Open access • Journal Article • DOI:10.1086/667996

Tournament Incentives in the Field: Gender Differences in the Workplace

Josse Delfgaauw, Robert Dur, Joeri Sol, Willem Verbeke

Institutions: Erasmus University Rotterdam

Published on: 02 May 2013 - Journal of Labor Economics (University of Chicago PressChicago, IL)

TI 2009-069/1
Tinbergen Institute Discussion Paper

Tournament Incentives in the Field: Gender Differences in the Workplace

Josse Delfgaauw[1,2]
Robert Dur[1,2,3,4]
Joeri Sol[1,2]
Willem Verbeke[1,5]

[1] *Erasmus University Rotterdam;*
[2] *Tinbergen Institute;*
[3] *CESifo;*
[4] *IZA;*
[5] *ERIM.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

Tournament Incentives in the Field: Gender Differences in the Workplace*

Josse Delfgaauw[†], Robert Dur[‡], Joeri Sol[§], and Willem Verbeke[¶]

April 2012

Abstract

We ran a field experiment in a Dutch retail chain consisting of 128 stores. In a random sample of these stores, we introduced short-term sales competitions among subsets of stores. We find that sales competitions have a large effect on sales growth, but only in stores where the store's manager and a sufficiently large fraction of the employees have the same gender. Remarkably, results are alike for sales competitions with and without monetary rewards, suggesting a high symbolic value of winning a tournament.

JEL-codes: C93, J16, M52.

Keywords: field experiment, gender differences, competition, sales contests, awards.

*We gratefully acknowledge comments by two Editors, several anonymous reviewers, Martin Bøg, Alison Booth, Nagore Iriberri, Jan van Ours, Imran Rasul, and seminar audiences at University of Amsterdam, Universitat Autonoma de Barcelona, Bar-Ilan University, Erasmus University Rotterdam, University of Frankfurt, University of Mannheim, Tilburg University, and the University of Zürich, and participants to the 20th Jerusalem School in Economic Theory at the Hebrew University, the conference Experiments in Organizations at Erasmus University Rotterdam, the 2009 EEA Meeting in Barcelona, and the 2010 MOVE Workshop on Gender Differences in Competitiveness and Risk Taking in Barcelona.

[†]Erasmus University Rotterdam and Tinbergen Institute. E-mail: delfgaauw@ese.eur.nl.

[‡]Erasmus University Rotterdam, Tinbergen Institute, CESifo, and IZA. E-mail: dur@ese.eur.nl.

[§]Erasmus University Rotterdam and Tinbergen Institute. E-mail: jsol@ese.eur.nl.

[¶]Erasmus University Rotterdam and ERIM. E-mail: verbeke@ese.eur.nl.

1 Introduction

Throughout the world, in business as well as in government, men are strongly overrepresented in top positions. For instance, in 2008, only 16% of all ministerial positions worldwide were held by women; similarly, among the world's 192 heads of government, there were only eight women (IPU 2008). In business, the situation is not much different. For example, in a large sample of publicly traded US firms, Bertrand and Hallock (2001) find that only 2.5% of the five highest-paid positions are held by women. Wirth (2001) reports similar patterns for other countries.

Traditional explanations for the small number of women in top positions are occupational sorting resulting from gender differences in ability or preferences (Polachek 1981) and gender discrimination (e.g. Snizek and Neil 1992). Inspired by evolutionary biology, recent experimental studies – starting with Gneezy et al. (2003) – suggest a third explanation: men are more strongly motivated by competitive incentives or more effective in competitive environments than women, thus impeding women in competitions for promotions or for new jobs.

By now, there is quite some empirical support for such gender differences. In a lab experiment, Gneezy et al. (2003) let participants solve computerized mazes and varied the competitiveness of the environment. They find that, while men and women perform equally well under individual piece rates, men perform much better than women under competitive incentives. Gneezy and Rustichini (2004) show that these gender differences are already present at a very young age. In a 40 meter dash, nine-year-old boys run much faster in a race than when they run alone. By contrast, while girls run as fast as boys when running alone, competition does not increase their running speed. In non-experimental settings, underperformance of women under competitive pressure is found in student admissions to schools (Jurajda and Münich 2011 and Örs et al. 2008) and in Grand Slam tennis (Paserman 2007). The recent field study by Lavy (2008), however, finds no gender differences in the effect of relative performance pay on high-school teacher's performance in Israel. Croson and

Gneezy (2009) provide a recent overview of the literature.[1]

This paper studies the effects of competition on performance by conducting a field experiment in a naturally occurring work environment. It finds interesting differences in responsiveness to competition related to gender interactions. A unique feature of our analysis is that we study competition among teams of employees, each headed by a professional manager. Using the variation in the gender composition of the teams as well as in the gender of the manager, we examine whether female-dominated and female-led teams respond differently to competitive incentives, which were introduced in a random sample of the geographically dispersed teams. Moreover, we explore possible interaction effects between the gender of the manager and the gender composition of the team. Studying gender differences in competition among manager-led teams is most relevant in the context of the sharp gender differences in holding executive-level positions discussed above. Reaching an executive-level position, be it in business or government, commonly requires winning several promotion or job competitions. These competitions are often decided by candidates' relative performance which (except for employees at the lowest hierarchical level) depends not merely on one's own effort or talent, but also crucially on the performance of the members of the team one leads.

More concretely, we ran a field experiment in a discount retail chain in The Netherlands specializing in shoes, sports apparel, and casual clothing. About half of the 128 stores are led by a female manager, while across stores the percentage of female employees ranges from 50% to 100%. In a randomly selected subset of stores, we introduced short-term sales competitions among stores. The selected stores were divided into pools of 5 and competed for a period of 6 weeks on the basis of percentage sales growth compared to the same period the year before. All employees of the store with the highest sales growth over 6 weeks received

[1] A closely related strand in the experimental literature studies self-selection into competitive environments. Datta Gupta et al. (2011), Dohmen and Falk (2011), Flory et al. (2010), and Niederle and Vesterlund (2007) find that men opt significantly more often for competitive compensation schemes than women. Gneezy et al. (2009) show that the reverse holds in a matrilineal society. Recent studies have shown that the gender gap in self-selection into competition by and large vanishes for girls attending single-sex schools (Booth and Nolen 2012) and when the tournament is among teams rather than among individuals (Dargnies 2011).

a bonus of 75 euro; employees of each pool's runner-up received 35 euro. The stores that took part in a competition received weekly feedback in the form of a poster that ranked each store in their pool on their cumulative sales growth figures.

We find that, on average, the tournaments increase percentage sales growth by about five percentage points. We find no significant difference in the effect of tournaments on sales growth between stores with a male manager and stores with a female manager, nor do we find that sales competitions have a larger effect on performance in stores with a higher fraction of male employees. However, this masks a remarkable interaction effect of these two gender variables on sales growth responsiveness: in stores with a male manager, the effect of competition increases in the share of male employees, while the reverse holds for female-led stores. These effects are substantial.[2]

We can think of three plausible mechanisms behind this result. First, the response of team members to competition may crucially depend on the way a competition is communicated and promoted by the team's manager. Both male and female managers may have succeeded in making the competition appeal to employees of their own sex, but less so to employees of the opposite sex. Alternatively, the team nature of the incentive scheme may drive the difference in response. A male (female) manager may be better in strengthening the team's internal cohesion or curtailing free-rider problems if many team members are male (female). Lastly, as managers and employees were not randomised over stores, teams' gender composition may be the result of endogeneous matching on unobservables, which may correlate with teams' responsiveness to competitive incentives. We elaborate on these interpretations after presenting the results in Section 4.

