UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

FACULTY OF BUSINESS ADMINISTRATION AND ECONOMICS

Working Paper Series

Working Paper No. 2017-02

Gender Diversity is Detrimental to Team Performance: Evidence from a Field Experiment

Bernd Frick, Anica Rose and André Kolle

January 2017

WIWI

Gender Diversity is Detrimental to Team Performance:

Evidence from a Field Experiment

Bernd Frick[a]

[a] University of Paderborn, Department of Management, Chair of Organizational, Media and Sports Economics, Warburger Str. 100, D-33098 Paderborn. bernd.frick@upb.de.

Anica Rose[a, *]

[a] University of Paderborn, Department of Management, Chair of Organizational, Media and Sports Economics, Warburger Str. 100, D-33098 Paderborn. anica.rose@upb.de.

* Corresponding author

André Kolle[a]

[a] University of Paderborn, Department of Management, Chair of Organizational, Media and Sports Economics, Warburger Str. 100, D-33098 Paderborn. andrekolle@arcor.de.

Abstract

We contribute to the inconclusive empirical research on the relationship between team gender diversity and team performance by investigating the returns to gender diversity in academia. Using a unique sample with 164 randomly formed student teams, we show that gender heterogeneity adversely affects team performance in a business strategy game. Both all-female and all-male teams outperform gender-heterogeneous teams in terms of financial success. We find evidence for the detrimental gender diversity effect to increase with task complexity. Our findings suggest that all-men and all-women teams do not differ in their strategic management behavior.

Keywords: Gender; Diversity; Teams; Performance; Business strategy game

APA: Diversity; Teams; Group performance

PsycINFO: 3020

JEL: J16, C93

1. Introduction

Due to the recent increase in women's participation in higher education on the one hand and their rising labor force participation on the other hand, the diversity of firms' workforces has increased considerably (Konrad, Prasad, & Pringle, 2006). Simultaneously, companies have increasingly turned to team production in order to satisfy customers demanding individually tailored and highly specialized products and services (Hamilton, Nickerson, & Owan, 2003). A major challenge, thus, is to staff teams with members from a heterogeneous pool of employees in a way such that team performance is maximized.

Given these developments, two strands of literature have rapidly expanded. First, a large body of empirical research on men's and women's preferences for risk, competition or social environments reveals significant gender differences in mental attitudes as well as decision-making approaches, which eventually translate into differences in performance outcomes (Andersen et al., 2013; Azmat & Petrongolo, 2014; Bertrand, 2011; Gneezy, Leonard, & List, 2009; Niederle, 2014). This research occasionally focuses on the impact of gender in competitive bilateral relationships suggesting that same gender pairings may cause rivalry and, therefore, lead to inefficiencies (Sutter et al., 2009). Accordingly, a second strand of literature focuses on the diversity-performance link in teams. Due to changes in legal rules (such as affirmative action programs or anti-discrimination laws) and the expected economic benefits arising from the employment of women (which may come indirectly in the form of reputation), gender is one of the most commonly studied team-level diversity attributes (Williams & O'Reilly, 1998).

From a theoretical perspective, the costs of and the returns to diversity are modeled as a trade-off between productive synergy effects stemming from team members' complementary and relevant skills, which may lead to constructive conflicts and better final decisions on the one hand, and the costs for communication and coordination that may hinder knowledge transfer on the other hand (Lazear, 1999). Cooperation is disrupted as soon as social-categorization processes lead to a separation between in- and out-group members (Tajfel, 1981; Turner et al., 1987). As a consequence, in-group favoritism and out-group stereotyping increases intragroup conflicts and, thus, negatively affects group processes and outcomes (van Knippenberg, de Dreu, & Homan, 2004). While research suggests that diversity in task-related characteristics (such as experience, tenure, and skills) particularly increases synergy potentials, social categorizations (which are usually detrimental to performance) are activated by individuals' readily observable (bio-) demographic attributes – gender in particular (Horwitz & Horwitz, 2007; Stangor et al., 1992). In line with this, meta-analytic evidence for a positive relationship between gender heterogeneity and team performance is rather limited (Kochan et al., 2003; Pitts & Wise, 2010). Most of the available studies conclude that team heterogeneity with respect to gender has either no statistically significant (Bowers, Pharmer, & Salas, 2000; Horwitz & Horwitz, 2007; Schneid et al., 2014; Webber & Donahue, 2001) or even a significant negative performance effect (Bell et al., 2011; Mannix & Neale, 2005; Milliken & Martins, 1996; Williams & O'Reilly, 1998).

While research using field data usually analyzes the effect of team diversity (for example in corporate boards) on organizational-level outcomes (Joecks, Pull, & Vetter, 2013; Kochan et al., 2003), few – if any – of these studies convincingly address the issue of reverse causality. In laboratory experiments, on the other hand, researchers can account for endogeneity problems more efficiently. Although laboratory evidence has recently been found to predict behavior in the field (Burks et al., 2016, Dai, Galeotti, & Villeval, 2016), the external validity may nevertheless be limited (Azmat & Petrongolo, 2014; List, 2011). As an example, gender dynamics of groups' decision-making processes in the laboratory might differ since outside of the laboratory teams usually have to focus on more complex team tasks simultaneously. Moreover, a reality for teams in organizations is that they usually spend a lot of time together. Such aspects are rather difficult to be designed in the lab. Field experiments can confront these problems as subjects typically do not know that they participate in an experiment and, thus, reveal their true preferences (Levitt & List, 2009; List, 2011). Hence, drawing on student data and academic tasks (such as group presentations, written assignments, or business simulation games) is an appropriate setting to study the effect of teams' gender composition on performance outcomes. Student teams are characterized by their temporary existence. Hence, findings may be generalized to similar organizational settings, such as for example project teams. Previous findings using student data suggest that teams consisting of women only are outperformed by teams with at least one male member (Apesteguia, Azmat, & Iriberri, 2012; Becker et al., 2006). Apesteguia et al. (2012) use field data from an online business simulation game played worldwide to investigate performance outcomes of different gender compositions of some 16,000 teams with three members each. They find that all-female teams are the worst performers and that teams are slightly – but not statistically significantly – found to perform best when consisting of two male and one female member. All-female teams were found to pursue inefficient pricing and investment decisions. Similarly, in a laboratory experiment, Becker et al. (2006) find that women-only teams perform worse as they are less aggressive than men (for example in their pricing strategy). The authors assume that in gender-mixed teams female members adapt to men's management styles, which is why not only all-male but also gender-balanced teams outperform all-female teams. Both studies, however, have particular methodological weaknesses. Apesteguia et al. (2012) cannot account for confounding issues stemming from endogeneity, since participants have not been randomly assigned to the teams. Hence, one cannot rule out the possibility of a non-random distribution of students' ability across teams. While Becker et al. (2006) overcome these limitations, the small number of observations (N=17) restricts their analyses to mean comparison tests and, thus, limits the generalizability of the results. Contrary to the findings discussed so far, Hoogendoorn, Oosterbeek and van Praag (2013) find that gender-balanced teams outperform male-dominated teams. They use data from a program, in which students are randomly assigned to 43 teams according to their gender with the aim to start their own business. Nevertheless, as the range of the share of women in a team is restricted and varies between 20 and 60 percent, the authors do not observe all-female or all-male teams. Moreover, they have only a small number of observations on female-dominated teams, which means that no conclusions concerning any symmetric effects of gender diversity can be drawn. Only few studies so far suggest symmetric effects, i.e. that gender-diverse teams perform better (i.e. are more efficient) than all-female and all-male teams in academia (Lee & Farh, 2004; Orlitzky & Benjamin, 2003; Rogelberg & Rumery, 1996; Umans, Collin, & Tagesson, 2008). Instead of business outcomes, these studies use grades for academic achievements in for example presentations or case studies as outcome variables. Such tasks, however, might have a lower impact than outcomes from simulation games as students do not work together over a longer time period in managing a business.

