Group Gender Composition and Economic Decision-Making

Karine Lamiraud, Radu Vranceanu

Group Gender Composition and Economic Decision-Making: Evidence from the *Kallystée* Business Game

RESEARCH CENTER
ESSEC Working Paper 1515
Last version

2017

Karine Lamiraud
Radu Vranceanu

ESSEC BUSINESS SCHOOL

L'esprit pionnier

Forthcoming: *Journal of Economic Behavior and Organization*, 2018

Group gender composition and economic decision-making: Evidence from the *Kallystée* business game

Karine Lamiraud & Radu Vranceanu
ESSEC Business School and THEMA
95021 Cergy, France
E-mail: Lamiraud@essec.edu ; Vranceanu@essec.edu

21/07/2017

Abstract

This paper analyses data collected in 2012 and 2013 at the ESSEC Business School from *Kallystée*, a proprietary mass-attendance business game. Company boards are simulated by teams of five students selected at random. The design manipulates the gender composition of the boards to allow for all possible gender combinations. Data show that all-men and mixed teams with four women perform significantly better than all-women teams. However, when controlling for the average tolerance to risk score of the teams, the performance advantage of all-men teams vanishes, while the team-specific economic performance of teams with four women is still positive and strong. Teams with four women take more risks than the team tolerance to risk score would predict, which suggests some form of team specific action bias or risk-shift.

Keywords: Team decision, gender studies, risk-taking, business game, performance, governance.
JEL Classification: D71; C93; L25; M14

1. Introduction

After the 2000-2001 "governance crisis", driven by new opinion trends and legal changes in some countries[1], female representation on corporate boards worldwide continues to increase (Azmat, 2014). Yet, the firm-level evidence of a positive link between executive board diversity and firms' economic performance is inconclusive (Adams and Ferreira 2009; Ahern and Dittmar, 2012; Matsa and Miller 2013; Joecks et al. 2013; Bertrand et al. 2014). More recently, in the wake of the Global Financial Crisis of 2007-2009, journalists and policymakers argued that more gender-balanced bank boards would have constrained the banks' taste for excessive risk in the early 2000s. Evidence from firm-level data in this respect has also been inconclusive (Bansak et al., 2011; Muller-Kahle et al. 2011; Adams and Ragunathan, 2012).

This paper uses evidence from a controlled experiment to address the question of whether a firm's economic performance is impacted by the gender composition of its executive board. It also studies whether improved performance resulting from specific gender compositions is the result of improved decision making processes or just a reflection of higher tolerance.

Data were collected in 2012 and 2013 at the ESSEC Business School (France) from a proprietary business game called *Kallystée.* The game was developed in the 1990s by Daniel Tixier and Raymond Gambini with the support of *l'Oréal Paris.*

The game simulates the managerial decision-making process at a large cosmetics company. The key observation unit is the "firm", represented by its "executive board" which comprises five students teamed together at random. Firms are then grouped at random into sets of five and assigned to a market, called "Universe". Firms have some ability to differentiate their products in a typical market environment of imperfect competition. Each board makes all decisions over several periods or rounds. The organization of the game including the decision steps, its key rules and the information structure are presented in the Appendix B. [2]

For the purpose of this study, the business school's administration allowed us to manipulate the gender composition of the five-member participant teams. We used a random selection process to create teams of five women (i.e., all-women teams), and teams with four (i.e., four women and one man), three, two, one and, finally, no women (i.e., all-male teams) (respectively referred to as 5W, 4W, 3W, 2W, 1W and 0W teams).

[1] Following the example of Norway, which introduced a 40% corporate board quota for women as early as 2003, 11 EU countries introduced legal instruments to promote gender diversity on company boards: Belgium, France, Italy, the Netherlands, Spain, Portugal, Denmark, Finland, Greece, Austria and Slovenia. See also European Commission (2012).

[2] A short movie, filmed in 2011, shows how the simulation is actually played (www.youtube.com/watch?v=ViED5L0geLI). A set of photos is also available online : http://behavioralresearchlab.essec.edu/research/ research-topics/results

In addition to measuring economic performance by standard indicators (total equity, profit, sales), the game environment allows us to observe teams' risk-taking behavior. Indeed, in this game, the early launch of a new, high-quality product appears to be a good proxy for a high-risk decision. This high-end, high-risk positioning strategy is also indirectly captured by firms' investment in R&D, which is a prerequisite for launching new products. We therefore have two good proxies to assess team risk-taking behavior based on gender composition.

First-year students with no training in business administration played the business game for three dedicated days, at the beginning of the first term (October-November). We collected data on 1100 students who played the game in 2012/2013 and 2013/2014. Our sample includes students enrolled in the Master in Management Program (MiM) (19.8 years old, on average, in our sample), and the Bachelor in Business Administration (BBA) Program (18 years old, on average).[3] Both types of students had a strong general educational background.

To measure the students' individual attitudes towards risk, we used the self-reported tolerance-to-risk score introduced by Dohmen et al. (2011).[4] Additional personal data were collected from the school's official records, including gender, age, parents' educational level, and administrative district of residence during high-school. As a proxy for academic ability, for each student we collected their grade in microeconomics, a compulsory first-year course with a homogenous syllabus across the two programs.

In brief, regression analysis highlighted that 0W and 4W teams appeared to perform significantly better than 5W teams, with economic performance assessed by three distinct variables: total equity, profits and sales. Controlling for team tolerance to risk (defined as the average tolerance to risk of the team members) dampens the estimated coefficient for the 0W dummy, suggesting that team tolerance to risk may be one important mechanism through which gender composition impacts economic performance. We then studied actual risk taking behavior in the game by analyzing the decision to launch a new product and the decision of how much to invest in R&D. Male-dominated teams and 4W teams behaved as typical "first movers", a strategy that, in this game, is quite successful.

Our analysis has many common elements with the paper by Apesteguia et al. (2012), who were the first to use data from a business game to study the impact of a decision team's gender

[3] The two degree programs have different admission tracks and different curricula. BBA students are recruited directly after high school, while the Master in Management program, which is part of the prestigious French "Grandes Ecoles" classification, recruits - for its first year classes - students who have undergone a very intensive and demanding two-year post-high school program in Math, History and Literature. These students must then pass a national competitive exam for admission to the program. Traditionally ESSEC admits approximately 400 such students each year.

[4] Economists often use the concept of *risk aversion* in connection with the Expected Utility Theory. In this framework, risk aversion represents an individual's distaste for risk as the curvature of his/her Von Neumann-Morgenstern utility function. *Tolerance to risk* can be seen as the opposite of risk aversion. This concept is preferred by psychologists who refer to a broader definition of an individual's attitude toward risk, which is most often contextualized.

composition on a firm's economic performance. They used the data provided by *StratX*, a private company that runs an online mass-attendance business game for *l'Oréal Paris*. In that game, firms' executive boards are simulated by teams of three students (instead of five in *Kallystée*) who self-select to participate as a team. The business decisions they must make are similar in nature to those in *Kallystée*, but the rules of the game are somewhat different. In particular, a corporate social responsibility (CSR) channel is explicitly introduced. The stated goal of each firm is to maximize shareholder value.[5] Furthermore, in *StratX* each firm (i.e., 3-person team) plays against the computer, not each other (unlike in *Kallystée)*. Accordingly, a company cannot take advantage of a possible mistake by a rival. Using data from the first round only (as many as 16,000 teams from 1,500 universities), the authors show that all-women teams (i.e., 3 women) had the poorest performance compared with mixed gender and all-male teams.[6] The best performers were mixed teams of two men and one woman.

Hoogendorn et al. (2013) also provide evidence that mixed-gender executive boards perform better. In that study, the authors collected data on real firms created and managed by students enrolled in the Entrepreneurship program of the Department of International Business Studies at the Amsterdam College of Applied Sciences. In 2008-2009, 550 students created 45 firms, with teams comprising between 9 and 16 students. Students were randomly assigned to a team; the average proportion of women in each team varied between 17 and 58%. The authors' key finding was that business teams with an equal gender mix performed better than those with a majority of males in terms of sales, profits and earnings per share. Teams with a majority of women seemed to perform better in terms of profits than the latter, but statistical tests were inconclusive. The authors acknowledged that mutual monitoring occurs more often in mixed gender teams than in more homogeneous teams (irrespective of gender), and that more intense monitoring has a positive impact on company performance. It is worth noting that the business legal structure (i.e., the "rules" of the game) included an interesting peer-punishing mechanism because students could decide to exclude a poorly performing member.

Finally, we wish to comment on a recent laboratory experiment related to our work. Berge et al. (2016) asked clients of the PRIDE microfinance program in Tanzania to participate in an experiment. Of the 229 participants, 128 were women and 101 were men. Participants had to complete a knowledge test in nutrition and business and also make an incentivized risky investment decision, both individually and as members of a 4-person team. In a typical between-subject design, teams were "all-men", "all-women" or "mixed" (2 men and 2 women). Contrary to results from the field

[5] In the game, the firm's share value is driven by past performance.