A number of recent studies argue that competition can motivate people not merely because of the chance of winning a monetary reward, but also because of non-pecuniary benefits

[2] Gneezy et al. (2003), Gneezy and Rustichini (2004), and Ivanova-Stenzel and Kuebler (2011) also study whether opponent's gender matters for performance under competition. We do not look into this issue, as teams have limited information, if any, on the gender composition of the stores they compete with. Casas-Arce and Martinez-Jerez (2009) analyze sales competitions among retailers organized by a commodities manufacturer. They do not study gender differences, but instead focus on the effect of the number of contestants in the tournament and on dynamic incentives.

such as perceived esteem, status, and social recognition (Auriol and Renault 2008, Besley and Ghatak 2008, Frey and Neckermann 2008, Moldovanu et al. 2007). Kosfeld and Neckermann (2011) show experimentally that a tournament with no more at stake than an award of zero material value can have a great impact on people's performance. Likewise, Blanes i Vidal and Nossol (2011) and Azmat and Iriberri (2010) find that simply providing information to subjects about their relative performance boosts performance substantially. Bandiera et al. (2009), however, find the opposite effect. In our experiment, parallel to the treatment described above, another subset of stores competed in tournaments with the same setup except for the absence of a monetary reward for winning. So, stores also competed in pools of five, for a period of six weeks, and received a weekly ranking of stores in their pool based on sales growth, but neither the manager nor the employees could earn a bonus. We find that tournaments without monetary rewards have a significantly positive effect on sales growth. The effect is of similar magnitude as the effect of tournaments with monetary rewards, suggesting a high symbolic value of winning a tournament. Gender differences in the effects of competition are also similar in both treatments.

We proceed as follows. In the next section, we describe the experimental set-up and the data. Section 3 describes the methodology of our empirical analysis and Section 4 reports the results. Section 5 concludes.

2 Experimental set-up and data description

The field experiment took place in 2007-2008 in a discount retail chain in The Netherlands, selling male and female clothing, shoes, and sports apparel. The chain consists of 128 geographically dispersed stores operating under one brand name and employing a total of 1574 people. Store employees earn a flat hourly wage slightly above the legal minimum hourly wage. Store managers earn about 45% more and part of their pay is performance-related. On average, slightly less than 5% of a manager's earnings is performance-related.

The company's management wished to intensify the use of incentives. In consultation with the management, we designed sales competitions among subsets of stores. We used stores' percentage growth in sales as compared to sales in the same period a year earlier as the performance measure. Percentage sales growth is a commonly used performance measure in this company and is one of the key determinants of store managers' performance pay. We decided to introduce relative performance incentives rather than incentives based on absolute targets, as sales are very volatile (see Figure 1). A large part of this volatility is caused by common shocks (weather, holidays, advertising campaigns on national television, etc.),[3] which renders relative performance pay attractive (Lazear and Rosen 1981, Green and Stokey 1983, and Nalebuff and Stiglitz 1983).

In the sales competitions, stores competed in pools of five during a period of six weeks. Stores received weekly feedback in the form of a poster containing cumulative sales growth figures for all five stores in their pool, ranked in descending order. Store managers were instructed to put up these posters in the store's canteen, where employees drink coffee and have lunch. The posters as well as the instructions were sent to the store managers through the company's usual channels; store managers and store employees did not know they took part in an experiment. Hence, our experiment can be classified as a natural field experiment (Harrison and List 2004). Store employees were not informed about the sales competitions by the company's management; it was up to the store managers to promote the competition.

Our study comprises two experimental treatments and an untreated control. First, in the 'bonus' treatment, stores compete for a monetary reward. The store manager and all employees of the winning store received a reward of 75 euro; the manager and employees of the runner-up received 35 euro.[4] Second, in the 'feedback' treatment, no monetary reward could be won. Apart from the presence or absence of a monetary reward, the bonus treatment and the feedback treatment were identical.

Our dataset covers a period of 84 weeks (starting in week 1 of 2007). Sales competitions

[3] A panel regression including only week fixed effects explains about 65% of the variation in stores' sales growth. The spike in week 71 in Figure 1 is a common shock, most likely resulting from weather conditions; details are available on request.

[4] The first prize was about 5 percent of an employee's monthly wage. Rewards were halved for part-time employees.

took place in two experimental periods of 6 weeks (in weeks 44 - 49 and weeks 71 - 76). Figure 2 gives an overview of all the events related to the experiment. In the first experimental round, all stores were assigned to one of the two experimental treatments, either bonus or feedback. In the second round, we included a control group of stores not taking part in a competition. We decided against a control group in the first round, because at that time we intended to focus our study on the effects of monetary rewards in tournaments. The second round gives us the opportunity to also assess the effects of tournaments per se.

Competition provides stronger incentives when contestants are more homogeneous (Lazear and Rosen 1981). Bearing this in mind, we used data on past sales performance to create relatively homogeneous pools of stores. The assignment procedure for the first round of the experiment was as follows. All stores were ranked according to percentage sales growth over the weeks 1 up to 37 compared to the same period the year before. The five stores with the highest sales growth were grouped into one pool and assigned to the bonus treatment; the next five were grouped into the next pool and assigned to the feedback treatment. This process was iterated consecutively until all 125 stores were grouped into 13 bonus treatment pools and 12 feedback treatment pools.[5]

The assignment procedure for the second experimental round was partly imposed by the company. For fairness reasons, the company obliged us to assign all stores who were in the feedback treatment during the first period to the bonus treatment in the second period.[6] We grouped these stores into new pools of five stores each, this time using sales performance in weeks 50 to 68 to create relatively homogeneous pools. The remaining stores were assigned either to the feedback condition or to the untreated control group according to a similar procedure as before: the five best-performing stores in weeks 50 to 68 were assigned to the control condition, the next five stores were grouped into a pool and became part of the feedback condition, and so on. To avoid confusion and diminish sabotage opportunities, we

[5] During the first round of the experiment, 3 stores were closed for renovation.

[6] The company wished, at a later point in time, to evaluate the experiment together with the store managers and feared that it would be considered unfair when some stores had never been assigned to the bonus condition.

replaced a store when two stores from the same city happened to be assigned to the same pool. In both periods, we made two of these adjustments.

