PERFORMANCE IN COMPETITIVE ENVIRONMENTS: GENDER DIFFERENCES*

URI GNEEZY
MURIEL NIEDERLE
ALDO RUSTICHINI

Even though the provision of equal opportunities for men and women has been a priority in many countries, large gender differences prevail in competitive high-ranking positions. Suggested explanations include discrimination and differences in preferences and human capital. In this paper we present experimental evidence in support of an additional factor: women may be less effective than men in competitive environments, even if they are able to perform similarly in noncompetitive environments. In a laboratory experiment we observe, as we increase the competitiveness of the environment, a significant increase in performance for men, but not for women. This results in a significant gender gap in performance in tournaments, while there is no gap when participants are paid according to piece rate. This effect is stronger when women have to compete against men than in single-sex competitive environments: this suggests that women may be able to perform in competitive environments per se.

I. INTRODUCTION

Allocations across genders of high profile jobs remain largely favorable to men, and are a major factor in the gender gap in earnings. For example, Bertrand and Hallock [2001] found that only 2.5 percent of the five highest paid executives in a large data set of U. S. firms are women (for a review on gender differences in wages, see Blau and Kahn [2000]). The numerous attempts to explain this fact can be classified into two broad categories. The first explanation rests on gender differences in abilities and preferences and hence in occupational self-selection [Polachek 1981]. The second class of explanations relates to discrimination in the workplace, which leads to differential treatment of men and women with equal preferences and abilities [Black and Strahan 2001; Goldin and Rouse 2000; Wennerås and Wold 1997].

In this paper we propose and experimentally test an *additional* explanation: women may be less effective than men in competitive environments. This fact will reduce the chances of success for women when they compete for new jobs, promotions,

* We thank Daniel Brodkey, Stefano Della Vigna, Gerhard Orosel, Nikita Piankov, Alvin Roth, Lise Vesterlund, and seminar participants at numerous universities. This work was financially supported by the Binational Science Foundation and by the German Israeli Foundation.

The Quarterly Journal of Economics, August 2003

etc. Similar arguments can be found in the evolutionary psychology literature,[1] and even in the popular press. They also form the basis for the recent wave of support for single-sex schooling. The basic argument is that girls, when shielded from competition with boys, have a higher chance of developing their skills and interests in science. Surprisingly, however, there has been no direct experimental test of this assertion: in psychological research, gender differences in competition or attitudes toward competition are hardly mentioned.[2]

To test whether men and women differ in their ability or propensity to perform in competitive environments, we run controlled experiments. They allow us to measure performance precisely, and to exclude any discrimination and any expectation of discrimination. We have groups of three men and three women perform the task of solving computerized mazes. In the benchmark treatment, the payoff to participants depends only on their own performance: each one is paid a fixed piece rate for every maze solved over a period of fifteen minutes. We find no statistically significant gender difference in performance. To study the effect of competition, we use a tournament: the size and composition of the group and the task are the same as before, but now only the participant who solves the largest number of mazes is paid proportionally to the output. The average performance of men increases, while that of women is not affected. As a result, men outperform women on average, and more so than in the noncompetitive environment.[3]

The tournament design differs from the piece rate condition in two ways: payment is uncertain, and it depends on the performance of others. A possible explanation of the observed gender difference is that women are more risk averse, so that if effort is costly, the introduction of uncertainty into payments will affect men and women differently.[4] We introduce uncertainty without

1. In one of the important books of this tradition [Daly and Wilson 1988, p. 161], it is stated "Intrasexual competition is far more violent among men than among women in every human society for which information exists."

2. See for instance, the special issue *Psychological Sex Differences* of the *American Psychologist* [March 1995], pages 145–171. In this issue even the contribution of the social psychologist [David Buss, "Origins through Sexual Selection," pages 164–168] never mentions different attitude to competition as a difference among genders.

3. In the noncompetitive treatment, men outperform women but not significantly so. The gender gap in average performance in mixed tournaments is significant, and significantly higher than in the noncompetitive treatment.

