MANAGEMENT SCIENCE
Vol. 58, No. 1, January 2012, pp. 78–93
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.1110.1348

The Impact of Gender Composition on Team Performance and Decision Making: Evidence from the Field

Jose Apesteguia
ICREA, Universitat Pompeu Fabra and Barcelona Graduate School of Economics, 08005 Barcelona, Spain,
jose.apesteguia@upf.edu

Ghazala Azmat, Nagore Iriberri
Universitat Pompeu Fabra and Barcelona Graduate School of Economics, 08005 Barcelona, Spain
{ghazala.azmat@upf.edu, nagore.iriberri@upf.edu}

We investigate whether the gender composition of teams affects their economic performance. We study a large business game, played in groups of three, in which each group takes the role of a general manager. There are two parallel competitions, one involving undergraduates and the other involving MBA students. Our analysis shows that teams formed by three women are significantly outperformed by all other gender combinations, both at the undergraduate and MBA levels. Looking across the performance distribution, we find that for undergraduates, three-women teams are outperformed throughout, but by as much as 0.47 of a standard deviation of the mean at the bottom and by only 0.09 at the top. For MBA students, at the top, the best performing group is two men and one woman. The differences in performance are explained by differences in decision making. We observe that three-women teams are less aggressive in their pricing strategies, invest less in research and development, and invest more in social sustainability initiatives than does any other gender combination.

Key words: gender; teams; performance; decision making
History: Received July 14, 2010; accepted February 11, 2011, by Brad Barber, Teck Ho, and Terrance Odean, special issue editors. Published online in *Articles in Advance* June 20, 2011.

1. Introduction

Gender differences and their impact on economic outcomes have attracted increasing attention, both in the media and in the economic literature. There is evidence for systematic differences in the origins of choice and behavior by gender, namely, in the preferences of men and women. Croson and Gneezy (2009, p. 1), in a comprehensive and exhaustive review of the work on gender differences in economic experiments, summarize the findings as follows: "We find that women are indeed more risk averse than men. We find that the social preferences of women are more situationally specific than those of men; women are neither more nor less socially oriented, but their social preferences are more malleable. Finally, we find that women are more averse to competition than are men."[1]

The gender difference in risk attitudes, social preferences, and preferences over competitive environments has important implications for the understanding of differences in economic and social outcomes. Two prominent issues that have recently been connected to gender differences in preferences are the gender gap in labor markets (Bertrand and Hallock 2001, Bertrand et al. 2010, Manning and Swaffield 2008) and the influence of women in human development (Miller 2008).

The studies on gender differences are most commonly done at the individual level, despite important decisions in modern economies often being made by groups or teams. Committees and boards, business-partners, and even industrial and academic research groups are only a few examples of group decision making in the real world. Interestingly, the recent financial crisis has brought media attention to the gender composition of boards and its influence on the firms' performance.[2] The extrapolation of the findings at the individual level to the group level is not, however, immediately apparent. It is well known that groups have their own idiosyncrasy. For example, a

[1] See also Andreoni and Versterlund (2001), Byrnes et al. (1999), Charness and Gneezy (2007), Croson and Buchan (1999), Gneezy et al. (2003), and Niederle and Vesterlund (2007).

[2] See, for example, Walker (2008) and Kristof (2009).

widely documented phenomenon is group polarization, whereby groups make more extreme decisions than the average of the individual views in the group (see Stoner 1968).[3] Therefore, the influence of gender on group performance and decision making deserves greater attention. This is the focus of this paper.

Dufwenberg and Muren (2006) are a prominent exception in the experimental economics literature because they study the influence of gender composition on group decisions. In a dictator game setting, they find that groups are more generous and equalitarian when women are the majority. They also find that the most generous groups are those with two men and one woman. In the field, Bagues and Esteve-Volart (2010) show that the chances of success of female (male) candidates for positions in the Corps of the Spanish Judiciary are affected by the gender composition of their evaluation committees. They find that female candidates have better chances of success with more males on the committees (see also Zinovyeva and Bagues 2010). Delfgaauw et al. (2009) study the interaction between the managers' gender and the gender composition of the workers. They find that sales competition is effective only in stores where the store's manager and a large fraction of the employees have the same gender. Finally, there are also empirical papers in finance that document a positive relationship between gender diversity in boardrooms and company performance. However, reverse causality is a pervasive problem in these studies, because companies that perform better are quite likely to be companies that also focus more on the gender diversity of their boards (see Carter et al. 2003, Farrell and Hersch 2005, Adams and Ferreira 2009).

In this paper, we explore the influence of the team gender composition on economic performance. We study a large online business game, the L'Oréal e-Strat Challenge, which is played by groups of three. Teams play the role of a general manager of a beauty-industry company, competing in a market composed of four other simulated companies. There are two identical competitions occurring in parallel, one involving undergraduate students and the other involving MBA students. The L'Oréal e-Strat Challenge was designed to simulate real business decisions; hence, teams must take decisions related to brand management, research and development (R&D), and corporate responsibility initiatives. The winning teams receive a 10,000 euros prize, plus a paid trip to Paris. Perhaps even more importantly, winning candidates have the possibility of being hired by L'Oréal. Our database consists of the last three editions of the L'Oréal e-Strat Challenge, from the years 2007, 2008, and 2009, yielding a total of 37,914 participants, organized into more than 16,000 teams from 1,500 different universities that are located in around 90 different countries around the world.

The L'Oréal e-Strat Challenge offers a unique setting to study the influence of the gender composition of teams on performance and on decision making. First, it is played worldwide, by a large number of individuals, coming from a large number of different institutions. Second, there are two separate competitions, involving two different subject pools, undergraduate and MBA students. Each subject pool constitutes relevant samples to study the influence of gender composition of teams. The former represents a subject pool that is close to the one typically used in the experimental economics literature. This may facilitate comparisons of established results at the individual level with new findings that emerge here at the group level. MBA students are also relevant because they represent a unique and important sample. These are subjects that, with a high probability, will play a key role in real-world business management. Hence, it is relevant to understand how these subjects interact in groups, conditioning on the gender composition in teams. Third, we can study the effect of the gender composition on performance and on other important aspects such as specific business decisions. Fourth, as mentioned, the game aims to simulate the business environment as close as possible to the real world. Finally, this study also offers an important advantage over existing empirical studies. In particular, reverse causality is less of a concern here. Teams are formed before their performance and, more importantly, remain fixed over the entire game.

Our analysis shows that teams formed by three women are significantly outperformed by any other gender combination, in both the undergraduate and MBA competitions. The magnitudes are sizable, about 0.18 of a standard deviation from the mean for the undergraduate students and about 0.17 for the MBA students. We show that these findings are robust after controlling for a number of important variables, such as the quality of the institutions, fields of study, and geographical areas.

When we extend our analysis to consider the distributional effects, we find that the performance of three-women teams shows interesting variations along the distribution. In the undergraduate competition, although the underperformance of three-women teams remains significant along the entire distribution, there is a marked decrease as we move to the right on the performance distribution. Among the lowest 10%, three-women teams are outperformed by as much as 0.47 of a standard deviation from the

[3] See also Sobel (2006) for a theoretical account and for references to the empirical literature. For other established differences between individuals and teams, see, e.g., Charness and Jackson (2007) and Charness et al. (2007).

mean, whereas at the top 20%, three-women teams are outperformed by only 0.09. We also find evidence that the optimal gender composition along the whole distribution is that of two men and one woman, although the differences are not statistically significant. In the MBA competition, the performance levels of all the gender combinations are higher than that of three-women teams along the entire distribution, although the differences are not always significant. Interestingly, at the top 10%, the team composed of two men and one woman shows to be the optimal gender combination.

After establishing the differences in performance, we seek to understand which decisions drive these differences. First, we find that three-women teams invest significantly less in R&D, both in the undergraduate and MBA competitions. A possible interpretation of these differences is that three-women teams are more conservative in their management vision. Second, teams differ in their decisions related to another crucial aspect for performance, namely, brand management. In both competitions, three-women teams show significantly lower profits. We identify an important difference in the pricing strategy that leads to these differences: three-women teams are pricing their products higher than any other gender combination. That is, these teams are significantly less aggressive in their pricing strategies, and this has consequences on sales, on profits, and ultimately on economic performance. Finally, we observe differences in decisions related to corporate responsibility. We find that in both competitions, three-women teams invest significantly more in social initiatives than does any other gender composition.

Although many of our results can be interpreted as being in line with established results in the literature on individual gender differences, we also find new results, which are idiosyncratic to the team level. First, although we do find that three-women teams perform worse than other teams, we do not find any evidence for a monotonic relationship between the number of men in the team and their performance, such that teams with one man and three men are not different. Moreover, we find evidence that the best gender composition of a team is two men and one woman. In particular, for MBA students our distributional analysis shows that for the top 10% of the distribution, the two men and one woman combination is highly significant.

In our setting, because teams are not exogenously formed, there are competing explanations for the findings we report. It is important to consider and identify the main driving force because the policy, social, and scientific conclusions are very different depending on the explanation. In the final section, we discuss at length the potential explanations.

The rest of this paper is organized as follows. In §2, we introduce the L'Oréal e-Strat Challenge. Section 3 is devoted to the presentation of the demographics in our data. Section 4 establishes the main result on the effect of gender composition on performance. Section 5 is devoted to the understanding of where the performance differences come from. In §6, we discuss alternative explanations for our findings.

