Differences in Team Work and Team Competition*

Radosveta Ivanova-Stenzel†
Humboldt University Berlin

Dorothea Kübler‡
Technical University Berlin and IZA, Bonn

July 9, 2009

Abstract

Diversity management has become part of the daily activities in many firms and organizations. When setting up team work, the question arises of how to group the employees into teams with respect to their gender. We conducted a real-effort experiment with wages based either on the team's performance, or on the outcome of a competition between teams. We find that performance does not simply depend on the incentive scheme, but rather on gender in conjunction with the incentive scheme. We observe a substantial gap between the performance of men and women, with men performing better than women, (i) when men and women are part of the same team and are paid according to joint output and (ii) when the competition is between teams of the same gender. This suggests that in team work, such combinations of incentive scheme and gender composition should be avoided, if the aim is to minimize the variability of performance. Moreover, the results show that there can be a tension between the objective to maximize overall performance and the potential goal of minimizing gender inequality.

Keywords: team incentives, gender, tournaments
JEL classification numbers: C72; C73; C91; D82.

---

*We are grateful to David Danz for his excellent advice. For helpful comments we thank Dirk Engelmann, Bernd Irlenbusch, Lydia Mechtenberg, Wieland Müller, Julia Schmid, participants at the GEW Meeting 2005 in Köln, the First Annual IZA Workshop on Behavioral and Organizational Economics 2005 in Bonn and the 5th Tilburg Workshop on Psychology and Economics 2006 as well as three anonymous referees, an associate editor and an editor of this journal. Financial support by the Deutsche Forschungsgemeinschaft (DFG) through SFB 649 "Economic Risk" is gratefully acknowledged.

†Department of Economics, Spandauer Str. 1, 10178 Berlin, Germany. E-mail: *ivanova@wiwi.hu-berlin.de*.

‡Department of Economics and Management, Straße des 17. Juni 135, 10623 Berlin, Germany. Email: *d.kuebler@ww.tu-berlin.de* (corresponding author).

1 Introduction

The difference in earnings between men and women is a well-known phenomenon that has been studied from many different perspectives. A number of factors have been identified as causing the earnings gap, such as differences in performance, differences in working hours and career paths, differences in pay for the same performance, lower average earnings for professions mainly exercised by women etc. However, even when restricting attention to one of these factors, many new questions arise. For example, differences in performance can have various reasons, such as education and ability, as well as the gender composition of the work place. The gender of one's colleagues may be one of the factors determining whether working conditions are experienced as hostile or as encouraging, as stimulating or stifling. We focus on this potential source of performance differentials between men and women.

We conducted a real-effort experiment to study the relevance of gender for performance in team work. We varied the gender composition of teams by employing teams consisting either of men or of women, as well as mixed teams. The participants in the experiment worked on a task alone, and afterwards the joint output of the team was determined. Thus, the subjects took decisions individually but their payoff depended on the performance of the whole team. The design reflects a situation where people work on their tasks separately, cannot communicate about intermediate results and cannot observe the effort level of the other team members during their work. This is the case when the team members perform different tasks for which they are specialists. The team members' performance in these different tasks then determines the joint output upon which individual wage payments are contingent.

From the perspective of economic theory (i.e., rational and self-interested subjects), the gender of the decision maker should not affect performance. And neither the gender of the other team members nor the gender of the competitors should matter. We hypothesize that in team work, however, a person's performance can be affected, for example, by the expectations he or she holds about the performance of the others in the team, whereby these expectations are correlated with gender. In addition, team work suffers from free-riding incentives which can be mitigated by competition, peer pressure and social norms, altruism, or loyalty among group members.[1] It is an open question how these motives are affected by the gender of the decision makers. We will discuss our results in the light of existing theories of gender roles and stereotypes, group conflict theory and evolutionary biology and sociobiology.

We designed the experiment to establish how the composition of a team affects performance under different incentive schemes. Apart from simple team pay, we studied competition between teams where the team with the highest performance wins a bonus. For the experiment, teams consisting only of men and teams consisting only of women were formed. In addition, we formed teams where the gender of its members was randomly drawn, henceforth called mixed teams.[2] In one treatment, we studied competition in a homogeneous environment, i.e., all-male [all-female] teams competed with all-male [all-female] teams. In a second treatment, male and female teams competed with each other, and in a third treatment, we studied competition between mixed teams. In this way, we were able to evaluate whether the gender composition of individual's own team and of the competing team matter for performance.

We sought to answer the following two questions with this experiment:

1. Does the composition of the team affect gender differences in performance (between-gender effect) for a given incentive scheme?

2. Does the composition of the team affect the performance of each gender alone (within-gender effect) for a given incentive scheme?

[1] The role of social pressure in team work is discussed by Kandel and Lazear (1992) and Huck, Kübler, and Weibull (2004). For loyalty and altruism in a model of team production see Ferreira (2002).

[2] Note that from the point of view of the participants, the gender of one's team member is uncertain in this treatment.

Our main results are as follows. Regarding the first question, we observed a difference in performance between men and women when men and women form mixed teams (i.e., when the gender of the other team member was uncertain) and when they are paid according to their joint output. Similarly, there is a gender difference in performance when comparing all-male teams competing with each other to all-female teams competing with each other. In both cases, men perform significantly better than women. There is no significant gender difference in performance for all other combinations of incentive scheme and team composition. The answer to the second question is negative, as we found no significant effect of team composition on the performance of each gender for a given incentive scheme. The remainder of the paper is organized as follows: In the next section we provide a brief overview of the relevant literature. In Section 3, the design and experimental procedures are introduced. In Section 4, we present and analyze the results of our study and relate them to previous findings from other studies. In Section 5, we discuss to what extent the results are in line with various economic, sociological and psychological theories. Section 6 contains the conclusion.

2 Review of the literature

The experimental literature on gender effects is extensive. One important aspect of gender experiments is whether participants are aware of the gender of the other players. Nowell and Tinkler (1994) conclude from previous studies that "gender differences would seem to be more forcefully expressed in environments in which the gender of the other subjects is known." In our experiments, participants can observe the gender of other participants, and we will therefore focus on studies of this type.[3]

[3] This excludes a number of papers. For example, in the work by Wiley (1973), Mason et al. (1991), Andreoni and Vesterlund (2001) and Eckel and Grossman (1998), subjects did not know the gender of their opponents. Some papers, for example, Cadsby and Maynes (1998), cannot be classified because it is not mentioned whether participants were able to observe the gender of others.