The company provided us with the weekly sales data of each store, presented in indexes for confidentiality reasons. We used these to calculate the percentage growth in sales as compared to sales in the same week a year earlier. We henceforth refer to this measure as weekly sales growth. We also received each store's personnel file before both experimental rounds, with information on gender, age, and tenure of the store's manager and employees. Descriptive statics are given in Table 1. Across all stores, average weekly sales growth was negative in the period we consider. The retail chain had slightly less female-led stores than male-led stores.[7] The average store had 12 employees (excluding the store manager), of which 85% was female. The average age and tenure of store managers was 39 years and 12 years, respectively. Some of the stores underwent a renovation, which made their appearance more modern without changing the range of products sold. Before the first experiment, six stores had been renovated; at the start of the second experimental period, an additional 9 stores had been renovated. As stores are closed during renovation, there are 268 missing store-week observations. In the analysis we control for the effects of renovation on subsequent sales growth by including a dummy variable which takes value 1 from the week in which a renovated store is reopened onwards.

Table 1 also reports the descriptive statistics within the two treatment groups and the control group to which stores were assigned in the second experimental period (see Figure 2). The three groups of stores hardly differ on observables. A randomization check using F-tests reveals that there are no statistically significant differences in the means of the observables between the three groups. Table 2 reports the descriptive statistics separated by store managers' gender, where we only include the 114 stores we use in analyzing the

[7] In both personnel files, for some stores information about the manager is missing, either because the store temporarily had no manager, or (in a single case) a store had two managers. In one store, a male manager was replaced by a female manager in between the two experimental periods. When analysing gender effects, we exclude these stores from the analysis. This leaves 114 stores. In five other stores, the manager was replaced by a manager with the same gender; excluding these stores from the analysis does not affect the results qualitatively.

gender effects (see footnote 7). Over the whole period, male-managed stores reached 0.4 percentage points higher weekly sales growth than female-managed stores, but the difference is not statistically significant. Male managers have significantly longer tenure than female managers, and run stores with significantly more employees. In the analysis below, we perform robustness checks where we control for these differences. Importantly, there is quite a lot of variation in the percentage of female employees, both for male- and female-led stores. Figure 3 depicts the distribution of the percentage of female employees in stores, separated by managers' gender. Gender of the manager is not significantly related to the gender composition of store employees. Note that there are no stores with a majority of male employees. This implies that our estimates of the effect of stores' gender composition are based on, and, hence, relevant for female-dominated teams.

3 Method

We estimate the effects of the competitions on sales growth using OLS panel estimation including week and store fixed effects. Let y_{st} be the sales growth of store s in week t. Further, let B^1_{st} be a dummy variable which is equal to one during the *first* experimental round when store s was assigned to the bonus treatment (rather than to the feedback treatment). Similarly, let B^2_{st} and F^2_{st} be dummy variables for whether in the *second* experimental round, store s was assigned to the bonus treatment and to the feedback treatment respectively (rather than to the control group). To assess the average effect of the treatments in both experimental periods, we estimate:

$$y_{st} = \alpha_s + \theta_t + \gamma B^1_{st} + \delta B^2_{st} + \mu F^2_{st} + \upsilon R_{st} + \varepsilon_{st} \qquad (1)$$

where α_s and θ_t are store fixed effects and week fixed effects, respectively, R_{st} is a dummy for whether store s had been renovated before week t, and ε_{st} is an error term. We cluster standard errors at the store level to correct for serial correlation within stores as well as for heteroscedasticity across stores (see Bertrand et al. 2004 for a discussion of the importance of correcting for serial correlation in differences-in-differences estimation).

Observe that we allow the effect of the bonus treatment relative to the feedback treatment to differ between the first and second experimental round, i.e., we do not restrict that $\gamma = \delta - \mu$. Loosely speaking, for each experimental period, we estimate differences-in-differences effects of the treatments, where we assume that in all non-experimental weeks, it is 'business-as-usual' for all stores. Hence, we do not allow for carry-over effects of treatments into the weeks following an experimental period. We have checked the robustness of this approach in two ways. First, none of our results is affected qualitatively if we exclude the first 8 weeks after either experimental period (weeks 50 - 57 and/or weeks 77 - 84) from our analysis. Second, all our results from the first experimental period carry over to an estimation which includes only the first 49 weeks (i.e., which excludes all weeks after the first period; see Figure 2). Similarly, we find qualitatively similar results for the treatment effects in the second experimental period if we include only the period after the first experimental period (week 50 onwards).[8]

Besides the average treatment effects, we investigate how these treatment effects depend on the gender of the store manager, the gender composition of the store's employees, the interaction between these two, and the store's team size. To study these issues, we add interaction effects to equation (1). In our preferred estimation of the interaction effects, we pool the bonus and feedback treatments; i.e., we investigate how the gender composition of the store affects the response to competition irrespective of whether a monetary prize could be won. In the appendix, we estimate these interaction effects separately for the bonus treatment and the feedback treatment, and find no significant differences between treatments.

Let E_t^2 be a dummy that takes value one for all observations in the second experimental period (weeks 71 - 76). The effect of variable X_s^2 on the effect of our treatments is estimated

[8] We cannot, however, identify any possible carry-over effects of first-period assignment to second-period treatment effects, as assignment in the second period was not completely random (see Section 2). Note, however, the long time lag between the first and the second round (20 weeks).

by:

$$y_{st} = \alpha_s + \theta_t + \delta(B^2_{st} + F^2_{st}) + \lambda X^2_s(B^2_{st} + F^2_{st}) + \psi X^2_s E^2_t + \upsilon R_{st} + \varepsilon_{st}. \tag{2}$$

where X^2_s is the value of the variable X for store s in the second experimental period as taken from the personnel file received just before this period. The inclusion of the interaction between E^2_t and X is necessary to obtain differences-in-differences estimates of the effect of X on the treatment effects. For reasons unrelated to the experiment, the relation between X and y_{st} could change over time. If we would not control for such time-variant effects of X, they would bias our estimates of how X influences the effect of the treatments on performance y_{st} (i.e., we would get a biased estimate of λ).[9]

4 Results

The first column in Table 3 gives the results of estimating (1). Focusing on the second round, we find that the bonus treatment and the feedback treatment both have positive average treatment effects on weekly sales growth. This average treatment effect is statistically significant for the feedback treatment, and is borderline significant for the bonus treatment with a p-value of 0.13.[10] The size of the effects is also economically significant, as stores in the bonus and feedback treatment achieve 4.8 and 6.9 percentage points additional sales growth, respectively.[11] The difference in the effects of the bonus treatment and feedback treatment in the second period is not significant: a Wald test on the restriction that the effects are equal ($\delta = \mu$) has a p-value of 0.32. In the first round of the experiment, stores in

[9]Obviously, in most of our regressions X^2_s is a vector of variables and, likewise, λ and ψ are vectors of coefficients. In equation (2), we have taken up X^2_s only in interaction with the treatments and with the experimental period, not as separate control variables. Across weeks, we have only two different values per store for these variables, taken from the two personnel files (see Figure 2), implying that we have hardly any variation in X_s over time. In the estimation, the gender of the manager is constant across the personnel files for all stores, implying that its effect is absorbed by the store-fixed effects. Including the other variables from the personnel files as separate control variables does not affect our results.

[10]Throughout the paper, p-values are based on two-sided tests.

