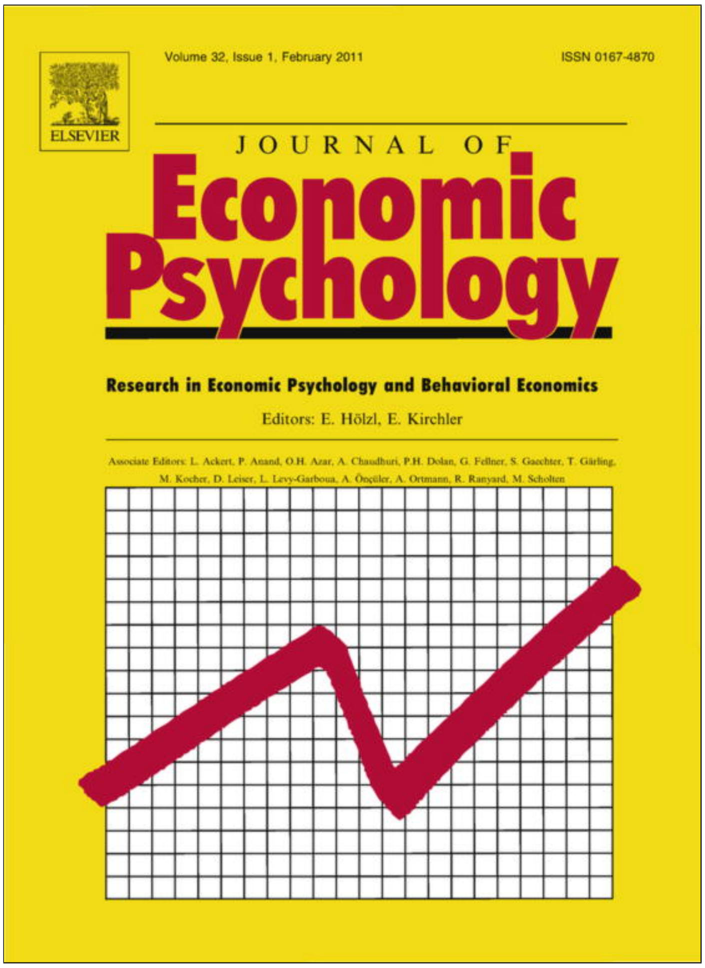

(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Journal of Economic Psychology 32 (2011) 797–808

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Gender differences in team work and team competition

Radosveta Ivanova-Stenzel [a,*], Dorothea Kübler [b]

[a] *Technical University Berlin, Department of Economics and Management, Straße des 17. Juni 135, D-10623 Berlin, Germany*
[b] *WZB (Social Science Research Center), Technical University Berlin and IZA, Reichpietschufer 50, 10785 Berlin, Germany*

A R T I C L E   I N F O

Article history:
Received 29 July 2010
Received in revised form 23 November 2010
Accepted 31 May 2011
Available online 23 June 2011

JEL classification:
C72
C73
C91
D82

PsycINFO classification:
3020
2260
2910
2970
3630
3660

Keywords:
Team incentives
Gender diversity
Tournaments

A B S T R A C T

We study whether the gender performance gap is affected by the gender composition of teams. A real-effort experiment is employed with wages based either on the team's performance, or on the outcome of a competition between teams. We find that, relative to a single-sex environment, gender diversity increases the gender performance gap with team pay whereas it decreases the gap with team competition. The results show that there can be a tension between the objective to maximize overall performance and to minimize gender inequality.

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, and lower average earnings for professions mainly exercised by women. However, even when restricting attention to one of these factors, many new questions arise. For example, there can be various reasons for differences in performance, such as education and ability, as well as the gender composition of the workplace. The gender of one's colleagues may be one of the factors determining whether working conditions are experienced as being hostile or as encouraging, as stimulating or stifling. We focus on this potential source of performance differentials between men and women which could contribute to the gender gap in earnings.

The organization of work in teams has become widespread. As the success of a team depends crucially on the interaction among the team members, the question of how diversity of the teams affects performance has become central. In particular,

* Corresponding author. Tel.: +49 030 314 29009; fax: +49 030 314 24968.
E-mail addresses: ivanova-stenzel@tu-berlin.de (R. Ivanova-Stenzel), kuebler@wzb.eu (D. Kübler).