In the following, we provide a brief summary of the literature that is relevant for our study. A detailed comparison of previous findings with our results is provided in Section 4, where the results are presented. The papers related to our study can be grouped into three categories.

First, cooperation is the focus of gender studies that use public good, prisoner's dilemma, ultimatum or dictator games. In Nowell and Tinkler's (1994) study of public goods games and in Ortmann and Tichy (1999) on the prisoner's dilemma game, all-female groups, all-male groups and mixed groups are investigated. Andreoni and Petrie (2004) study the effect of beauty and gender in a repeated public goods game, where subjects are identified with the help of photographs. Both Andreoni and Petrie (2004) and Nowell and Tinkler (1994) focus on the impact of group composition on overall cooperation rates. Andreoni and Petrie (2004) observe a beauty premium in that contributions are higher in the presence of a beautiful participant, whereas Nowell and Tinkler (1994) find that women contribute more than men in the public goods game. Only Ortmann and Tichy (1999) study the differences in behavior between men and women *within* a treatment, which is one of the main questions of our study. They find that the cooperation rate of women is higher than of men initially, independent of the gender composition, but this difference disappears in the course of the experiment. Experimental tests of ultimatum games (Eckel and Grossman [2001]) and dictator games (Dufwenberg and Muren [2006]) examine whether behavior is influenced by the decision maker's own gender and the (known) gender of the other player. Both studies find that female proposers give more money than male proposers.

Second, a number of experimental studies focus on gender effects in competitive environments. Gneezy, Niederle and Rustichini (2003) and Gneezy and Rustichini (2004) study the interaction of gender and competition. Their experiments are based on individual compensation (i.e., tournament-style incentives versus piece rates), and they find that competition leads to a higher performance of men but not of women. In a similar framework, Gupta, Poulsen and Villeval (2005) as well as Niederle and Vesterlund (2007) focus on the endogenous choice of the payoff scheme, observing that significantly more men choose tournaments over piece rates compared to women. A number of psychological studies also investigate the performance of men and women or boys and girls in various competitive environments, see Moely, Skarin and Weil (1979) as well as Conti, Collins and Picariello (2001). They observe that girls perform better when competing with boys than with girls. However, none of these papers examines the role of gender in team work. A third group of papers considers competition among teams as a way to increase incentives compared to team pay. Field and laboratory experiments have been used to demonstrate that competition between groups can increase the performance within groups significantly. As formulated by Bornstein and Ben-Yossef (1994), "the most recurrent and explicit hypothesis of the intergroup conflict literature is that intergroup conflict increases intragroup cooperation." (p.53) Team work (as an instance of intragroup cooperation) and team competition (as an example of intergroup conflict) have been studied in a number of articles. Nalbantian and Schotter (1997) compare team competition to team work in an experiment where effort choice creates a monetary cost. A real-effort experiment is employed by van Dijk, Sonnemans and van Winden (2001). Finally, Erev, Bornstein and Galili (1993) use field experiments to compare the performance under team pay to the performance under team competition. The effect of intergroup competition has also been studied in the context of the step-level public goods game with and without communication (Bornstein [1992]), the prisoner's dilemma game (Bornstein and Ben-Yossef [1994]), the minimum-effort game (Bornstein, Gneezy, Nagel [2002]), and price competition (Bornstein and Gneezy [2002]). All papers find that competition among groups significantly increases effort (or lowers prices) relative to the non-competitive incentive scheme. However, none of the papers on intergroup competition deals with issues of gender.

3 Experimental design and procedure

In order to test for gender effects under different incentive schemes in an appropriate environment, we employed a real task in the experiment. The participants had to solve as many memory games as possible within 15 minutes.[4] We used memory games posted on the web, consisting of ten pairs of cards. At the start of the game, 20 cards with their faces down are shown. Two cards have to be clicked on. Then a "Check It"-button has to be pressed, and if the cards match, they are automatically removed. Otherwise, a "Pick again"-button has to be pressed and two more cards can be selected. The game ends when all pairs have been matched.[5]

In the experiment, when a participant had solved a game completely, he or she raised her hand, we recorded the solved game on a personal record sheet and opened a new game for this participant. The website offers 10 different games (with pictures of dinosaurs, flowers, colors, bugs, animals, musical instruments, etc.). In order to prevent confusion caused by recalling games played previously, we opened a new game with different cards every time. In all sessions, we used the same order of games. The first game, however, differed between sessions and was either dinosaurs or flowers. We varied the first game in order to prevent the results being influenced by the content chosen in the first memory task. We pooled the data from the session with different starting games as no significant differences were found.[6]

Each team consisted of two members. We varied the composition of teams with respect to gender. In addition, two different incentive schemes were used: revenue sharing and

[4] We chose the memory game for a number of reasons. First, a more realistic setting might import interpretations and perceptions which are neither controlled for nor relevant considering the specific question explored. Also, solving memory games captures some key elements of real work – it is an absorbing and demanding task that requires full concentration. Furthermore, output is exactly measurable. It also enables comparison with previous experimental studies that have already provided a number of stylized facts concerning gender-specific behavior in memory tasks.

[5] The website can be found at http://www.funbrain.com/match/.

[6] The p-value of the Mann-Whitney U test on the number of solved games is 0.94 for men and 0.23 for women.

team competition. Payoffs in the revenue sharing treatments (RS) were computed by adding up the number of games solved by the two team members (called "points" in the instructions), dividing the sum by two and paying out the resulting number in euros to each team member. In the team competition treatments (TC), first the number of games solved by both team members was determined. Then, the number of games solved by a team was compared to the number of games solved by a randomly selected second team. The team that had solved more games received a bonus of 4 points (=euros), and for the losing team we subtracted 4 points from the number of games solved jointly. Each team member then received points equal to the number of games solved in the team plus or minus the bonus, divided by two. If both teams had solved the same number of games, no points were subtracted or added.