[6] This statistically significant result is driven by the extremely poor performance of the first decile of 3W teams along the performance measure.

and quasi-field experiments, the female-only teams outperformed male-only and mixed teams in the knowledge test, although men tended to perform better than women in the individual test. Similarly, when making joint risk-decisions, all-women teams took more risks than their all-men and mixed counterparts, even though there were no gender differences in risk-taking at the individual level.

Despite similarities to Apesteguia et al. (2012) (see above), our study differs in several respects. First, the data set is different. In particular, it comes from a business game that incorporates a strategic dimension by allowing teams to compete against each other. Second, while their performance analysis only used cross-sectional data at round two, we analyzed firms' performance over all rounds, in a standard panel regression model. Third, as in the field study by Hoogendorn et al. (2013), students in our sample were randomly allocated to teams, conditional on gender. This ruled out any would-be self-selection bias. Kuhn and Villeval (2015) show that team selection depends on incentives, with the best-performing men being attracted to highly competitive compensation systems. Such a mechanism can bias the results, for instance if the competition logic of the business game attracts relatively better-performing men than women. In addition to randomization, controlling for academic ability by means of our proxy variable allows us to further reduce this risk. Finally, we collected data on individual tolerance to risk and investigated whether team tolerance to risk had a mediating effect on the relationship between gender composition and firm performance.

The paper is organized as follows: the next section presents the data. Section 3 introduces the empirical methodology and presents the results. The last section presents our conclusions.

2. The dataset

The database was constructed by merging firm-level data from *Kallystée* with individual-level data (collected from an individual survey administered during the game) and individual administrative data collected from the school. We ensured that students remained anonymous by replacing names with numerical identifiers at the beginning of the analysis.

2.1. Individual data

A total of 1100 students participated in the simulation in 2012 and 2013. Out of the whole sample, 49.50% of students were female (49.14% and 50.13% in the MiM and BBA programs, respectively).[7] The mean age was 19.25 years (19.8 for MiM students and 18.0 for BBA students). For 85.3% of students, their father had a post-secondary educational level (85.9 and 83.8% in the MiM and BBA programs, respectively). For 82% of students, their mother had a post-secondary educational level (83.1 and 80.7% in the MiM and BBA programs, respectively).

[7] In the last 15 years the proportion of women in the total student population of ESSEC Business School has been close to 50%.

As a proxy for academic ability we collected each student's grade in first-year microeconomics.[8] At ESSEC, the teaching of microeconomics is compulsory and draws on students' skills in abstract reasoning and computation. MiM and BBA students take similar microeconomics classes (same syllabus, same set of professors teaching the classes, grade based on a written test at the end of the class). The mean grade in the first-year microeconomics class was 10.19 (out of 20), and was quite similar between male (10.05) and female students (10.32) (p = 0.20).

It has to be acknowledged that this measure for academic ability is probably only imperfectly correlated to the ability to run a firm, and ideally should have been complemented with a specific relevant validated measure.[9] Another limitation of using first-year grades in microeconomics as a proxy measure, is that because students in the two academic years and two programs took different exams, the academic ability measure certainly had more associated noise.[10] However, data presented in Table 1A in the Appendix show that the average grade in microeconomics was stable over years and across programs. Furthermore, the difference between female and male average grades was never significant. This consistency in the descriptive statistics suggests that our proxy reasonably captures some measure of individual academic ability.

Individual tolerance to risk was measured through an internet-administered survey that all individual participants had to complete. They received the link to the survey on their personal computers 10 minutes after the end of the fourth round of the game and had two hours to answer on a voluntary basis before making their decisions for the fifth round. Questions were elementary, involved no strategic dimension, and anonymity was guaranteed. Ideally, we should have collected data on the individual tolerance to risk before, during and after the game. Unfortunately the organization of the game – with a large number of students (up to 400), over three days only, and implemented on a very tight schedule – restricted us to using a single survey.[11] Students were asked the following question: *How do you see yourself? Are you generally a person fully prepared to take risks or do you avoid taking risks?* Please tick a box on this 10-point scale, where box number 1 is "Unwilling to take risks" and box number 10 is "Fully prepared to take risks". This measure of tolerance to risk was introduced by Dohmen et al. (2011), who also showed that this measure correlates well with lottery-based measures of risk aversion (see also Both and Nolan (2012) and

[8] It was not possible to use the students' ESSEC admission scores as a measure of academic ability because MiM and BBA tracks have very different admission processes and requirements. Using the average grade during the first year of study as a proxy for students' ability would also have been difficult for the same reason (MiM and BBA tracks are quite different).

[9] Hoogendorn et al. (2013) introduced a multiple survey on relevant skills for entrepreneurs to capture such managerial ability.

[10] Furthermore, the grade reflects not only the ability to develop abstract reasoning and mobilize computational skills, but also the effort made to succeed in the exam. We cannot rule out the possibility that a small number of students obtained poor grades simply because they chose not to study.

[11] Hoogendorn et al. (2013) use the more robust three-step survey method, its implementation being facilitated by an entrepreneurial project that lasts for the whole academic year.

Vieider et al. (2014)). A benefit from using the fourth round break to have students complete questionnaires was that we could also collect data on how participants evaluated their team's decision compared with what their individual decision would have been.

The survey response rate was 95.5% for the MiM program (in both the 2012/2013 and 2013/2014 sessions) and 72.7% for the BBA program (74.7% and 70.3% in the 2012/2013 and 2013/2014 sessions, respectively). The percentage of missing answers did not significantly differ between male respondents (14%) and female respondents (11.5%) ($p = 0.22$). Furthermore, there were no significant differences according to team gender composition at the firm level, with the percentage of missing answers ranging from 11.6% to 15.9% ($p = 0.60$).

On average, women appeared to be significantly ($p<0.001$) more risk-averse than men, with mean tolerance to risk scores of 5.78 and 6.43 for women and men, respectively. The significant difference between men and women remained stable over programs and years as can be seen in Table 2A in the Appendix. The general pattern of distribution (median, percentiles 25th and 75th, min, max) was also stable across programs and years.

From observing the way in which students filled in the questionnaire, we can claim that individuals answered the risk questionnaire independently from the other members of their team. However, it might still be the case that the self-reported tolerance to risk score at the fourth round of the game might have been influenced by the team's gender composition or by the student's experience during the first rounds (in particular, how the team had performed up to the end of the fourth round).

We thus ran five robustness checks to test whether the individual self-reported measure of tolerance to risk were not contaminated by firm history and gender profile, which would lead to a biased within-firm correlation of the tolerance to risk measure.

First, individual reports on risk by male and females might have been influenced by their group composition. For instance, a man placed in a group of women might report a higher tolerance to risk than the same man in an all-men group. Data presented in Table 3A in the Appendix reveal quite a linear relationship between the team tolerance to risk and the number of females on a team, showing that such compression/disturbance is not present in our data. The average tolerance to risk decreases by the same quantity (around 0.1) when the number of female participants out of five participants increases by one. 0W and 1W teams had a significantly higher tolerance to risk than 5W teams ($p < 0.05$). 4W teams had the lowest average tolerance to risk but were not significantly different from their 5W counterparts in terms of average tolerance to risk. These team patterns are consistent with differences between males and females as assessed independently from the group composition. Table 4A in the Appendix indicates that everything else being equal, the tolerance to risk of a male is 0.63 point higher than the tolerance to risk of a female.

Second, we computed the average risk tolerance measure for males and females, by team gender composition. If the average tolerance to risk for males (respectively females) is independent of the gender composition of the teams to which these males (respectively females) belong to, this would suggest that individual risk measures were not influenced by the teams' gender composition. Data in table 5A in the Appendix reveal no differences in the average tolerance to risk of males (and respectively females) contingent on group composition

As an additional robustness check, we decomposed the standard deviation of the tolerance to risk variable into "within (firm)" and "between (firm)" components. This was computed over the whole sample and also by gender composition of the firms. Results presented in Table 6A in the Appendix show that there was less between-firm than within-firm variation. Furthermore, the within standard deviation was stable across teams with various gender compositions, which confirms that individuals' responses were not influenced by the team's gender composition.

Third, we performed three regressions on individual self-reported tolerance to risk scores with respect to various independent variables (age, grade in microeconomics, parents' educational level, type of program attended, academic year and place of residence). The first model is a standard OLS regression with pooled data. The second is a random-effects GLS panel regression, in which observations were structured by individuals and firms. The third model is a firm fixed-effects regression. The second and third models control for possible correlations between observations within a given company. Results displayed in Table 4A in the Appendix show consistent results between the three models, suggesting that there is no significant correlation in the self-reported tolerance to risk scores between individuals of the same company regardless of the "history" of the firm. This is corroborated by the Breusch and Pagan LM test of independence (p-value = 0.275) in the random-effects model and by the F-test following the firm fixed-effects model (Prob > F = 0.2091).

Fourth, in our data, the correlation of individual risk measures within firms was equal to 0,038 (the standard error of the intra-class correlation estimate is 0,017). This represents the share of the overall variance in the tolerance to risk which is explained by the within firm variance. As this share is almost close to zero, it suggests that individual risk measures within firms are no more similar than individual risk measures across different firms. As a complementary test, we computed the correlation of individual risk measures within groups of individuals that would have been randomly put together ex-post. The average within group correlation computed on 50 different random allocation of individuals into groups of five was equal to 0,037, which is almost identical to the measure obtained in the actual sample; this corroborates the absence of a contagion of tolerance to risk within firms.