Taking into account the available evidence and the studies' limitations (with respect to the formation of teams, the research designs, the number of observations, and/or the conceptual understanding of gender diversity), the relationship between team gender diversity and team performance particularly in academic environments remains an open question. In this paper, we aim to contribute to the discussion and analyze whether and how the gender composition of 164 undergraduate student teams affects the strategic decision-making behavior as well as the performance in a computer-based business strategy game. We conjecture that our research design has a number of distinct advantages: First, we observe the behavior of tomorrow's workforce (and even prospective managers) in a natural setting. The teams we observe take the role of executives, who are responsible for the operation and management of a manufacturing company and compete against other student teams in a particular market. Team performance is measured by each company's final share price and their probability of filing for bankruptcy. Due to the close proximity between virtual business simulations and real-life scenarios, business strategy games have become an integral part of today's education of business and economics students in order to enhance their strategic and operational skills and, thus, to prepare students for real-life business decisions (Faria et al., 2009). Second, our dataset is unique in that we observe teams over a period of eight consecutive weeks (Apesteguia et al. (2012) also observe teams over several periods, but restrict their analysis to data from the first round only). We account for the fact that task assignments become more challenging in advanced periods and, thus, study the impact of gender diversity at different levels of task complexity – an issue that has often been neglected in empirical research so far (Schneid et al., 2014). Third, we are able to glance into the "black box" and to study the influence of team gender diversity on various intragroup decisions that, in turn, determine the teams' final outcomes. This allows us to study not only the differences in the decision making between gender-homogeneous and gender-heterogeneous teams, but also to analyze possible differences between all-men and all-women teams.

Our analyses yield quite surprising results: Gender diversity (measured by Blau's (1977) index) affects both measures of team performance statistically significant and adversely: A 0.1-unit increase in a team's gender diversity decreases the final share price by 16 percent and increases the bankruptcy probability by 2.7 percentage points. These effects remain robust when controlling for the diversity in the team members' individual skills (such as their individual course performance) as well as for other team features (for example team and market size). More precisely, we find that gender-mixed teams are outperformed by pure men- as well as pure women-teams and that team performance is worse when gender composition reaches parity. Further results indicate the moderating effect of task complexity: The negative effect of gender diversity is largest as soon as tasks become more challenging and, thus, require a higher level of intragroup cooperation in order to solve the task. More detailed analyses of intragroup decisions reveal that while female-dominated and male-dominated teams do not differ in their decision-making strategies, gender-diverse teams particularly fail to invest in the company's sales activities. Hence, the costs from intragroup biases seem to outweigh the benefits from variety.

2. Research design

2.1 Institutional setting

We use data from a computer-based management game that is part of an undergraduate course in organizational economics at a medium-sized German university. During the lecture part of the course, students are taught basic principles in organizational economics that they have to use and apply to real-life cases in a written final exam at the end of the semester. The weekly lecture is accompanied by a business strategy game ("TOPSIM – General Management"), in which teams of three to seven students manage a fictitious company and compete against five to seven other teams in a simulated market. A student's final course grade is composed of the individual performance in the exam and the team's performance in the simulation game. According to the software provider, different versions of the business simulation are widely used for teaching purposes at universities inside and outside of Germany, but are also used for training purposes by 70 percent of the 30 largest listed corporations in Germany. In the version we have used, teams work together over a period of eight weeks managing a company that produces photocopiers. Table 1 provides an overview of the challenges teams have to cope with in the particular periods: While in periods 1 and 2 they have to handle only a single product on one market, the subsequent periods are characterized by an increase in the degree of complexity, i.e. the teams are confronted with the decisions of re-launching old or introducing new products and entering foreign markets. According to students' feedback, the need for more sophisticated strategies implies that teams are under enormous pressure in these later periods, which then requires a higher level of team cooperation. [Insert Table 1 here]

Teams have to develop and adapt business strategies based on the basic endowments in the different functional areas of their company and based on the economic situation in the respective market. Both aspects are determined by the software at the beginning of each period and are identical for all teams in a particular market. Companies' progress is a function of how teams as well as the competitors react to the market conditions. In each period, teams have to decide on a number of variables driving company performance: the sales price(s) of the product(s), marketing and sales activities, R&D expenses, investment in environmental facilities or process optimizations, and financial resources. A manual helps students to develop appropriate strategies. Teams that are able to cope with these various challenges and to make rational decisions after carefully weighting all opportunities and risks have the highest success probability. As strategy development and decision making is very complex, it requires logical thinking and creativity as well as intensive communication and interaction of the teams' members. However, team meetings and task assignments in teams are not institutionalized or externally determined by the instructor. That is, teams internally agree upon where, how often and for how long they meet to debate their strategies. Thus, in some teams decisions may be worked out together in group meetings, while in others, team members specialize on different functional areas and discuss their suggestions at a weekly summit. As a consequence, we have no reliable information on the teams' internal interaction and communication processes and cannot even be sure that all team members participate to the same extent in the decision-making processes. However, as in real-life the teams first have to decide on internal processes. Due to issues of favoritism and stereotyping, it might be more challenging in gender-mixed teams to decide upon a team leader, a company strategy or a specific organization of the team. This can then detract the focus on the essential task and lead to less efficient decisions. At the end of each period, each group transmits its decision to a course administrator, who processes all the data and simulates each team's business performance of that particular period. The applied software uses different algorithms to calculate each company's periodic share price based on how the company as well as their competitors decided on the above-described variables given the predetermined endowment and economic circumstances. The group then receives a full report documenting how the team performed as well as an updated economic forecast for the subsequent period in order to make new business decisions.

2.2 Team assignments

At the beginning of each semester, course participants are randomly assigned to a team by using random number tables. In our dataset, we observe, on average, more than 30 teams per semester. The teams are, in turn, randomly assigned to a market by the simulation software, which does not allow more than eight companies in a market. Hence, teams are formed before their actual performance takes place and remain fixed in their composition over the entire eight-week period. We are, thus, confident to avoid any causality problems and issues originating from self-selection into teams or markets. The initial conditions are identical for each team on a market, while initial market conditions depend on market size (i.e. number of competitors), which we account for in our empirical analyses. The market developments, however, are totally independent from each other, i.e. companies compete only with other companies in their particular market. Following the random assignment of students to teams they are no longer allowed to withdraw from the course, nor are they permitted to switch teams. Nevertheless, students can simply stop attending class and team meetings. Unfortunately, we lack exact information on these early dropouts, even though teams are explicitly encouraged to report members who do not participate in decision making at all. In a few instances of obvious free-riding, the course administrator helped to solve conflicts. Nevertheless, more subtle forms of conflict may have occurred. Apart from that, tentative evidence suggests that the number of unreported dropouts is very small and can, therefore, be neglected.