Incentive scheme	Composition of teams	# Sessions	# Participants
Revenue sharing (RS)	Single-sex teams:	4	48
	Male	2	24
	Female	2	24
	Mixed teams	4	48
	All	8	96
Team competition (TC)	Single-sex teams:	4	48
	Male vs. male	2	24
	Female vs. female	2	24
	Female teams vs. male teams	4	48
	Mixed teams	4	48
	All	12	144
All		20	240

Table 1: Experimental treatments

We ran two different revenue sharing treatments: single-sex (all-male or all-female) and mixed teams (one man and one woman). In addition, we ran three different treatments with team competition: competition between single-sex (all-male or all-female) teams; competition of female against male teams; and competition between mixed teams. Table 1 summarizes our treatments. With each session consisting of 12 participants, 240 subjects took part in the experiment (RS-treatments: 96; TC-treatments: 144).

Most participants were students of economics, business administration, or industrial engineering at Humboldt University or Technical University Berlin. They had been invited by E-mail to participate in an experiment that would last no longer than one hour. Sessions actually took about 45 minutes. Communication among participants was not allowed during the whole session.

The participants in each session were seated in two groups of six persons at opposite walls of the computer lab. When participants entered the lab, the chairs were positioned so that the two groups faced each other. Only after the participants had read the instructions did we tell them to turn to the wall with the computer terminals. This ensured that all participants had enough time to observe the gender composition on both sides of the room. Participants were informed that they had been randomly matched with a person sitting at the same side of the room.[7] These two participants formed a team. In the sessions with teams consisting only of men or women, only male or female participants were seated on either side of the room. Thus, the information that they were randomly paired with a participant from their side of the room implied that this was a person of their own gender. We never explicitly mentioned gender in order to test for the relevance of gender without directly pointing out this aspect of the situation to the participants. In the treatment with mixed teams, three men and three women were seated on each side of the room.[8] Thus, participants knew that they would be paired either with a man or a woman.

In the treatments with team competition, we additionally informed participants that the competing team consisted of participants placed at the opposite side of the room. For competition between men or women only, we invited either only men or only women to the experiment. For competition between male and female teams, we invited six men and

[7]See Appendix B for a translated version of the instructions. The original instructions (in German) are available upon request from the authors.

[8]Before entering the room, subjects picked a seat number from a stack of cards. The stack had been manipulated in such a way that the correct distribution of men and women in the room was always ensured.

six women and placed them at the opposite sides of the room. For competition between mixed teams, we again invited six men and six women, but placed them at both sides of the room, i.e., three men and three women were seated on each side of the room. Thus, participants could see each other and they could figure out the possible gender compositions of both their own and the competing team.

At the end of the experiment, all participants were informed about the number of games solved by their team, their own payoff, and, in the team competition treatments, about whether their team received a bonus of 4 points, no bonus or whether their team payoff was decreased by 4 points. During the experiment, subjects were not able to observe the performance of the other member of their own team as well as of the competing team. There was a show-up fee of 3 euros. Average earnings in the experiment were 8.14 euros (including the show-up fee).

We ran 20 sessions (8 for the RS treatments and 12 for TC treatments). The number of games solved by each subject represents one independent observation, yielding a total of 240 independent observations (96 for RS and 144 for TC).

4 Results

Before addressing the two main questions of our paper, namely within- and between-gender effects of the composition of teams for a given incentive scheme, we give an overview of the aggregate effects of gender and incentives on performance. Table 2 reports descriptive statistics of the experimental outcomes for the two incentive schemes and the gender composition of the teams.

First, we focus on the overall performance of men and women in the memory game. Several psychological gender studies show that women outperform men in the memory game (see, for example, Tottenham et al. [2003], McBurney et al. [1997]). In our experiment, we also observe a strong gender effect on performance, but in the opposite direction, as men solved significantly more games than women (RS: MWU: $p = 0.006$; TC: MWU: $p = 0.0001$)[9] under both payoff schemes. This difference is highly significant also when aggregating over all treatments (MWU: $p = 0.000$). Of course, differences in performance between women and men depend on the specific task.[10] Our task reveals a clear asymmetry in that men solve on average more games than women. In the psychology experiments, no monetary incentives were used, and performance in these studies was measured on the basis of the number of turns of cards needed to solve a game. Thus, the difference to our findings might be due to the interaction of gender with these features of the experiments. In any case, given the result from our experiment that men outperform women and the opposite result from the psychological studies, there is no clear evidence that the innate ability of men and women to solve memory games is different. For the subsequent analysis, we will therefore ascribe any systematic differences in performance to effort choices rather than to innate ability. Effort can be affected by the objective monetary incentives, but also by gender roles, stereotypes and differences in expectations, for example.

Another aggregate result concerns the effect of the incentive scheme on overall performance. The average results in the TC treatments (5.22) as well as for the male subgroup (5.64) are slightly higher than in the RS treatments (5.10 for the pooled data and 5.40 for the male subgroup). However, none of these differences is significant (MWU: $p > 0.26$). For women, there is no difference between performance under the two incentive schemes. A number of studies find a positive effect of competition on performance (Nalbantian and Schotter [1997], van Dijk, Sonnemans and van Winden [2001], Bornstein and Ben-Yessef [1993], Bornstein, Gneezy and Nagel [2002], Gneezy, Niederle and Rustichini [2003], and Erev, Bornstein and Galili [1993]). However, each of these studies differs from ours in

[9]Throughout the paper, all Mann-Whitney U tests (MWU) are two-tailed.

[10]For example, in Gneezy, Niederle and Rustichini (2003), where subjects had to solve mazes, men also performed better than women. In Niederle and Vesterlund (2005), no gender difference can be found for the task of adding numbers.

several aspects.[11] Also, none of them controls for the composition of teams with respect to gender as a possible determinant of performance.

Incentive scheme	Composition of the teams	Men mean (std. dev.)	Women mean (std. dev.)	Both mean (std. dev.)
Revenue sharing	Single-sex teams	5.17 (0.816)	4.79 (1.215)	4.98 (1.041)
	Mixed teams	5.63 (1.096)	4.83 (1.129)	5.23 (1.171)
	All	5.40 (0.984)	4.81 (1.161)	5.10 (1.110)
Team competition	Single-sex teams	5.67 (1.167)	4.63 (1.279)	5.15 (1.321)
	Male vs. female teams	5.71 (1.268)	4.79 (1.103)	5.25 (1.263)
	Mixed teams	5.54 (1.503)	5.00 (1.063)	5.27 (1.317)
	All	5.64 (1.303)	4.81 (1.146)	5.22 (1.292)
All		5.54 (1.187)	4.81 (1.147)	5.18 (1.222)

Table 2: Descriptive statistics. Number of games solved

In our experimental setup, competition does not increase performance on the aggregate level. However, this does not exclude performance effects of the incentive scheme *in conjunction* with the composition of the team.[12] These effects are the main focus of the paper, and they are studied in the following.