The fifth and last check investigates the possible correlation between the self-reported tolerance to risk score and the team's performance during the first rounds of the game. We regress the individual tolerance to risk as measured after the round four: (1) on team performance prior to completing the risk survey (the equity level at round four was here used as a covariate as it best summarizes to performance achieved at the end of round 4); (2) on the round specific performance at each of the previous rounds (namely profits at round 1, round 2, round 3 and round 4). Results are displayed in columns 4 and 5 of Table 4A in the Appendix. The performance variables are not significant, thus suggesting that the tolerance to risk measure has not been affected by the team experience in the game prior to answering the survey.

Altogether, these robustness checks suggest that we can reasonably accept that the individual tolerance to risk measure was not affected by the gender composition of the teams or individuals' early experience in the game.

It could also be claimed that the individual tolerance to risk measure could have been affected by the gender composition of competitors (i.e. other firms within the same universe). If individuals know the gender composition of their competition boards, they can adjust their declared tolerance to risk according to their beliefs about how different types of teams will behave. However, in our study 87% of universes had a similar structure in terms of the gender composition of firms (either a 0W/ 1W/ 2W/ 4W/ 5W or a 0W/ 1W/ 3W/ 4W/ 5W structure), which would mitigate the risk of such contagion.

2.2. Firm-level data

The business simulation was run on the ESSEC servers. Our database comprised firm-level data for 44 Universes (markets) with five companies in each market, for a total of 220 firms. There were 140 firms in the MiM program and 80 firms in the BBA program (involving 1100 students).

The gender composition of companies in our sample, resulting from the random allocation of the students to firms contingent on their gender, is displayed in Table 1, which presents the overall data and data split by academic year and educational program. The percentages of 0W, 1W, 2W, 3W, 4W and 5W firms were, respectively 17.7%, 20.9%, 13.2%, 11.8%, 17.3% and 19.1%. The distribution of firms by gender composition was not significantly different between academic years (p = 0.34) or between programs (p = 0.65).

	All		Academic Year				Program			
			2012		2013		MiM		BBA	
	n	%	n	%	n	%	n	%	n	%
Number of women in the firm: 0	39	17,73	19	16,52	20	19,05	24	17,14	15	18,75
Number of women in the firm: 1	46	20,91	25	21,74	21	20,00	28	20,00	18	22,50
Number of women in the firm: 2	29	13,18	20	17,39	9	8,57	20	14,29	9	11,25
Number of women in the firm: 3	26	11,82	15	13,04	11	10,48	20	14,29	6	7,50
Number of women in the firm: 4	38	17,27	18	15,65	20	19,05	24	17,14	14	17,50
Number of women in the firm: 5	42	19,09	18	15,65	24	22,86	24	17,14	18	22,50
Number of observations	220		115		105		140		80	

Table 1. Number and distribution of firms with respect to gender composition

Furthermore, as the data presented in the Appendix show, the background characteristics of firms (i.e., average age of firm members, percentage of firm members whose father's educational level was higher than secondary education, percentage of firm members whose mother's educational level was higher than secondary education, average grade in microeconomics) did not differ significantly between firms with different gender compositions (0W, 1W, 2W, 3W, 4W and 5W) (see Tables 3A and 7A in Appendix). This confirms that the randomization procedure was applied properly.

The literature in experimental economics emphasizes that in individual decisions, women tend to display greater aversion to risk than men (Schubnaert et al. 1999; Eckel and Grossman, 2008a; 2008b; Croson and Gneezy, 2009; Charness and Gneezy, 2012), yet these differences are highly dependent on the context and on the task used to elicit the tolerance to risk (Filippin and Crosseto, 2016). The transformation of individual risk preferences into a team-based measure raises even more questions, as within-team interactions can alter answers. Several years ago, psychologists found evidence in favor of a "risk shift" phenomenon, according to which a team tends to adopt a riskier course of action than that which individuals in the team would take if deciding alone (inter alia: Wallach et al., 1962; Wallach and Kogan, 1965; Clark, 1971). Decision theorists analyzed various voting and bargaining models, and showed that each has its own predictions. Experimental economists analyzed team decision-making in terms of choosing risky alternatives using predetermined voting rules, and found that teams were either more risk averse than individuals composing the team, or presented no difference between individual and team attitudes (Zang and Casari, 2012; Brunette et al. 2015; Harrison et al. 2012). Morone and Temerario (2016) found that if no decision rule is imposed, then no statistical difference is observed between individual and team risk aversion.

From observing how students work in *Kallystée*, it turns out that consensus is the dominant decision process, which places individuals in a decision context close to that studied by Morone and

Temerario (2016). We will thus assume that the average tolerance to risk of the individual members of a given team is a good proxy for the tolerance to risk of that team.[12]

Missing individual values for female (respectively male) respondents were replaced by the average values computed for female (respectively male) respondents in the same academic year (2012/2013 or 2013/2014) and in the same program (MiM or BBA).[13]

3. Results

3.1. The economic performance equation

In standard economic literature, the performance of a corporation is related to its return (from the accounting perspective) or market value (from the financial perspective). In this game, students are allowed to make their own guesses about what they should achieve, being informed from the outset that their firm is publicly listed, that they are running the firm on behalf of their shareholders and that the market value is defined as the firm's total equity. It is not difficult to guess that maximizing total equity (i.e., the cumulated operating profit) should be the overriding goal of the executive board.[14]

The rules of *Kallystée* state that in each market (Universe) there is only one award winner. Given the context (the company is publicly listed), the winner is the firm with the highest equity.[15] At the end of the game, winning teams present their firm's performance and strategy in a four-minute talk to all their colleagues, professors and the *l'Oréal* representatives. They also receive a bundle of high-end, expensive l'Oréal products. In this respect, the incentives to perform well as a team are quite strong.

As a main performance indicator, we used the total equity as recorded by the firm at each round. However, in real life, professional managers can pursue other goals than simply maximizing the market value of the firm, namely one-period profits (if they are "slaves of the short run") or total sales (if they are concerned with "market power"). Our executive boards may have been victims of

---

[12] Risk-taking decisions could respond to differences in the within-group variance in tolerance to risk. However, in our data, the standard deviation of the average tolerance to risk variable was very stable across groups with very different gender compositions (Appendix - Table 3A), which suggests that tolerance to risk of groups did not differ in their within-group variance. We also tested the impact of within-group variance on tolerance to risk, together with the average tolerance to risk measure in the various sets of regressions. The within-group variance in tolerance to risk was never significant, and therefore no longer considered in the analyses presented in Tables 2-4.

[13] As a robustness check, we also computed an average tolerance to risk measure including only those students who answered the tolerance to risk questionnaire. Our results were not affected. Neither were they affected when we excluded groups with missing values.

[14] By contrast, in the *StratX* game, as analyzed by Apesteguia et al. (2012), there is an additional Corporate Social Responsibility goal that conflicts with profit maximization.

[15] The winner is decided by consensus at the very end of the simulation, in a discussion between the game administrator and all five teams of the Universe. Should two firms be very close to each other in terms of total equity, other criteria such as market share can be taken into account.

the same managerial biases. Accordingly, we followed Hoogendoorn et al. (2013) and considered one-period profits and sales as possible performance measures.[16]

As already mentioned, some noise in the data is inherent given the three-day experiment on a tightly schedule. We attempted to address this difficulty by collecting a relatively large number of observations and choosing an estimation technique – three-level hierarchical regression for panel data – which allowed us to contain estimation biases.

Firms' decisions were observed at several rounds (resulting in variables being correlated with one another for a given firm), and firms were nested within Universes that may have had specific features (e.g. universes with different structure in terms of the gender composition of the teams). The 140 companies from the MiM played the game for 8 rounds, while the 80 companies from the BBA played 5 rounds. Hence our dataset was composed of 1,520 firm-level observations.

With R_{tiu} as the performance measure at time t for firm i in universe u, we estimated the following three-level hierarchical model:

$$R_{tiu} = g'_{tiu}\alpha + x'_{tiu}\beta + \mu_u + v_{iu} + \varepsilon_{tiu} \qquad (1)$$

where g_{tiu} is a vector of gender composition dummies and x_{tiu} is a vector of covariates. These include the percentage of team members whose father (respectively mother) had a post-secondary educational level, the average age of the firm's members, the average team members' grade in microeconomics, a dummy for the academic year 2013, and a dummy for the MiM program. Error terms were broken down into normally distributed universe-specific effects μ_u, firm-specific effects v_{iu} and time-specific chance events ε_{tiu}.

We then estimated the augmented equation:

$$R_{tiu} = g'_{tiu}\alpha + x'_{tiu}\beta + a_{iu}\gamma + \mu_u + v_{iu} + \varepsilon_{tiu} \qquad (2)$$

where a_{iu} is the average tolerance to risk of the team.

In this second model, the vector of coefficients α captures the partial effect of gender composition *after controlling for the team tolerance to risk.* If α coefficients in model (2) are attenuated relative to corresponding coefficients in model (1), this would hint that team tolerance to risk is a mediating factor in explaining the overall team gender composition effect.