2.3 Incentives

The goal of the business strategy game is to maximize the firm's share price at the end of the eight-week period. Nevertheless, teams may underperform (i.e. make inappropriate business decisions) or may be driven out of the market by competitors prior to the official end of the game. Outperformed teams end up with a share price of zero and are excluded from any further periods. The final share price translates into a team grade with the best possible grade going to each market's best performing company. The remaining companies are graded relative to the best performers. In other words, the highly competitive setting incentivizes teams not to cooperate with competitors. The team grade constitutes between 30 and 50 percent (depending on the year, in which the class took place) of an individual's overall course grade. The remaining part of the course grade (either 50 or 70 percent) is then based on the individuals' results in the final written exam. In order to avoid any strategic behavior in the final test, the grades from the business simulation are made available to students after the written exam. Obviously, students whose team went bankrupt anticipate a worse grade than students managing surviving firms. Therefore, the former are more likely to skip the final exam. In such cases, we lack information on these individuals' performance. Overall, the course grade makes up 1/18 of students' entire degree (10 out of 180 ECTS credits). We therefore believe that the importance of the team outcome for the individuals' final grades is large enough to motivate students to exert appropriate levels of effort in the team task.

3. Data and descriptive analyses

During the observation period (2009 to 2013), 755 students participated in the course, which results in five repeated cross-sections of data with 166 different teams. Yet two of the 166 teams had to be excluded from the analyses as all members withdrew from the final examination and, therefore, did not complete the class. Thus, we use 164 team-level observations in our econometric analyses.

3.1 Performance measures

The team performance measures in our analyses are the firm's final share price and whether a particular firm went bankrupt or not. Information on the share price was retrieved from firms' reports after the last particular period had been simulated. On average, firms yield a final share price of 228.83 € (see Table 2). In order to report elasticities for our continuous dependent variable and to simultaneously include firms that dropped out early and, thus, had a share price of zero, we calculate the log of the share price as follows: LG ($PRICE_i$+1). Bankruptcy is denoted by a dummy variable that takes the value of one if a firm could not generate profits any longer, and is zero otherwise. Firms were considered to be bankrupt as soon as their equity capital fell below zero in any of the periods played. These firms were immediately excluded from the game and assigned a share price of zero in order to avoid "sabotage activities". In our sample, 30 percent of the firms had to file for bankruptcy during the course of the eight-week period. In addition to the teams' final performance, Table 2 displays the average share prices of the weekly population at the beginning of the respective period. The decreasing number of observations reflects the disqualification of those teams that went bankrupt in the previous period. [Insert Table 2 here]

3.2 Team demographics

From the course registration, we have information on the students' names, field of study and degree pursued. We use names to infer a student's gender and ethnic origin. On average, 38.4 percent of the students in our sample are female and 20.9 percent of the sample have a migration background. Two research assistants independently classified names into 'German' and 'non-German' sounding. The field of study provides information on students' majors. The majority of course participants (on average 75.9 percent) come from the business and economics faculty, while the course may also be attended by students from other business-related fields of study including business education, engineering, and computer science. The distribution of degrees strongly depends on the year of observation. While in 2009 about one third of the students were registered in a diploma degree program, in 2013, as a result of the Bologna reforms, all course participants pursued a Bachelor's degree.

In our analyses, we are primarily interested in the effect of teams' gender composition on team performance. In order to fully capture the gender composition effects, we measure gender diversity in two different ways. According to the scaling of gender as a dummy variable, we calculate Blau's index of gender diversity (Blau, 1977) as: $GENDIV=1-\sum_{i=1}^{K} p_k^{\,2}$, where p_k denotes the share of team members p belonging to category k. As for gender k=2 (i.e., male and female, respectively), the Blau index ranges from 0 to 0.5. For the ease of interpretation, we standardize the Blau index by dividing it by its theoretical maximum (k-1)/k. Consequently, the Blau index has a minimum of 0 (complete male or female homogeneity), a maximum of 1 (complete gender heterogeneity), and a mean of 0.72 (sd=0.32). According to this definition, however, maximum gender heterogeneity can only be achieved in teams with an even number of members. In order to test for a non-linear relation between female team members and team performance, we further use the share of women in a team (*S_FEMALE*) and also include the variable's quadratic term. Descriptive statistics reveal that there are statistically significant (p=0.0003) performance differences between gender homogeneous and heterogeneous (i.e. at least one member of the opposite sex) teams: The average final share price of the completely homogeneous teams (mean=361.08) is 73 percent higher than the average share price of the heterogeneous teams (mean=208.35). Moreover, Figure 1 illustrates the differences in the average share prices after each period, which are statistically significantly different between gender homogeneous and gender heterogeneous teams from period 4 until period 8. [Insert Figure 1 here]

In addition to the variables of theoretical interest, we include other covariates reflecting the demographic composition and diversity in a team. Similar to gender diversity, ethnic origin and field of study were included as standardized Blau indices and – alternatively – as percentage shares (and their quadratic terms). We further use data from the final exam to measure the teams' diversity in individual performance. Due to data protection and security regulations, we are unable to observe the overall academic performance of the students on the basis of their current grade point averages. As the relative importance and the design of the exam (multiple choice and "open questions") have varied over the observation period, students' individual performances have been standardized, that is, the points reached in the final exam were divided by each year's maximum number of points obtainable. Overall, 676 students participated in the final exam and, on average, achieved 54.5 of the maximum points (sd=0.13, min=0.18, max=0.99). We use the within-team standard deviation in individual performance as a proxy for ability diversity. In order to compare ability diversity across groups of different size, we normalize the sample standard deviation and divide the formula by its theoretical maximum: (u-l)/2, where u stands for the upper and l reflects the lower limit of the observed individual performances. Consequently, the minimum value still equals 0, meaning that team members are completely homogeneous with respect to their ability. The maximum value is 1 and reflects complete separation in the sense that 50 percent of the individuals are at the upper and 50 percent at the lower bound (Harrison & Klein, 2007). As around 10 percent of the students skipped the final exam, we had to exclude these before calculating for each team its diversity in individuals' abilities. Apart from heterogeneity in individual performance, we control the individuals' different preferences and effort levels by including independent variables measuring the share of team members belonging to the top (bottom) ten percent performers in the final exam in a given year. Here, we assume that performance in the final exam is correlated with the overall study performance, the latter of which we are unable to observe. Moreover, we control for team size in order to account for cooperation and communication problems that might be more severe in larger teams. Similarly, market size as well as the number of bankrupt competitors are used as proxies for market conditions – even though economic conditions are determined by the software's algorithm. Since a large number of bankrupt competitors might, however, also be driven by the presence of particularly incompetent teams, we weight each bankruptcy by the length of time elapsed since the start of the game (i.e. the period of bankruptcy). Thus, we assume that groups exiting early (indicated by low levels of *BANKCOMP_PERIOD*) are less able in responding to the game's challenges than teams which go bankrupt in later periods.

Finally, we use year dummies to account for any unmeasured differences such as changes in the administrator of the business game and the course lecturer as well as changes in the composition of the final grade. Descriptive statistics of all independent variables included in our analyses are displayed in Table 3. By definition, the shares in demographic characteristics are strongly correlated with the respective diversity measures (see Table 4). In order to avoid spurious results due to multicollinearity, we do not include these variables in our regressions simultaneously. [Insert Table 3 and Table 4 here]

3.4 Empirical strategy

We test our hypotheses using OLS and probit regressions. In all estimations we cluster standard errors at market-level to allow for unobserved heterogeneity across markets. In our full empirical model the share price is estimated as follows:

$$LG\ (PRICE_i+1)=\beta_1+\beta_2\cdot DIV_i+\beta_3\cdot ABILITY_i+\beta_4\cdot EXT_i+\varepsilon_i,$$

where LG ($PRICE_i$+1) is the natural logarithm of the share price of team i. DIV is a vector of diversity measures, ABILITY is a vector of variables denoting diversity in group ability as well as the share of top and bottom 10 percent performers, and EXT is a vector of further control variables, such as team and market size, the weighted number of bankrupt competitors, and the year of observation.