Before starting with the analysis of gender differences in performance conditional on the composition of the teams, some remarks on the methodology are warranted. We set out to

[11] In van Dijk, Sonnemans and van Winden (2001) and Gneezy, Niederle and Rustichini (2003), for example, single individuals compete instead of teams as in our case. Nalbantian and Schotter (1997) do not use a real effort task but have participants pick a number for their effort choice, where higher numbers are associated with a higher monetary cost. Erev, Bornstein and Galili (1993) consider team competition in a real-effort environment as we do, but participants were able to constantly observe the performance of the competing team as well as of the other members of their own team during the experiment. In addition, Nalbantian and Schotter (1997) and Erev et al. (1993) split the teams (consisting of six respectively four participants), who first work together under revenue sharing, into two competing teams. Thus, by introducing competition between teams the team size is decreased, leaving it open whether the increase in performance can be ascribed to competitive pressure or to the smaller team size.

[12] The nonparametric Kruskal-Wallis test shows indeed the existence of significant effects among the different treatments ($p = 0.0024$).

investigate first whether there are gender differences in performance for a given incentive scheme and a given composition of the team (between-gender effect). Then, we studied whether the performance of women and men depends on the composition of the team for a given incentive scheme (within-gender effect).

In addition to the non-parametric Mann-Whitney U test, we performed OLS regression using the whole dataset in order to analyze gender differences in performance depending on the composition of the teams and controlling for interaction effects. Note that our independent variables, gender, team composition and incentive scheme, can each take only one of two values. A description of the regression procedure can be found in Appendix A. As we had to make more than one pairwise comparison, there is a problem that a higher number of tests makes a type-I error more likely. In order to avoid this problem, we used a post-hoc test. As we were only interested in a subset of all possible pairwise comparisons, we applied the Bonferroni correction for multiple testing.[13] In particular, we needed four pairwise comparisons in the case of revenue sharing, and nine pairwise comparisons in the treatments with competition between the teams. Therefore, choosing $\alpha = 0.05$ results in the following adjusted levels of significance: $\alpha_B^{RS} = 0.0125$ and $\alpha_B^{TC} = 0.006$.

4.1 Gender differences with revenue sharing

We will first analyze the performance of men and women in the treatments with revenue sharing. Assuming that the utility of participants is linear in money, the expected number of games solved by the other team member should not affect their own performance. On the other hand, gender may be relevant because motives such as solidarity, courtesy, competitiveness, or gender stereotypes can come into play. Also, social norms regarding helping behavior between men and women might change a subject's motivation to

[13] For this test, the α-level of each individual test is adjusted downwards. Thus, even if more than one test is performed, the probability of finding an incorrect effect continues to be α, i.e., $\alpha_B = \frac{\alpha}{c}$, where c is the number of pairwise comparisons.

contribute when participating in a mixed team compared to a single-sex male or female team.

As shown in Figure 1 and Table 2, for revenue sharing (RS) there is no significant difference between male and female performance for single-sex teams (5.17 vs. 4.79; MWU: $p = 0.190$; OLS: $p = 0.271$)[14]. However, in mixed teams men solve on average 5.63 games, as compared to 4.83 for women (MWU: $p = 0.012$; OLS: $p = 0.021$).[15] These results lead to:

Observation 1 *With revenue sharing, the performance of men and women does not significantly differ in the case of single-sex teams. However, in mixed teams men perform significantly better than women.*

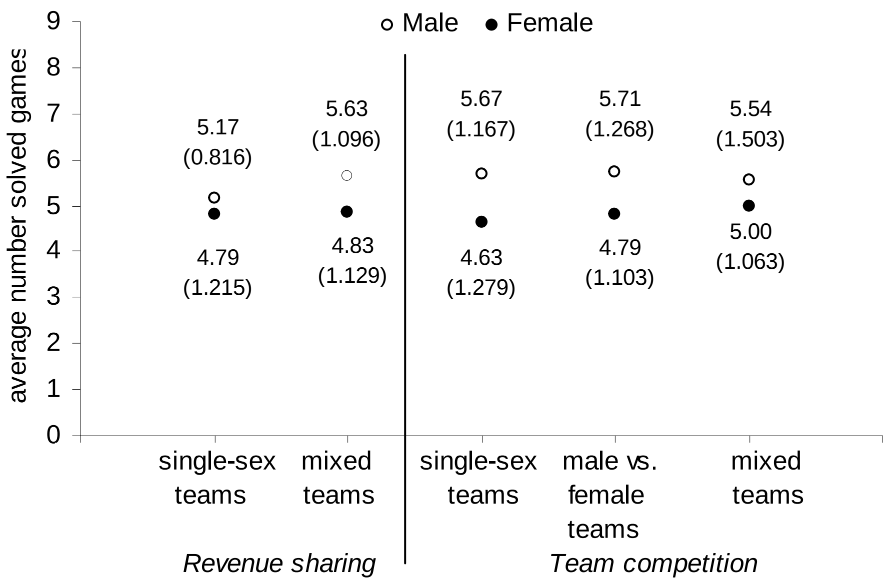

Figure 1: Average performance of men and women in all treatments (standard deviations in parantheses)

Our results corroborate the findings from studies based on games involving cooperation. For example, Ortmann and Tichy (1999) in their study of the prisoner's dilemma game observe that the differences between cooperation rates of men and women are significant

[14]For the details of the hypothesis testing in the context of the OLS regression, see Appendix A.

[15]Given that (i) with the robust non-parametric MWU test, the null hypothesis of no difference in performance can be rejected, and (ii) the Bonferroni correction is perceived as too conservative, these test results can be interpreted as evidence for a significant difference in performance.

in mixed-sex treatments and insignificant in single-sex treatments.[16] However, in their experiment this difference decreases after a number of rounds and disappears if one controls for past experience. In their experiment consisting of mixed groups with respect to gender, Andreoni and Petrie (2004) find that male and female earnings differ significantly. In particular, women make more money than men when no information about previous contributions is provided.[17] This is consistent with our results if positive effort costs in the real task are assumed.