Table 2 presents the relationship between gender composition of teams and economic performance (as expressed by three different measures: equity level, sales and profit). We chose the all-women team as the reference to facilitate comparison of results with Apesteguia et al. (2012),

[16] Total sales, net profit (or losses) and total equity of *all firms* are displayed on a large electronic board after each round (i.e., they are common knowledge).

who also adopted this convention. Estimates were performed using the *gllamm6* routine in *Stata*, a method for fitting a wide range of multilevel models.[17]

First, note that firm-level and universe-level random effects are statistically significant, which validates our estimation strategy of using a hierarchical random effects model with firm and universe effects.

Output estimates of Equation 1 based on equity level (Table 2, first column) show that all-men (0W) and 4W teams performed significantly better than all-women teams. The performance gap between all-women teams and the 1W, 2W and 3W teams was not statistically significant. In contrast, in the data processed by Apesteguia et al. (2012), all-women teams performed significantly worse than all other teams.

There is another important contrast with the findings of Apesteguia et al. (2012). In their analysis, mixed teams with a majority of men (i.e., two men and one woman) performed best. In our analysis, mixed teams with a majority of women (4 W) appear to be the best performers. This result is more in line with Hoogendoorn et al. (2013), who also reported the strong performance of equally mixed gender teams and suggested that, for their sample, teams with a majority of women might perform even better.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Equity Level		Sales		Profit		Equity level (round 1)	
	Coef	Coef	Coef	Coef	Coef	Coef	Coef	Coef
Number of women in the firm: 0	1181***[a]	443[b]	2077***	986	242**	131	181**	22
Number of women in the firm: 1	362	262	413	325	102	92	62	29
Number of women in the firm: 2	284	183	-328	-428	20	9	39	15
Number of women in the firm: 3	342	328	564	367	102	95	47	20
Number of women in the firm: 4	1291**[a]	1270**[b]	2052***	2031***	202**	196**	198**	178**
Number of women in the firm: 5	*ref*	*ref*	*ref*	*ref*	*ref*	*ref*	*ref*	*ref*
Average grade in Microeconomics	254***		205***	231***	22**	23**	*56***	60**
% of firm members whose father 's educational level : > secondary education	349*	337*	182	172	-35	-35	65	64
% of firm members whose mother 's educational level : > secondary education	302*	262*	-876	-875	-73	-70	64	60
Average age in the firm	124	119	-670	-631	34	34	121	125
Year 2013	-793**	-731***	-493*	-470*	-231**	-224**	113	123
MiM program	1764**	1772***	693***	676***	145	144	143**	135**
Average risk self-assessment		806***		1163***		82***		80**
Variances (covariances) of random-effects								
level 1	5184685	5116560	1228499	1228499	1416389	1416492		
	(201983)	(197190)	(476357)	(476357)	(55544)	(55546)		
level 2 (firm)	6226242	6483597	1169959	1169959	330299	328037		
	(626474)	(564633)	(1024379)	(1024379)	(55844)	(55755)		
level 3 (universe)	311600	385175	116589	144197	64127	66580		
	(32578)	(28634)	(12276)	(12559)	(50367)	(51240)		
Number of observations	1520	1520	1520	1520	1520	1520	220	220

[a] Wald test of H0: the coefficients of OW and 4W dummies are equal;p = 0,42

[b] Wald test of H'0: the coefficients OW and 4W dummies are equal;p = 0,03

All regressions include time (round) dummies (except for the two last regressions which inlude observations from round 1 only)

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively

[17] For an introduction to these Generalized Linear Latent and Mixed Models (GLLAMM), see the text by Rabe-Hesketh et al. (2004).

Table 2. Association between team gender composition and economic performance

Column 2 in Table 2 presents output estimates of Equation 2; the set of control variables includes the team (average) tolerance to risk. First, we notice that team tolerance to risk has a significant positive impact on economic performance. This result is not surprising: if teams with a higher tolerance to risk score take more risks, then the economic performance improves. Second, when controlling for team tolerance to risk (increasing with the number of men in the team), all-men (0W) teams no longer perform significantly better than all-women (5W) teams (i.e., the coefficient becomes insignificant). Furthermore, the estimated coefficient of the 0W dummy is significantly reduced ($p < 0.01$). This suggests that all-men teams perform better mostly because they are, on average, more tolerant to risk.

However, after controlling for team tolerance to risk, 4W teams still performed significantly better than 5W teams. As a matter of fact, the coefficient for the 4W team is not attenuated when including the new control variable.[18] This finding hints at some hidden "performance enhancing mechanism" that might have been at play in 4W teams.

Estimations of equation 1 and 2 based on sales (columns 3-4 in Table 2) and profit (columns 5 and 6 in Table 2) corroborate the former results (based on total equity): after controlling for risk aversion, 4 W teams still perform significantly better than 5W teams while 0W teams no longer perform significantly better than 5W teams.

Hoogendoorn et al. (2013) explained the strong performance of mixed teams by increased peer monitoring; other factors such as better team dynamics or preference shifts could be responsible for this abnormal performance (Azmat, 2014). Unfortunately, our data did not allow us to study the psychological determinants of this outcome. However, a complementary analysis helped us indicate to what extent this "abnormally" good performance may be connected to risk taking (see section 3.2 below).

Other results in Table 2 show that equity levels significantly increase with the average academic ability of the team members (their grades in microeconomics being a proxy for academic ability). Equity levels were higher in firms whose members had parents with post-secondary educational levels. The average age of individuals on the teams was not significantly associated with increased performance.

As a robustness check, we also ran equations 1 and 2 using equity level as the dependent variable and using only first-round observations (Columns 7 and 8 in Table 2). Coefficients were

[18] The coefficient of the 4W dummy variable is now significantly different from the coefficient of the 0W dummy variable (see Wald test in Table 2). They were similar when we did not control for the tolerance to risk variable.

smaller but the results were consistent with those obtained in columns 1 and 2, confirming that the observed differences in performance were not driven by differences in learning ability.

We also ran a performance regression model using, as a measure of performance, the equity rank of the firm within each market of five firms (instead of the equity level). The results obtained from a categorical data regression model (with the firm rank as the dependent variable) were not qualitatively different from those presented in columns 1 and 2 of Table 4.

3.2. Association between gender composition and risk-taking behavior

In this section we use both a direct and an indirect measure of actual risk-taking behavior to study the impact of team gender composition on actual risk-taking behavior.

Launching a new product is a challenging decision in *Kallystée*. It requires a realistic forecast of future sales, many calculations to determine the expected unit cost, and a clever pricing strategy. In general, teams launched at most two new products over the eight (five) rounds. In the MiM 2012-2013 sample, at the second round, 75 % of the firms introduced a new product to the market (of an average quality index of 4.7).

In the early rounds, launching a new product comes with a strategic risk because a competitor can simultaneously bring an identical product to the market. In this case, both firms would incur losses because they must share the market and might not be able to cover the development cost of the new product. Hence, launching a new complex product at round 2 is clearly a risky decision. Launching a third product early in the game can also be seen as a risky decision for the same reasons invoked above, but the strength of the test is weaker than for the second product (only firms that have already launched a second product at round 2 can subsequently launch a third product). This definition of “risk-taking” was largely shared by participants. In the survey (see above) administered at the fourth round, we asked participants what was a risky decision in this game in their opinion. The majority referred to "launching a complex product quickly". Data show that all teams ended up with at least one new product.

We therefore created two dummy variables as direct measures of risk-taking behavior: the first took the value 1 if the firm launched a product with a quality index higher than 7 at round 2 (note that the firm had already been producing the quality index 3 product). The second variable took the value 1 if the firm launched a third product with a quality index higher than 7 at round 3.

Table 3 presents the output of the regressions with the two dummies as dependent variables and the team composition dummies as main independent variables. The results in the first two columns refer to Model 1, where the dependent variable is “launch a complex new product in period 2”, and results in the last two columns refer to Model 2, where the dependent variable is “launch a third complex product in period 3”. Note that for these cross-sectional estimations our sample was reduced to 220 observations (one observation per firm) and the structure of the regression was a two-level hierarchical model with a universe specific error term.

The (a) version does not include the average tolerance to risk, the (b) version does. In both models (1) and (2), team tolerance to risk appeared to be an important driver of the decision to launch a new product.