Analogously, we estimate a series of probit regressions with the bankruptcy dummy as the dependent variable. Here, we estimate the following equation:

$$BANKRUPTCY_i=\beta_1+\beta_2\cdot DIV_i+\beta_3\cdot ABILITY_i+\beta_4\cdot EXT_i+\varepsilon_i,$$

where $BANKRUPTCY_i$ takes the value one, if team i goes bankrupt (and is zero otherwise), while all covariates are the same as above.

4. Results

4.1 Baseline results

Table 5 reports the results from OLS regressions on the log of the final share price using different model specifications and diversity measures. The results are based on the full sample of 164 teams. Columns (1) to (2) use the different diversity indices discussed above. The baseline model (1) is gradually expanded to include team ability measures, team and market size as well as the weighted number of bankrupt competitors and year dummies (2). Since Blau's (1977) diversity index does not indicate whether homogeneous teams are composed of female or male participants only, in columns (3) and (4) we use the share of the demographic variables and their quadratic terms to account for possible curvilinear effects on team performance. While column (3) leaves out controls, column (4) includes all covariates as described above. [Insert Table 5 here]

The results in the first two columns of Table 5 demonstrate the significance of the negative effect of gender diversity on team performance. The effect is robust to the inclusion of various control variables. Columns (1) and (2) show that the final share price decreases, if the standardized Blau index of gender diversity increases. More precisely, a 0.1-unit increase in gender diversity ceteris paribus reduces the final share price by 16 percent. As argued above, the Blau indices' intervals depend on team size: In teams with only three members, all team members either have the same sex (*GENDIV*=0) or there is one team member of the opposite gender (*GENDIV*=0.89). Other combinations as well as a 0.1-unit change in the Blau index do not exist. Nevertheless, in teams with more members, intervals between Blau indices become smaller. As an illustration, if we would observe a combination of four male (female) and three female (male) students instead of five male (female) and two female (male) students, this would lead to a 0.1-unit difference in the Blau index. Nevertheless, unit-changes are typically greater; hence, negative effects are even stronger.

The significant u-shaped effect of the share of female team members depicted in columns (3) and (4) of Table 5 reveals that the detrimental gender diversity effect is symmetrical. In other words, gender-mixed teams are outperformed by all-male as well as all-female teams: Performance is estimated to reach its minimum if the share of women falls in a range between 0.48 and 0.51, all other things equal. The u-shaped form of the share of female team members on team performance, hence, indicates that diversity effects are most harmful in gender-balanced teams.

A closer look at the remaining independent variables does not reveal any robust and statistically significant effects for any of the alternative diversity measures (migration background, subject, ability). However, there is some evidence of a u-shaped relationship between the share of team members with a migration background and final share prices as well as the share of business and economics students and team performance. The share of top performers in a team, team size, and year do not have any consistent effects across the different model specifications. The positive and significant effect of the share of low performers might indicate that students who anticipate a bad performance in the exam put more effort into the game in order to minimize the risk of failing the course. The negative effect of market size, which indicates a poorer performance of teams in larger markets, suggests detrimental effects of increased competition. A similar statistically significant and negative effect appears for the weighted number of bankrupt competitors, i.e. the more firms go insolvent in later periods, the poorer the survivors' performance. At first, this result may seem surprising as the probability of success should increase as the number of competitors goes down. We attribute this counter-intuitive finding to the poorer economic conditions in these markets, which are more likely to have caused the early exits than for example the bankrupt teams' lack of abilities.

In addition to the teams' final share prices, Figure 2 reports the predicted probabilities of filing for bankruptcy over the course of the business simulation on the basis of probit regressions. The positive and significant effect of gender diversity on the left-hand side underlines the results presented above. Gender diversity not only harms firms' share prices, but also increases their failure rates. Marginal effects after probit regressions reveal that in an average team with a Blau index of *GENDIV*=0.72, a 0.1-unit increase in gender diversity is associated with a 2.71 percentage points higher probability of going bankrupt. Again supporting the results presented in Table 5, the figure's right-hand side depicts that the share of women is inversely u-shaped related to a team's probability of going bankrupt, i.e. that both purely male and purely female teams outperform teams with (nearly) equal shares of men and women. [Insert Figure 2 here]

In order to investigate the sensitivity of the results reported in Table 5, we estimated the impact of gender diversity on team performance conditional on firm survival. Therefore, we reduced the sample to the 114 teams, whose share prices exceeded zero at the end of the business simulation. The regression estimations in Table 6 indicate that the negative impact of gender diversity on firm performance persists, if firms filing for bankruptcy are excluded. Nevertheless, the detrimental effect is now slightly weaker than in Table 5. Moreover, it again appears that both, men-only and women-only teams perform better than teams, in which male and female students are equally represented. [Insert Table 6 here]

Summarizing, we find that gender diversity is detrimental to team performance in terms of lower share prices and higher failure rates. In addition, the non-linear relationship between the share of women and team performance suggests that gender-balanced teams are outperformed by all other gender combinations and that completely homogeneous teams seem to perform best.

4.2 Periodic analyses

So far, the final share price has been used to capture team outcomes. Since this measure might not reflect the teams' overall performance appropriately, the following analyses draw on teams' periodic share prices. Periodic data are particularly suitable to test the dynamic effects of gender diversity, i.e. if diversity effects differ with regard to changes in the decision-making requirements or an increase in pressure. Table 7 reports separate OLS estimations with the natural log of each period's share price as the dependent variable. The decreasing number of team-level observations reflects the exclusion of those teams that went bankrupt in the previous period: Only five of the 164 teams went bankrupt in periods 1 and 2, while there were 37 failures in periods 3 to 6 as well as another eight bankruptcies in the last two rounds. In the end, 114 teams survived the complete game (69 percent). The periodic estimation results indicate that for the survivors of the current period the negative effect of gender diversity on team performance is not statistically significant during the first two periods. This changes at the beginning of period 3, in which business decisions become more complex. Furthermore, gender diversity has the most harmful consequences for the teams' performance in period 7, in which members have to handle two different products. In the last period the detrimental effect decreases, probably because teams are already familiar with the decisions that have to be taken. [Insert Table 7 here]

According to these findings, changes in the game structure, especially the need to develop new and particularly complex strategies at advanced stages, exacerbate the detrimental effects of gender diversity. In those periods, communication and cooperation in order to adequately analyze market conditions and make the right decisions is of growing importance, but adversely affected by social categorizations, which seem to persist over the course of the eight weeks.

4.3 Looking into the "Black Box"

Our findings indicate that gender diversity negatively impacts the teams' performance and is the worst when both genders are (nearly) equally represented. Moreover, task complexity seems to moderate these negative results. So far, we do, however, not know which intragroup decision-making processes drive these outcomes. In the remainder of this empirical section, we aim to look into the "black box" to better understand the differences in the strategic behavior of gender-homogeneous and gender-mixed teams. Therefore, in line with Apesteguia et al. (2012), we treat the teams' weekly management decisions as outcome variables: This comprises the sales prices for both products (P1 and P2), corporate identity expenses (CI), the employment of additional sales persons (SF), investments in R&D, environmental facilities, process optimizations as well as expenditures for loans. Due to the software's algorithm, which determines the final share prices on the basis of the teams' decisions as a response to pre-determined endowments but also on the basis of companies' reactions to changing market conditions, we refrain from analyzing which of the management decisions determine the teams' periodic share prices.