2. The Game: L'Oréal e-Strat Challenge

2.1. Overview of the Game

The L'Oréal e-Strat Challenge is one of the biggest online business simulation games. It was designed by StratX for L'Oréal. The game was launched in 2000, and since then there have been more than 250,000 participants from more than 2,200 institutions spread all over the world. It is open to all students in their final or penultimate year of undergraduate study, or studying an MBA, registered at a university anywhere in the world. Undergraduate and MBA students participate in separate competitions of the same game. All team members must attend the same university and must provide the following information: name, official ID, age, gender, university, field of study, and country of origin. Teams that do not comply with these requirements are discarded.

The game is played by teams of three members. Each of the teams plays the role of a general manager of a beauty-industry company, competing in a market composed of four other simulated companies. That is, the participating teams do not compete with one another in the same market. The game was designed to simulate real business decisions. In turn, teams make decisions that are related to R&D, brand management, and corporate responsibility initiatives. The rules of the game are clearly stated in the detailed instructions. A proper understanding of the instructions requires a good deal of time and effort.[4] The game is comprised of six rounds, plus a final round that is in a different format. The main performance variable is the *stock price index* (*SPI*), which measures the market value of the company, as a consequence of team's decisions, as well as the decisions taken by the competing (simulated) firms. As such, the SPI is not only determined by current profits alone but also by broader management decisions that may be exploited in the future, such as investment in R&D.

The initial conditions are identical for all participating teams. Subsequently, after decisions are taken in round one, the SPI is computed and only the best 1,700 participating teams in terms of their SPI are

[4] The instructions from the 2008 edition are available from the authors upon request.

selected to pass to the second round, taking into account country and zone quotas. The winning teams in the final round, one undergraduate team and one MBA, are awarded with a 10,000 euros prize each. Perhaps more importantly, they have the opportunity to meet high profile professionals at L'Oréal, and some of them are offered employment opportunities.

In this paper, for an evenhanded comparison across teams, we will use data from the first round only. The starting situation in round one is exactly the same for all teams; therefore, the decisions and associated performances in the first round are fully comparable across teams. This is not the case for other rounds because teams' performance depends on their game history.

2.2. Management Decisions

We distinguish between three different types of decisions that teams must undertake. These decisions generate what we refer to as *midway* outcome variables that in turn affect the final, and most important, outcome variable, the SPI.

First, teams make decisions regarding the investment in R&D. Teams have an R&D department, where researchers discover two new formulae. These formulae should be interpreted as innovations that, if developed, can be used to create new brands or improve the existing ones. Teams must then make two main decisions. First, teams must decide whether to invest in zero, one, or two formulae. Second, if they decide to invest in any formula, they must specify the amount that they wish to spend. These decisions form the midway outcome variable, namely, total R&D investment.

Second, teams must manage their brands. All teams start with the same two brands. Brands differ in their characteristics to target a specific customer profile. Participants know whether the brand is targeting high earners; affluent families, medium income families, low-income singles, or low-income families. In each edition, there are different brand targets for each of the two brands. The two main decisions that teams must make are the price and the production level for each of the two brands. These, in turn, influence the main midway outcome variables regarding the brand management, which are composed of sales, revenues, production costs, and inventory costs. Finally, these variables determine profits.

Third, teams must decide how much to invest in social and environmental initiatives. The former includes initiatives such as having health programs or continuous learning plans for employees. The environmental initiatives include actions such as using renewable raw materials, reducing water consumption, or having safety and health compliant plants. Teams' investment in these initiatives determine the *social sustainability* and *environmental sustainability* indices, which are the main two midway outcome variables in this area.

Overall, the decisions made in all three of these areas affect the market value of the company, and this is incorporated in the main performance variable, the SPI.

2.3. Data and the Relation Between Managerial Decisions and Performance

Our database consists of the last three editions of the L'Oréal e-Strat Challenge from the years 2007, 2008, and 2009. This includes a total of 37,914 participants from 1,500 different universities, located in about 90 different countries around the world.[5]

Success in the L'Oréal E-Strat Challenge, represented by high values in SPI, is determined by the midway outcome variables and ultimately by teams' management decisions. In this section, we elaborate on the relationship between decision variables, midway outcome variables, and the final performance variable, the SPI. This will facilitate the understanding and interpretation of the team differences in performance. We look at the associations across variables in two ways.

First, we show that there exists a relationship between the midway outcome variables and the SPI. We use a simple regression analysis, where we relate the final performance measure, the SPI, to the midway decision variables, separately for the two competitions, undergraduate and MBA students. The results are shown in columns (1) and (2) of Table 1. For both competitions, all the midway outcome variables are positively and significantly related to SPI. We also estimate these regressions for each of the editions separately and we find qualitatively the same results; see Table A.3 in the online appendix, available on the authors' websites (http://www.econ.upf.edu/~{apesteguia, azmat, iriberri}). When we run regressions separately for each of the midway outcome variables to SPI, shown in Table A.4 in the online appendix, we see that there are large differences in the importance of each variable. The values for the R-squared show that, not surprisingly, most of the variation in SPI is explained by the

[5] About 55% of the registered teams do not participate in the game, such that there is attrition. More importantly, we find that there are no significant differences in attrition across different gender combinations. For the undergraduate competition, the attrition rates are 54%, 55%, 53%, and 55% for teams that are all women, one man and two women, two women and one man, and three men, respectively. For the MBA competition, the corresponding attrition rates are 59%, 57%, 56%, and 55%. In both cases, a chi-square test cannot reject the null hypothesis of independence between the gender composition and attrition rate variables. This suggests that reverse causality is less of a concern in our data set.

Table 1 Explaining the SPI

	SPI			
	UG (1)	MBA (2)	UG (3)	MBA (4)
R&D inv.	0.340***	0.445***	5.870***	6.404***
	[0.121]	[0.131]	[0.340]	[0.505]
Profits	0.360***	0.271***	0.199***	0.481***
	[0.0221]	[0.0717]	[0.0377]	[0.112]
SSI	0.119**	0.151***	96.36**	23.71
	[0.0454]	[0.0498]	[39.43]	[111.2]
ESI	0.182***	0.181***	−12.31	−17.13
	[0.0202]	[0.0427]	[7.272]	[12.87]
R&D inv.^{2}			−0.129***	−0.150***
			[0.0105]	[0.0143]
Profits^2			0.000189***	−0.0000092
			[3.32e−05]	[4.64e−05]
SSI^2			−0.0895**	−0.0216
			[0.0373]	[0.105]
ESI^2			0.0111	0.0152
			[0.00661]	[0.0117]
R&D inv.^{3}			0.000766***	0.000979***
			[8.17e−05]	[0.000112]
Profits^3			−4.17e−08***	−5.55E−09
			[7.99e−09]	[4.65e−09]
SSI^3			2.77e−05**	6.56E−06
			[1.17e−05]	[3.30−05]
ESI^3			−3.30E−06	−4.45E−06
			[2.00e−06]	[3.54e−06]
Constant	−146.8*	21.34	−30.068**	−2.408
	[80.91]	[161.7]	[14.447]	[39.448]
Year FE	Yes	Yes	Yes	Yes
Observations	7,531	2,889	7,531	2,889
R-squared	0.668	0.523	0.807	0.739

Notes. The standard errors in all columns are clustered at the year and zone level. *R&D investment* and *Profits* are measured in units of 100,000.
*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

variation in profits and to a lesser extent by the variation in investment in R&D and the expenditure on social and environmental initiatives. Furthermore, when we extend this simple framework and allow for a nonlinear relationship, we see that the fit improves. Columns (3) and (4) in Table 1 include polynomials up to the third degree, where we see that the R-squared improves a great deal, explaining up to 80% of the variation in SPI.[6] This implies that the midway outcome variables are the key variables in determining the SPI.

Second, we examine the ex post decisions and performance of the top and bottom performing teams. Table 2 reports the mean for each of the decisions, midway outcomes, and the SPI, separately for the top and bottom 10% teams. Columns (3) and (6) show the p-values for the one-way ANOVA test of equality of means across the top and bottom performing teams' decisions and outcome variables. We can clearly see that there are sizeable and highly significant differences in the decisions and in the outcome variables of these two groups, both in the undergraduate and MBA competitions.

As for the midway outcome variables, top and bottom performing teams differ in all of them, with the exception of total costs and social sustainability index. The top 10% teams invest more in R&D; they have significantly higher sales, revenues, and profits, and they hold significantly fewer inventories. Finally, the top 10% teams have significantly higher environmental sustainability index.

As for the specific decision variables, top and bottom performing teams again differ significantly. Top performing teams have on average more formulae, and their pricing and production strategies are also systematically different. For high- and medium-income brands, the top 10% set significantly lower prices, whereas for singles and low-income brands, the top 10% set significantly higher prices. Finally, for all brands, the top performing teams produce significantly more.

3. Demographics

We now look at the main demographic variables, shown in Table 3. In the undergraduate competition there is a total of 12,759 women and 14,525 men, and in the MBA competition there are 3,934 women and 6,697 men. Participation, by gender, in the undergraduate competition is comparable (47% women and 53% men), whereas in the MBA competition, men are more prevalent (37% and 63%). These proportions are representative of the real-life gender ratios in undergraduate and MBA studies.[7]

The L'Oréal e-Strat Challenge is played by teams of three people. We classify teams into four categories: all women; one man and two women; two men and one woman; all men. We denote the team composition by Mx, where x is the number of men in a team and $(3-x)$ is the number of females. In the undergraduate competition, the distribution of teams by gender composition is 19%, 27%, 30%, and 24% for M0, M1, M2, and M3 teams, respectively; in the MBA competition, it is 11%, 23%, 33%, and 33%.