0167-4870/$ - see front matter
doi:10.1016/j.joep.2011.05.011

when setting up team work in a firm, the question arises of how to group the employees into teams. Should they be of the same gender, culture, nationality, age, race or should they differ in some or all of these characteristics? Economic theory has focused on heterogeneity of team members with respect to ability, but it does not provide any clues as to the issue of how different cultural backgrounds or the gender composition of teams affects outcomes. The empirical literature on diversity in teams yields mixed evidence for the effect of diversity on performance (for an overview see Milliken & Martins, 1996). For example, Hamilton, Nickerson, and Owan (2003) show that team diversity with respect to ability has a significant positive impact on productivity. Adams and Ferreira (2009) find a negative relationship between gender diversity in the board room and firm performance. In general, the results of empirical studies on directly observable attributes like gender suggest that an increase in diversity is more likely to have a negative effect on team performance.

The experimental approach allows for a systematic analysis of the specific effects of diversity in team work and helps exploring the exact mechanisms by which diversity affects outcomes. Although there exists a large experimental literature on gender effects, almost all studies on incentive contracts consider mainly environments where subjects are paid based on their individual performance. 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 two-person teams consisting either of men or of women, as well as mixed teams. Each participant 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 team members perform different tasks in which they specialize. The team members' performance in these different tasks then determines the joint output upon which individual wage payments are contingent.

Assuming rational and self-interested agents, 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. On the other hand, we hypothesize that in team work, the productivity of teams can be influenced by competition, peer pressure and social norms, or loyalty among group members.[1] So far, it has not been studied systematically how these motives interact with the gender of the decision makers.

We designed the experiment to establish how the composition of a team with respect to gender affects performance under different incentive schemes. Apart from the case when the members of the team are paid solely according to their joint output (i.e., revenue sharing), we studied competition between teams where the team with the highest performance wins a bonus. Team work and team competition have been studied in a number of experimental studies.[2] The general finding in this literature is 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 addresses how competition affects the gender gap in performance.

In one treatment of our experiment, 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 teams where the gender of its members was randomly drawn, henceforth called mixed teams.[3] In this way, we were able to evaluate whether the gender composition of an 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?

Our main results are as follows. With revenue sharing, we observed a significant 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). Similarly, there is a between-gender effect when the competition is in a homogenous environment (i.e., when all-male teams compete with each other and all-female teams compete with each other). Thus, the effect of gender diversity on the performance gap between men and women depends on the incentive scheme. Compared to single-sex settings, gender diversity widens the performance gap with revenue sharing. The opposite is true for team competition. Gender diversity (mixed teams or women competing against men) reduces the performance gap. We found no significant difference in the performance of each gender alone for a given incentive scheme (within-gender effect). Our results demonstrate that there can be a tension between the objective to maximize overall performance and to minimize gender inequality.

We discuss possible explanations of our results in the light of sociological and psychological theories of gender and competition. Gender stereotypes paired with salience of gender can explain the behavioral patterns in the environment with

[1] The role of social pressure in team work is discussed by Kandel and Lazear (1992) and Huck et al. (2004).

[2] 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). Erev, Bornstein, and Galili (1993) use field experiments to compare theperformance 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 & Ben-Yossef, 1994), the minimum-effort game (Bornstein, Gneezy, & Nagel, 2002), and price competition (Bornstein & Gneezy, 2002).

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

team pay. Sociobiological differences in competitiveness together with salience of gender can rationalize the observed behavior when competition between teams is introduced.

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 existing economic, sociological and psychological theories. Section 6 concludes.

2. Review of the experimental literature on gender effects

An important aspect of the experimental literature on gender 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 experiment, participants can observe the gender of the other participants, and we will therefore focus on studies of this type.[4]

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 two categories: (i) studies that deal with gender differences in cooperative environments (comparable to our treatments with revenue sharing) and (ii) studies that investigate gender effects in competitive environments.