To sum up, our results reveal a significant gender difference in performance in mixed teams which is not present in single-sex teams. The difference in behavior in mixed teams is consistent with the results from previous gender experiments.

4.2 Gender differences with team competition

Economic theory predicts that an individual's own performance and the performance of his or her opponents in a tournament affect the optimal effort choice (see Nalbantian and Schotter [1997]). Thus, for example, teams consisting of women only who think (correctly) that they solve on average less memory games than men might decrease their contribution when competing against a male team. A similar effect could be due to the stereotype that women perform poorly in a competitive environment. On the other hand, solidarity among women might increase the women's performance when competing against men as opposed to women.

As shown in Table 2 and Figure 1, competition among single-sex teams leads to a significant gender gap in mean performance when the competing teams are of the same sex (5.67 vs. 4.63 in favor of the men; MWU: $p = 0.006$; OLS: $p = 0.002$). The difference in performance when male and female teams compete against each other (the average number of games solved is 5.71 and 4.79, respectively) becomes marginally significant due

[16] The design is similar to ours in that subjects know the gender of the other player in the single-sex treatments, and in the mixed treatments, they know that they are either paired with a man or a woman.

[17] However, if information about past behavior is provided, men contribute more and also induce others to make higher contributions, which leads to higher payoffs for men than for women.

to the Bonferroni correction (MWU: $p = 0.010$; OLS: $p = 0.007$). But it should be noted that the difference is larger than in mixed teams where there are no significant differences between male and female performance (5.54 vs. 5.00; MWU: $p = 0.201$; OLS: $p = 0.112$). These results lead to

Observation 2 *With competition of male versus male and female versus female teams, men perform significantly better than women. While this difference is still marginally significant when male teams compete with female teams, the performance of men and women does not significantly differ in the case of competition between mixed teams.*

Our results are in line with findings from previous studies on tournaments based on individual performance. For example, Niederle and Vesterlund (2007) find no significant gender gap in performance in tournaments where the group of competitors consists of both women and men. In a field study with boys and girls running on a track, Gneezy and Rustichini (2004) observe that the gender gap in performance in single-sex tournaments is noticably higher when compared to mixed tournaments. Thus, they also find a significant effect of the gender of the competitor(s) on the performance. Finally, a number of studies by psychologists based on game-playing by children show that girls are less competitive when playing against girls than against boys (see Moely, Skarin and Weil [1979] and the studies cited therein, as well as Conti, Collins and Picariello [2001]). In contrast, Gneezy, Niederle and Rustichini (2003) only find a significant difference in individual performance in mixed treatments, whereas the gender gap becomes insignificant when competition is among men or women only.

To sum up, the results reveal a significant gender difference in performance when competition between teams is in a homogeneous (male or female) environment. This difference becomes smaller and almost disappears as soon as men and women interact with each other as members of the same or the competing team. These findings underline the insight gained from tournaments based on individual performance that "the puzzle that remains concerns the more subtle effects of competition in homogeneous and heterogeneous groups" (Gneezy and Rustichini [2004], p. 380). Our results add another dimension as they show that the composition not only of the individual's own team, but also of the competing team matters.

4.3 Performance of men and women dependent on the composition of the team

We will now focus on the performance of each gender with respect to the composition of the team. When comparing pairwise between, for example, men in RS with single-sex teams and men in RS with mixed teams (MWU: $p = 0.053$; OLS: $p = 0.178$), or women in TC between single-sex teams with competition among women only and women in TC between mixed teams (MWU: $p = 0.237$; OLS: $p = 0.271$), we do not find any significant differences. For both incentive schemes, none of the comparisons leads to a significant difference when keeping gender constant and changing only the gender composition.

Observation 3 *The composition of the team has no significant effect on the performance of each gender for a given incentive scheme.*

This result is in contrast to findings from experiments with incentive schemes based on individual performance. Gneezy, Niederle and Rustichini (2003) find that women perform better in tournaments when competing against women than against men. In the same setup, but with endogenous choice of the payoff scheme, Gupta, Poulsen, and Villeval (2005) show that men are better performers when competing against men than when competing against women. However, they do not find a similar effect for women whose performance is not affected by the gender composition, as in our experiment.

Our data set also allows us to answer the question whether the performance of each gender depends on the combination of the composition of the teams and the incentive scheme. For example, Figure 1 and Table 2 might suggest that men perform worse when the benefits are shared with another man in a cooperative environment than when there is either competition or women are present. Indeed, the OLS regression analysis yields a weakly significant difference for this case ($p = 0.082$). Investigating further possible comparisons of this type, however, does not reveal any other significant differences.

Altogether, the results suggest that the within-gender performance effects of the treatments are rather small, but in some treatments men and women are affected in opposite directions, leading to significant between-gender differences.[18]

5 Discussion

In this section, we discuss possible explanations of our results in the light of sociological and psychological theories of gender and competition.

First, consider gender stereotypes which can be responsible for differences in behavior of women and men. For example, women might have internalized the belief that they are less able or that they need to be helped by men. The stereotype that women are less able to play computer games can explain the overall lower performance of women because women who expect to be less able might put in less effort. However, the lower expectations of women alone cannot explain the differences in the gender gap between the treatments.[19] But suppose that the presence of both men and women strengthens this stereotype, compared to a situation where only men or only women are present. This is suggested by stereotype threat theory (Steele [1997]). It posits that women experience a higher pressure than men when they are in a situation in which a negative stereotype about them applies. It is the threat that their own behavior makes the stereotype more

[18] The statistical analysis of differences between the gender gap in different treatments reveals no siginificant results. However, one should be aware that the test of interactions between treatments and gender dummies has a lower power than the test of a simple gender effect. In the latter case, the test statistic consists of only two independent means whereas the test statistic in the former case is based on four independent means.

[19] Similarly, even if women were in fact less able than men to play the memory game on the computer, this would not be sufficient to explain the differences in the gender gap between the treatments. For a discussion of differential ability see also Section 4.

plausible which then results in a lower performance. Thus, stereotype threat theory can be interpreted as saying that women should be less threatened in a situation where they carry a negative stereotype if men are absent. This would imply a higher performance of women when they are among themselves than when interacting with men.[20] Then, the stereotype of a lower ability of women could account for the results in the revenue sharing treatments where male and female performance differ more in mixed teams than in single-sex teams. However, under this hypothesis, the performance of women in all-female teams competing among themselves should be better than in the team competition treatments where men are present, which is clearly not the case.