[11]Lack of data on the absolute value of sales and profit margins implies that we cannot establish whether this increase in sales outweighs the cost of the tournaments. However, the company's management was content with these results.

the bonus treatment perform slightly better than stores in the feedback treatment, however this difference is again insignificant.[12] Taken together, we cannot reject the hypothesis that average treatment effects of the bonus treatment and the feedback treatment are equally large. In other words, we find no evidence that the financial reward in the bonus treatment led to additional sales growth on top of the effect of the tournament that was also present in the feedback treatment.

Next, we analyse whether the response to competition depends on the gender of the store manager or on the gender composition of the store's team of employees. The first column of Table 4 reports the results of estimating (2) with X_s^2 only including a female manager dummy. We find a small and statistically insignificant effect. Hence, we find no evidence that, on average, the effect of competition on sales growth differs between stores with a male manager and those with a female manager. Similarly, we find small and statistically insignificant interaction effects if X_s^2 only includes the percentage of female employees in a store, as reported in the second column of Table 4. Thus, across all stores, we find no evidence that the gender composition of store employees influences the effects of competition. However, interacting the store manager's gender and the gender composition of store employees reveals an interesting pattern. The third column of Table 4 gives the results of estimating (2), where the treatments are interacted with both a female manager dummy and the percentage of female employees, as well as interacted with the interaction between the female manager dummy and the percentage of female employees. Thus, we allow for different effects of the gender composition of the stores' personnel on the effect of competition in stores with a male manager compared to stores with a female manager.[13]

Remarkably, we find that the sign of the effect of the percentage of female employees

[12] Furthermore, we cannot reject the hypothesis that $\gamma = \delta - \mu$ in (1), i.e., that the differences between the effects of the bonus and the feedback treatments are equal in the first and second round (p-value is 0.18).

[13] In the estimations, the percentage of female employees is mean-centered. Table 4 only reports the coefficients from the second round of the experiment, as we did not include a control group in the first round of the experiment, see Section 2. Note also that we find significant time-variant effects of the percentage of female employees on store's sales growth in column 3 (variables "% Female employees x Experimental round 2" and "Female Manager x % Female employees x Experimental round 2"). This underlines the importance of controlling for such time-variant effects, as otherwise these effects would be picked up by our estimates of the gender interactions with the treatment effects.

on a store's responsiveness to competition depends on the gender of the store manager. In stores with a male manager, the effect of competition on sales growth significantly decreases in the share of female employees with a marginal effect of -0.404 percentage point sales growth. An increase of one standard deviation in the percentage of women employed in a store, or about 12.5 percent points, leads to a decrease of about 5 percentage points in the treatment effect for male-led stores. By contrast, in female-led stores the responsiveness to competition increases in the percentage of female employees, with a marginal effect of $-0.404 + 0.886 = 0.482$ percentage point sales growth. A Wald test shows that this effect differs significantly from zero (p-value is 0.03). The magnitude of this effect is the same as in male-led stores: an increase of one standard deviation in the percentage of female employees increases the effect of competition by 6 percentage points.

The estimated treatment effects for various manager/employee combinations are depicted in Figures 4 and 5. Figure 4 gives the point estimates and the 95% confidence intervals of the effect of competition in the second round for male-managed and female-managed stores separately. Figure 5 depicts the same but with 90% confidence intervals. Both figures clearly show that the competition has been most effective in raising sales growth in male-led stores with a relatively high percentage of male employees. Figure 5 shows that the effect of competition is statistically significant in male-managed stores as long as the percentage of women employed does not exceed 80%. In female-managed stores, the pattern is reversed: the estimated impact of competition strongly increases with the percentage of female employees. The effect of competition in these stores is significant when at least 90% of the employees is female.

We have checked the robustness of our findings by controlling for several other variables. First, the results reported in Table 4 are not affected if we control for the interaction of the treatment effect with managers' tenure or with managers' age (i.e. if X_s^p in (2) includes managers' tenure or managers' age). Similarly, neither interacting employees' average age with the treatment effect nor interacting the treatment effect with employees' average tenure affects our results. Lastly, none of our results is affected qualitatively when weighing employees by their full-time equivalent.

Overall, our findings give a nuanced picture of gender differences in manager-led team performance under competition. It is not gender per se that affects performance under competition, but rather the match between the team's manager and the gender composition of the team: competition positively affects performance when the manager and a sufficiently high percentage of employees have the same gender. As mentioned in the Introduction, we can think of three plausible mechanisms behind this result. First, the response of team members to competition may crucially depend on the way a competition is communicated and promoted by the team's manager. In our experiments, we deliberately left a lot of discretion to team managers on how to use the competition as an incentive device. In particular, both the announcement of the competitions and the weekly posters were only sent to the store managers, not to the employees. It was up to the team managers to make the competitions appealing to their employees. Managers may have succeeded in making the competitions appealing to team members of their own sex, but less so to team members of the opposite sex.[14] This interpretation is well in line with evidence from management studies showing that when working for a manager of the opposite sex, employees find their duties and responsibilities much more ambiguous than when working for a manager of the same sex (Tsui and O'Reilly 1989, McNeilly and Russ 2000). Relatedly, a number of studies in organizational psychology have shown that, as compared to employees with opposite-sex managers, those with same-sex managers are more likely to develop high-quality leader-member exchange relationships (LMX) – a widely used measure of manager-employee mutual support, trust, and obligation – which may in turn facilitate communication.[15] Lastly, experimental evidence

[14] The idea that dissimilarity in personal attributes such as gender can deteriorate communication in organizations dates back to at least March and Simon (1958), who argue that dissimilarity may give rise to 'language incompatibility' and less frequent communication. There is pervasive evidence for 'homophily': the tendency that people interact more frequently with people with similar rather than dissimilar characteristics (McPherson et al. 2001, Borgatti and Foster 2003, Reagans 2005). While we find strong effects of gender similarity, we find no effect of similarity in age or tenure on stores' responsiveness to competition. This suggests that it is indeed gender similarity rather than similarity in general that drives our results.

[15] See e.g. Duchon et al. (1986), Pelled and Xin (2000), and Varma and Stroh (2001). Wayne et al. (1994) discuss a number of reasons for why these differences may arise.

using a subject pool of both students and banking executives finds that female participants tend to feel more comfortable supervising a female person than a male person in a challenging task, while male participants expect fewer conflicts with a male subordinate and perceive males to be more competent in a challenging task (Mai-Dalton and Sullivan 1981).

Alternatively, as team composition was not randomized in our experiment, the teams' gender composition may be the result of endogeneous matching on unobservables. If these unobservable characteristics are correlated with teams' responsiveness to competition, the pattern we find may arise without there being a causal link between teams' gender composition and performance under competition. For instance, suppose that relatively competitive managers have a preference for supervising employees of their own gender. These managers will self-select into (or gather) a team with more employees of their own gender than less competitive managers, and they will respond more strongly to tournament incentives. In this scenario, our results are driven by unobserved managerial characteristics rather than by team composition. Our experimental design does not allow us to discriminate between these mechanisms, so that this remains an important open question we hope to address in future work.