[6] We also try other nonlinear specification, such as using a logarithm model and by including interactions between variables. However, these specifications were dominated by the polynomial model. The results for these specifications can be found in Table A.5 in the online appendix.

[7] For undergraduates, according to the *World Development Indicators* database (World Bank 2008), the average worldwide ratio of female to male enrollments in tertiary education is 105.3. For MBA students, Bertrand et al. (2010) report that the U.S. average of MBA degrees earned by women in the last two decades is about 40%.

Table 2 Comparison of Top and Bottom Performing Teams' Decision and Outcomes

	Undergraduates			MBA Students		
	Top 10% mean (1)	Bottom 10% mean (2)	*p*-value (3)	Top 10% mean (4)	Bottom 10% mean (5)	*p*-value (6)
SPI	1,098.55	763.08	0.00	1,100.06	792.36	0.00
No_formula	1.56	1.49	0.00	1.56	1.41	0.00
R&D investment	2,843,832	2,515,079	0.00	2,891,775	2,451,551	0.01
Price high-income brand	28.93	30.51	0.00	28.86	30.15	0.00
Price medium-income brand	16.52	17.74	0.00	16.47	17.84	0.00
Price singles brand	12.55	12.08	0.00	12.47	11.98	0.10
Price low-income brand	7.41	6.61	0.00	7.38	6.65	0.00
Production high-income brand	6,013,914	6,988,092	0.00	6,945,655	6,060,997	0.00
Production medium-income brand	8,133,023	9,975,688	0.00	1.00E + 07	8,371,768	0.00
Production singles brand	9,467,281	1.01E + 07	0.00	1.00E + 07	9,829,152	0.28
Production low-income brand	1.47E + 07	1.53E + 07	0.00	1.53E + 07	1.46E + 07	0.00
Sales	2.26E + 07	1.95E + 07	0.00	2.26E + 07	2.00E + 07	0.00
Inventories	177,317.6	1,615,838	0.00	160,306.1	1,307,136	0.00
Revenues	2.80E + 08	2.26E + 08	0.00	2.80E + 08	2.34E + 08	0.00
Cost	4.75E + 07	4.75E + 07	1.00	4.75E + 07	4.81E + 07	0.09
Profits	2.32E + 08	1.79E + 08	0.00	2.33E + 08	1.86E + 08	0.00
SSI	1,022.13	1,021.16	0.48	1,017.67	1,019.03	0.51
ESI	1,072.84	1,050.03	0.00	1,079.32	1,049.38	0.00

Notes. In 2007, one brand is a low-income brand, and the other is a high-income brand. In 2008, one brand is a low-income brand, and the other is a brand directed to singles. In 2009, one brand is a brand directed to singles, and the other is a medium-income brand. For further description, see §5.2. The *p*-values for the one-way ANOVA test of equality of means across the average decisions and outcomes between the top and bottom 10% teams are shown in columns (3) and (6).

With respect to the demographic variables, undergraduate and MBA students differ on a number of expected dimensions. MBA students are older, study more business-related subjects, and are more likely to study in foreign institutions. At both the undergraduate and MBA levels, the four different types of teams look very similar in terms of their characteristics. We do see that at the undergraduate level, M0 teams are formed by students with less science-oriented studies. Also, M0 teams are slightly younger than M3 teams both at the MBA and undergraduate levels. Finally, it is interesting to note that there is slightly more field diversity in mixed teams both in the MBA and in the undergrad teams. In the analysis that follows, we will control for all of these characteristics.

4. Does the Gender Composition of Teams Matter for Performance?

4.1. The Overall Effect

We start our analysis by looking at the main performance variable, namely, the SPI. In what follows, we will use standardized SPI.[8] To understand whether the gender composition of the team has any effect on performance, we estimate the following equation by ordinary least squares, separately for the undergraduate and MBA competitions:

$$Y_i = \alpha + \beta_1 \mathrm{M1}_i + \beta_2 \mathrm{M2}_i + \beta_3 \mathrm{M3}_i + X'\theta + \delta_j + \gamma_t + \varepsilon_{ijt}, \quad (1)$$

where the dependent variable Y_i denotes the standardized SPI of team i. The gender composition of the teams is captured by the variables M1, M2, and M3, where Mi takes the value of 1 when the team has i man and $(3 - i)$ women and 0 otherwise. The omitted category, to which these variables are compared, refers to M0 teams. A vector of variables that control for the mean age, field of study, field diversity, country diversity, and institution diversity is denoted by X.[9] In addition, we control for geographical areas (i.e., zones), δ_j, and year fixed effects, γ_t. Finally, we cluster the standard errors at the zone and year level.

Table 4 reports the results from estimating Equation (1). Columns (1) and (2) show the effect of gender composition of teams on SPI for the undergraduate and MBA competitions, respectively, without any controls. We find that the three-women teams are significantly outperformed by any other gender composition, both at the undergraduate and MBA levels. In the undergraduate competition, we see that the teams with one man, two men, and three men significantly outperform three-women teams by 0.15,

[8] SPI values are standardized to a distribution with zero mean and a unit standard deviation by subtracting the mean SPI from the SPI and dividing this difference by the standard deviation of the SPI, for each competition in a given year. In the same way, we standardize the dependent variables we study in §5.

[9] See the notes in Table 3 for the definitions of these variables.

Table 3 Demographics by Gender Composition of Teams

	Undergraduates				MBA Students			
	3 women	2 women–1 man	1 woman–2 men	3 men	3 women	2 women–1 man	1 woman–2 men	3 men
SPI	960.85	977.60	981.20	973.88	963.68	977.19	986.46	984.02
	(120.96)	(99.89)	(99.40)	(108.23)	(107.56)	(104.29)	(95.24)	(98.15)
Mean age	22.348	22.690	22.922	22.992	26.094	26.916	27.651	27.385
	(1.787)	(2.178)	(2.059)	(2.144)	(2.905)	(3.115)	(3.635)	(3.753)
Business (*team*)	0.650	0.689	0.670	0.614	0.877	0.897	0.901	0.906
	(0.477)	(0.463)	(0.470)	(0.486)	(0.328)	(0.303)	(0.298)	(0.291)
Economics (*team*)	0.422	0.442	0.407	0.419	0.279	0.270	0.299	0.292
	(0.494)	(0.496)	(0.491)	(0.493)	(0.449)	(0.444)	(0.458)	(0.455)
Sciences (*team*)	0.138	0.224	0.291	0.285	0.062	0.091	0.096	0.089
	(0.345)	(0.417)	(0.454)	(0.451)	(0.242)	(0.288)	(0.295)	(0.285)
Other fields (*team*)	0.046	0.027	0.013	0.015	0.028	0.009	0.009	0.000
	(0.210)	(0.164)	(0.114)	(0.122)	(0.167)	(0.099)	(0.096)	(0.029)
Central Europe (*institution*)	0.073	0.072	0.091	0.124	0.127	0.139	0.136	0.143
	(0.260)	(0.258)	(0.288)	(0.329)	(0.334)	(0.346)	(0.343)	(0.350)
South Europe (*institution*)	0.052	0.060	0.077	0.101	0.091	0.087	0.094	0.091
	(0.223)	(0.238)	(0.266)	(0.302)	(0.288)	(0.283)	(0.292)	(0.287)
Eastern Europe (*institution*)	0.110	0.105	0.103	0.122	0.073	0.058	0.048	0.040
	(0.313)	(0.307)	(0.305)	(0.328)	(0.260)	(0.234)	(0.215)	(0.197)
Africa (*institution*)	0.063	0.067	0.093	0.121	0.062	0.049	0.065	0.059
	(0.244)	(0.251)	(0.291)	(0.327)	(0.242)	(0.216)	(0.248)	(0.235)
South America (*institution*)	0.039	0.052	0.062	0.088	0.041	0.051	0.054	0.067
	(0.194)	(0.222)	(0.242)	(0.283)	(0.200)	(0.221)	(0.227)	(0.251)
North America (*institution*)	0.038	0.037	0.044	0.033	0.258	0.163	0.126	0.101
	(0.191)	(0.189)	(0.205)	(0.179)	(0.438)	(0.369)	(0.332)	(0.301)
East Asia (*institution*)	0.453	0.447	0.354	0.201	0.187	0.242	0.215	0.068
	(0.497)	(0.497)	(0.478)	(0.401)	(0.391)	(0.428)	(0.411)	(0.252)
South Asia (*institution*)	0.165	0.154	0.168	0.203	0.154	0.201	0.252	0.425
	(0.371)	(0.361)	(0.374)	(0.402)	(0.361)	(0.401)	(0.434)	(0.494)
Area others (*institution*)	0.002	0.003	0.002	0.002	0.002	0.006	0.004	0.003
	(0.054)	(0.056)	(0.054)	(0.047)	(0.051)	(0.078)	(0.065)	(0.058)
Field diversity	0.545	0.735	0.741	0.595	0.454	0.566	0.582	0.490
	(0.706)	(0.756)	(0.781)	(0.728)	(0.640)	(0.700)	(0.720)	(0.685)
Country diversity	0.120	0.144	0.171	0.142	0.507	0.491	0.470	0.376
	(0.384)	(0.417)	(0.449)	(0.410)	(0.751)	(0.745)	(0.728)	(0.684)
Institution diversity	0.284	0.322	0.358	0.341	0.762	0.803	0.775	0.732
	(0.792)	(0.824)	(0.854)	(0.845)	(1.130)	(1.174)	(1.159)	(1.158)
Number teams 2007	623	913	1,032	893	167	335	501	444
Number teams 2008	600	867	923	787	119	277	377	442
Number teams 2009	484	692	739	543	97	197	289	299
Total number of teams	1,707	2,472	2,694	2,223	383	809	1,167	1,185