Table 8 displays the effect of gender diversity (first row) and – alternatively – the impact of the share of women (second and third row) on the standardized values of the different management decisions. Even when controlling for team diversity and ability, team and market size, as well as year and period, the first row indicates that an increase in the teams' gender diversity leads to an under-investment in the company's image (CI). The second and third rows suggest that – due to the statistically significant negative effect of the linear variable *S_FEMALE* and the positive effect of its squared term – both all-male and all-female teams invest more in CI. [Insert Table 8 here]

Thus, gender-homogeneous teams – irrespective of whether they are composed of women or men – seem to develop more adequate strategies. Due to the small number of male-only and female-only teams we refrain from comparing the two groups' decisions. Instead, we differentiate between female-dominated (at least 60 percent women) and male-dominated teams (not more than 40 percent women). Even though we find significant differences in both groups' investments in environmental facilities (see Table 9), there are no statistically significant differences in the remaining strategic decisions that were made. Our results, therefore, indicate that gender-heterogeneous teams do not underperform as a result of task-related differences between men and women. [Insert Table 9 here]

5. Discussion and conclusions

Given that diversity is ubiquitous in educational contexts, this paper attempts to establish an empirical link between team gender diversity and team performance using a unique dataset including 164 student teams. In our field experiment, the teams not only have to manage various functional areas of a company but also have to react to changing market conditions influenced by crisis situations and/or competitors' activities. Over the course of eight weeks, they constantly have to optimize their operating processes, increase the company's productivity or reduce production costs. In order to cope with these challenges, teams need to develop efficient decision-making processes and strategies.

We find strong evidence that gender diversity is detrimental to team performance and that gender-diverse teams are outperformed by both female-only and male-only groups as well as all other gender combinations. In more detail, a 0.1-unit increase in gender diversity decreases the teams' final share prices by 16 percent and increases teams' bankruptcy probability by 2.7 percentage points. In line with Lazear's (1999) theory of a global firm, we conjecture that knowledge transfer in teams might be hindered in more gender-diverse groups. The u-shaped relationship between the share of women and team performance indicates that team performance decreases until gender composition reaches parity and then increases again. Hence, intragroup biases seem to be triggered by the categorization of team members into in- and out-group members on the basis of the social attribute "gender". Harrison and Klein (2007) argue that the "us-them" perspective is particularly elicited once there are two symmetrically distributed oppositions at each end of the continuum. An environment, in which there are two equally sized sub-units is more likely to trigger inefficient biases than a team setting, in which one gender dominates the other in terms of numbers. Those who belong to the underrepresented gender will adapt or subordinate to the majority. With each additional sub-unit member, however, a direct confrontation between both sub-units becomes more likely (Harrison & Klein, 2007; Pelled, Eisenhardt, & Xin, 1999). Our findings from periodic analyses further suggest that gender diversity and, consequently, intragroup biases are most detrimental when team tasks become more complex. Hence, as soon as the consequences of the teams' decisions become more severe, intragroup cooperation seems to suffer the most.

Using information on the teams' weekly management decisions leads us to conclude that gender-heterogeneous teams fail to optimize their respective companies' sales planning. However, we fail to find robust differences in all-female and all-male teams' strategic behavior concerning for example midway outcome variables (with investments in environmental facilities as an exception). Our results are, hence, in line with Paola, Gioia, and Scoppa (2013), who also find that male and female students do not differ in their performance in competitive situations. This can be explained by men's and women's similar educational attainment. As an example, team members must have completed the same introductory business classes during their studies and have to take part in the accompanying lecture. Thus, male as well as female students have the relevant skills to solve their tasks; these skills, however, are not necessarily complementary. Since male and female students do make optimal decisions as long as they do not have to work with each other, task-related gender differences seem to be completely absent. Consequently, two other explanations seem (equally) plausible to explain why gender-heterogeneous teams fail to use their full potential and, thus, yield less efficient solutions than homogeneous teams: On the one hand, differences in men's and women's working attitudes might evoke intragroup biases and conflicts. As an example, women might prefer to work on all business areas simultaneously and, thus, to meet more often, whereas men might prefer working individually on allocated sub-tasks and to meet rather sporadically

(Azmat & Petrongolo, 2014). While both approaches can yield efficient solutions, they might turn out to be incompatible. Moreover, leaders might be determined faster in single gender teams, whereas the social pressure to share in the decision making is perceived to be stronger in mixed-gender environments. Hence, mixed teams are likely to suffer from coordination problems that gender homogenous teams are either able to avoid or to overcome. On the other hand, one might argue that men still have prejudices against women in the sense that they do not want to work with female colleagues, as they assume them to be less able to manage a company. As a consequence, stereotyping would lead to an exclusion of women from communication rounds and decision-making processes, which simultaneously increases discontent among the female out-group and increases the conflict potential. According to Haile (2012), women are unhappy when working with men in the sense that their job satisfaction and affective well-being decreases by some 27 percentage points in gender-diverse workforces. Both explanations trigger polarization and lead to inefficient communication and interaction structures among team members and, thus, hinder gender-mixed teams to reap the benefits of their potential (which is, of course, likely to be as large as the potential of male-only and/or female-only teams). Thus far, however, we are not able to unambiguously identify the intragroup processes that lead to social categorizations, which, in turn, cause conflicts and negatively affect team outcomes.

Our findings have various implications for the formation of temporary teams (at least in higher education institutions in Germany). Although our results seem incompatible with the findings reported in previous studies using data from educational contexts – suggesting that gender-diverse teams are more efficient than all-female (Becker et al., 2006; Apesteguia et al., 2012) or male-dominated teams (Hoogendoorn et al., 2013) – we are confident that our research design avoids several drawbacks of previous studies. First of all, our results are not only plausible (men and women perform equally good as long as they work with other persons of the same gender), but we can also completely rule out problems of endogeneity: Since in our field experiment teams have always been formed randomly, we are able to eliminate self-selection on the basis of sympathy or ability. In previous studies, teams rather formed themselves (Apesteguia et al., 2012). Second, using the final share price instead of the share price of the first round only seems to be a more appropriate performance measure, since over the course of time the effects of intragroup diversity are likely to change (something Apesteguia et al. (2012) failed to do). Since our dataset allows investigating the dynamic effects of gender diversity, we can analyze the performance of (student) teams over time, i.e. as tasks become more complex and market conditions change. Third, and in contrast to previous studies, our analyses yield robust findings when using different measures of gender diversity (Blau's (1977) diversity index as well as the share of women). The data we use is not restricted to "predominantly" male/ female and "gender equal" teams, but we also observe purely male and female teams, i.e. we are able to use the whole range of possible gender compositions.

In the end, some caution is warranted when it comes to the generalization of our findings, since there are a couple of setting-specific peculiarities that might have mediated the possible benefits of gender diversity. First, the teams we observe only have to work together for a limited amount of time. Presumably, the goal of each individual is to complete the class successfully and students might, therefore, not invest too much time and effort into maintaining social contacts. This temporal restriction might also apply to project teams in other organizations, but in a company the necessity or even obligation to cooperate again with prior team members is likely to be higher. Second, the course grade alone might not exhibit an incentivizing effect on students. Even though we control for as many as possible observable individual-level characteristics, we are unable to control for differences in students' individual preferences regarding course objectives and effort levels as well as grades. However, since we fail to find statistically significant differences between those years, in which the group grade made up 50 instead of 30 percent of the final course grade, we are confident that these factors are of minor importance only. One could, however, also increase students' stakes by introducing financial rewards for the best teams over and above the top grades going to the top performers to check whether the detrimental gender diversity effects still persist. Third, in this paper we use data from one German university. Other national or organizational cultures might exhibit a particular moderating influence, suggesting the use of similar data from universities in other countries to see whether our results can be replicated for more "gender equal" as well as for more "gender unequal" societies (Schneid et al., 2014). Finally, we have no systematic, but only anecdotic evidence from individual members about internal group processes. More information on intragroup procedures is required to properly distinguish between conflicts triggered by gender differences in working attitudes or by stereotyping behavior. Therefore, it might be helpful to expand the dataset and to distribute questionnaires at different points during the game or to let students keep logbooks, from which information on group meetings or actively involved decision makers can be retrieved. The individuals' attitudes towards cooperation might further be tested in an experiment prior to the start of the strategy game. In line with Burks et al. (2016), who use a form of the Prisoner's Dilemma to categorize subjects at a trucker training program as free riders, conditional cooperators or unconditional cooperators, one could determine each team member's other-regarding behavior and, thus, predict the level of cooperation within the team.