A third possible mechanism behind our results might be that male managers are better at reducing free-rider behavior in a team with many male employees, and likewise for female managers with female-dominated teams. We address free-rider behaviour by using the number of workers employed in the store, which varies between 5 and 20 employees across stores. Note, though, that we do not have an ideal set-up to analyze free-rider effects, as store size is not randomized. When there is an unobserved, systematic difference between small and large stores that affects the responsiveness to competitive incentives, this is reflected in the estimates, so that these are to be interpreted with caution.[16] The fourth column of Table 4 gives the results of including the number of employees interacted with the competition dummy, i.e., the pooled treatments. We find no indication for free-riding behavior. The

[16] For instance, free-rider effects are mitigated when managers with better team-building capabilities are more likely to be assigned to larger stores.

effect of competition does not depend on the number of employees.[17] Also, the inclusion of the number of employees hardly affects the estimates of the gender effects on performance under competition. Finally, interacting the number of employees with the gender composition of stores yields results that do not support the interpretation that free-rider behavior is reduced in stores where the manager and a large part of the store's employees are of the same gender; details are available upon request.

5 Concluding remarks

We have studied how teams led by a professional manager respond to competitive incentives. Overall, we find strong effects, even when there is no monetary reward to winning the competition. Further, our results suggest that the gender of the manager and the gender composition of the team jointly affect performance under competition. Male-led teams are more responsive to competition when a larger fraction of the team members is male. By contrast, female-led teams respond more strongly to competition when the fraction of female members is larger. We have discussed three plausible mechanisms behind this result. Future research should shed light on the relative importance of these mechanisms.

A Appendix

In our estimations of the effects of the gender of the manager and the gender composition of the team of employees on the response to competition, we pooled the bonus treatment and the feedback treatment, see equation (2) and Table 4. This appendix reports the results of estimating these interaction effects separately for the bonus treatment and the feedback treatment. Let E_t^1 be a dummy that takes value one for all observations in the first experimental period (weeks 44 - 49) and, as defined before, let E_t^2 be a dummy that takes value one in the second experimental period (weeks 71 - 76). The effect of variable X_s^p on the

[17] The same conclusion is drawn when the gender interaction terms are excluded from the estimation (so that X_s^2 in (2) only includes the number of employees).

effect of our treatments is estimated by:

$$y_{st} = \alpha_s + \theta_t + \gamma B^1_{st} + \eta X^1_s B^1_{st} + \kappa X^1_s E^1_t + \delta B^2_{st} + \mu F^2_{st} + \lambda X^2_s B^2_{st} + \pi X^2_s F^2_{st} + \psi X^2_s E^2_t + \upsilon R_{st} + \varepsilon_{st} \tag{A1}$$

where X^p_s is the value of the variable X for store s in experimental period $p \in \{1, 2\}$ as taken from the personnel files received just before period p. Table A1 reports the results of estimating (A1). Note that in the first experimental period, all stores were either in the bonus treatment or in the feedback treatment. Hence, the top half of Table A1 is a direct comparison of the gender interaction effects with the treatment effects between the bonus treatment and the feedback treatment during the first period. We find small and statistically insignificant differences in how the gender composition of stores affects performance under competition. The bottom half of Table A1 shows the gender interaction effects with the two treatments relative to the control group during the second period. Comparing the estimates of the interaction effects between the bonus treatment and the feedback treatment (i.e. comparing λ and π as used in (A1)), we find again no significant differences between the treatments. Comparing the estimated effects of variable X^p_s on the effect of the bonus treatment relative to the feedback treatment between the first and second experimental period (η versus $\lambda - \pi$), Wald tests show that these differences are not statistically significant.

References

[1] Auriol, Emmanuelle, and Régis Renault (2008), Status and Incentives, *RAND Journal of Economics*, vol. 39(1), pp. 305-326.

[2] Azmat, Ghazala, and Nagore Iriberri (2010), The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students, *Journal of Public Economics,* vol. 94(7-8), pp. 435-452.

[3] Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2009), Team Incentives: Evidence from a Field Experiment, *Journal of the European Economic Association,* forthcoming.

[4] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), How Much Should We Trust Differences-in-Differences Estimates?, *Quarterly Journal of Economics*, vol. 119(1), pp. 249-275.

[5] Bertrand, Marianne, and Kevin F. Hallock (2001), The Gender Gap in Top Corporate Jobs, *Industrial and Labor Relations Review*, vol. 55(1), pp. 3-21.

[6] Besley, Timothy, and Maitreesh Ghatak (2008), Status Incentives, *American Economic Review*, vol. 98(2), pp. 206–211.

[7] Blanes i Vidal, Jordi, and Mareike Nossol (2011), Tournaments without Prizes: Evidence from Personnel Records, *Management Science,* vol. 57(10), pp. 1721-1736.

[8] Booth, Alison L., and Patrick J. Nolen (2012), Choosing to Compete: How Different are Girls and Boys?, *Journal of Economic Behavior and Organization,* vol. 81, pp. 542-555.

[9] Borgatti, Stephen P. and Pacey C. Foster (2003). The Network Paradigm in Organizational Research: A Review and Typology, *Journal of Management,* vol. 29(6), pp. 991-1013.

[10] Casas-Arce, Pablo, and F. Asis Martinez-Jerez (2009), Relative Performance Compensation, Contests, and Dynamic Incentives, *Management Science,* vol. 55(8), pp. 1306-1320.

[11] Croson, Rachel, and Uri Gneezy (2009), Gender Differences in Preferences, *Journal of Economic Literature*, vol. 47(2), pp. 448-474.

[12] Dargnies, Marie-Pierre (2011), Men Too Sometimes Shy Away from Competition: The Case of Team Competition, *Management Science,* forthcoming

[13] Datta Gupta, Nabanita, Anders Poulsen, and Marie-Claire Villeval (2011), Male and Female Competitive Behavior: Experimental Evidence, *Economic Inquiry*, forthcoming.

[14] Dohmen, Thomas, and Armin Falk (2011), Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender, *American Economic Review,* vol. 101(2), pp. 556-590.

[15] Duchon, Dennis, Stephen G. Green, and Thomas D. Taber (1986), Vertical Dyad Linkage: A Longitudinal Assessment of Antecedents, Measures, and Consequences, *Journal of Applied Psychology*, 71(1), pp. 56-60.

[16] Flory, Jeffrey A., Andreas Leibbrandt, and John A. List (2010), Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Gender Differences in Job-Entry Decision, NBER Working Paper Number 16546.

[17] Frey, Bruno S., and Susanne Neckermann (2008), Awards - A View from Psychological Economics, *Journal of Psychology*, vol. 216(4), pp. 198-208.

[18] Gneezy, Uri, Kenneth L. Leonard, and John A. List (2009), Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society, *Econometrica*, vol. 77(5), pp. 1637-1664.

[19] Gneezy, Uri, Muriel Niederle, and Aldo Rustichini (2003), Performance in Competitive Environments: Gender Differences, *Quarterly Journal of Economics*, vol. 118(3), pp. 1049-1074.

[20] Gneezy, Uri, and Aldo Rustichini (2004), Gender and Competition at a Young Age, *American Economic Review*, vol. 94(2), pp. 377-381.