Notes. *Mean age* is the average age of the team members. Fields of study: *Business* (*team*), *Economics* (*team*), *Sciences* (*team*), and *Other fields* (*team*) are dummy variables representing categories for the fields of study of the individuals in the team. A given category takes a value of 1 if the field of study of any of the three individuals of the team belongs to that category and 0 otherwise. Table A.1 in the online appendix reports the classification of fields of study in the four categories. Geographical areas are dummy variables taking a value of 1 if the country where the institution is located belongs to the respective geographical area and 0 otherwise. Institutions that are unclassified geographically are collected in *Area others* (*institution*). Table A.2 in the online appendix reports the assignment of countries to the geographical areas. *Field diversity* takes a value of 0, of 1, or of 2 if the maximum number of team members with fields of study belonging to the same category is 3, 2, or 1, respectively. *Country diversity* takes a value of 0, of 1, or of 2 if the maximum number of team members with the same country of origin is 3, 2, or 1, respectively. *Institution diversity* takes a value of 0, of 1, of 2, or of 3, if the number of team members originally from a country different to the country of the institution is 0, 1, 2, or 3, respectively. Standard errors are reported in parentheses.

0.19, and 0.15 of a standard deviation from the mean, respectively. The corresponding values at the MBA level are 0.15, 0.24, and 0.23, respectively.

Columns (3) and (4) in Table 4 include the control variables X as well as the year and zone fixed effects. The differences persist. We find that teams with one man, two men, and three men outperform three-women teams by approximately 0.15, 0.21, and 0.18 of a standard deviation from the mean, respectively, in the undergraduate competition. The

Table 4 Gender Composition of Teams on SPI

	Standardized SPI							
	UG (1)	MBA (2)	UG (3)	MBA (4)	UG (5)	MBA (6)	UG (7)	MBA (8)
M1	0.153***	0.152*	0.153**	0.12	0.138**	0.151*	0.146***	0.110*
	[0.0546]	[0.0780]	[0.0556]	[0.0723]	[0.0530]	[0.0836]	[0.0313]	[0.0616]
M2	0.192***	0.243***	0.207***	0.197***	0.211***	0.199***	0.196***	0.168***
	[0.0459]	[0.0643]	[0.0497]	[0.0647]	[0.0491]	[0.0691]	[0.0314]	[0.0600]
M3	0.146***	0.231***	0.176***	0.173***	0.162***	0.163**	0.182***	0.162***
	[0.0273]	[0.0678]	[0.0316]	[0.0602]	[0.0349]	[0.0707]	[0.0331]	[0.0615]
Constant	−0.134***	−0.192***	−0.26	−0.584***	−0.1	−0.746***	−0.152	−0.425
	[0.0385]	[0.0562]	[0.273]	[0.167]	[0.252]	[0.179]	[0.211]	[0.266]
Controls	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Zone FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
School rank	No	No	No	No	Yes	Yes	Yes	Yes
School FE	No	No	No	No	No	No	Yes	Yes
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Observations	9,099	3,545	8,998	3,482	7,535	2,435	8,998	3,482

Notes. The standard errors in all columns, except in columns (7) and (8), are clustered at the year and zone level. The dependent variable, SPI, is standardized to a distribution with zero mean and a unit standard deviation. The excluded team gender category is M0 (i.e., all-women teams).
*, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

corresponding values at the MBA level are 0.12, 0.20, and 0.17. These differences are significant at conventional levels, except in the MBA case for the relationship between M1 teams and M0 teams, which is significant only at the 11% level. Furthermore, there are no significant differences among teams of one man, two men, and three men, suggesting that there is not a monotonic relationship between the number of men and team performance. Therefore, it is the case that the main difference is between three-women teams and all other gender combinations. However, it is worth mentioning that from the magnitudes, we see that there is some suggestive evidence that teams with one woman and two men are the best performing teams; however, we do not find statistical significance for this. When we estimate these regressions for each of the editions separately, and we find qualitatively similar results (see Table A.6 in the online appendix).

With respect to the control variables, it is important to control for year fixed effects because the different editions have some variations. Other variables, such as age at the MBA level, are also important in explaining the differences in SPI. Given that the impact of experience on performance may differ for women and men, we interact age with team's gender composition, shown in Table A.7 in the online appendix.[10] We see that for both MBA and undergraduate students, the coefficients on the gender composition variables remain significant and positive, showing that age is not driving the results on the underperformance of three-women teams.

We next check for the robustness of the overall effect. In particular, we study the influence of the quality of the institution attended by the team members. One potential explanation for three-women teams being outperformed is that all-women teams, when compared with the other team compositions, are attending a university or business school that is of a poorer quality, thus reflecting a low ability level of the team members. We address this point in two ways. First, we use measures of institutional quality that are external to the L'Oréal e-Strat Challenge. Namely, we use the 2009 Ranking Web of World Universities as a measure of the quality of the school for the undergraduate competition and the 2009 *Financial Times* Ranking of MBAs for the MBA competition.[11] These rankings contain around 85% of the universities and 70% of the business schools in our database. Columns (5) and (6) in Table 4 report the results when we include the ranking of the schools. The results are robust to the inclusion of this additional control, and the coefficients and significance levels are very similar. Second, we control for the institutional quality by including institution fixed effects. Because we observe many of the same institutions over the years, by adding the fixed effect, we can control for the

[10] First, we interact the gender composition variables with the deviation of the average age in the team from the average age in the corresponding sample. Second, we interact the gender composition variables with the deviation of the maximum age of the team from the average age in the corresponding sample.

[11] http://www.webometrics.info/top6000.asp (accessed 2009) and http://rankings.ft.com/businessschoolrankings/global-mba-rankings (accessed 2009).

quality of each institution as well as for any other school-specific characteristic. Columns (7) and (8) in Table 4 report the results. Once more, we find that three-women teams are outperformed by teams of any other gender composition. Furthermore, the magnitude and the significant levels remain the same.[12]

Therefore, we conclude that the underperformance of three-women teams is present at both the undergraduate and MBA competitions. Also, interestingly, looking at the point estimates, there is some suggestive evidence that the best performing gender combination is the mixed team with two men and one woman, both at the undergraduate and MBA levels, although it is not statistically significant at conventional levels.

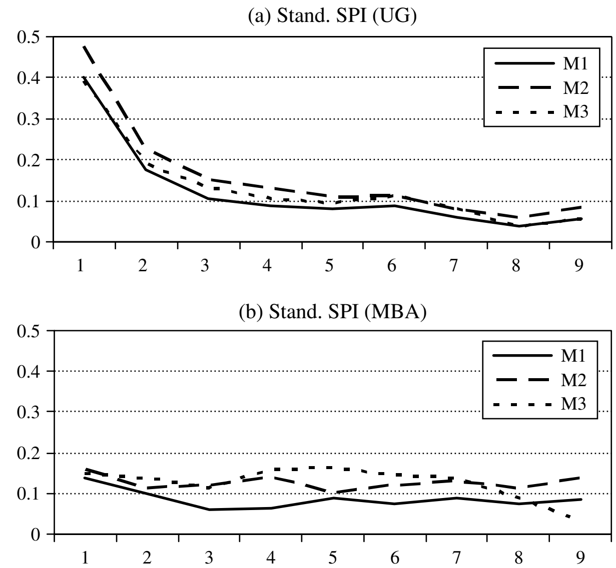
4.2. Distributional Analysis

Our estimation analysis so far has focused on the mean effect of the influence of teams' gender composition on SPI. However, it is also important to understand how the effect of the team's gender composition on SPI varies at different points of the performance distribution. In order to do so, we estimate quantile regressions using Equation (1). The results are shown in Figures 1(a) and 1(b), where we plot the coefficients for the different gender compositions, M1, M2, and M3, relative to the omitted category, M0, for each quantile, for undergraduates and MBA students, respectively. Thus the distance between the coefficients, with respect to the horizontal axis, reflects the distance with respect to M0 teams. Table 5 reports the point estimates for the 10th, 20th, 50th, 80th, and 90th quantiles.

We start by analyzing the undergraduate case. Figure 1(a) and panel A of Table 5 show that M0 teams are significantly dominated by teams with any other gender combination throughout the entire performance distribution. Interestingly, we see large disparities in the magnitudes of these effects. Most notably, the largest differences come from the bottom of the performance distribution, and they decrease monotonically along the distribution. Whereas for teams whose performance is at the bottom 10% of the distribution, the M0 teams are outperformed by 0.40, 0.47, and 0.39 of a standard deviation of the mean by M1, M2, and M3 teams, respectively, for teams whose performance is at the top 10% of the distribution, the M0 teams are outperformed by less than 0.09 of a standard deviation.