4. For the psychology literature see Byrnes, Miller, and Schafer [1999]. In general, the results seem mixed, with a possibly higher degree of risk aversion among women.

competition in the next treatment. Again, only one participant is paid (as in the tournament), but this participant is now chosen at random. We do not find statistically significant gender differences in this treatment, and the gender gap in performance is lower than in the tournament.

All the above experiments were conducted in mixed groups of three men and three women. But women might perform differently in single-sex groups than in mixed groups. As mentioned above, this argument is used to support single-sex schools or classes. The basic argument is that girls in a single-sex environment have a higher chance of developing their skills and interests in science. Two main types of reasoning may support this. The first is that if girls are less inclined to compete than boys, the environment in single-sex schools for girls might be less competitive than in mixed sex schools, and hence be more suitable for girls. The second argument is that girls do not dislike competition: they only do not compete against boys. Hence, in single-sex schools girls will be more competitive, and education more effective than in mixed schools.[5]

To discern the effect of competition per se on the performance of women, we measure the performance of women and men in single-sex tournaments. We conduct tournaments as before, except that now each group of six competitors consists of either only women or only men. We find that the performance of women is significantly higher in single-sex tournaments than in the noncompetitive treatment. Hence there are competitive environments in which women's performance increases. Furthermore, the gender gap in performance is not different from the one in noncompetitive environments and is smaller than the gender gap in mixed tournaments.

Why do we observe this difference in reaction to competition? The behavior of men and women in a competitive environment may differ because of differences in skill, talent, and beliefs. A competitive environment may produce differences in behavior as subjects adjust their best choices to different strategic environments. In particular, if subjects believe (even if incorrectly) that men are more skillful at solving mazes than women, and effort is

5. See Schuss [2001] and Solnick [1995] and for overviews Harwarth, Maline, and DeBra [1997] and the AAUW Educational Foundation 1998 report "Separated by Sex: A Critical Look at Single-Sex Education for Girls."

costly, one would expect males to put in more effort than females in the winner-takes-all tournaments.

A different explanation is based on the argument that preferences over outcomes (that is, over individual effort, payment, and performance) are not independent of the institutional setup in which they are obtained and in particular of the competitive nature of the institution. The crucial element in this argument is that male's and female's preferences are affected differently by changes in the institution (its competitiveness, gender composition, etc.).

II. Experimental Design

We first establish an environment in which we can measure the effects of different incentive schemes on performance. For this purpose we conduct experiments in which participants have to solve a real task.[6]

II.A. The Task

Participants were told that the task they have to perform is to solve mazes.[7] The maze game has five levels of difficulty, from 1 = easy to 5 = hard. Each participant was asked to solve one maze of difficulty level 2 in order to get familiar with the task. After each participant finished one maze, the final part of the instructions was distributed. Only this final part differed between treatments.

The game is solved by operating the arrows on the keyboard, tracking a marker through a maze appearing on the screen. Participants were allowed to use only the arrows to move the cursor. The game was considered solved when the marker reached the end of the maze. The skill required to solve the problem requires a moderate amount of familiarity with a computer, plus the ability to look forward in the maze to detect dead ends. After finishing a maze, participants were asked to use the mouse in order to click "OK" and "New maze," and then start the new maze using only the arrows. They were instructed not to use any other function. After finishing a maze, they were asked to record this in a table. The experimenter confirmed that they

6. Many experiments directed at models of effort do not involve a real task, instead they model effort as a cost.

7. The mazes can be found at http://games.yahoo.com/games/maze.html.

marked the table correctly, and these records become the data of the experiment.

II.B. The Subjects

The experiment was conducted at the Technion, Haifa, Israel. Students at the Technion receive a degree in engineering, and it is a very competitive institution. Therefore, we do not expect the women in our subject pool to be especially intimidated when competing with men. Students were recruited through posters on campus that promised money for participating in a one-hour experiment. Students were asked to call a phone number, which was written on the poster. When they called, an answering machine replied asking them to leave their phone number, and they were told that they would be contacted later. Six participants were invited by phone to each session. Using this procedure, we could set up groups of the desired gender composition: either three females and three males or six females or six males. This fact was never explicitly pointed out to the participants. Although gender was not explicitly discussed at any time, participants could see each other in the laboratory and hence could determine the gender composition of their group. In each session after all six students entered the computerized experimental lab, they received a standard introduction. Each participant was told that (s)he will be paid 20 shekels for showing up.[8] Participants then received the instructions for the experiment. They were allowed to ask questions privately. In each treatment the experiment was replicated ten times with different participants. Hence in each treatment we have thirty men and thirty women. Across treatments we use different participants; hence we compare performance across individuals. Overall, we conducted 54 sessions with 324 participants.