[21] Green, Jerry R., and Nancy L. Stokey (1983), A Comparison of Tournaments and Contracts, *Journal of Political Economy*, vol. 91(3), pp. 349-364.

[22] Harrison, Glenn, and John A. List (2004), Field Experiments, *Journal of Economic Literature*, vol. 42(4), pp. 1009-1055.

[23] IPU (2008), *Women in Politics: 2008*, poster published by Inter-Parliamentary Union and United Nations Division for the Advancement of Women.

[24] Ivanova-Stenzel, Radosveta and Dorothea Kuebler (2011), Gender Differences in Team Work and Team Competition, *Journal of Economic Psychology,* vol. 32, pp. 797–808.

[25] Jurajda, Stepan, and Daniel Münich (2011), Gender Gap in Admission Performance under Competitive Pressure: Admissions to Czech Universities, *American Economic Review,* vol. 101(3), pp. 514-518.

[26] Kosfeld, Michael, and Susanne Neckermann (2010), Getting More Work for Nothing? Symbolic Awards and Worker Performance, *American Economic Journal: Microeconomics*, vol. 3, pp. 86-99.

[27] Lavy, Victor (2008), Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-Based Pay Tournaments among Teachers, *Economic Journal,* forthcoming.

[28] Lazear, Edward P., and Sherwin Rosen (1981), Rank-Order Tournaments as Optimum Labor Contracts, *Journal of Political Economy*, vol. 89(5), pp. 841-864.

[29] Mai-Dalton, Renate R., and Jeremiah J. Sullivan (1981), The Effects of Manager's Sex on the Assignment to a Challenging or a Dull Task and Reasons for the Choice, *Academy of Management Journal*, vol. 24(3), pp. 603-612.

[30] March, James G., and Herbert A. Simon (1958), *Organizations*, New York: John Wiley.

[31] McNeilly, Kevin M., and Frederick A. Russ (2000), Does Relational Demography Matter in a Personal Selling Context?, *Journal of Personal Selling and Sales Management*, vol. 20(2), pp. 279-288.

[32] McPherson, Miller, Lynn Smith-Lovin, and James M. Cook (2001), Birds of a Feather: Homophily in Social Networks, *Annual Review of Sociology*, vol 27, pp. 415-444.

[33] Moldovanu, Benny, Aner Sela, and Xianwen Shi (2007), Contests for Status, *Journal of Political Economy*, vol. 115(2), pp. 338–363.

[34] Nalebuff, Barry J., and Joseph E. Stiglitz (1983), Prices and Incentives: Towards a General Theory of Compensation and Competition, *Bell Journal of Economics*, vol. 14(1), pp. 21-43.

[35] Niederle, Muriel, and Lise Vesterlund (2007), Do Women Shy Away from Competition? Do Men Compete Too Much?, *Quarterly Journal of Economics*, vol. 122(3), pp. 1067-1101.

[36] Örs, Evren, Frédéric Palomino, and Eloïc Peyrache (2008), Performance Gender-Gap: Does Competition Matter?, CEPR Discussion Paper No. 6891.

[37] Paserman, Daniele M. (2007), Gender Differences in Performance in Competitive Environments: Evidence from Professional Tennis Players, mimeo, Hebrew University.

[38] Pelled, Lisa Hope, and Katherine R. Xin (2000), Relational Demography and Relationship Quality in Two Cultures, *Organization Studies*, vol. 21(6), pp. 1077-1094.

[39] Polachek, Solomom W. (1981), Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure, *Review of Economics and Statistics*, vol. 63(1), pp. 60-69.

[40] Reagans, Ray (2005), Preferences, Identity, and Competition: Predicting Tie Strength from Demographic Data, *Management Science*, vol. 51(9), pp. 1374-1383.

[41] Snizek, William E., and Cecily C. Neil (1992), Job Characteristics, Gender Stereotypes and Perceived Gender Discrimination in the Workplace, *Organization Studies*, vol. 13(3), pp. 403-427.

[42] Tsui, Anne S., and Charles A. O'Reilly III (1989), Beyond Simple Demographic Effects: The Importance of Relational Demography in Superior-Subordinate Dyads, *Academy of Management Journal*, vol. 32(2), pp. 402-423.

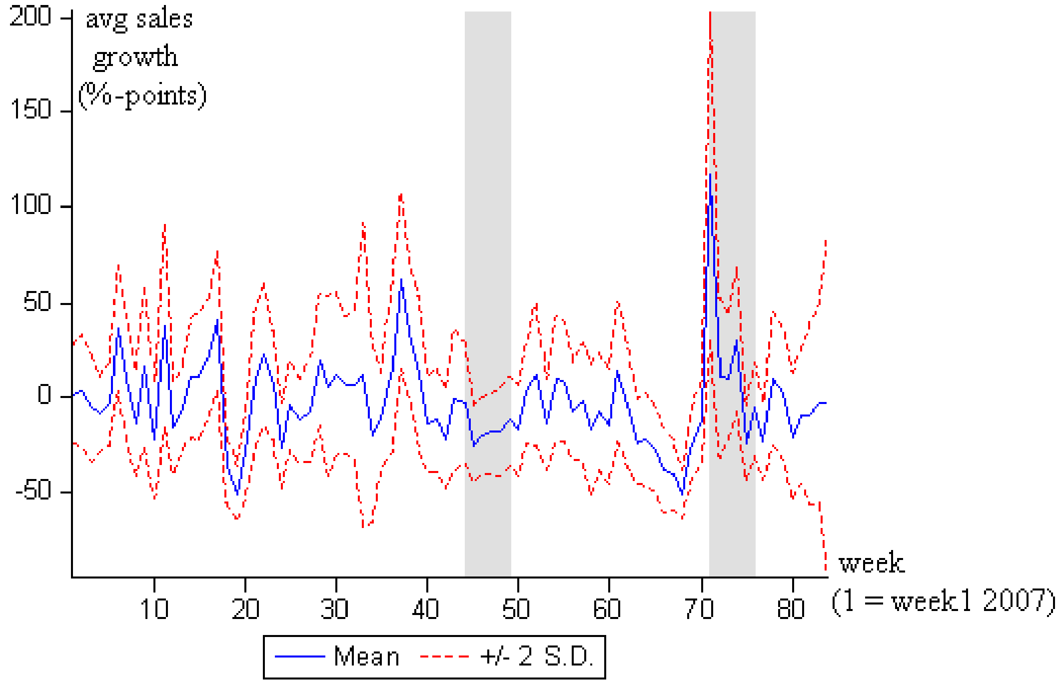
[43] Varma, Arup, and Linda K. Stroh (2001), The Impact of Same-Sex LMX Dyads on Performance Evaluations, *Human Resource Management*, vol. 40(4), pp. 309-320.

[44] Wayne, Sandy, Robert C. Liden, and Raymond T. Sparrowe (1994), Developing Leader-Member Exchanges, *American Behavioral Scientist*, vol. 37(5), pp. 697-714.