Another relevant stereotype might be that women are less competitive than men. This should lead to systematically lower overall performance levels of women in the team competition treatments compared to the revenue sharing treatments, which is not what we observe. Also the stereotype threat that women are less competitive than men is not supported by our data. In single-sex teams with team competition, the women perform significantly worse than the men although the stereotype threat should be milder compared to the other treatments with team competition where men are present.

A hypothesis that runs contrary to stereotype threat theory for women is that the salience of gender in treatments where both men and women are present (mixed teams with revenue sharing, male vs. female teams as well as mixed teams with team competition) makes men and women exert more effort than in the treatments where only men or only women are present. Higher effort levels of men and women can indeed be observed in revenue sharing with mixed teams compared to single-sex teams and in team competition with male vs. female teams compared to single-sex teams (while this is only partly true for mixed teams). Thus, salience of gender might have an effect on the performance in our setup. However, salience cannot explain our main result, namely that there is a significant difference between the performance of men and women in some treatments but

[20] See Gneezy, Niederle and Rustichini (2003) for this interpretation.

not in others, as it does not predict differential reactions of men and women to certain treatment conditions.

In the literature on evolutionary biology and sociobiology (see Knight [2002]), competition is seen as the natural environment for the male species (which often finds itself in fierce competition for females) whereas women typically do not compete among themselves and therefore do not react to the competitive environment with an increase in effort. This leads to the prediction that competition enhances the performance of men relative to the performance of women. Thus, the theory offers a possible explanation for the observed gender gap in the treatments with competition only among women or only among men (and for the marginally significant difference in the treatment with competition between male vs. female teams). Moreover, our results show that this male-female performance gap in competitive environments is somewhat softened when mixed teams are formed. Possibly, the gender roles are more muted in a situation where gender of the other players is unknown. However, as we find no significant effect of competition on overall male performance, the evidence for the sociobiological explanation is not that strong. Furthermore, it cannot account for the observed gender effects in the revenue sharing treatments, where competition plays no role.

There is a view that, possibly due to a different socialization, women see themselves as part of a network of social relations whereas men frame their environment as a system of hierarchical relationships (see Gilligan [1982]). From this could be inferred that women will cooperate more than men and that women react more to the context they are placed in. However, in our experiment women do not contribute more in sessions with revenue sharing than men. Moreover, the performance of women is not significantly affected by any of the treatment conditions. Thus, the predictions for our game derived from Gilligan's theory are not in line with the experimental results.

Group conflict theory puts forward the hypothesis that intergroup conflict leads to more intragroup cooperation (see Bornstein and Ben-Yossef [1994] among others). The game played by the team in our experiment has the structure of a social dilemma as there is an individual incentive to free ride on the performance of the other team member. Thus, cooperation can be increased by competition among the groups. From this follows that overall performance is predicted to increase with team competition compared to the revenue sharing treatments, which is not the case in our experiment. Thus, group conflict theory is at odds with the experimental results, which might be due to the specific task and the design we have employed. For example, if the participants have some intrinsic motivation to play the memory game, the strength of the monetary incentives may be less important for performance. In addition, in our experiment there is no intermediate stage at which team members are informed about their performance relative to the competing team, which possibly weakens the effect of competition. Finally, the one-shot nature of the game we employed does not allow participants to learn that it might pay off to exert more effort in competition.

To sum up, gender stereotypes, salience of gender as well as evolutionary biology and sociobiology provide some clues towards an explanation of our findings. However, none of these theories alone can fully predict the behavioral patterns observed in our experiment. This might be due to the fact that in our treatments gender, competition as well as incentives to free-ride and to cooperate interact in a non-trivial way.

6 Conclusions

This paper reports on an experiment designed to study the relevance of gender for performance in team work. We focus on gender effects with respect to the composition of a team. We set up an environment in which the effects of changes in the gender composition of a team on individual performance under different incentive schemes were measured.

The results of the experiment suggest that gender plays a role in the context of team work. We find that the gender composition of the team accounts for gender differences in performance, both with team work and with team competition. In particular, the data reveal that the performance of men is significantly higher than the performance of women in revenue sharing with mixed teams and in team competition with male vs. male compared to female vs. female teams. Thus, the effect of gender diversity on the performance gap between men and women depends on the incentive scheme: compared to single-sex settings, the performance gap is widened with revenue sharing, but it is reduced in team competition. Possibly, this is due to gender effects in the team environment interacting with differences in the competitiveness of men and women.

In the case of revenue sharing, the difference between men and women is mainly due to increased effort by the men. This indicates that there can be a tension between the firm's objective to maximize overall performance and the possible goal to increase gender equality.[21] Similarly, when considering all our treatments, there is no clear winner in the two dimensions of maximizing overall performance and minimizing the gender performance gap. However, team competition with mixed teams yields the maximum overall output and leads to a gender gap that is only slightly higher than the smallest observed gender gap with single-sex teams under revenue sharing. Thus, competition between mixed teams emerges as a relatively good solution for a firm aiming at high performance levels but also at some degree of gender equality.

What additional evidence is necessary to explain our findings based on the theories discussed above? One important element of some of the theories are expectations and beliefs about the behavior of others and about oneself. Their elicitation in further experiments might help to gain direct evidence of gender differences in beliefs, which can yield further insights regarding the role of gender in team work. Also, more evidence is needed to evaluate whether our findings can be generalized, for example, by employing other tasks or by increasing the team size. Moreover, field studies can be a powerful method to understand human behavior, and the investigation of gender effects in teams seems a topic

[21] For example, a firm might aim at equal pay of women and men in comparable positions for reasons of public relations.

that is well-suited for the field. As pointed out by Gneezy and List (2006), one important advantage of field experiments in the context of labor markets is the duration of the task. While lab experiments usually do not last for more than two hours, real labor market interactions can last for months and years. Therefore, we believe that the role of gender in team work has more facets than the ones highlighted by our study. Our experiment should thus be seen as a first step towards a better understanding of gender effects in team work.