	(1)		(2)	
	(a)	(b)	(a)	(b)
	Coef	Coef	Coef	Coef
Number of women in the firm: 0	0,966**	0,562**	0,313	0,227
Number of women in the firm: 1	0,657*	0,552*	0,294	0,204
Number of women in the firm: 2	0,633	0,53	-0,079	-0,067
Number of women in the firm: 3	0,527	0,454	0,253	0,203
Number of women in the firm: 4	0,445	0,45	0,627**	0,636**
Number of women in the firm: 5	*ref*	*ref*	*ref*	*ref*
Average grade in microeconomics	0,097	0,102	0,007	0,012
% of firm members whose father 's educational level : > secondary education	-0,241	-0,273	2,137**	2,109**
% of firm members whose mother 's educational level : > secondary education	0,911	0,905	-0,622	-0,591
Average age in the firm	0,026	0,025	0,082	0,094
MiM program	-0,083	-0,068	-0,544	-0,571
Year 2013	0,111	0,152	-0,857	-0,848
Average risk self-assessment		0,498**		0,155**
Nmber of observations	220	220	220	220

(1) Launch of a second complex product in period 2 (yes/no)
(2) Launch of a third complex product in period 3 (yes/no)
(a) without average risk self-assessment (b) including average risk self-assessment
*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively

Table 3. Association between team gender composition and launching a new high quality product (random effects probit models)

In the realistic decision-making context of our business simulation, all-women teams appeared to take the fewest risks, although the difference was not statistically significant with teams where the number of women was higher than two. This result contrasts with findings from the laboratory experiment by Berge et al (2016), where all-women teams took more risks than all men and mixed teams. Furthermore, when controlling for team tolerance to risk, 0W and 1W teams in model 1, and 4W teams in model 2 seemed to engage in risk taking action beyond what the average tolerance to risk would indicate.

Although launching new products is tantamount to taking risks, the motivation behind implementing such a course of action is not obvious. Certainly, a higher team tolerance to risk would prompt greater risk-taking. If teams take more risks than the average tolerance to risk score would suggest, this may suggest either a risk-shift phenomenon or *action bias* (described below). However, it is impossible to disentangle these two possibilities.

Research on *action bias* can be traced back to the seminal paper by Kahneman and Tversky (1982). They argued that negative outcomes are judged more severely when caused by action than when caused by inaction. As a consequence, people tend to prefer the *status quo* or develop an *omission bias*. A subsequent body of literature showed that this bias is strongest in environments where inaction is the norm (see the surveys in Anderson, 2003; Feldman and Albarracín, 2016). More recently, several papers revealed that, in an environment where action is the norm, faced with the same expected negative outcome from action or inaction, people would choose the former. This is referred to as *action bias* (Ritov and Baron, 1994; Patt and Zeckhauser, 2000; Feldman and Albarracín, 2016). An empirical analysis of soccer by Bar-Eli et al. (2007) tends to corroborate these predictions revealing that when considering a large number of penalty kicks, the goalkeeper tends to jump to the side when it would be better to stay in the center of the goals.

Our business game has similarities to the world of sports. In particular, the pedagogical context is favorable to action: students were there to "take decisions" in order to win the competition. Given the complex nature of the rules of the game, one cannot rule out that some teams were driven into risk-taking action by emotions connected to "doing something" and to boredom avoidance (i.e., action bias) rather than taste for risk (i.e., risk-shift). This possible action bias may have been impacted by the gender composition of the teams. Unfortunately, as highlighted above, our data do not enable us to disentangle these two possibilities.

Because firms aiming to launch new products must first invest in R&D, and because the volume of R&D is related to the quality of the product to be launched, R&D investment is a good *indirect* measure of risk-taking behavior.

Table 4 presents the estimates of an "investment equation" investigating the relationship between the teams' gender composition and the amount of R&D investment. Estimates in columns 1 and 2 used the full panel dataset (with and without team tolerance to risk). Because investment naturally declines near the end of the game, the last column presents estimates for the first 5 rounds only for MIM and 2 rounds for BBA.

Following the pattern revealed by the product launch analysis, all-men and 1W teams carried out significantly more R&D than all-women teams. We also note that 4W teams invested significantly more in R&D than all-women teams (in line with their decision to launch a third high-quality product in the game's third round).

Both the analyses of product launching and of R&D investment suggest that 4W teams (as well as teams with a majority of men) appeared to take risks beyond what their average tolerance to risk as a team would suggest. Again, this may be the result of either a risk-shift phenomenon, or an action bias specific to these teams. In turn, this first-mover attitude of 4W teams probably justifies their "abnormal" performance, as revealed by the performance analysis in the previous section.

	(1) Coef	(2) Coef	(3) Coef
Number of women in the firm: 0	22,886**	19,638**	18,861**
Number of women in the firm: 1	20,915**	17,649**	17,593**
Number of women in the firm: 2	16,366*	13,112	13,0754
Number of women in the firm: 3	15,028	12,788	12,804
Number of women in the firm: 4	15,968**	17,190**	17,928**
Number of women in the firm: 5	*ref*	*ref*	*ref*
Average grade in microeconomics	2,064*	2.355*	2,344*
% of firm members whose father 's educational level : > secondary education	25,804*	23,717*	29,773*
% of firm members whose mother 's educational level : > secondary education	9,499	10,53	7,484
Average age in the firm	10,977	11,308	9,798
Year 2013	7,567	9,487	5,459
MiM program	15,720	15,521	15,357
Average risk self-assessment		6,975**	6,483***
Variances (covariances) of random-effects			
level 1	11215 (406)	11193 (406)	15218 (733)
level 2 (firm)	11727 (423)	11881 (442)	13954(564)
level 3 (universe)	1985(98)	1867 (90)	2231 (104)
Number of observations	1520	1520	860

(1) Full sample
(2) Full sample (without average risk self assessment)
(3) rounds 1-5 for MimM and rounds 1-2 for BBA
All regressions include time (round) dummies)
*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively

Table 4. Association between team gender composition and R&D investments (3-level hierarchical linear regression model)

As mentioned above, tolerance to risk was measured using an online test administered at the fourth round of the game. At that time, students had already made the most critical decisions, particularly about whether to launch new products and what the quality of the products should be. In the same survey, subjects were asked to answer the following question: "with respect to decisions made so far by your team in the *Kallystée* business game, did the team take more risks than you would have taken on your own?". They could answer on an increasing scale from 1 to 5, where values 1, 3 and 5 were "less risk", "same risk" and "more risk" (5), respectively. We collected 958 answers with 35% percent of the population giving scale values of 4 and 5.

We built an indicator variable taking the value 1 for scale answers 4 and 5 (the individual declared that the team took more risks than he/she would have personally taken) and 0 for the remaining answers. Table 5 presents the estimation output of a probit model, with this indicator as the dependent variable and the team composition dummies as covariates, as well as other controls.

	Coef
Female	0,386***
Grade in microeconomics	-0,002
Age	-0,011
Father 's educational level: > secondary education	-0,031
Mother 's educational level: > secondary education	0,329***
Attends MiM program (versus BBA program)	-0,092
Year 2013	-0,281***
Number of women in the firm: 0	0,172
Number of women in the firm: 1	-0,005
Number of women in the firm: 2	0,156
Number of women in the firm: 3	-0,076
Number of women in the firm: 4	0,283**
Number of women in the firm: 5	*ref*
Number of observations	958

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively

Table 5. Probability of individuals declaring that the team took more risks than they would have personally taken by themselves

Among the team composition dummies, 4W teams distinguish themselves by a significant positive coefficient. Individuals in 4W teams were significantly more likely (than 5W teams) to declare that the team took more risks than they would have taken on their own. This is consistent with the risk-shift / action bias assumption expressed above. This result holds even when controlling for the respondent’s gender, with our results showing that female respondents were more likely to declare that the team took more risks than they would have taken on their own.

4. Conclusion

In general, laboratory experiments, field experiments and quasi-field experiments such as the observation of human behavior in business games can offer significant insights into how team interaction may lead to higher performance in gender-diverse teams. The accumulated evidence on this important topic is relatively scarce, and confidence in early results can improve only by replicating existing studies and performing a substantial number of new analyses. Using data collected in 2012 and 2013 from *Kallystée*, a proprietary business game developed for the ESSEC Business School, this paper contributes to this literature by analyzing the relationship between the gender composition of a decision team and its economic performance and risk taking behavior.

Our analysis of economic performance corroborates but also qualifies to some extent the early findings by Apesteguia et al. (2012), based on a different business game. They showed that all-women teams performed significantly worse than teams of other gender combinations. We also poor performance of all-women teams, but this poor performance was not significantly different from the performance of some mixed teams. As in Apesteguia et al (2012), in our analysis, mixed teams performed best. However, unlike their analysis, where mixed teams with a majority of men performed best, in our data the best performers were mixed teams with a strong majority of women (four women in a five-member team). Thus, our results concerning the gender effect on performance are closer to those obtained by Hoogendoorn et al. (2013) from a field experiment in which business students had to create real small businesses. In their data, the best performance in terms of equity, profits or sales was achieved by mixed teams with 50% to 60% women, a result they explain by the higher peer monitoring observed in mixed teams.

Moving beyond sheer performance analysis, we carried out a joint analysis of team economic performance and tolerance to risk. The performance analysis revealed that all-men teams and 4W teams performed significantly better than all-women teams. However, when controlling for team tolerance to risk, the performance premium of all-men teams vanished, while the performance premium of all other gender combination teams diminished. It turns out that team tolerance to risk has an important mediating effect on economic performance. Even when controlling for average tolerance to risk, four women teams maintained their performance edge compared with all-women teams.

This outstanding performance of 4W teams appears to be grounded in their risk taking behavior. As shown in the second part of our study, 4W teams take more risks than the average tolerance to risk of their team would indicate, as if these teams develop some specific form of risk-shift or action bias that drives them to use aggressive first-mover strategies. They launch new products and invest heavily in R&D, in a decision pattern similar to that of teams with a majority of men. An opinion survey administered during the game corroborated this finding, with a significant number of individuals belonging to 4W teams declaring that the team took more risks than they would have taken by themselves. Such strong performance of mixed (gender) but unbalanced (in number) teams raises new research questions about gender interaction in teams.