[45] Wirth, Linda (2001), *Breaking Through the Glass Ceiling: Women in Management,* Geneva: International Labour Office.

Tables and Figures

Figure 1: Weekly percentage sales growth as compared to the same week the year before



Experimental periods are shaded

Figure 2: Overview of the experimental set-up and timing of events

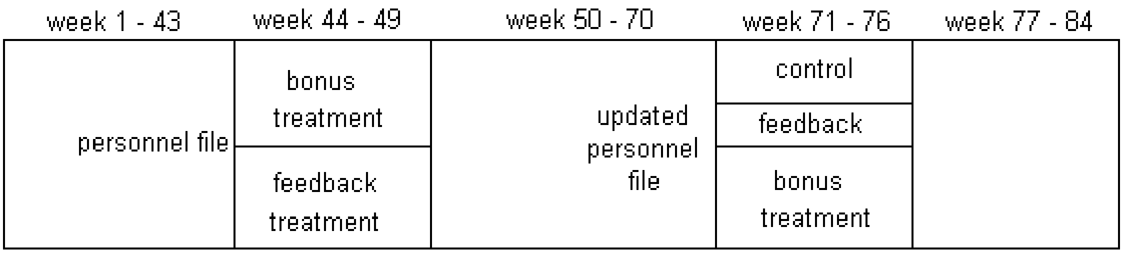


Figure 3: Distribution of the percentage of female employees by store managers' gender

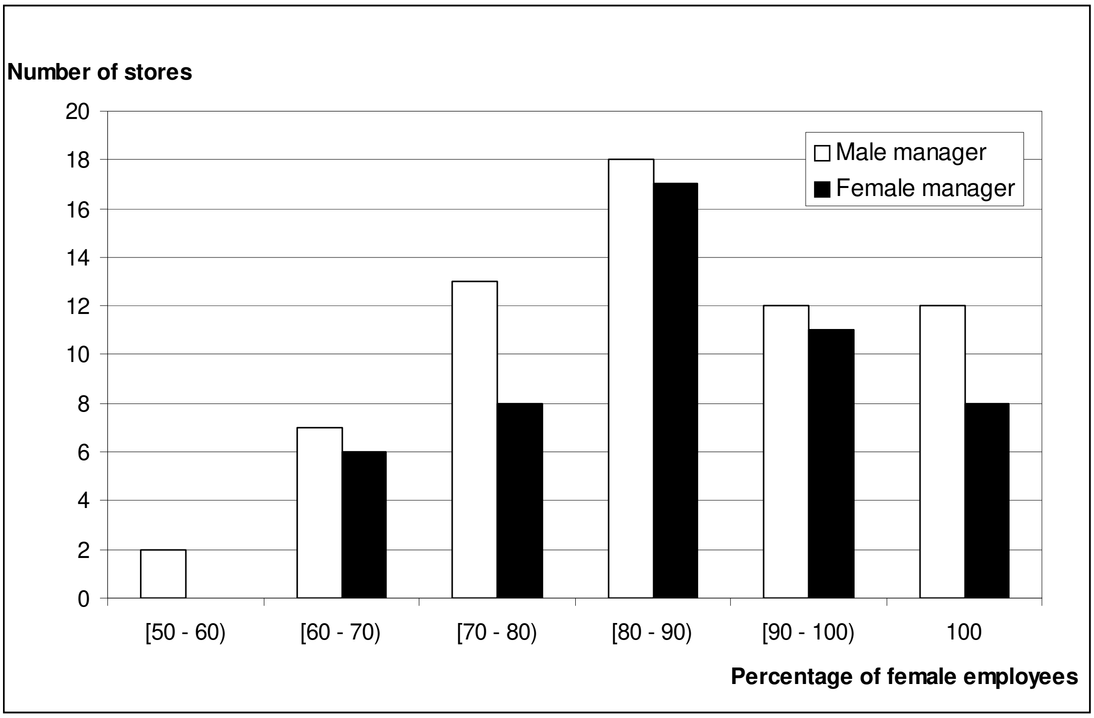


Table 1: Descriptive statistics

	All stores		Second period assignment					
			Bonus treatment		Feedback treatment		Control group	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Weekly sales growth (percentage points)	-2.44	29.87	-2.62	29.21	-2.22	30.90	-2.32	30.09
Female manager[a]	0.44		0.48		0.36		0.43	
Age manager[a]	38.7	9.3	39.2	10.3	38.2	7.6	38.1	9.1
Tenure manager[a]	11.9	8.7	12.9	9.3	11.7	8.7	10.7	7.8
Number of employees	11.8	3.2	11.8	3.4	11.2	2.7	12.2	3.5
Percentage of female employees	84.2	12.6	85.1	12.3	81.9	13.9	84.7	12.1
Average age of employees	25.2	4.0	26.0	4.5	24.7	3.3	24.4	3.3
Renovated store	0.13		0.17		0.07		0.13	
Number of stores	128		60		30		38	

Figures are the averages of the values in the two personnel files (see Figure 2).

F-tests show that none of the differences in means between the two treatment groups and the control group is statistically significant at the 10% level.

[a] The first personnel file contained information on 122 store managers, and the second personnel file had information on 119 store managers; see footnote 10. The missing observations are not correlated with either of the treatments.

Table 2: Descriptive statistics by store managers' gender

	Gender manager			
	Male		Female	
	Mean	Std	Mean	Std
Weekly sales growth (percentage points)	-2.07	29.52	-2.46	30.79
Age	39.4	9.5	37.8	9.1
Tenure*	14.9	9.5	8.3	6.4
Number of employees*	12.6	3.2	10.7	3.0
Percentage of female employees	83.7	13.0	85.8	11.5
Average age of employees	26.0	3.6	24.6	3.7
Renovated store	0.17		0.10	
Number of stores	64		50	

This table includes only the 114 stores used in the analysis of the gender effects, see footnote 10.

Apart from sales growth, the figures are the averages of the values in the two personnel files.

* Difference in means for male and female managers statistically significant at the 5% level.

Table 3: The effect of competition on sales growth

	Dependent variable: Sales growth	
Independent variables	(1)	
First round:		
Base category = Feedback treatment		
Bonus treatment	0.995	(0.986)
Second round:		
Base category = Control group		
Bonus treatment	4.796	(3.133)
Feedback treatment	6.902	(3.494)**
Renovated store	8.315	(3.554)**
Period fixed effects	YES	
Store fixed effects	YES	
Number of stores included	128	
Observations	10484	
R^2	0.693	
Log likelihood	-44292.3	

***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

% Female employees is mean-centered.

Standard errors clustered at store level in parentheses.