Acknowledgements

We are grateful to the participants at the “17th Colloquium on Personnel Economics in Cologne, Germany as well as to the participants at the “XV. Symposium zur ökonomischen Analyse der Unternehmung” in Regensburg, Germany, and the “39th Workshop der Kommission Organisation im Verband der Hochschullehrer für Betriebswirtschft” in Zurich, Switzerland, for helpful comments and suggestions.

References

Andersen, S., Ertac, S., Gneezy, U., & List, J. A. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics, 95(4),* 1438-1443.

Apesteguia, J., Azmat, G., & Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science, 58(1),* 78-93.

Azmat, G., & Petrongolo, B. (2014). Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics, 30(C),* 32-40.

Becker, O., Feit, T., Hofer, V., Leopold-Wildburger, U., Pope, R., & Selten, R. (2006). Männer schöpfen Märkte besser aus – Empirische Evidenz anhand des Unternehmensplanspiels SINTO-Markt. *Perspektiven der Wirtschaftspolitik, 7(4),* 445-458.

Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L., & Briggs, A. L. (2011). Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *Journal of Management, 37(3),* 709-743.

Bertrand, M. (2011). New perspectives on gender. In O. Ashenfelter, & D. Card (Eds.), *Handbook of Labor Economics* (pp. 1543-1590). Amsterdam: Elsevier.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*. New York: Free Press.

Bowers, C. A., Pharmer, J. A., & Salas, E. (2000). When member homogeneity is needed in work teams: A meta-analysis. *Small Group Research, 31(3),* 305-327.

Burks, S. V., Nosenzo, D., Anderson, J., Bombyk, M., Ganzhorn, D., Götte, L., & Rustichini, A. (2016). Lab measures of other-regarding preferences can predict some related on-the-job behavior: Evidence from a large scale field experiment. *IZA Discussion Paper No. 9767*. Bonn: Institute for the Study of Labor.

Dai, Z., Galeotti, F., & Villeval, M. C. (2016). Cheating in the lab predicts fraud in the field: An experiment in public transportations. *IZA Discussion Paper No. 9702*. Bonn: Institute for the Study of Labor.

Faria, A. J., Hutchinson, D., Wellington, W. J., & Gold, S. (2009). Developments in business gaming: A review of the past 40 years. *Simulation Gaming, 40(4),* 464-487.

Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica, 77(5),* 1637-1664.

Haile, G. A. (2012). Unhappy working with men? Workplace gender diversity and job-related well-being in Britain. *Journal of Labor Economics, 19(3),* 329-350.

Hamilton, B. H., Nickerson, J. A., & 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.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review, 32(4),* 1199-1228.

Hoogendoorn, S., Oosterbeek, H., & van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science, 59(7),* 1514-1528.

Horwitz, S. K., & Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytical review of team demography. *Journal of Management, 33(6),* 987-1015.

Joecks, J., Pull, K., & Vetter, K. (2013). Gender diversity in the boardroom and firm performance: What exactly constitutes a 'critical mass'? *Journal of Business Ethics, 118(1),* 61-72.

Kochan, T., Bezrukova, K., Ely, R., Jackson, S., Joshi, A., Jehn, K., Leonard, J., Levine, D., & Thomas, D. (2003). The effects of diversity on business performance: Report of the diversity research network. *Human Resource Management, 42(1),* 3-21.

Konrad, A., Prasad, P., & Pringle, J. K. (2006). *Handbook of workplace diversity*. London: Sage.

Lazear, E. P. (1999). Globalisation and the market for team-mates. *Economic Journal, 109(454),* C15-C40.

Lee, C., & Farh, J. L. (2004). Joint effects of group efficacy and gender diversity on group cohesion and performance. *Applied Psychology: An International Review, 53(1),* 136-154.

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review, 53(1),* 1-18.

List, J. A. (2011). Why economists should conduct field experiments and 14 Tips for pulling one off. *Journal of Economic Perspectives, 25(3),* 3-16.

Mannix, E., & Neale, M. A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest, 6(2),* 31-55.

Milliken, F. J., & Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review, 21(2),* 402-433.

Niederle, M. (2014). Gender. *NBER Discussion Paper No. 20788*. Cambridge: National Bureau of Economic Research.

Orlitzky, M., & Benjamin, J. D. (2003). The effects of sex composition on small-group performance in a business school case competition. *Academy of Management Learning and Education, 2(2),* 128-138.

Paola, M. D., Gioia, F., & Scoppa, V. (2015). Are females scared of competing with males? Results from a field experiment. *Economics of Education Review, 48(C),* 117-128.

Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the black box: An analysis of workgroup diversity, conflict and performance. *Administrative Science Quarterly, 44(1),* 1-28.

Pitts, D. W., & Wise, L. R. (2010). Workforce diversity in the new millennium: Prospects for research. *Review of Public Personnel and Administration, 30(1),* 44-69.

Rogelberg, S. G., & Rumery, S. M. (1996). Gender diversity, team decision quality, time on task, and interpersonal cohesion. *Small Group Research, 27(1),* 79-90.

Schneid, M., Isidor, R., Steinmetz, H., Kabst, R., & Weber, H. (2014). Der Einfluss der Teamdiversität auf Teamleistung: Eine Metaanalyse. *Die Betriebswirtschaft, 74(3),* 183-210.

Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology, 62(2),* 207-218.

Sutter, M., Bosman, R., Kocher, M. G., & van Winden, F. (2009). Gender pairing and bargaining – beware of the same sex! *Experimental Economics, 12(3),* 318-331.

Tajfel, H. (1981). *Human groups and social categories*. Cambridge: University of Cambridge Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford: Blackwell.

Umans, T., Collin, S.-O., & Tagesson, T. (2008). Ethnic and gender diversity, process and performance in groups of business students in Sweden. *Intercultural Education, 19(3),* 243-254.

Van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology, 89(6),* 1008-1022.

Webber, S., & Donahue, L. (2001). Impact of highly and less job-related diversity on work group cohesion and performance: A meta-analysis. *Journal of Management, 27(2),* 141-162.

Williams, K. Y., & O'Reilly, III, C. A. (1998). Demography and diversity in organizations: A review of 40 years of research. In B. M. Staw, & L. L. Cummings (Eds.), *Research in Organizational Behavior* (pp. 77-140). Greenwich: JAI Press.

Table 1

Periodic developments.

Period	Developments
1	Introduction of Product 1 on Market 1.
2	Business as usual.
3	Optional development of a successor product for Product 1–old.
4	- Relaunch of Product 1–old possible. - Production and introduction of Product 1–new (if development results are appropriate in Period 3). - Introduction of products on new market (Market 2 with foreign currency).
5	Development of new product (Product 2).
6	Introduction of Product 2 on Market 1 (if product features are appropriate).
7	Introduction of Product 2 on Market 2.
8	Business as usual.

Note. Periodic endowments are determined by the software and are identical for all markets.

Table 2

Summary statistics of dependent variables.