In general, the evidence on gender effects in social dilemma games is mixed (see Croson & Gneezy, 2009). Ortmann and Tichy (1999) study the differences in behavior between men and women *within* a treatment with a given gender composition, which is the main focus of our study. They find that the cooperation rate of women in the prisoners's dilemma game is initially higher than that of men, independent of the gender composition. Similarly, Nowell and Tinkler (1994) observe that regardless of the treatments (whether all men, all women, mixed) women contribute more than men. On the other hand, in public good games Sell (1997) finds evidence that in mixed groups with respect to gender men contribute more than women. In a design where the pictures of other participants were shown to the players, Andreoni and Petrie (2008) find that women's earnings were higher than that of men, which they ascribe to the stereotype that women are more helpful than men. In a similar vein, studies on gender differences in ultimatum and dictator games find that female proposers give more money than male proposers (Dufwenberg & Muren, 2006; Eckel & Grossman, 2001). However, all experiments of these studies differ from our real-effort environment in which subjects might focus more on performance than on the distribution of money.

Second, there are very few papers that examine the role of gender in team competition. The focus of these papers is not on the gender differences in performance but rather on the choice between tournaments and piece rates (Dargnies, 2009; Healy & Pate, in press). They find that the gender gap with respect to the choice of the incentive scheme is significantly reduced by the team setup.

The interaction of gender and competition has been studied in more depth for individual compensation (i.e., tournament-style incentives vs. piece rates). Gneezy, Niederle, and Rustichini (2003) and Gneezy and Rustichini (2004) find that competition leads to a higher performance of men but not of women. In a similar framework with individual compensation, Niederle and Vesterlund (2007) and Gupta et al. (2005) study the endogenous choice of the payoff scheme, observing that significantly more men choose tournaments over piece rates compared to women. Booth and Nolen (2009) show that the difference in the willingness to compete between women and men disappears for women from single-sex schools. Finally, a number of psychological studies 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.

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 min.[5] We used memory games posted on the web, consisting of ten pairs of cards. At the start of the game, 20 cards are shown with their faces down. 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.[6]

In the experiment, when a participant had solved a game completely, she raised their 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

[4] This excludes a number of papers. For example, in the work by Wiley (1973), Mason, Philips, and Redington (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.

[5] 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 ofeal 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 (see, for example, McBurney, Gaulin, Devineni, & Adams, 1997; Tottenham, Saucier, Elias, & Gutwin, 2003).

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

Table 1
Experimental treatments.

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

of dinosaurs, flowers, colors, bugs, animals, musical instruments, etc.). In order to prevent confusion caused by recalling games played previously, we opened a different game 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 sessions with different starting games as no significant differences were found.[7]

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 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 the most games received a bonus of four points (=euros), and for the losing team we subtracted four points from the number of games solved jointly by its members. 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.

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 employed 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 min. Communication among participants was not allowed during the course of the 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 on the same side of the room.[8] 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. In the treatment with mixed teams, three men and three women were seated on each side of the room.[9] 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 positioned on 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 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 on both sides of the room, i.e., three men and three women were seated on each

[7] 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.
[8] See Appendix B for a translated version of the instructions. The original instructions (in German) are available upon request from the authors.
[9] 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.

side of the room. Since participants could see each other, 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 four points, no bonus or whether their team payoff was decreased by four points. During the experiment, subjects were not able to observe the performance of the other member of their own team or that 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

We set out to 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).

Before starting with the analysis of gender differences in performance in the different treatments, some remarks on the methodology we used for our data analysis are warranted. As we had to make more than one pairwise comparison, the likelihood that one or more of these comparisons are significant just due to chance (Type I error) increases. Post-hoc tests help in avoiding this bias. As we were only interested in a subset of all possible pairwise comparisons, we applied the Bonferroni correction for multiple testing. For this test, the α-level of each individual test is adjusted downwards taking into account the number of pairwise comparisons. 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). 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 $\alpha_B^{TC} = 0.006$.

In addition to the non-parametric Mann–Whitney *U* test,[10] we performed OLS regressions 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. Table 2 reports descriptive statistics of the experimental outcomes for the two incentive schemes and the gender composition of the teams.

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 as well as chivalry might change a subject's motivation to contribute when participating in a mixed team compared to a single-sex male or female team.

As shown in Fig. 1, 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: $p = 0.271$).[11] However, in mixed teams men solve on average 5.63 games, as compared to 4.83 for women (MWU: $p = 0.012$; OLS: $p = 0.021$).[12] 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. Gender diversity within a team leads to a significant increase of the gender gap in performance.