[12] We also consider a different way of controlling for the team's field of study. We construct dummy variables that identify every possible combination of different fields of study; e.g., *EBS* is a dummy variable that takes the value of 1 when the team is composed of an economics student, a business student, and a science student and 0 otherwise. The results are both qualitatively and quantitatively the same.

Figure 1 Distributional Analysis for (a) Undergraduates and (b) MBA Students



Notes. 10% to 90% quantile. Table 5 reports the coefficients for the 10%, 20%, 50%, 80%, and 90% levels, and includes the significance levels. All quantiles control for observable characteristics (as listed in Table 4) as well as control for year and zone fixed effects. The standard errors in all columns are clustered at the year and zone level. The dependent variable, SPI, is standardized to a distribution with zero mean and a unit standard deviation for all quantiles. The team gender categories, M1, M2, and M3, are all compared with the excluded category, M0.

These results are informative for three reasons. First, we see that the underperformance of three-women teams is persistent throughout the entire distribution. Second, there is a great deal of heterogeneity in the disparity. In particular, it is important to stress that the high performing three-women teams are much more similar to teams of any other gender combination. Third, the point estimates suggest that the mixed team, composed of two men and one woman, has the highest performance levels all over the distribution. This is in line with our findings in §4.1, when studying the overall effect. However, these differences are not significant at conventional levels, so they provide only suggestive evidence.

Figure 1(b) and panel B in Table 5 report the results for the MBA students. The coefficients for all the gender combinations, with respect to M0, are positive along the entire distribution, suggesting that three-women teams are underperforming. Interestingly, unlike the undergraduates, the differences are less robust across the distribution and are often insignificant, especially in the bottom half of the distribution. Furthermore, at the top 10 percentile, we see that the only gender composition performing significantly better than M0 teams is the team composed of two men and one woman. This again provides evidence in favor of gender diversity at the top of the performance distribution.

Table 5 Gender Composition of Teams on SPI: Distributional Analysis

	Panel A: Undergraduates					Panel B: MBA students				
	Stand. SPI 10%	Stand. SPI 20%	Stand. SPI 50%	Stand. SPI 80%	Stand. SPI 90%	Stand. SPI 10%	Stand. SPI 20%	Stand. SPI 50%	Stand. SPI 80%	Stand. SPI 90%
Quantiles	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
M1	0.400***	0.175***	0.0797***	0.0375*	0.0580***	0.138	0.065	0.0889**	0.0727	0.0441
	[0.0885]	[0.0431]	[0.0293]	[0.0223]	[0.0213]	[0.119]	[0.0613]	[0.0420]	[0.0481]	[0.0711]
M2	0.476***	0.224***	0.108***	0.0614***	0.0859***	0.16	0.141**	0.103**	0.113**	0.138**
	[0.0887]	[0.0429]	[0.0291]	[0.0221]	[0.0212]	[0.115]	[0.0591]	[0.0405]	[0.0465]	[0.0681]
M3	0.392***	0.189***	0.0938***	0.0383	0.0555**	0.149	0.159***	0.164***	0.0871*	0.0314
	[0.0920]	[0.0447]	[0.0306]	[0.0233]	[0.0219]	[0.118]	[0.0601]	[0.0411]	[0.0468]	[0.0681]
Constant	−2.114***	−0.728***	−0.24	0.420***	0.963***	−2.207***	−0.681***	−0.231	0.528***	0.924***
	[0.594]	[0.279]	[0.183]	[0.135]	[0.129]	[0.476]	[0.246]	[0.170]	[0.191]	[0.262]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zone FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,997	8,997	8,997	8,997	8,997	3,481	3,481	3,481	3,481	3,481

Notes. The standard errors in all columns are clustered at the year and zone level. The dependent variable, SPI, is standardized to a distribution with zero mean and a unit standard deviation. The excluded team gender category is M0 (i.e., all-women teams). This table corresponds to Figures 1(a) and 1(b).

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

5. Understanding the Differences in Performance: The Decision Analysis

We have shown that all-women teams are significantly outperformed by teams with any other gender composition. In this section, we proceed to understand these differences in performance by analyzing the managerial decisions that teams undertake. In particular, we study team decision making on R&D, brand management, and social and environmental responsibility initiatives. In the analysis that follows, we estimate Equation (1) for each of the three decision variables, including all the controls as in columns (3) and (4) in Table 4.

5.1. Investments in R&D

We analyze whether teams with different gender compositions make significantly different decisions regarding the number of formulae to develop and the investment in R&D. The results are shown in Table 6. Columns (1) and (2) in Table 6 show the estimates for the number of formulae. Both at the undergraduate and MBA levels, all gender combinations have significantly more formulae than do M0 teams. Columns (3) and (4) in Table 6 show the estimates for the standardized R&D investment. Because there are significant differences in the developed number of formulae, we look at the R&D investment controlling for the number of formulae. The results in the table show that all teams invest more in R&D than do the three-women teams, even after controlling for the number of formulae.

Given these results, we conclude that the underperformance of three-women teams is, in part, explained by their behavior related to R&D; that is, women teams invest too little in R&D (see Tables 1 and 2). A possible interpretation of these differences is that all-women teams are more risk averse in their management vision. That is, all-women teams seem to

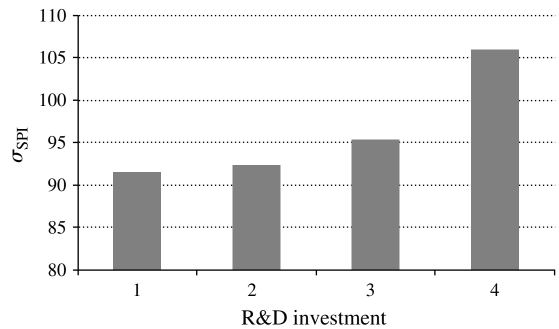
Table 6 Gender Composition of Teams on R&D Decisions

	Number of formulae		Stand. R&D investment	
	UG	MBA students	UG	MBA students
	(1)	(2)	(3)	(4)
M1	0.0350**	0.0669*	0.0738***	0.142**
	[0.0156]	[0.0331]	[0.0236]	[0.0531]
M2	0.0510**	0.0600**	0.118***	0.196***
	[0.0199]	[0.0236]	[0.0263]	[0.0585]
M3	0.0533***	0.0651**	0.0809**	0.199***
	[0.0187]	[0.0275]	[0.0388]	[0.0489]
Number of formulae			0.902***	0.863***
			[0.0344]	[0.0436]
Constant	1.212***	1.139***	−1.702***	−1.949***
	[0.0933]	[0.1000]	[0.159]	[0.234]
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Zone FE	Yes	Yes	Yes	Yes
Cluster	Yes	Yes	Yes	Yes
Observations	7,561	2,910	7,561	2,910

Notes. The standard errors in all columns are clustered at the year and zone level. The dependent variable, *R&D investment*, in columns (3) and (4), is standardized to a distribution with zero mean and a unit standard deviation. The excluded team gender category is M0 (i.e., all-women teams).

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

Figure 2 Variation in SPI Conditional on R&D Investment



Notes. The first bin of the bar chart corresponds to R&D investment less than 1,000,000; the second bin corresponds to R&D investment between 1,000,000 and 2,500,000; the third bin corresponds to R&D investment between 2,500,000 and 5,000,000; and the fourth bin corresponds to R&D investment more than 5,000,000.

overweight the cost associated to R&D decisions with respect to the potential improvement in the ultimate value of the firm. Indeed, R&D decisions can be interpreted as risky decisions. Looking at the variance of performance conditional in R&D investment, we see that as investment in R&D increases, the variance produced in performance increases significantly, suggesting riskier choice (see Figure 2).[13]

The literature on gender differences at the individual level has consistently documented differences in risk preferences, that women tend to be more risk-averse than men (see Eckel and Grossman 2003, Croson and Gneezy 2009). In this sense, all-women teams being more risk averse is in line with the findings at the individual level. Interestingly, we also see that mixed teams are not significantly different from all-men teams, such that having one man or three men does not show significant differences in the team's R&D investment.

5.2. Brand Management

We now analyze the impact of teams' gender composition on the midway outcome variables that are directly determined by brand management. We start with an analysis at the aggregate level, looking at (standardized) variables such as profits, revenues, costs, sales, and inventories. We then break down the aggregate analysis to study each of the brands separately.

The main outcome variable related to brands is profits. Accordingly, we first analyze whether the level of total profits earned by teams varies across the different gender compositions. In columns (1) and (2) of Table 7, we report the profits at the aggregate level separately for undergraduates and MBA students. Both at the undergraduate and MBA levels, we find that every gender composition achieves significantly higher profits than do M0 teams. When we separate the profits into revenues and production costs, we see that the difference is largely related to differences in revenues but not in production costs. The M1, M2, and M3 teams attain significantly higher revenues than do M0 teams, but there are no significant differences in production costs. Consistent with these results, Table 7 also shows that all teams produce more than the M0 teams, but they also sell more, resulting in lower inventory costs. These differences highlight that the underperformance of M0 teams is also related to their brand management. We see that M0 teams are choosing worse selling strategies than are the rest of teams.

To understand the differences in their selling strategy, we now turn our attention to the analysis of brand management at the brand-type level. Consumers are divided into five different segments, which differ in size, price sensitivity, and preferences. Teams are provided with this information in their instruction manuals. The five segments, ordered by their income (highest to lowest) and price sensitivity (lowest to highest), are (i) high earners, (ii) affluent families, (iii) medium-income families, (iv) singles, and (v) low-income families. Accordingly, brands differ in terms of the type of consumers to which they are targeted. In the three editions of the game in our database, there were four different brand types: (a) high income (edition 2007), (b) medium-income families (2009), (c) singles (2008 and 2009), and (d) low income (2007 and 2008).