Variable	Definition	Obs.	Mean	Std. Dev.	Min	Max
PRICE	Final share price at the end of the game (in €)	164	228.83	199.17	0	604.3
BANKRUPTCY	1=Team goes bankrupt, 0=Otherwise	164	0.30	-	0	1
PRICE_1	Share price at the end of period 1 (in €)	164	82.48	27.02	0	139.1
PRICE_2	Share price at the end of period 2 (in €)	163	112.67	44.88	0	222.2
PRICE_3	Share price at the end of period 3 (in €)	159	142.88	58.94	0	309.8
PRICE_4	Share price at the end of period 4 (in €)	157	195.31	88.95	0	439.5
PRICE_5	Share price at the end of period 5 (in €)	155	194.74	124.07	0	498.8
PRICE_6	Share price at the end of period 6 (in €)	137	197.30	123.66	0	472.6
PRICE_7	Share price at the end of period 7 (in €)	122	254.95	140.52	0	570.1
PRICE_8	Share price at the end of period 8 (in €)	116	323.53	159.15	0	604.3

Note. The decreasing number of observations for the periodic share prices results from bankrupt teams that were excluded from further periods and, thus, refers to the weekly population at the beginning of the respective period.

Table 3

Summary statistics of independent variables.

Variable	Definition	Mean	Std. Dev.	Min	Max
Team diversity (DIV)					
S_FEMALE	Share of women	0.38	0.24	0	1
GENDIV	Gender diversity (Standardized Blau index)	0.72	0.32	0	1
S_MIGBACK	Share of members with migration background	0.21	0.19	0	0.80
MIGDIV	Ethnic diversity (Standardized Blau index)	0.52	0.40	0	1
S_BUSADM	Share of members studying business administration as a major	0.76	0.21	0	1
MAJORDIV	Major diversity (Standardized Blau index)	0.80	0.26	0	1
Team ability (ABILITY)					
ABILITYDIV	Ability diversity (Standardized sample standard deviation)	0.24	0.11	0.03	0.54
S_TOPPERF	Share of top 10 percent individual performers	0.09	0.14	0	0.60
S_LOWPERF	Share of bottom 10 percent individual performers	0.08	0.14	0	0.75
External controls (EXT)					
TEAMSIZE	Number of team members	4.55	0.81	3	7
MARKETSIZE	Number of competitors	7.01	0.76	6	8
BANKCOMP_PERIOD	Weighted number of bankrupt competitors	11.54	6.37	0	23
2009	1=2009, 0=Otherwise	0.19	-	0	1
2010	1=2010, 0=Otherwise	0.19	-	0	1
2011	1=2011, 0=Otherwise	0.17	-	0	1
2012	1=2012, 0=Otherwise	0.24	-	0	1
2013	1=2013, 0=Otherwise	0.21	-	0	1

Note. Summary statistics refer to 164 team-level observations.

Table 4

Correlation matrix.

Variable		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1)	LG_PRICE	1																		
(2)	BANKRUPTCY	-0.98	1																	
(3)	S_FEMALE	-0.11	0.1	1																
(4)	GENDIV	-0.19	0.15	0.56	1															
(5)	S_MIGBACK	-0.07	0.06	0.24	0.21	1														
(6)	MIDIV	-0.12	0.1	0.23	0.24	0.9	1													
(7)	S_BUSADM	-0.12	0.1	0.25	0.15	0.18	0.23	1												
(8)	MAJORDIV	-0.05	0.04	0.09	0.12	0.14	0.14	-0.37	1											
(9)	ABILITYDIV	0.06	-0.07	0.02	0.1	0.09	0.1	-0.07	0.15	1										
(10)	S_TOPPERF	0.04	-0.03	0.05	0.08	0.06	0.13	0.07	0.04	0.25	1									
(11)	S_LOWPERF	0.12	-0.12	0.05	0.06	0.04	0.06	0.07	0.03	0.39	-0.16	1								
(12)	TEAMSIZE	-0.16	0.14	-0.01	0.17	0.1	0.16	-0.06	0.1	0.05	-0.07	-0.04	1							
(13)	MARKETSIZE	-0.15	0.11	-0.06	0.03	-0.08	-0.04	-0.24	0.04	0.01	0.02	-0.13	0.3	1						
(14)	BANKRUPT_COMP	-0.27	0.29	0	0	0.03	0.05	0.08	-0.03	0.06	0.13	-0.06	0.04	0.32	1					
(15)	2009	-0.08	0.05	-0.08	0.1	-0.12	-0.11	-0.17	0.18	0.03	-0.2	0.09	0.46	0.49	0.03	1				
(16)	2010	-0.1	0.09	0.02	-0.1	0.03	0.05	0.01	-0.15	-0.02	-0.03	-0.1	-0.12	0.49	0.37	-0.23	1			
(17)	2011	-0.03	0.05	0.08	0.02	0.08	0.01	-0.09	0.16	0.15	0.31	-0.1	0.07	-0.01	0.16	-0.22	-0.22	1		
(18)	2012	0.08	-0.09	0.03	0.1	0.06	0.15	0.06	-0.05	-0.1	0.18	-0.16	0.01	-0.22	-0.14	-0.27	-0.27	-0.25	1	
(19)	2013	0.12	-0.09	-0.05	-0.13	-0.05	-0.11	0.17	-0.12	-0.05	-0.24	0.26	-0.4	-0.7	-0.38	-0.25	-0.25	-0.24	-0.29	1

Note. Correlations refer to 164 team-level observations.

Table 5

OLS regression estimates with LG_PRICE as the dependent variable.

LG_PRICE	(1)	(2)	(3)	(4)
S_FEMALE			**-5.169***	**-5.490***
			(2.530)	**(2.968)**
S_FEMALE^2			**5.333***	**5.344***
			(2.654)	**(2.839)**
GENDIV	**-1.472****	**-1.587****		
	(0.616)	**(0.712)**		
S_MIGBACK			-3.795*	-4.487**
			(2.117)	(2.117)
S_MIGBACK^2			6.595*	7.820*
			(3.697)	(4.051)
MIGDIV	-0.485	-0.517		
	(0.497)	(0.477)		
S_BUSADM			-6.756**	-7.847**
			(2.773)	(3.258)
S_BUSADM^2			4.143*	4.969*
			(2.241)	(2.466)
MAJORDIV	-0.163	-0.481		
	(0.661)	(0.602)		
ABILITYDIV		1.149		0.607
		(2.172)		(2.304)
S_TOPPERF		1.392		1.776
		(1.730)		(1.863)
S_LOWPERF		2.408**		2.929**
		(1.109)		(1.124)
TEAMSIZE		-0.450		-0.441
		(0.303)		(0.335)
MARKETSIZE		-0.690**		-0.681**
		(0.307)		(0.280)
BANKCOMP_PERIOD		-0.128***		-0.120***
		(0.00981)		(0.0149)
2010		-0.126		-0.185
		(0.402)		(0.513)
2011		-0.473		-0.580
		(0.401)		(0.456)
2012		-0.409		-0.320
		(0.507)		(0.499)
2013		-1.830*		-1.781*
		(0.907)		(0.926)
Constant	5.376***	14.10***	7.664***	16.36***
	(0.498)	(3.711)	(0.841)	(3.967)
Observations	164	164	164	164
R^2	0.043	0.170	0.063	0.192

Notes. Table reports OLS coefficients. Robust standard errors clustered on 24 markets in parentheses. Further polynomials of *S_FEMALE* have been tested, but were not found to improve the model fit. Base year: 2009. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6

OLS regression estimates with LG_PRICE as the dependent variable – conditional on firm survival.