Our results corroborate some of the findings from studies based on games involving cooperation. In the prisoner's dilemma game, the differences between cooperation rates of men and women were found to be significant in mixed-sex treatments and insignificant in single-sex treatments (although cooperation of women tended to be higher than that of men).[13] In an experiment consisting of mixed groups with respect to gender, Andreoni and Petrie (2008) find that male and female earnings differ significantly. In particular, women make more money than men when no information about previous contributions is provided.[14] This is consistent with our results if positive effort costs in our real task are assumed. Eckel and Grossman (2001) find that men are chivalrous when playing ultimatum games in that they are more likely to accept offers from

---

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

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

[12] 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.

[13] See Ortmann and Tichy (1999). 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.

[14] 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.

Table 2
Descriptive statistics: number of solved games.

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)

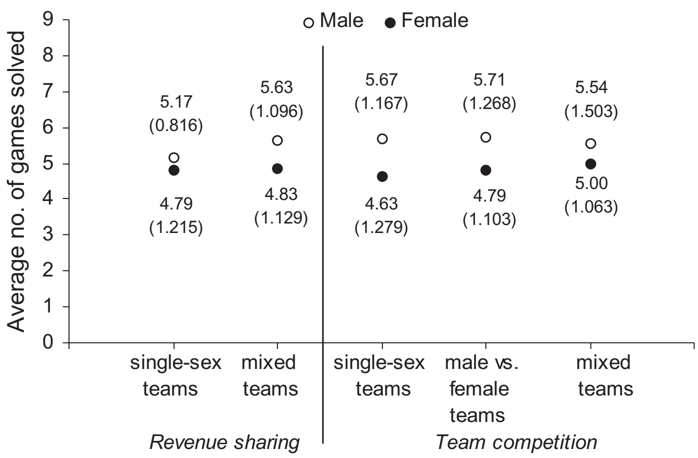

Fig. 1. Average performance of men and women in all treatments (standard deviations in parentheses).

women. Dufwenberg and Muren (2006) observe that women receive higher donations than men in dictator games. These observations are in line with our finding that men exert higher effort (i.e., behave nicely towards women) when paired with a woman in a team.

Note that when aggregating over all treatments, men solved significantly more games than women under both payoff schemes (RS: MWU: $p = 0.006$; TC: MWU: $p = 0.0001$). However, several psychological gender studies show that women outperform men in the memory game (see, for example, McBurney et al., 1997; Tottenham et al., 2003). Although the studies based on the memory task differ with respect to the design (monetary incentives, performance measures, etc.), we take these contradictory results as inconclusive evidence on the ability of men and women to solve memory games. Therefore, in our experiment, we would interpret differences in performance as due to the interaction between the incentive structure and the gender composition of the teams and not to differences in ability.

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 & Schotter, 1997).[15] 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.

[15] A number of studies find a positive effect of competition on performance (e.g., Bornstein & Ben-Yossef, 1994; Bornstein et al., 2002; Erev et al., 1993; Gneezy et al., 2003; Nalbantian & Schotter, 1997; van Dijk et al., 2001).

In our experimental setup, competition does not increase performance on the aggregate level.[16] However, we observe performance effects of incentive schemes *in conjunction* with the composition of the team.[17] As shown in Fig. 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 to the Bonferroni correction (MWU: $p = 0.010$; OLS: $p = 0.007$). And the gender gap in performance vanishes 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 in a single-sex environment (male vs. male and female vs. female teams), there is a significant gap in performance between men and women. This difference gradually decreases and becomes insignificant with an increase in gender diversity, i.e. when moving to male teams competing with female teams and to competition between mixed teams.

Our results are in line with findings from previous studies on tournaments based on individual performance. 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 higher when compared to mixed tournaments. Niederle and Vesterlund (2007) find no significant gender gap in performance in tournaments when men and women compete with each other. 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 et al. (1979) and the studies cited therein, as well as Conti et al., 2001). However, it should be noted that the opposite is observed by Gneezy et al. (2003) who only find a significant gender gap in individual performance in mixed treatments, whereas this 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 cooperate and compete with each other. 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 & Rustichini, 2004, p. 380). Our results add another dimension to this observation by showing that the composition not only of one's own team, but also of the competing team matters.