References

[1] Andreoni, J. and Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. Quarterly Journal of Economics, 116, 293-312.

[2] Andreoni, J. and Petrie, R. (2008). Beauty, gender and stereotypes: evidence from laboratory experiments. Journal of Economic Psychology, 29, 73-93.

[3] Bornstein, G. (1992). The free-rider problem in intergroup conflicts over step-level and continuous public goods. Journal of Personality and Social Psychology, 62 (4), 597-606.

[4] Bornstein, G. and Ben-Yossef (1994). Cooperation in intergroup and single-group social dilemmas. Journal of Experimental Social Psychology 30, 52-67.

[5] Bornstein, G. and Gneezy, U. (2002). Price competition between teams. Experimental Economics 5, 29-38.

[6] Bornstein, G., Gneezy, U., and Nagel, R. (2002). The effect of intergroup competition on group coordination: an experimental study. Games and Economic Behavior 41, 1-25.

[7] Cadsby, C. B. and Maynes, E. (1998). Gender and free riding in a threshold public goods game: experimental evidence. Journal of Economic Behavior and Organization, 34, 603-620.

[8] Conti, R., Collins, M. A., and Picariello, M. L. (2001). The impact of competition on intrinsic motivation and creativity: considering gender, gender segregation and gender role orientation. Personality and Individual Differences, 30, 1273-1289.

[9] Dufwenberg, M. and Muren, A. (2006). Generosity, anonymity, gender. Journal of Economic Behavior and Organization, 61, 42–49.

[10] Eckel, C. C. and Grossman, P. J. (1998). Are women less selfish than men? Evidence from dictator experiments. Economic Journal 108, 726-735.

[11] Eckel, C. C. and Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. Economic Inquiry, 39, 171-188.

[12] Erev, I., Bornstein, G., and Galili, R. (1993). Constructive intergroup competition as a solution to the free rider problem: a field experiment. Journal of Experimental Social Psychology, 29, 463-478.

[13] Gilligan, C. (1982). In a different voice: psychological theory and women's development. Harvard University Press, Cambridge, MA.

[14] Gneezy, U. and List, J. A. (2006). Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments. Econometrica, 74, 1365-1384.

[15] Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: gender differences. Quarterly Journal of Economics, 118, 1049-1074.

[16] Gneezy, U., and A. Rustichini (2004). Gender and competition at a young age. American Economic Review (Papers and Proceedings), 377-381.

[17] Huck, S., Kübler, D., and Weibull, J. (2004). Social norms and economic incentives in firms. University of Tilburg Working Paper Series. Mimeo.

[18] Kandel, E. and Lazear, E. (1992). Peer pressure and partnerships. Journal of Political Economy, 100, 801-817.

[19] Knight, J. (2002). Sexual stereotypes. Nature, 415, 254-56.

[20] Mason, C. F., Philips, O. R., and Redington, D. B. (1991). On the role of gender in a noncooperative game. Journal of Economic Behavior and Organization, 15, 215-235.

[21] McBurney, D. H., Gaulin, S. J. C., Devineni, T., and Adams, C. (1997). Superior spatial memory of women: stronger evidence for the gathering hypothesis. Evolution and Human Behavior, 18, 165-174.

[22] Moely, B. E., Skarin, K., and Weil, S. (1979). Sex differences in competition-cooperation behavior of children at two age levels. Sex Roles 5/3: 329-342.

[23] Nalbantian, H. R. and Schotter, A. (1997). Productivity under group incentives: an experimental study. American Economic Review, 87, 314-341.

[24] Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? Quarterly Journal of Economics, 122, 1067-1101.

[25] Nowell, C. and Tinkler, S. (1994). The influence of gender on the provision of a public good. Journal of Economic Behavior and Organization, 25, 25-36.

[26] Ortmann, A. and Tichy, L. K. (1999). Gender differences in the laboratory: evidence from prisoner's dilemma games. Journal of Economic Behavior and Organization, 39, 327-339.

[27] Steele, C. M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. American Psychologist, 52, 613-629.

[28] Tottenham, L. R., Saucier, D., Elias, L., and Gutwin, C. (2003). Female advantage for spatial location memory in both static and dynamic environments. Brain and Cognition, 53, 381-383.

[29] van Dijk, F., Sonnemans, J., and van Winden, F. (2001). Incentive systems in a real effort experiment. European Economic Review, 45, 187-214.

[30] Wiley, M.G. (1973). Sex roles in games. Sociometry, Vol. 36, No. 4, 526-541.

Appendix A: OLS Regression

For the analysis of the data, we used the following regression equation:

$$\begin{aligned} nr\_games\_solved \;=\; & \alpha_0 + \alpha_1 * D_{\text{gender}} + \alpha_2 * D_{\text{comp}} + \alpha_3 * D_{\text{tmbrss}} + \\ & + \alpha_4 * D_{\text{gender}} D_{\text{comp}} + \alpha_5 * D_{\text{gender}} D_{\text{tmbrss}} + \\ & + \alpha_6 * D_{\text{comp}} D_{\text{tmbrss}} + \alpha_7 * D_{\text{gender}} D_{\text{comp}} D_{\text{tmbrss}} + \\ & + \alpha_8 * D_{\text{comp}} D_{\text{tmbrss}} D_{\text{trtmt4}} + \alpha_9 * D_{\text{gender}} D_{\text{comp}} D_{\text{tmbrss}} D_{\text{trmt4}} + \varepsilon, \end{aligned}$$

where

$$D_{\text{gender}} = \begin{cases} 1 & \text{, if woman} \\ 0 & \text{, if man} \end{cases}$$

$$D_{\text{comp}} = \begin{cases} 1 & \text{, if team competition (TC)} \\ 0 & \text{, if revenue sharing (RS)} \end{cases}$$

$$D_{\text{tmbrss}} = \begin{cases} 1 & \text{, if single-sex team} \\ 0 & \text{, else} \end{cases}$$

$$D_{\text{trtmt4}} = \begin{cases} 1 & \text{, if TC between male and female teams} \\ 0 & \text{, else} \end{cases}$$

$D_{\text{gender}}, D_{\text{comp}}, D_{\text{tmbrss}}, D_{\text{trtmt4}}$, and ε are $(n \times 1)$ vectors
$n = 240$: number of observations
$\alpha_0, \alpha_1, ..., \alpha_9$: (1×1) constants