LG_PRICE	**(1)**	**(2)**
S_FEMALE		**-1.351*****
		(0.460)
S_FEMALE^2		**1.631****
		(0.582)
GENDIV	**-0.377*****	
	(0.129)	
S_MIGBACK		-0.731
		(0.486)
$S\_MIGBACK^2$		1.368
		(0.826)
MIGDIV	-0.109	
	(0.110)	
S_BUSADM		-1.365
		(1.258)
S_BUSADM^2		0.568
		(0.901)
MAJORDIV	0.00259	
	(0.159)	
ABILITYDIV		-1.351***
		(0.460)
CONTROLS	INCL.	INCL.
Constant	7.520***	8.329***
	(0.734)	(0.968)
Observations	114	114
R^2	0.191	0.250

Notes. Table reports OLS coefficients. Robust standard errors clustered on 24 markets in parentheses. Controls include team ability measures, team and market size, and further external controls. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 7

OLS regression estimates with periodic share prices (LG_PRICE_1 – LG_PRICE_8) as dependent variables.

	Period 1	**Period 2**	**Period 3**	**Period 4**	**Period 5**	**Period 6**	**Period 7**	**Period 8**
GENDIV	-0.0476	-0.102	**-0.259***	**-0.306***	**-0.433*****	**-0.475*****	**-0.483*****	**-0.374*****
	(0.0798)	(0.136)	**(0.137)**	**(0.148)**	**(0.145)**	**(0.111)**	**(0.140)**	**(0.130)**
MIGDIV	-0.0475	0.00227	-0.0771	-0.00744	-0.0722	-0.224*	-0.140	-0.124
	(0.0588)	(0.112)	(0.104)	(0.134)	(0.145)	(0.121)	(0.109)	(0.109)
MAJORDIV	0.0241	0.0112	-0.0174	-0.0201	0.0629	0.210	0.150	0.0233
	(0.0768)	(0.111)	(0.124)	(0.140)	(0.131)	(0.165)	(0.189)	(0.155)
BANKRUPT	-4.235***	-4.012***	-4.447***	-5.932***	-5.611***	-5.632***	-5.482***	-4.028*
	(0.0834)	(0.503)	(0.148)	(0.746)	(0.124)	(0.129)	(0.473)	(2.050)
BANK*GENDIV	0	-0.607	0	0.933	0.478**	0.453***	0.175	-1.657
	(.)	(0.642)	(.)	(0.937)	(0.175)	(0.156)	(0.579)	(2.178)
CONTROLS				INCL.				
Constant	4.031***	6.748***	7.109***	6.610***	6.929***	5.809***	5.626***	6.350***
	(0.613)	(1.086)	(1.101)	(1.080)	(0.861)	(0.553)	(0.553)	(0.476)
Observations	164	163	159	157	155	137	122	116
R^2	0.504	0.709	0.505	0.507	0.905	0.908	0.792	0.700

Notes. Table reports OLS coefficients. Robust standard errors clustered on 24 markets in parentheses. Controls include team ability measures, team and market size, and year. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 8

OLS regression estimates with teams' periodic management decisions as dependent variables.

	P1	P2	CI	SF	R&D	ENVIRON-MENT	PROCESS OPT.	LOAN
GENDIV	0.0357	0.0173	**-0.372****	-0.338	-0.0630	0.250	-0.320	-0.0992
	(0.0680)	(0.0968)	**(0.146)**	(0.213)	(0.102)	(0.167)	(0.231)	(0.0974)
S_FEMALE	0.195	0.0604	**-1.418****	-0.932	-0.167	0.817	-1.559	-0.457
	(0.257)	(0.394)	**(0.522)**	(0.598)	(0.505)	(0.644)	(0.921)	(0.442)
S_FEMALE^2	-0.223	-0.0448	**1.810*****	1.168*	0.194	-0.636	2.093*	0.371
	(0.290)	(0.412)	**(0.504)**	(0.606)	(0.572)	(0.719)	(1.186)	(0.553)
Observations	1,173	1,173	1,173	1,173	1,173	1,173	1,173	1,173

Notes. Table reports OLS coefficients with either *GENDIV* or *S_FEMALE* and *S_FEMALE^2* as the independent variables. Dependent variables are standardized (zero mean and one-unit standard deviation). Robust standard errors clustered on 24 markets in parentheses. Controls include team ability measures, team and market size, year and period dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9

Management decisions of female- versus male-dominated teams.

	P1	P2	CI	SF	R&D	ENVIRON-MENT	PROCESS OPT.	LOAN
FEM_DOM	-0.0207	-0.0256	0.0565	-0.00802	0.0343	**0.218***	0.101	0.0378
	(0.0455)	(0.0897)	(0.122)	(0.0995)	(0.0712)	**(0.124)**	(0.163)	(0.0983)
Observations	848	848	848	848	848	848	848	848

Notes. Table reports OLS coefficients. Dependent variables are standardized (zero mean and one-unit standard deviation). Robust standard errors clustered on 24 markets in parentheses. Controls include team ability measures, team and market size, year and period dummies. Reference category: male-dominated teams (N=581). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

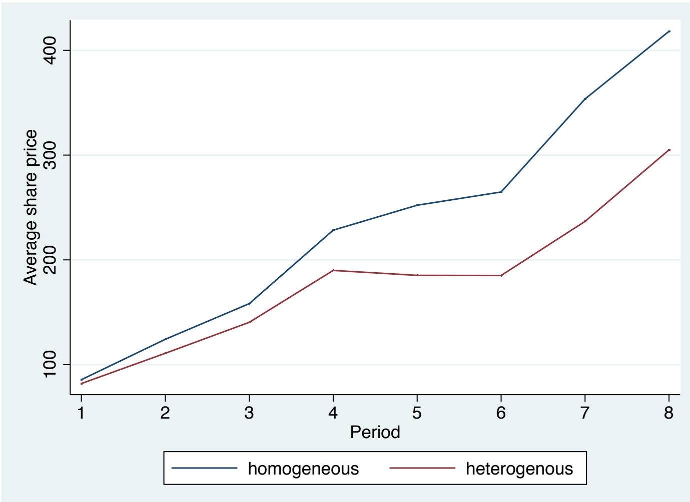

Fig. 1. Mean period share prices by gender composition.

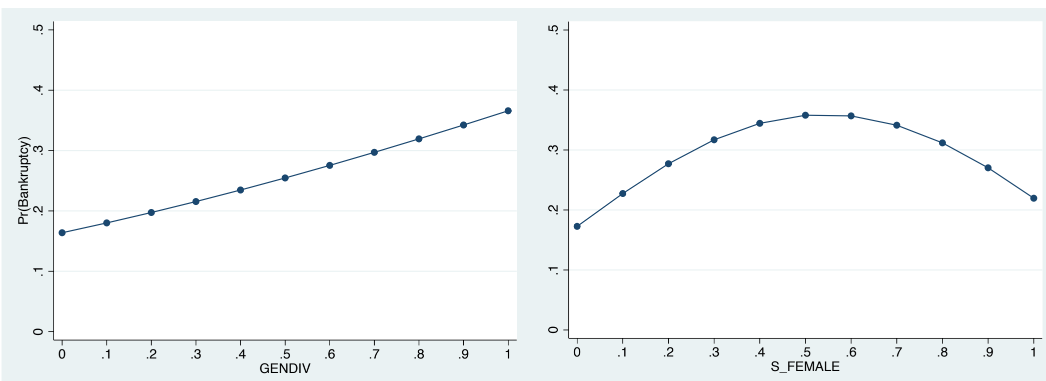

Fig. 2. Average marginal effects of GENDIV and S_FEMALE on BANKRUPTCY probability.