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

When keeping gender constant and changing only the gender composition for a given incentive scheme, we do not find any significant differences in the performance of both men and women.[18] With respect to the revenue-sharing treatments, this result is in line with the findings in Ortmann and Tichy (1999) for the prisoner's dilemma game where the cooperative behavior of women was independent of the gender composition of the groups. The results from the team competition treatments, however, are in contrast to findings from experiments based on individual performance. For example, Gneezy et al. (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. We hypothesize that the different results can be explained by the fact that in our setup with competition between teams there is an interplay among different aspects of the situation. First, women and men might differ in their response to team competition (as suggested by the work of Dargnies (2009) and Healy & Pate (in press)). Second, the teams setup might induce the participants to put less weight on the competitive aspect than on the cooperative aspect of the situation. Third, this reaction might depend on the gender composition of both, the own and the competing team.

Our dataset 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, Fig. 1 might suggest that men perform worst when the benefits are shared with another man in a cooperative environment compared to when there is either competition or when 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 too small to become significant with the given number of observations, but in some treatments men and women are affected in opposite directions, leading to significant between-gender differences as reported in the previous subsections.[19]

[16] 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.

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

[18] For example, men in RS with single-sex teams and men in RS with mixed teams (MWU: $p = 0.053$; OLS: $p = 0.178$); women in TC between single-sex teams with competition among women only and women in TC between mixed teams (MWU: $p = 0.237$; OLS: $p = 0.271$).

[19] The statistical analysis of differences between the gender gap in different treatments reveals no significant 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.

5. Discussion

In this section, we aim at explaining the observation that gender diversity widens the performance gap in the simple team-pay environment while it decreases the gap with team competition. Given the complexity of the games used in the experiment and in particular the interplay between cooperation and competition in the treatments with team competition, one single theory will not be sufficient to explain the behavioral patterns in all different decision situations. Thus, we will evaluate our results in the light of several sociological and psychological theories of gender and competition.

First, some specifics of our experimental task in comparison to the literature should be mentioned that can be responsible for the overall lower performance of women compared to men. In our computerized memory game task, speed was more important than the number of correct or false tries, as in previous psychology experiments where women outperformed men. Also, the task of playing a computerized memory game possibly activated gender differences with respect to computer games. Finally, the revenue sharing treatment was not framed as a situation about helping behavior or cooperation, which might have elicited a higher performance of women.

To explain the gender gap in mixed teams with revenue sharing compared to single-sex teams, gender stereotypes that are activated by the experimental setup can be decisive. For example, if there is a gender stereotype that women are less able to play computer games than men, this can explain the overall lower performance of women because women who expect to be less able might put in less effort and might also react less to the monetary incentives. Moreover, a common gender norm for men is chivalry which has been related to the gender stereotype of women as passive and unable to act for themselves.[20] Chivalry would predict higher effort of men in mixed teams compared to all-male teams.

An alternative or complementary explanation for the observed higher gap in performance in mixed teams is based on the salience of gender in treatments where both men and women are present. In general, salience makes men and women exert more effort in teams with gender diversity than in teams where only men or only women are present. We indeed observe higher effort levels of men and women in revenue sharing with mixed teams compared to single-sex teams. At the same time, salience is not sufficient to explain any differential reactions of men and women. However, when paired with gender stereotypes and gender norms such as women being less able to play computer games and men being chivalrous, the relatively stronger increase in performance of men relative to women when gender is salient can be explained. It can also explain the findings of significant gender differences in social dilemma situations in environments with gender diversity as observed for example by Ortmann and Tichy (1999) as well as Andreoni and Petrie (2008). They observe a higher gender gap in mixed groups but in their case women cooperate more than men. This is in line with our explanation as long as the gender stereotypes activated in the game and reinforced by salience go into the opposite direction, such as that women are supposed to be helpers.

The introduction of competition between teams adds a new dimension to the complexity of the decision situation, i.e., the interplay between cooperation and competition. Thus, the role of salience, gender stereotypes and gender norms can be muted or even offset. For example, the role of chivalry might be limited in a competitive environment. In addition, chivalry does not make a clear-cut prescription when there are women in one's own and in the competing team. The literature on evolutionary biology and sociobiology (see Knight, 2002) offers some insights on competition and gender. While competition is seen as the natural environment for the male species, 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, in particular in a single-sex environment where other effects such as salience are absent. This offers a possible explanation for the observed gender gap in the treatments with competition between single-sex teams (and for the marginally significant difference in the treatment with competition between male vs. female teams). Our results show that the male–female performance gap in competitive environments is smaller when mixed teams are formed. Possibly, the sociobiological gender roles specific to the competitive environment are more muted in a situation with gender diversity.