Columns (3)–(10) in Table 7 report the analysis at the brand-type level. We see that for undergraduates, the differences identified at the aggregate level, in terms of profits, revenues, sales, and inventory costs, are concurrent only for brand types (b) and (c); the intermediate brand types. When we analyze the other brands, the high-income and low-income brand types, there is almost no difference across teams. We next consider the differences in teams' pricing strategies. We see that with brand types (b) and (c) that M0 teams choose significantly higher prices than do all other gender composition teams. This pricing strategy results in significantly lower revenues (and profits) for the M0 teams; in turn, this also explains why M0 teams have significantly higher inventory costs. We can interpret such a pricing strategy by M0 teams as being less aggressive than the rest of the teams'. Again, the literature on gender differences at the individual level offers a plausible explanation for these findings. In particular, it has been shown that women and men have different attitudes toward competition. Not only do women seem to dislike competition

[13] Looking at the mean investment by team gender composition, the mean investment by M0 teams is in the second bin of Figure 2, whereas all other teams' investments are in the third bin.

Table 7 Gender Composition of Teams on Brand Management

	Aggregate		Brand a: High-income		Brand b: Medium-income		Brand c: Singles		Brand d: Low-income	
	UG	MBA	UG	MBA	UG	MBA	UG	MBA	UG	MBA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Stand. profits										
M1	0.0972**	0.0674	−0.0105	−0.011	0.153**	0.144	0.139***	0.0502	0.05	0.0562
	[0.0393]	[0.0637]	[0.0608]	[0.120]	[0.0575]	[0.0823]	[0.0458]	[0.0458]	[0.0359]	[0.0723]
M2	0.147***	0.190**	0.00579	0.244*	0.256***	0.138	0.184***	0.0934	0.0871***	0.116
	[0.0469]	[0.0683]	[0.0897]	[0.117]	[0.0559]	[0.151]	[0.0393]	[0.0715]	[0.0279]	[0.0782]
M3	0.142***	0.111**	0.107*	0.134	0.209**	0.152	0.138**	0.0339	0.0632	0.081
	[0.0328]	[0.0529]	[0.0469]	[0.0918]	[0.0677]	[0.113]	[0.0556]	[0.0660]	[0.0445]	[0.0569]
Stand. revenues										
M1	0.0944**	0.0665	−0.0498	0.00337	0.0898	0.0931	0.0522	0.0423	−0.0036	0.0387
	[0.0384]	[0.0632]	[0.0574]	[0.0875]	[0.0755]	[0.0635]	[0.0405]	[0.0313]	[0.0305]	[0.0453]
M2	0.147***	0.185**	−0.0283	0.237*	0.185*	0.0751	0.0854*	0.0739	0.0164	0.107
	[0.0486]	[0.0672]	[0.0859]	[0.110]	[0.0868]	[0.122]	[0.0462]	[0.0692]	[0.0305]	[0.0719]
M3	0.146***	0.109**	0.0425	0.109	0.125	0.123	0.0716	0.0215	0.0245	0.0573
	[0.0349]	[0.0513]	[0.0449]	[0.0706]	[0.107]	[0.0868]	[0.0549]	[0.0316]	[0.0385]	[0.0391]
Stand. production costs										
M1	0.0254	0.026	0.0111	−0.0025	0.0285	0.173	0.0603*	−0.0182	−0.0162	0.0248
	[0.0394]	[0.0745]	[0.0554]	[0.131]	[0.0962]	[0.136]	[0.0337]	[0.0924]	[0.0303]	[0.109]
M2	0.0622	0.0565	0.0384	0.174	0.132	0.0883	0.100*	−0.0986	−0.0291	0.0748
	[0.0520]	[0.0663]	[0.0686]	[0.195]	[0.0873]	[0.0744]	[0.0554]	[0.0956]	[0.0409]	[0.117]
M3	0.0825*	0.0317	0.136**	0.044	−0.0138	0.1	0.118*	−0.0744	−0.00653	0.0695
	[0.0451]	[0.0630]	[0.0587]	[0.147]	[0.0638]	[0.112]	[0.0603]	[0.0967]	[0.0469]	[0.0964]
Stand. sales										
M1	0.111***	0.121*	0.00223	0.00155	−0.0247***	−0.0244**	−0.0268***	−0.0183**	0.0448	0.0968
	[0.0348]	[0.0600]	[0.00971]	[0.0158]	[0.00634]	[0.00842]	[0.00637]	[0.00710]	[0.0306]	[0.0781]
M2	0.154***	0.199***	0.00356	−0.0248	−0.0357***	−0.0182	−0.0341***	−0.0209*	0.0719**	0.158*
	[0.0411]	[0.0655]	[0.0112]	[0.0142]	[0.00586]	[0.0177]	[0.00594]	[0.0119]	[0.0301]	[0.0812]
M3	0.159***	0.135**	−0.0106*	−0.00632	−0.0286**	−0.0247*	−0.0318***	−0.0138	0.0733*	0.120*
	[0.0337]	[0.0532]	[0.00549]	[0.0119]	[0.00998]	[0.0129]	[0.00916]	[0.0108]	[0.0404]	[0.0627]
Stand. prices										
M1	—	—	−0.0347	0.0349	−0.213***	−0.211***	−0.116***	−0.0374	−0.0187	−0.0302
	—	—	[0.0458]	[0.147]	[0.0631]	[0.0517]	[0.0379]	[0.0760]	[0.0377]	[0.0524]
M2	—	—	0.0105	−0.00254	−0.162***	−0.141	−0.122***	−0.0704	−0.0143	0.012
	—	—	[0.0258]	[0.132]	[0.0438]	[0.0860]	[0.0420]	[0.0727]	[0.0336]	[0.0682]
M3	—	—	−0.0248	0.102	−0.189**	−0.183*	−0.141***	−0.0846	−0.0374	−0.00671
	—	—	[0.0244]	[0.105]	[0.0590]	[0.0823]	[0.0369]	[0.0554]	[0.0403]	[0.0507]
Stand. production										
M1	0.0660*	0.0586	−0.0213	−0.0217	0.112*	0.114*	0.102**	0.139**	0.023	0.00197
	[0.0347]	[0.0544]	[0.0868]	[0.0864]	[0.0566]	[0.0589]	[0.0354]	[0.0529]	[0.0356]	[0.0867]
M2	0.119***	0.134**	−0.0236	0.182	0.208***	0.0997	0.165***	0.124*	0.0495	0.0804
	[0.0389]	[0.0538]	[0.0826]	[0.108]	[0.0505]	[0.0938]	[0.0428]	[0.0669]	[0.0391]	[0.0736]
M3	0.107**	0.0956**	0.0448	0.0705	0.138	0.111*	0.141**	0.0773	0.0459	0.073
	[0.0417]	[0.0450]	[0.0403]	[0.0757]	[0.0787]	[0.0579]	[0.0511]	[0.0612]	[0.0512]	[0.0582]
Stand. inventory costs										
M1	−0.101**	−0.127	−0.0017	−0.00997	−0.216***	−0.227	−0.149***	0.0196	−0.037	−0.148
	[0.0439]	[0.0842]	[0.0758]	[0.164]	[0.0573]	[0.163]	[0.0450]	[0.0869]	[0.0450]	[0.0928]
M2	−0.102**	−0.127	0.0145	−0.129	−0.224***	−0.141	−0.150***	−0.0182	−0.0399	−0.127
	[0.0392]	[0.0787]	[0.0549]	[0.127]	[0.0357]	[0.158]	[0.0353]	[0.0770]	[0.0401]	[0.0926]
M3	−0.121***	−0.091	−0.0858*	0.00943	−0.233***	−0.238	−0.133**	−0.0346	−0.0473	−0.0692
	[0.0414]	[0.0679]	[0.0455]	[0.105]	[0.0434]	[0.141]	[0.0599]	[0.0658]	[0.0511]	[0.0835]
Observations	8,997	3,481	3,367	1,386	2,458	882	5,630	2,095	6,539	2,599

Notes. All columns control for observable characteristics as in columns (3) and (4) in Table 4, as well as controls for year and zone fixed effects. The standard errors in all columns are clustered at the year and zone level. The dependent variables in all columns are standardized to a distribution with zero mean and a unit standard deviation. The excluded team gender category is M0 (i.e., all-women teams).

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

Table 8 Gender Composition of Teams on Corporate Responsibility

	Stand. SSI		Stand. ESI	
	UG (1)	MBA (2)	UG (3)	MBA (4)
M1	−0.0708**	−0.0900*	0.0202	−0.0181
	[0.0329]	[0.0450]	[0.0259]	[0.0953]
M2	−0.126***	−0.0701	0.0261	0.0135
	[0.0345]	[0.0533]	[0.0315]	[0.0680]
M3	−0.0888**	−0.160***	−0.0216	−0.0306
	[0.0381]	[0.0487]	[0.0335]	[0.0736]
Constant	0.33	0.696***	−0.0649	−0.207
	[0.204]	[0.218]	[0.211]	[0.271]
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Zone FE	Yes	Yes	Yes	Yes
Cluster	Yes	Yes	Yes	Yes
Observations	8,855	3,405	8,855	3,405

Notes. The standard errors in all columns are clustered at the year and zone level. The dependent variables in all columns are standardized to a distribution with zero mean and a unit standard deviation. The excluded team gender category is M0 (i.e., all-women teams).