From a policy perspective, our preliminary findings support the idea of increasing the gender diversity of corporate boards, as the best performance is achieved by mixed teams, although not without taking more risks than the reference team. Certainly, our results should be interpreted with a significant degree of caution when looking for policy implications and generalizations. In particular, the fact that the business game was played by participants without managerial experience might be a limitation. It would be interesting to see if the same results held for experienced managers playing the game.

Acknowledgments

This research was carried out within the research framework of the LABEX MME-DII (http://labex-mme-dii.u-cergy.fr/). We are extremely grateful to Daniel Tixier for sharing the game database with us, for supporting us, and for his advice during the implementation of the experiment. We are grateful to ESSEC Business School's administration, and mainly to Françoise Rey, the former ESSEC Grande Ecole Director (2012-2013) and Jean-Marc Xuereb, the game coordinator, for their support in implementing our project. We would like to thank three anonymous referees and the Associate Editor Ragan Petrie; participants in the *Annual Conference of the ASFEE*, Mai 24-25, 2014, Besançon, France; participants in the *Annual Congress of the European Economic Association*, Toulouse, August 25-29, 2014; and Estefania Santacreu-Vasut, Seeun Jung and Oana Peia for their remarks which helped us to improve the paper. Finally, we thank Delphine Dubart and Valérie Fournier for their help with data collection and processing.

References

Adams R, Ferreira D (2009), Women in the boardroom and their impact on governance and performance, *Journal of Financial Economics* 94 (2): 291-309.

Ahern K, Dittmar A (2012), The changing of the boards: The impact on firm valuation of mandated female board representation, *Quarterly Journal of Economics* 127 (1): 137-197.

Anderson C. J (2003), The psychology of doing nothing: forms of decision avoidance result from reason and emotion, *Psychological Bulletin*, 129 (1): 139-67.

Azmat G (2014), Gender diversity in teams, *IZA World of Labor*, 29: 1-10.

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.

Bansak C, Graham M, Zebedee A (2011), The effects of gender composition of senior management on the economic fallout, *Applied Economics Letters* 18 (16): 1603-1607.

Bar-Eli, M, Azar, O H, Ritov, I, Keidar-Levin, Y, Galit S, (2007), Action bias among elite soccer goalkeepers: The case of penalty kicks. *Journal of Economic Psychology*, 28, 5: 606-621

Bertrand M, Black S, Jensen S, Lleras-Muney A (2014), *Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labor Market Outcomes in Norway*. *IZA Working Paper* 8266.

Berge, L I O, Juniwaty, K S, Sekei, L H (2016), Gender composition and group dynamics: Evidence from a laboratory experiment with microfinance clients, *Journal of Economic Behavior and Organization*, Online at: doi:10.1016/j.jebo.2016.07.015

Booth A, Nolen P (2012), Gender differences in risk behaviour: does nurture matter? *Economic Journal* 122 (558): F56-F78.

Brunette, M, Cabantous, L, Couture, S (2015). Are individuals more risk and ambiguity averse in a group environment or alone? Results from an experimental study. *Theory and Decision*, 78 (3), 357-376.

Byrnes J, Miller D, Schafer W (1999), Gender differences in risk taking: A meta-analysis, *Psychological Bulletin* 125 (3): 367.

Charness G, Gneezy U (2012), Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1): 50-58.

Clark R (1971), Group-induced shift toward risk: A critical appraisal, *Psychological Bulletin* 76 (4): 251-270.

Croson R, Gneezy U (2009), Gender differences in preferences. *Journal of Economic Literature* 47 (2): 1–27.

Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner G, (2011) Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association* 9 (3): 522-550.

Eckel C, Grossman P (2008a), Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior and Organization* 68 (1): 1-17.

Eckel C, Grossman P (2008b), Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results* 1: 1061-1073.

European Commission (2012), Women on Boards: Commission proposes 40% objective. *Ref.* IP/12/1205 on 14/11/2012, Online at: http://europa.eu/rapid/press-release_IP-12-1205_en.htm.

Feldman, G, Albarracín D, (2016) "Norm theory and the action-effect: The role of social norms in, regret following action and inaction." *Journal of Experimental Social Psychology*, On line.

Filippin A, Crosetto P (2016), A reconsideration of gender differences in risk attitudes. *Management Science, Online at:* http://dx.doi.org/10.1287/mnsc.2015.2294.

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.

Harrison, G W, Morton I L, Rutström E E. Tarazona-Gomez M (2012). Preferences over Social Risk, *Oxford Economic Papers* 65 (1): 25-46.

Holt C, Laury S (2002), Risk aversion and incentive effects, *American Economic Review* 92 (5): 1644-1655.

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.

Kuhn P, Villeval MC (2015), Are women more attracted to teamwork than men. *Economic Journal*, 125 (582): 115–140.

March J G, Shapira Z. (1987) Managerial perspectives on risk and risk taking, *Management Science* 33 (11): 1404-1418.

Matsa, A, Miller, A R, 2013, A female style in corporate leadership? Evidence from quotas. *American Economic Journal: Applied Economics*, 5 (3): 136-69.

Moore E, Eckel C (2003), Measuring ambiguity aversion. Unpublished manuscript, Department of Economics, Virginia Tech. (revised 2006).

Morone, A, Temerario, T. "Group preferences over social risk: does (group) size matter? (2016), *Econstore /Zew* nr. 10419.

Muller-Kahle M, Lewellyn K (2011), Did board configuration matter? The case of US subprime lenders. *Corporate Governance: An International Review* 19 (5): 405-417.

Patt A and R Zeckhauser (2000), *Journal of Risk and Uncertainty* (2000) 21, 1: 45-72.

Ritov I and J Baron (1994). Judgments of compensation for misfortune: the role of expectation. *European Journal of Social Psychology*, 24: 525–539.

Schubert R, Brown M, Gysler M, Brachinger HW (1999), Financial decision-making: are women really more risk averse? *American Economic Review* 89 (2): 381-385.

Rabe-Hesketh S, Skrondal A, and Pickles A (2004), GLLAMM Manual. *U.C. Berkeley Division of Biostatistics Working Paper* 160, Online at: http://www.bepress.com/ucbbiostat/paper160

Vieider, F M, Lefebvre, M, Bouchouicha R, Chmura T, Hakimov R, Krawczyk M and Martinsson P (2015), Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries, *Journal of the European Economic Association*, 13, 3: 421-452.

Wallach M, Kogan N, (1965) The roles of information, discussion, and consensus in group risk taking. *Journal of Experimental Social Psychology* 1 (1): 1-19.

Wallach M, Kogan N, Bem D, (1962) Group influence on individual risk taking. *Journal of Abnormal and Social Psychology* 65 (2): 75-86.

Zhang, J, Casari M. (2012): How groups reach agreement in risky choices: an experiment. *Economic Inquiry* 50 (2): 502-515.

Appendix A. Complementary data analysis

(Online)

	All			Academic Year						Program					
				2012			2013			MiM			BBA		
	Males	Females	p	Males	Females	p	Males	Females	p	Males	Females	p	Males	Females	p
Mean	10,05	10,32	0,20	10,16	10,44	0,30	9,92	10,20	0,40	10,01	10,28	0,30	10,12	10,38	0,30
Standard Deviation	3,65	3,42		3,39	3,07		3,78	3,57		3,55	3,40		3,68	3,59	
Median	10	11		11	11		10	10		10	10		11	11	
P25	8	8		9	9		8	8		8	8		9	9	
P75	12	13		13	13		12	12		12	12		13	13	
Min	0	0		0	0		0	0		0	0		0	0	
Max	19	19		19	19		19	18		19	17		19	19	
Number of observations	554	542		302	270		252	272		356	343		198	199	

Table 1A. Distribution of students' grade in microeconomics exam variable

	All			Academic Year						Program					
				2012			2013			MiM			BBA		
	Males	Females	p	Males	Females	p	Males	Females	p	Males	Females	p	Males	Females	p
Mean	6,43	5,78	<0,01	6,56	5,91	<0,01	6,28	5,62	<0,01	6,39	5,81	<0,01	6,50	5,74	<0,01
Standard Deviation	1,70	1,67		1,62	1,67		1,69	1,71		1,66	1,68		1,69	1,64	
Median	7	6		7	6		7	6		7	6		7	6	
P25	6	5		6	5		5	5		5	5		6	5	
P75	8	7		8	7		8	7		8	7		8	7	
Min	1	1		1	1		1	1		1	1		1	1	
Max	10	10		10	10		10	10		10	10		10	10	
Number of observations	477	482		259	243		218	239		335	334		142	148	