OLS for number of games solved

Number of games solved	Coefficient	Std. Err.	t	$P > \lvert t \rvert$
constant	5.625	0.240	23.433	0.000
D_{gender}	-0.792	0.339	-2.332	0.021
D_{comp}	-0.083	0.339	-0.245	0.806
D_{tmbrss}	-0.458	0.339	-1.350	0.178
$D_{\text{gender}} D_{\text{comp}}$	0.250	0.480	0.521	0.603
$D_{\text{gender}} D_{\text{tmbrss}}$	0.417	0.480	0.868	0.386
$D_{\text{comp}} D_{\text{tmbrss}}$	0.583	0.480	1.215	0.226
$D_{\text{gender}} D_{\text{comp}} D_{\text{tmbrss}}$	-0.917	0.679	-1.350	0.178
$D_{\text{comp}} D_{\text{tmbrss}} D_{\text{trtmt4}}$	0.042	0.339	0.123	0.902
$D_{\text{gender}} D_{\text{comp}} D_{\text{tmbrss}} D_{\text{trtmt4}}$	0.125	0.480	0.260	0.795

Note: The results of the regression are qualitatively the same when dropping D_{trtmt4}.

Hypothesis testing

The null hypothesis for all tests is that there are no differences in performance.

Note that the coefficient of the dummy variable D_{gender} captures only the gender difference under the condition of revenue sharing in mixed teams ($D_{\text{comp}} = 0, D_{\text{tmbrss}} = 0$). All other reported results are supported by linear combinations of dummy coefficients in the regression. Here are three examples:

1. The null hypothesis for the test of the gender gap in mean perfomance for single-sex teams with revenue sharing (RS) is

$H_0 : E[nrgames|D_{\text{gender}} = 0, D_{\text{comp}} = 0, D_{\text{tmbrss}} = 1] =$
$E[nrgames|D_{\text{gender}} = 1, D_{\text{comp}} = 0, D_{\text{tmbrss}} = 1].$

2. The null hypothesis for the test of the gender gap in mean perfomance for team competition (TC) between teams of the same sex is

$H_0 : E[nrgames|D_{\text{gender}} = 0, D_{\text{comp}} = 0, D_{\text{tmbrss}} = 1, D_{\text{trtmt4}} = 0] =$
$E[nrgames|D_{\text{gender}} = 1, D_{\text{comp}} = 0, D_{\text{tmbrss}} = 1, D_{\text{trtmt4}} = 0].$

3. The null hypothesis for the test of the gender gap in mean perfomance for team competition (TC) between male and female teams is

$H_0 : E[nrgames|D_{\text{gender}} = 0, D_{\text{comp}} = 0, D_{\text{tmbrss}} = 1, D_{\text{trtmt4}} = 1] =$
$E[nrgames|D_{\text{gender}} = 1, D_{\text{comp}} = 0, D_{\text{tmbrss}} = 1, D_{\text{trtmt4}} = 1].$

Appendix B: Instructions

[for Revenue Sharing (RS); translated from German]

Welcome to the experiment! Please read the instructions carefully. If you don't understand something, please raise your hand. We will come to you and answer your question individually. The instructions are the same for all participants.

At the beginning of the experiment you will be randomly matched with a person **from the same side of the room**. Of course your anonymity will be guaranteed throughout the whole experiment. This means that the other participants will not get to know your true identity. The same holds for all participants.

Your task is to solve as many memory games as possible within 15 minutes. As soon as you have solved a game successfully, please announce this by raising your hand. We will then come to you, update your score and start a new game.

Your payoff is determined as follows:

The number of memory games that you solve will be added to the number of solved memory games by the other participant of your group. Each finished game yields one point (= 1 euro). Each member of the group gets paid half of the joint score.

Example: You have solved X1 and the other member of your group has solved X2 memory-games. Then each of you will be paid $\frac{1}{2}$(X1+X2).

At the end of the experiment you will be informed about the number of memory games solved by you and by the other member of your group and about your payoff.

At the beginning you will receive an initial endowment of 3 euros.

If you don't understand something, please raise your hand. We will come to you and answer your question individually.

[for Team Competition (TC); translated from German]

Welcome to the experiment! Please read the instructions carefully. If you don't understand something, please raise your hand. We will come to you and answer your question individually. The instructions are the same for all participants.

At the beginning of the experiment you will be matched with a person **from the same side of the room**. Of course, anonymity will be guaranteed throughout the whole experiment. This means that the other participants will not get to know your true identity. The same holds for all participants.

Your task is to solve as many memory games as possible within 15 minutes. As soon as you have solved a game successfully, please announce this by raising your hand. We will then come to you, update your score and start a new game.

Your payoff is determined as follows:

The number of memory games that you solve will be added to the number of memory games solved by the other participant of your group. Each solved game yields one point (= 1 euro). To determine your payoff, the joint score of your group will be compared with the score of another group that consists of persons **at the opposite side of the room** and is randomly matched with your group. In the case that the score of your group is higher, your group will receive a premium of 4 points and the other group will have 4 points subtracted. Should your group's score be lower than that of your opponents, your group will have 4 points subtracted, while the other one will receive 4 points as a premium. If the scores are the same, a premium will not be paid nor will points be subtracted. In any case, the payoff of each member of your group amounts to half of the final score of the group.

To put it differently, the payoffs are computed in the following way:

You have X1 and the other member of your group has solved X2 memory games. In the other randomly selected group, one player has solved Y1 and the other has solved Y2 games.

- If X1+X2 > Y1+Y2 holds, your group will receive a premium of 4 points and the payoff of each group member is (X1+X2+4)/2.
- If X1+X2 < Y1+Y2 holds, your group will have 4 points subtracted. Therefore each group member receives a payoff equal to (X1+X2-4)/2.

- If X1+X2=Y1+Y2 holds, a premium will not be paid nor will points be subtracted.

The payoff of each group member is (X1+X2)/2.

At the end of the experiment you will be informed about how many memory games you and the other participant of your group have solved, and whether you have received a premium or whether points have been subtracted, as well as your final payoff.

At the beginning you will receive an initial endowment of 3 Euros. In the case of losses, these will be subtracted from your initial endowment.

If you don't understand something, please raise your hand. We will come to you and answer your question individually.