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

more than men do, but under competition the performance of men is improved relative to that of women (see Gneezy et al. 2003, Gneezy and Rustichini 2004, Niederle and Vesterlund 2007).[14] Arguably, setting the price is a form of competition, and hence that M0 teams are not competitive enough can be related to the findings above.

The analysis at the MBA level, when disaggregated by brand types, does not show consistent and clear significant differences. This is likely to be the result of a reduction in the number of observations, and hence the significance levels are lower. However, the magnitudes and signs are comparable to those found at the aggregate level.

5.3. Corporate Responsibility

We now study whether the gender composition of the team has any effect on the social and environmental responsibility decisions, as measured by the standardized SSI and ESI indices. Table 8 reports the results. With regard to social initiatives, we find that three-women teams invest significantly more in social initiatives than does any other gender composition, both at the undergraduate and MBA level. These differences are as large as 0.12 and 0.16 of a standard deviation from the mean in undergraduate and MBA competitions, respectively. All comparisons are significant, except in the M2 case for MBA students where the coefficient goes in the same direction than all others, but it is not significant at conventional levels. In columns (3) and (4), on the other hand, the gender composition of the teams does not appear to influence decisions related to environmental initiatives.

Hence, gender composition seems to matter for the type of decisions taken regarding social initiatives. In §2.3 we observed that SSI is positively related to SPI. However, we also observed that the influence on SPI of the social sustainability initiatives is of an order of magnitude lower than is profits. This shows that although M0 teams invest significantly more in SSI, it has little impact on the final and main outcome variable, on SPI. From Table 1, the polynomial regressions show that there are diminishing returns to spending on SSI, which suggests that M0 may be overinvesting in SSI.

The literature on gender differences at the individual level provides mixed evidence on differences with regard to altruism, values, and social preferences (see Croson and Gneezy 2009). Although there are papers reporting that women are significantly more altruistic than men are (see Eckel and Grossman 1998, Bolton and Katok 1995, Andreoni and Vesterlund 2001) and that women have more social oriented values than men have (see Adams and Funk 2011), other papers also show that the behavior of women in social contexts is highly sensitive to the details of the context (see Ben-Ner et al. 2004, Houser and Schunk 2009). Our findings related to corporate responsibility decisions of teams seem to be mixed, in line with the literature. In particular, whereas in the case of social initiatives all-women teams invest significantly more than do all other teams, in the case of environmental initiatives there are no significant differences.

6. Discussion

We have shown that three-women teams are outperformed by teams of any other gender combination and that there is not a monotonic relationship between the number of men and team performance. Because teams are formed endogenously, there are competing explanations for our findings. It is important to discuss these explanations because the policy, social, and scientific conclusions one could draw from our results are clearly very different depending on what drives them.

We distinguish between two broad competing explanations. First, participating women and men may be different in terms of their unobservable characteristics, such as ability, expectations about the reward structure implicit in the game, or the way in which they sort into teams. Second, three-women teams may have worse team dynamics than do teams of any other gender composition, such that, despite

[14] When team performance is defined as the sum of two individuals' independent performance, it has been shown that gender differences under competition are reduced (see Dargnies 2010, Healy and Pate 2011).

Table 9 Demographics by Individuals Conditioning on the Gender Composition of Teams

	Women in teams composed of			Men in teams composed of		
	3 women	2 women–1 man	1 woman–2 men	3 men	2 men–1 woman	1 man–2 women
			Panel A: Undergraduates			
Mean age	21.968	22.181	22.264	22.578	22.672	22.591
	(1.846)	(2.256)	(2.100)	(2.318)	(2.313)	(2.377)
Business	0.528	0.534	0.526	0.469	0.496	0.485
	(0.499)	(0.499)	(0.499)	(0.499)	(0.500)	(0.500)
Economics	0.299	0.268	0.249	0.285	0.259	0.291
	(0.458)	(0.443)	(0.432)	(0.452)	(0.438)	(0.455)
Sciences	0.093	0.117	0.152	0.207	0.202	0.171
	(0.291)	(0.322)	(0.359)	(0.405)	(0.401)	(0.377)
Other fields	0.080	0.080	0.074	0.038	0.043	0.052
	(0.271)	(0.272)	(0.261)	(0.192)	(0.204)	(0.223)
Central Europe	0.073	0.072	0.092	0.124	0.092	0.072
	(0.261)	(0.259)	(0.289)	(0.330)	(0.289)	(0.258)
South Europe	0.053	0.060	0.077	0.102	0.077	0.061
	(0.223)	(0.238)	(0.267)	(0.302)	(0.266)	(0.240)
East Europe	0.111	0.105	0.104	0.122	0.104	0.105
	(0.314)	(0.307)	(0.305)	(0.328)	(0.305)	(0.307)
Africa	0.065	0.068	0.094	0.122	0.094	0.068
	(0.246)	(0.251)	(0.291)	(0.327)	(0.291)	(0.251)
South America	0.039	0.052	0.063	0.088	0.063	0.052
	(0.194)	(0.222)	(0.243)	(0.283)	(0.242)	(0.223)
North America	0.038	0.038	0.044	0.033	0.044	0.038
	(0.191)	(0.190)	(0.205)	(0.180)	(0.205)	(0.190)
East Asia	0.453	0.447	0.355	0.202	0.355	0.447
	(0.498)	(0.497)	(0.479)	(0.402)	(0.479)	(0.497)
South Asia	0.165	0.154	0.169	0.204	0.169	0.154
	(0.372)	(0.361)	(0.374)	(0.403)	(0.375)	(0.361)
Area others	0.003	0.003	0.003	0.002	0.003	0.003
	(0.054)	(0.057)	(0.054)	(0.047)	(0.054)	(0.057)
Foreign at institution	0.093	0.107	0.120	0.111	0.116	0.104
	(0.291)	(0.309)	(0.325)	(0.315)	(0.320)	(0.305)
Number individuals ed. 2007	1,869	1,826	1,032	2,678	2,064	913
Number individuals ed. 2008	1,800	1,734	923	2,360	1,846	866
Number individuals ed. 2009	1,453	1,384	738	1,629	1,478	691
Total number individuals	5,122	4,944	2,693	6,667	5,388	2,470
			Panel B: MBA students			
Mean age	25.634	26.261	26.606	26.957	27.416	26.964
	(3.226)	(3.537)	(3.813)	(4.152)	(4.181)	(3.833)
Business	0.767	0.784	0.788	0.782	0.764	0.755
	(0.423)	(0.411)	(0.409)	(0.413)	(0.425)	(0.430)
Economics	0.162	0.139	0.129	0.155	0.163	0.163
	(0.369)	(0.346)	(0.335)	(0.362)	(0.370)	(0.370)
Sciences	0.031	0.042	0.042	0.050	0.051	0.054
	(0.174)	(0.201)	(0.201)	(0.217)	(0.221)	(0.227)
Other fields	0.040	0.035	0.041	0.013	0.021	0.027
	(0.196)	(0.183)	(0.199)	(0.114)	(0.145)	(0.163)
Central Europe	0.127	0.140	0.135	0.144	0.136	0.140
	(0.333)	(0.347)	(0.341)	(0.351)	(0.343)	(0.347)
South Europe	0.091	0.088	0.094	0.091	0.094	0.088
	(0.288)	(0.283)	(0.292)	(0.288)	(0.292)	(0.283)
Eastern Europe	0.074	0.059	0.049	0.041	0.049	0.058
	(0.262)	(0.235)	(0.216)	(0.197)	(0.216)	(0.234)
Africa	0.063	0.051	0.066	0.060	0.067	0.049
	(0.242)	(0.219)	(0.248)	(0.237)	(0.250)	(0.217)
South America	0.042	0.051	0.055	0.068	0.054	0.052
	(0.200)	(0.219)	(0.228)	(0.252)	(0.225)	(0.222)

Table 9 (Continued)

	Women in teams composed of			Men in teams composed of		
	3 women	2 women–1 man	1 woman–2 men	3 men	2 men–1 woman	1 man–2 women
			Panel B: MBA students			
North America	0.258	0.164	0.129	0.100	0.128	0.164
	(0.438)	(0.371)	(0.335)	(0.300)	(0.334)	(0.371)
East Asia	0.190	0.241	0.215	0.068	0.216	0.242
	(0.392)	(0.428)	(0.411)	(0.252)	(0.412)	(0.429)
South Asia	0.152	0.201	0.254	0.425	0.252	0.201
	(0.359)	(0.401)	(0.435)	(0.494)	(0.434)	(0.401)
Area others	0.003	0.006	0.004	0.003	0.004	0.005
	(0.051)	(0.078)	(0.065)	(0.058)	(0.065)	(0.070)
Foreign at institution	0.246	0.269	0.251	0.242	0.255	0.257
	(0.431)	(0.444)	(0.434)	(0.428)	(0.436)	(0.437)
Number individuals ed. 2007	501	670	501	1,332	1,002	335
Number individuals ed. 2008	357	554	377	1,325	754	277
Number individuals ed. 2009	291	394	289	897	578	197
Total number individuals	1,149	1,618	1,167	3,554	2,334	809

Notes. Business, *Economics*, *Sciences*, and *Other fields* are dummy variables representing categories for the fields of study of the individual. Each category takes a value of 1 if the field of study of the individual belongs to that category and 0 otherwise. Table A.1 in the online appendix reports the classification of fields of study in the four categories. Geographical areas are dummy variables taking a value of 1 if the country of origin of the individual belongs to the respective geographic area, and 0 otherwise. Individuals not reporting a geographic area are collected in *Area others*. Table A.2 in the online appendix reports the assignment of countries to the respective geographic areas. *Foreign at institution* is a dummy variable giving a value of 1 when the country of origin of the individual does not coincide with the country where the educational institution is located and a value of 0 otherwise. Standard errors are reported in parentheses.

women and men being comparable in all the relevant characteristics (observable and unobservable), three women perform worse within a team than do mixed teams or all-men teams. This could be due to different reasons. For example, different teams may use different aggregation rules of individual preferences, or they may adopt different mechanisms of internal organization.