Table 2A. Distribution of individual tolerance to risk measure

	% of firm members whose father 's educational level > secondary education		% of firm members whose mother 's educational level > secondary education		Average Age		Average grade in microeconomics		Average tolerance to risk	
	mean	std	mean	std	mean	std	mean	std	mean	std
Number of women in the firm: 0	0,84	0,23	0,85	0,18	19,26	0,90	10,32	2,30	6,41	0,95
Number of women in the firm: 1	0,85	0,18	0,83	0,18	19,07	1,06	10,34	2,09	6,28	0,95
Number of women in the firm: 2	0,84	0,19	0,80	0,20	19,35	0,91	10,54	1,82	6,21	0,92
Number of women in the firm: 3	0,84	0,20	0,84	0,13	19,33	0,75	10,06	1,76	6,11	0,98
Number of women in the firm: 4	0,87	0,17	0,82	0,19	19,06	0,83	10,32	2,32	5,76	0,96
Number of women in the firm: 5	0,85	0,17	0,86	0,15	19,02	1,04	10,20	1,52	5,86	0,95
Number of observations	220		220		220		220		220	

Table 3A. Characteristics of created firms according to gender composition

	(1)	(2)	(3)	(4)	(5)
	OLS	Random-effects GLS	Fixed-effects model	OLS	OLS
	Coef	Coef	Coef	Coef	Coef
Female	-0,636***	-0,634***	-0,633***	-0,630***	-0,630***
Grade in Microeconomics	-0,027*	-0,027*	-0,021*	-0,027*	-0,027*
Age	0,014	0,014	0,019	0,015	0,014
Father 's educational level: > secondary education	0,148	0,143	0,139	0,146	0,142
Mother 's educational level: > secondary education	-0,055	-0,052	-0,048	-0,047	-0,056
Attends MiM program (versus BBA program)	-0,049	-0,051		-0,044	-0,048
Year 2013	-0,275***	-0,272***		-0,260***	-0,258***
Equity level at round 4 / 10000				0,208	
Profit at round 1 / 10000					1,088
Profit at round 2 / 10000					0,442
Profit at round 3 / 10000					-0,166
Profit at round 4 / 10000					0,413
sigma_u		0,227[a]	0,861[b]		
sigma_e		1,67	1,67		
rho		0,019	0,21		
Number of observations	959	959	959	959	959

*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

[a]Breusch and Pagan Lagrangian multiplier test for random effects. Chi2(1)=0,36. Prob>chi2 = 0.2877

[b]F test that all u_i=0: F(219, 734) = 1.09 ; Prob > F = 0.2091

The regressions also included dummies for place of residence

Dependent variable : tolerance to risk (1: unwilling to take risks…….10: Fully prepared to take risks)

Table 4A. Regressions on individual tolerance to risk

	Males		Females	
	Mean	std	Mean	std
Number of women in the firm: 0	6,39	1,63		
Number of women in the firm: 1	6,48	1,77	5,83	1,58
Number of women in the firm: 2	6,41	1,62	5,84	1,69
Number of women in the firm: 3	6,38	1,70	5,73	1,60
Number of women in the firm: 4	6,45	1,88	5,76	1,69
Number of women in the firm: 5			5,81	1,67
Number of observations	477		482	

Table5A. Average "tolerance to risk" for males and females, by team gender composition

	Overall std	Between std	Within std
All observations	1,72	0,90	1,48
Number of women in the firm: 0	1,63	0,75	1,47
Number of women in the firm: 1	1,75	0,94	1,52
Number of women in the firm: 2	1,64	0,91	1,42
Number of women in the firm: 3	1,73	0,96	1,43
Number of women in the firm: 4	1,75	0,76	1,48
Number of women in the firm: 5	1,67	0,85	1,45

Table6A. Decomposition of the standard deviation of the individual "tolerance to risk" variable into between and within standard deviation

	average % of firms members whose father's educational level > secondary education	average % of firms members whose father's educational level > secondary education	average age	average grade in microeconomics
Nb of women in the firm = 0 versus Nb of women in the firm = 1	0,71	0,61	0,40	0,75
Nb of women in the firm = 0 versus Nb of women in the firm = 2	0,60	0,31	0,65	0,65
Nb of women in the firm = 0 versus Nb of women in the firm = 3	0,79	0,51	0,95	0,46
Nb of women in the firm = 0 versus Nb of women in the firm = 4	0,69	0,46	0,21	0,99
Nb of women in the firm = 0 versus Nb of women in the firm = 5	0,79	0,95	0,38	0,58
Nb of women in the firm = 1 versus Nb of women in the firm = 2	0,75	0,53	0,27	0,88
Nb of women in the firm = 1 versus Nb of women in the firm = 3	0,98	1,00	0,41	0,22
Nb of women in the firm = 1 versus Nb of women in the firm = 4	0,48	0,81	0,75	0,75
Nb of women in the firm = 1 versus Nb of women in the firm = 5	0,97	0,53	1,00	0,32
Nb of women in the firm = 2 versus Nb of women in the firm = 3	0,88	0,37	0,71	0,23
Nb of women in the firm = 2 versus Nb of women in the firm = 4	0,44	0,68	0,11	0,65
Nb of women in the firm = 2 versus Nb of women in the firm = 5	0,81	0,22	0,25	0,44
Nb of women in the firm = 3 versus Nb of women in the firm = 4	0,51	0,70	0,16	0,46
Nb of women in the firm = 3 versus Nb of women in the firm = 5	0,95	0,60	0,52	0,60
Nb of women in the firm = 4 versus Nb of women in the firm = 5	0,50	0,40	0,61	0,58

Table 7A. Test of comparison in background characteristics between teams with different gender composition (Wilcoxon rank-sum tests)

Appendix B. The *Kallystée* Business Simulation Game, a primer

The business simulation game Kallystée was developed in the 1990s by Daniel Tixier and Raymond Gambini with the support of *l'Oréal Paris*. As with other business games, *Kallystée* simulates the fundamentals of the market dynamics of a large company, with an established brand name, facing competition from a small number of similar rivals. In *Kallystée*, the virtual company is part of the cosmetics sector. The strong involvement of *l'Oréal Paris* brings a touch of the real business world to the game.

The simulation has several learning goals, which include: (1) providing first-year students, who have very limited practical experience, with a taste of real-life business decision-making; (2) introducing fundamental concepts that will be further developed in specialized classes (e.g., return on assets, balance sheets, interest rates, demand elasticity, advertising, etc.); (3) increasing student awareness of how important firms' strategic interactions are for decision making; and finally, (4) creating strong team-building skills by sharing a common academic experience (participation in the game by all first- year students is mandatory).

In general the game lasts for three entire dedicated days. It alternates periods of decision-making with periods of training and marketing-related lectures.

The key observation unit is the "firm", represented by its "executive board," which comprises five students teamed together at random. Each team makes all decisions over several periods or rounds. Firms are then teamed at random into sets of five and assigned to a "Universe". The composition of each Universe is kept constant during the simulation. Universes are independent of each other. Each Universe reflects a market for skin cream, with five suppliers and its own computer-simulated demand (for each product variety). At the end of the game, a reward is provided to the best firm in each Universe (according to a process described in the next section). Students receive an individual grade to reward their participation and involvement; their presence is strictly monitored (which probably explains why attendance is close to 100% throughout the three days).

The practical organization of the game is logistically demanding. On the first day, the 300 to 400 first year students are organized into teams and provided with the game's rules and general instructions.[19] For each of the 5-firm Universes, a "supervisor" (a professor or a final-year student) is appointed. This supervisor clarifies the most difficult concepts and monitors the teams in terms of team and individual participation. Decisions and interactions are computerized.

[19] Rules and definitions are provided in a 66-page document, available online at : http://behavioralresearchlab.essec.edu/research/research-topics/results.

Teams then move to the school's gymnasium ("The Dome"), transformed into a large workspace for the purposes of the game. Each firm has its own decision desk, where students meet, discuss, analyze and record their decisions using their laptops. Each firm's computers are connected to the ESSEC server. Once decisions are recorded, the simulation is run on the server under the direct control of the administrators. Desks of firms belonging to the same Universe are placed as far as possible from each other.

Within a given Universe, students' names and allocation to teams is public information. Teams are forbidden to communicate their decisions or strategies to other teams, and any collusive behavior is subject to heavy sanctions (poor grades, elimination from the game). The game is conducted in an open space where students can circulate freely, meet during breaks, and have informal discussions about the game. Firms compete each against each other, not against a computer. In this environment, students have little incentive to share information with their rivals.

At the very beginning of the game, students are required to work on a test decision that will be discussed with the administrators but will not be recorded. Then, they move to the regular rounds of the game. On average, the decision time per round is approximately two hours, followed by one hour for the administrators to run the simulation and display the results.

The decision horizon comprises eight successive rounds for the MiM program and five rounds for the BBA students. At the beginning of the game (t=1), all firms have a similar "history", i.e., they are identical in all respects, having a similar balance sheet and a similar stock of inventories. In particular, they have a product ("skin cream") of a relatively modest quality index in stock. During all decision-making rounds, students must make a large number of decisions. For each product, they must make five specific decisions (procurement volume, selling price, commercial discount, referencing budget, and product-specific advertising expenses) and an additional 11 general decisions at the firm level (including general advertising and brand management, sales staff, R&D investment, short-term loans, long-term loans, trade credit, etc.).