Table 4: The effect of competition on sales growth: Gender differences and team size effects

	Dependent variable: Sales growth							
Independent variables	(1)		(2)		(3)		(4)	
Second round: **Base category = Control group**								
Competition (=Bonus+Feedback)	4.306	(3.475)	4.173	(2.374)*	3.555	(3.183)	3.995	(3.779)
Female Manager x Competition	-0.778	(4.554)			-0.206	(4.391)	-0.964	(5.346)
Female Manager x Experimental round 2	3.611	(3.731)			3.411	(3.551)	4.021	(4.751)
% Female employees x Competition			-0.113	(0.183)	-0.404	(0.212)*	-0.413	(0.213)*
% Female employees x Experimental round 2			0.246	(0.147)*	0.408	(0.169)**	0.413	(0.170)**
Female Manager x % Female employees x Competition					0.886	(0.304)***	0.867	(0.300)***
Female Manager x % Female employees x Experimental round 2					-0.548	(0.230)**	-0.533	(0.224)**
Number of employees x Competition							-0.351	(0.754)
Number of employees x Experimental round 2							0.255	(0.706)
Renovated store	7.659	(3.742)**	7.774	(3.862)**	7.750	(3.760)**	7.774	(3.862)**
Period fixed effects	YES		YES		YES		YES	
Store fixed effects	YES		YES		YES		YES	
Number of stores included	114		114		114		114	
Observations	9326		9326		9326		9326	
R^2	0.704		0.704		0.705		0.705	
Log likelihood	-39290.4		-39299.7		-39290.4		-39290.1	

***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

% Female employees is mean-centered.

Standard errors clustered at stores level in parentheses.

Figure 4: Gender differences in the estimated effect of competition on sales growth depicted with a 95 percent confidence interval of the estimates

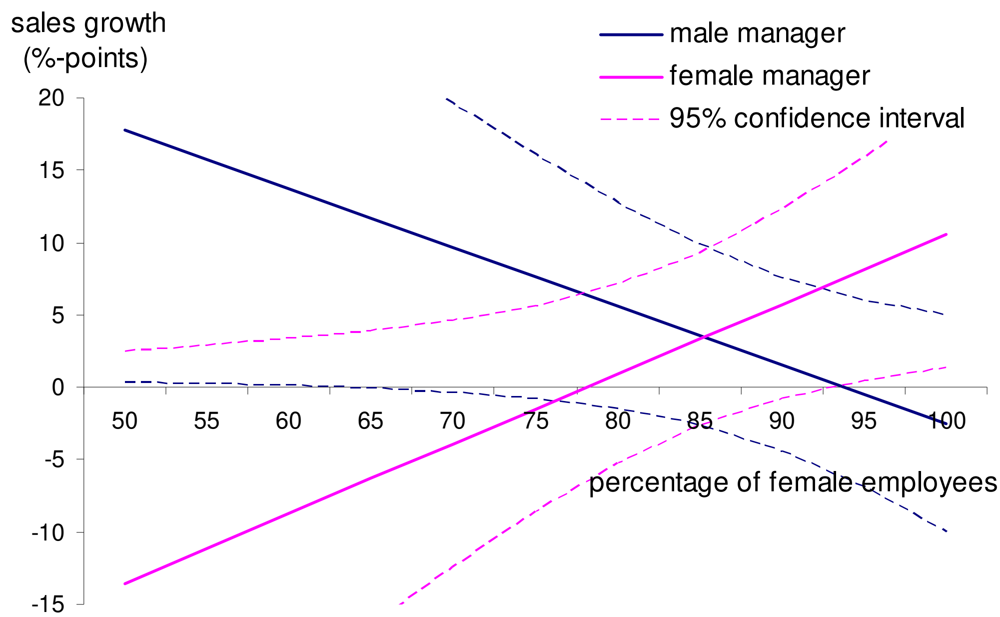


Figure 5: Gender differences in the estimated effect of competition on sales growth depicted with a 90 percent confidence interval of the estimates

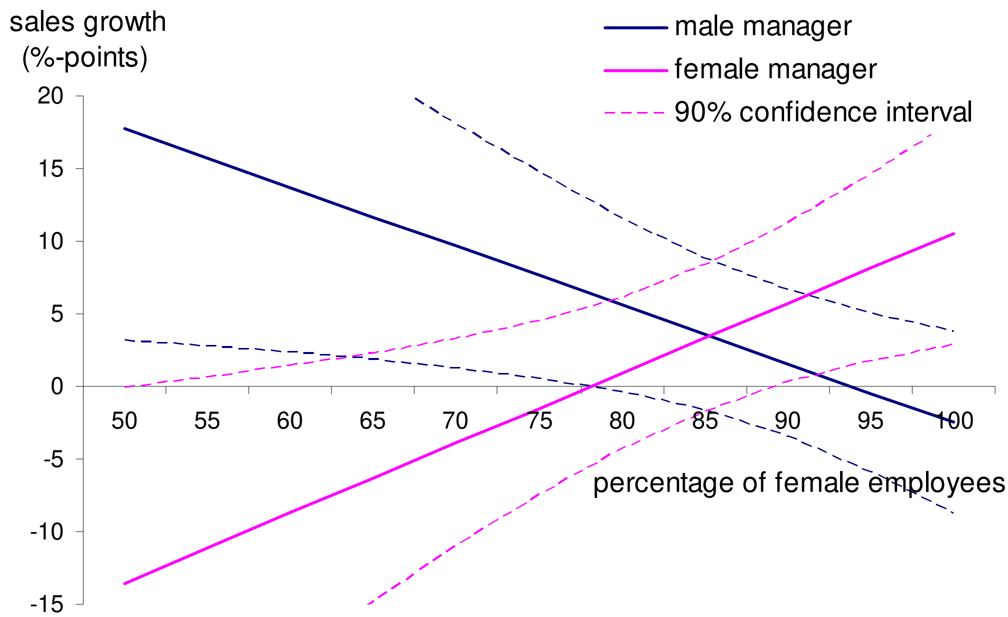


Table A1: The effect of competition on sales growth: gender differences by treatment

Independent variables	(1)	
First round:		
Base category = Feedback treatment		
Bonus treatment	1.763	(1.667)
Female Manager x Bonus treatment	-1.590	(2.075)
Female Manager x Experimental round 1	-1.057	(1.292)
% Female employees x Bonus treatment	-0.058	(0.119)
% Female employees x Experimental round 1	0.168	(0.078)**
Female Manager x % Female employees x Bonus treatment	0.172	(0.155)
Female Manager x % Female employees x Experimental round 1	-0.187	(0.100)*
Second round:		
Base category = Control group		
Bonus treatment	3.537	(3.491)
Feedback treatment	3.346	(3.290)
Female Manager x Bonus treatment	-1.832	(4.633)
Female Manager x Feedback treatment	4.098	(5.968)
Female Manager x Experimental round 2	3.234	(3.563)
% Female employees x Bonus treatment	-0.384	(0.261)
% Female employees x Feedback treatment	-0.416	(0.229)*
% Female employees x Experimental round 2	0.413	(0.170)**
Female Manager x % Female employees x Bonus treatment	0.831	(0.347)**
Female Manager x % Female employees x Feedback treatment	1.019	(0.463)**
Female Manager x % Female employees x Experimental round 2	-0.547	(0.231)**
Renovated store	7.866	(3,781)**
Period fixed effects	YES	
Store fixed effects	YES	
Number of stores included	114	
Observations	9326	
R^2	0.705	
Log likelihood	-39284.8	

***, **, * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

% Female employees is mean-centered.

Standard errors clustered at store level in parentheses.