We start by considering the differences between women and men regarding unobservable characteristics. First, it could be the case that the distributions of ability between those men and those women that decide to participate in the game are different, and this translates into different skills in teams by gender composition. Because we do not directly observe individual ability, we cannot rule out this possibility. However, we can use observable characteristics, such as age, which may reflect experience; field of study; and quality of the university attended, as proxies for individual ability. Note that when controlling for all of these factors, including the university fixed effects (see Table 4 and the discussion in §4), we continue to find that all-women teams are outperformed by any other gender compositions, both at the undergraduate and MBA competitions. Controlling for these variables is somehow suggestive that the differences in ability between women and men are not the main driving force. Of course, arguably, our observable characteristics are not extensive enough to rule out this explanation entirely.

Second, all-women teams might have different expectations about the reward structure implicit in the L'Oreal e-Strat Challenge than the other teams do. Although we do not have control over the participants' expectations about the reward structure, it is important to note that all participants are provided with the same instructions, so in principle the information about the rules and reward structure should not be different.

Third, men and women might have different ways of sorting into teams that is reflected in their performance. For example, low ability women may be more likely to sort into all-women teams than are low ability men into all-men teams. Alternatively, it could be that shy women are less likely to form a team with men. In this case, women in all-women teams will tend to be less social. If social skills are correlated with attributes that matter in the game (e.g., more social people are more self-confident and perform better), then the effect of all-women teams could be capturing the impact of social skills. Again, as in the previous case, we can evaluate whether individuals differ on observable characteristics, depending on the team they sort into, to imperfectly test for this hypothesis. From Table 9 we see that, overall, women and men in the different teams look remarkably similar. At the MBA level, there are some small differences; e.g., women who sort into teams with more men are slightly older. However, when we control for these differences in the analysis in §4, our main findings hold. Nevertheless, more work is needed to study gender differences in team formation.

Finally, if men and women are comparable in all the relevant characteristics, then our results can be explained by three-women teams having worse team dynamics than mixed teams or all-male teams. Our findings also suggest that mixed teams show the highest performance, such that, in light of this second competing explanation, diversity in terms of gender would be positively related to good team dynamics.

We hope that this paper will promote future empirical research to understand the influence of gender on teams' performance. In particular, understanding the causal effect behind the underperformance of three-women teams needs further research designed to identify the influence of different competing explanations. This will be of great relevance when understanding the functioning of organizations.

Acknowledgments
The authors are grateful to L'Oréal and StratX for their collaboration and assistance in this study. They thank two anonymous referees, Manuel Arellano, Miguel A. Ballester, Manel Baucells, Rachel Croson, Vicente Cuñat, David Dorn, Gabrielle Fack, Robin Hogarth, Pedro Rey-Biel, and Kurt Schmidheiny for helpful comments. Ozan Eksi and Jacopo Ponticelli provided excellent research assistance. Financial support by the Spanish Commission of Science and Technology [ECO2008-06395-C05-01, ECO2008-01768, ECO2009-12836, ECO2010-09555-E, ECO2009-11213, SEJ2007-64340], Fundación Rafael del Pino, the Barcelona Graduate School of Economics research network, and the Government of Catalonia is gratefully acknowledged.

References
Adams, R. B., D. Ferreira. 2009. Women in the boardroom and their impact on governance and performance. *J. Financial Econom.* **94**(3) 291–309.

Adams, R. B., P. Funk. 2011. Beyond the glass ceiling: Does gender matter? Working paper, Universitat Pompeu Fabra, Barcelona, Spain.

Andreoni, J., L. Vesterlund. 2001. Which is the fair sex? *Quart. J. Econom.* **116**(1) 293–312.

Bagues, M. F., B. Esteve-Volart. 2010. Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *Rev. Econom. Stud.* **77**(4) 1301–1328.

Ben-Ner, A., F. Kong, L. Putterman. 2004. Share and share alike? Intelligence, socialization, personality, and gender-pairing as determinants of giving. *J. Econom. Psych.* **25**(5) 581–589.

Bertrand, M., K. Hallock. 2001. The gender gap in top corporate jobs. *Indust. Labor Relations Rev.* **55**(1) 3–21.

Bertrand, M., C. Goldin, L. F. Katz. 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *Amer. Econom. J.: Appl. Econom.* **2**(3) 228–255.

Bolton, G. E., E. Katok. 1995. An experimental test for gender differences in beneficent behavior. *Econom. Lett.* **48**(3–4) 287–292.

Byrnes, J., D. Miller, W. Schafer. 1999. Gender differences in risk taking: A meta-analysis. *Psych. Bull.* **75**(3) 367–383.

Carter, D., B. Simkins, W. Simpson. 2003. Corporate governance, board diversity, and firm value. *Financial Rev.* **38**(1) 33–53.

Charness, G., U. Gneezy. 2007. Strong evidence for gender differences in investment. Working paper, University of California, Santa Barbara, Santa Barbara. http://ssrn.com/abstract=648735.

Charness, G., M. Jackson. 2007. Group play in games and the role of consent in network formation. *J. Econom. Theory* **136**(1) 417–445.

Charness, G., L. Rigotti, A. Rustichini. 2007. Individual behavior and group membership. *Amer. Econom. Rev.* **97**(4) 1340–1352.

Croson, R., N. Buchan. 1999. Gender and culture: International experimental evidence from trust games. *Amer. Econom. Rev.* **89**(2) 386–391.

Croson, R., U. Gneezy. 2009. Gender differences in preferences. *J. Econom. Literature* **47**(2) 1–27.

Dargnies, M. P. 2010. Team competition: Eliminating the gender gap in competitiveness. Working paper, Social Science Research Center Berlin (WZB), Berlin.

Delfgaauw, J., R. Dur, J. Sol, W. Verbeke. 2009. Tournament incentives in the field: Gender differences in the workplace. Working paper, Erasmus University Rotterdam, Rotterdam, The Netherlands.

Dufwenberg, M., A. Muren. 2006. Gender composition in teams. *J. Econom. Behav. Organ.* **61**(1) 50–54.

Eckel, C., P. J. Grossman. 1998. Are women less selfish than men? Evidence from dictator games. *Econom. J.* **108**(448) 726–735.

Eckel, C., P. J. Grossman. 2003. Differences in the economic decisions of men and women: Experimental evidence. C. Plott, V. L. Smith, eds. *Handbook of Experimental Results*. New York, Elsevier, 509–519.

Farrell, K., P. Hersch. 2005. Additions to corporate boards: The effect of gender. *J. Corporate Finance* **11**(1–2) 85–106.

Gneezy, U., A. Rustichini. 2004. Gender and competition at a young age. *Amer. Econom. Rev.* **94**(2) 377–381.

Gneezy, U., M. Niederle, A. Rustichini. 2003. Performance in competitive environments: Gender differences. *Quart. J. Econom.* **118**(3) 1049–1074.

Healy, A., J. Pate. 2011. Can teams help to close the gender competition gap? *Econom. J.* Forthcoming.

Houser, D., D. Schunk. 2009. Social environments with competitive pressure: Gender effects in the decisions of German schoolchildren. *J. Econom. Psych.* **30**(4) 634–641.

Kristof, N. D. 2009. Mistresses of the universe. *New York Times* (February 7), http://www.nytimes.com/2009/02/08/opinion/08kristof.html.

Manning, A., J. Swaffield. 2008. The gender gap in early-career wage growth. *Econom. J.* **118**(530) 983–1024.

Miller, G. 2008. Women's suffrage, political responsiveness, and child survival in American history. *Quart. J. Econom.* **123**(3) 1287–1327.

Niederle, M., L. Vesterlund. 2007. Do women shy away from competition? Do men compete too much? *Quart. J. Econom.* **122**(3) 1067–1101.

Sobel, J. 2006. Information aggregation and group decisions. Working paper, University of California, San Diego, San Diego.

Stoner, J. A. F. 1968. Risky and cautious shifts in group decisions: The influence of widely held values. *J. Experiment. Soc. Psych.* **4**(4) 442–459.

Walker, L. 2008. Crisis gives women a shot at top corporate jobs. *Financial Times* (October 18), http://www.ft.com/cms/s/0/4797515a-9cae-11dd-a42e-000077b07658.html#axzz1NI2SDAWF.

World Bank. 2008. *World Development Indicators*. Online database. World Bank, Washington, DC.

Zinovyeva, N., M. Bagues. 2010. Does gender matter for academic promotion? Evidence from a randomized natural experiment. FEDEA Working Paper 2010-15, Fundación de Estudios de Economía Aplicada, Madrid.