Teams can launch new products, knowing that they are allowed to manage at most three products simultaneously. For each new product, the management team can choose an integer between 1 and 15 as a quality index, knowing that the quality index of the existing product is 3. Investment in R&D is needed to launch any new product in the next round. However, once the development phase is completed, the firm can buy the product from an external manufacturer that uses its design. The firm must pay a higher price to acquire a higher quality product.

According to a standard principle in financial economics, companies that take higher risks can expect a better average return. This assumption was introduced in the simulation through a relationship between total demand and the number of product varieties in that market. Accordingly, each new product will push up total market demand and pull the profits of all firms upwards. This relationship is spelled out clearly in the rules of the game.

Figure 1 presents the evolution of the average number of products and of their quality index for the MiM sample, as analyzed in the main text. At the second round, the average number of products was lower than two, showing that not all firms had launched a second product. After round six, firms no longer launched new products, expecting that they would not have enough time left to reap any benefits. The average quality index rose over time.

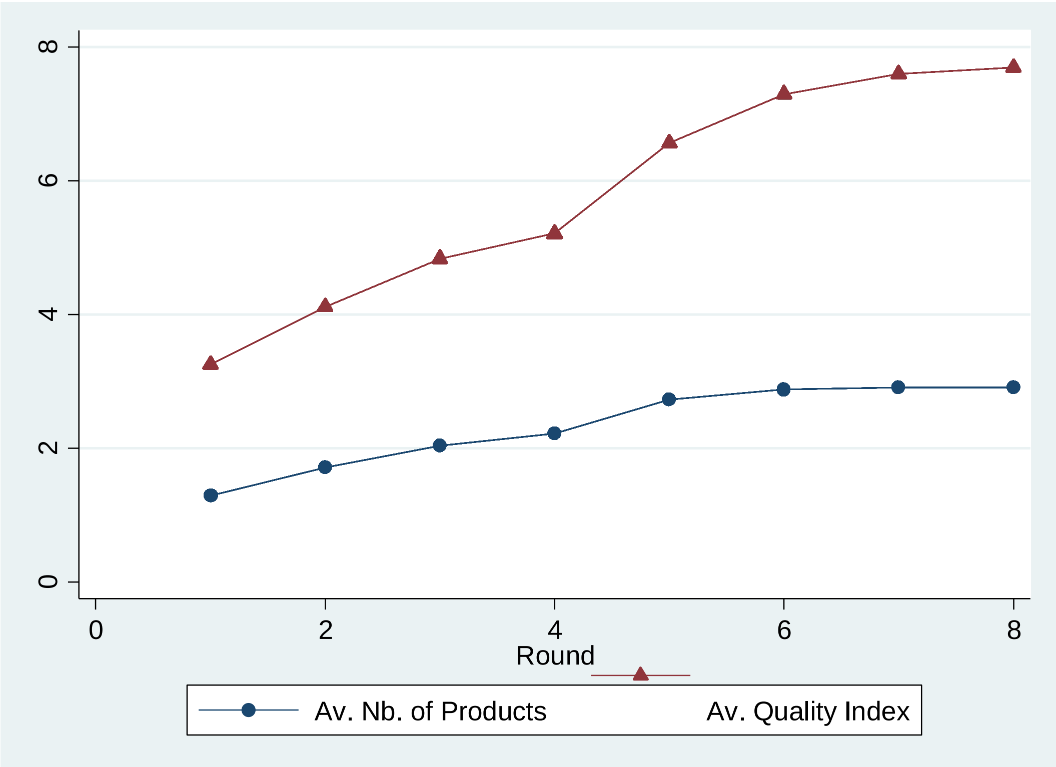


Figure 1. Average number of products and average quality index over time (MiM 2012-2013 sample).

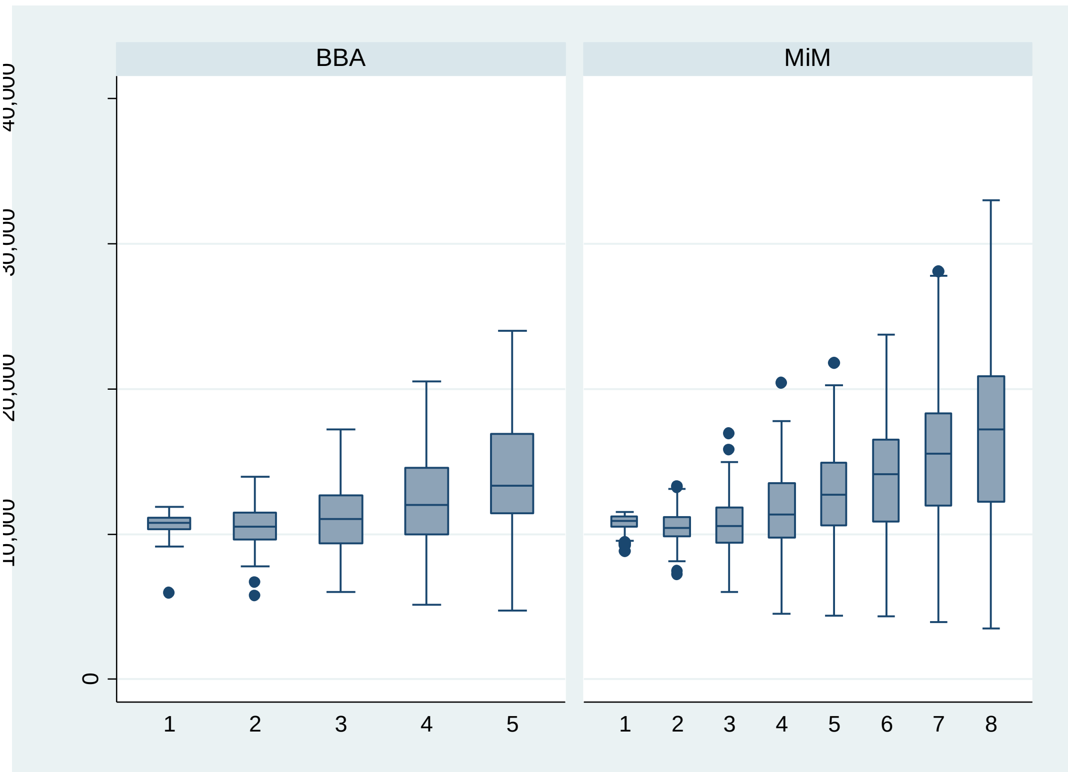
At the end of each round, teams are provided with comprehensive feedback about the consequences of their decisions. A key piece of information is the number of items sold: given the posted price and various fixed and variable costs, this ultimately determines profits and losses. After each round, firms also learn what products were brought to market by their rivals, as well as their sales, price, total marketing spending, and market share. Total sales, net returns (profit or losses) and total equity of *all firms* are displayed on a large electronic board after each round.

For each product, the demand addressed to one specific firm depends not only on total market demand but also on the price and marketing strategies of this firm relative to the prices and marketing strategies of its four rivals. The game allows for a large number of advertising levers (at least six), all having a positive and different impact on demand via the “attractiveness” of the product, but which all come at a cost which must be taken into account. Launching a new product with an optimal combination of general and specific advertising to create its brand identity, and selling it at the “right price”, are the keys to success.

As mentioned above, *Kallystée* involves several financial decisions. Launching a new product requires investment in R&D, and several sources of funding are available. Firms that choose the best financing combination will have lower financial costs. The management of liquidity is also important, and experience shows that this is a difficult topic for first-year students. In the game, poor financial management can be the reason for large losses.

At the beginning of the game, all firms have the same equity. Losses reduce equity, and net-of-tax profits are reinvested in the firm. No dividend is distributed. Thus, total equity increases over time with cumulated profits (and is reduced by losses). Should total equity ever fall to zero, the firm is declared bankrupt.

Figure 2 displays the evolution of the firms’ total equity as recorded during the business games analyzed in our paper. The left-hand graph applies to BBA students, who play the game for five rounds, the right-hand graph to MiM students, who play the game for eight rounds. The box plot representation highlights that the students’ economic performance is quite diverse, with some firms doing very well, while others are close to bankruptcy. The performance gap widens over time.



The box indicates the 75% percentile as the upper hinge, the 25% percentile as the lower hinge and the median; the other elements are the upper / lower adjacent lines, dots are outliers.

Figure 2. The evolution of total equity, BBA (80 firms) and MiM (140 firms)

ESSEC Business School
Avenue Bernard Hirsch
CS 50105 Cergy
95021 Cergy-Pontoise Cedex
France
Tél. +33 (0)1 34 43 30 00
Fax +33 (0)1 34 43 30 01
www.essec.fr

ESSEC Executive Education
CNIT BP 230
92053 Paris-La Défense
France
Tél. +33 (0)1 46 92 49 00
Fax +33 (0)1 46 92 49 90
www.executive-education.essec.fr

ESSEC Asia-Pacific
100 Victoria Street
National Library # 13-02
Singapore 188064
Tél. +65 6884 9780
Fax +65 6884 9781
www.essec.edu/asia

Contact:

Research Center
+33 (0)1 34 43 33 58
research.center@essec.fr

ISSN 1291-9616

ESSEC BUSINESS SCHOOL

affilié à la CCI PARIS ILE-DE-FRANCE