We also observe that women in team competition with mixed teams or women competing against men tend to increase effort relative to competition between all-female teams. Based on the theories reported above, this finding can be explained. That competition in the presence of men makes women spend more effort can be due to the salience of gender which makes women work harder, such that the gender gap vanishes. Also, if chivalry is less relevant in a competitive environment, the gender stereotype of passive women is likely to be weaker as well. The loss of relevance of gender stereotypes in mixed competition can also account for the results of Gneezy and Rustichini (2004) who find that girls compete harder in mixed-sex tournaments than when they are among themselves. In contrast, the finding of Gneezy et al. (2003) that the performance of women is higher in single-sex competition than in mixed-sex competition cannot be easily accommodated with the theories above. Sociobiology would clearly predict a low performance of women when they compete among themselves, independent of whether they compete in teams or as individuals. Gneezy et al. (2003) ascribe the different results to the differences of their setup compared to the task used by Gneezy and Rustichini (2004), where there is continuous feedback on (relative) performance. Our design with teams is different from both studies due to the cooperative aspect in team work.

[20] Behavior is labeled as "chivalrous" if it is "characterized by pure and noble gallantry, honor, courtesy, and disinterested devotion to the cause of the weak or oppressed" (Oxford English Dictionary, 1971). For the finding that chivalry is linked to gender stereotypes about women as virtuous and lacking agency, see Altermatt, Cohen, and Johnson (2004). For a conceptualization of chivalry as a norm see Pruitt, Carnevale, and Van Slyck (1986).

Overall, the results from various studies indicate that the activation of gender stereotypes depends crucially on the environment.

Finally, the only within-gender effect we observe is that the performance of men is highest when women are present or with competition. This is not surprising in the light of the theories discussed above. The presence of women activates chivalry in cooperative environments whereas competition is a natural environment for men to increase their performance. Without any one of the two, effort levels of men are significantly lower.

To sum up, as both cooperative and competitive elements interact with gender roles and stereotypes especially when teams compete with each other, the determinants of behavior are multi-dimensional and complex. However, gender stereotypes, salience of gender as well as evolutionary biology and sociobiology together can rationalize our findings.

6. Conclusions

This paper reports on an experiment designed to study the relevance of gender for performance in team work. In two-person teams with non-interdependent tasks, we focus on gender differences with respect to the composition of a team under two different incentive schemes. Participants learned about the gender of the other players indirectly. This entails that they knew the gender composition of the own (and competing) teams for sure only in the single-sex teams. The results of the experiment suggest 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, gender diversity widens the performance gap with revenue sharing, but reduces it in team competition. We hypothesize that this is due to gender effects in the team environment interacting with differences in the competitiveness of men and women. In particular, gender stereotypes, salience of gender and gender differences with respect to competition can be seen as complementary to organizing the findings.

Given our results, there is a conflict between the firm's objective to maximize overall performance and the potential goal to increase gender equality.[21] For example, in the case of revenue sharing, the difference between men and women is mainly due to increased effort by the men with mixed teams. In order to satisfy both objectives, 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. It 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.

More evidence is needed to evaluate whether our findings can be generalized, for example, by employing other tasks, by increasing the team size, or by studying different cultures. We believe that the role of gender in team work has many 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.

Acknowledgments

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, the First Annual IZA Workshop on Behavioral and Organizational Economics, the 5th Tilburg Workshop on Psychology and Economics, the 5th International Meeting on Experimental and Behavioral Economics as well as two anonymous referees of this journal. Financial support by the Deutsche Forschungsgemeinschaft (DFG) through SFB 649 "Economic Risk" and SFB Transregio 15 ("Governance and The Efficiency of Economic Systems"), is gratefully acknowledged.

Appendix A. OLS regression

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

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

where

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

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

[21] For example, a firm might aim at equal chances of promotion for women and men in comparable positions. If women perform worse than men, men are more likely to be promoted based on their performance.

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

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

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

OLS for number of games solved

Number of games solved	Coefficient	Std. err.	t	$P > \|t\|$
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}.

A.1. 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 performance 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 performance 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 performance 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 do not understand something, please raise your hand and 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 **on 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 min. As soon as you have solved a game successfully, please announce this by raising your hand and 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 start of the experiment you will receive an initial endowment of 3 euros.

If you do not understand something, please raise your hand and 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 do not 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 **on 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 **on 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 four points and the other group will have four points subtracted. Should your group's score be lower than that of your opponents, your group will have four points subtracted, while the other one will receive four 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 four 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 start of the experiment you will receive an initial endowment of 3 euros. In the case of losses, these will be subtracted from your initial endowment.

If you do not understand something, please raise your hand. We will come to you and answer your question individually.

References

Adams, R. B., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics, 94*, 291–309.

Altermatt, T. W., Cohen, D., & Johnson, B. A. (2004). Chivalry and stereotypes about women. *Mimeo*.

Andreoni, J., & Petrie, R. (2008). Beauty, gender and stereotypes: Evidence from laboratory experiments. *Journal of Economic Psychology, 29*, 73–93.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics, 116*, 293–312.

Booth, A. L., & Nolen, P. J. (2009). Choosing to compete: How different are girls and boys? IZA Discussion Paper No. 4027.

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.

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

Bornstein, G., & Gneezy, U. (2002). Price competition between teams. *Experimental Economics, 5*, 29–38.

Bornstein, G., Gneezy, U., & Nagel, R. (2002). The effect of intergroup competition on group coordination: An experimental study. *Games and Economic Behavior, 41*, 1–25.

Cadsby, C. B., & Maynes, E. (1998). Gender and free riding in a threshold public goods game: Experimental evidence. *Journal of Economic Behavior and Organization, 34*, 603–620.

Conti, R., Collins, M. A., & 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.

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

Dargnies, M.-P. (2009). Team competition: Eliminating the gender gap in competitiveness. *Mimeo*.

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

Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men? Evidence from dictator experiments. *Economic Journal, 108*, 726–735.

Eckel, C. C., & Grossman, P. J. (2001). Chivalry and solidarity in ultimatum games. *Economic Inquiry, 39*, 171–188.

Erev, I., Bornstein, G., & 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.

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

Gneezy, U., & Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review (Papers and Proceedings)*, 377–381.

Gupta, N. D., Poulsen, A., & Villeval, M. C. (2005). Male and Female Competitive Behavior: Experimental Evidence, IZA Discussion Paper No. 1833.

Hamilton, B., Nickerson, J., & Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy, 111*(3), 465–497.

Healy, A., & Pate, J. (in press). Can teams help to close the gender competition gap? *Economic Journal*, doi:10.1111/j.1468-0297.2010.02409.x.

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

Kandel, E., & Lazear, E. (1992). Peer pressure and partnerships. *Journal of Political Economy, 100*, 801–817.

Knight, J. (2002). Sexual stereotypes. *Nature, 415*, 254–256.

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

McBurney, D. H., Gaulin, S. J. C., Devineni, T., & Adams, C. (1997). Superior spatial memory of women: Stronger evidence for the gathering hypothesis. *Evolution and Human Behavior, 18*, 165–174.

Milliken, F. J., & Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *The Academy of Management Review, 21*, 402–433.

Moely, B. E., Skarin, K., & Weil, S. (1979). Sex differences in competition—cooperation behavior of children at two age levels. *Sex Roles, 5*(3), 329–342.

Nalbantian, H. R., & Schotter, A. (1997). Productivity under group incentives: An experimental study. *American Economic Review, 87*, 314–341.

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

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

Ortmann, A., & Tichy, L. K. (1999). Gender differences in the laboratory: Evidence from prisoner's dilemma games. *Journal of Economic Behavior and Organization, 39*, 327–339.

Pruitt, D. G., Carnevale, P. J. D., & Van Slyck, B. F. M. (1986). Gender effects in negotiation: Constituent surveillance and contentious behavior. *Journal of Experimental Social Psychology, 22*, 264–275.

Sell, J. (1997). Gender, strategies, and contributions to public goods. *Social Psychology Quarterly, 60*, 252–265.

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

van Dijk, F., Sonnemans, J., & van Winden, F. (2001). Incentive systems in a real effort experiment. *European Economic Review, 45*, 187–214.

Wiley, M. G. (1973). Sex roles in games. *Sociometry, 36*(4), 526–541.