III. Competitive and Noncompetitive Environments

We want to establish whether men and women differ in their ability or propensity to perform in environments in which they have to compete against one another. As a benchmark measure of gender differences in performance, we use a noncompetitive environment, a piece rate scheme. To test whether there is a gender differentiated impact of competition on performance, the second

8. At the time of the experiment, $1 = 4.1 shekels.

treatment measures performance in a mixed tournament. If men are more “competitive” than women, we expect an increase in the gender differences in performance when moving from a noncompetitive piece rate scheme to a competitive mixed tournament.

However, a tournament schedule differs in two ways from a piece rate. First, payment depends on the performance of the other participants, and second, payment is uncertain. In order to attribute gender differences in tournaments to the competitiveness of the environment, we have to investigate whether the introduction of uncertain payments per se might have a differential impact on the performance of men and women.[9] We therefore consider an incentive scheme where payment is uncertain, though independent of the performance of others.

III.A. Treatment 1: Piece Rate Payment

Participants were told that they have fifteen minutes to solve mazes. Their reward consists of 2 shekels for every maze they solved. Participants would not know how much other participants earned (i.e., how many mazes they solved).

III.B. Treatment 2: Competitive Pay: Mixed Tournament

Participants have fifteen minutes to solve mazes. Only the participant who solved the most mazes will be paid 12 shekels for every maze he or she solved. In case of a tie, the winners shared the payment equally. The other participants in the group did not receive any payment additional to the show-up fee. Participants would not know how much other participants earned (i.e., how many mazes they solved and hence the identity of the winner of the tournament).

III.C. Treatment 3: Random Pay

Participants have fifteen minutes to solve mazes. They are told in advance that only one of them (though they do not know which one) would be paid 12 shekels for every maze he or she solved. This participant was chosen at random at the end of the experiment, other participants in the group did not receive any payment additional to the show-up fee. Participants would not know how many mazes other participants solved.

9. There may be gender differences in risk aversion, or any other kind of aversion toward uncertainty. See, for example, Rabin [2000] for a discussion on risk aversion on small gambles.

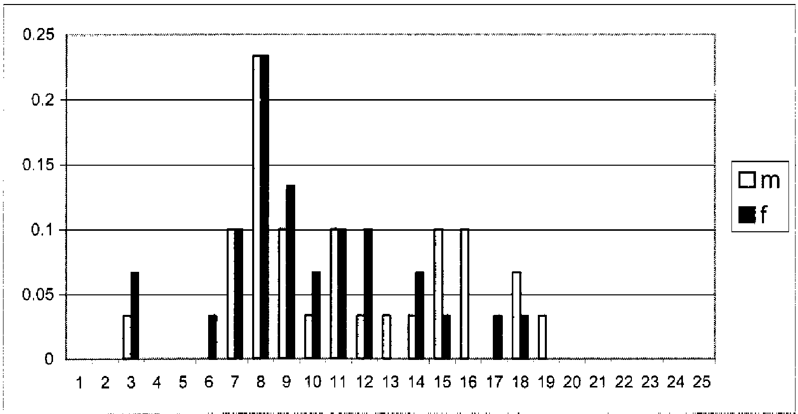

FIGURE I
Number of Mazes Solved under Piece Rate
The horizontal axis corresponds to the number of mazes solved, and the height of each bar reports the proportion of the 30 male and 30 female participants, respectively, who solved that many mazes.

Figure I and Figure II present the number of mazes solved by each gender for the piece rate treatment and the mixed tournament, the competitive environment; the performance under ran-

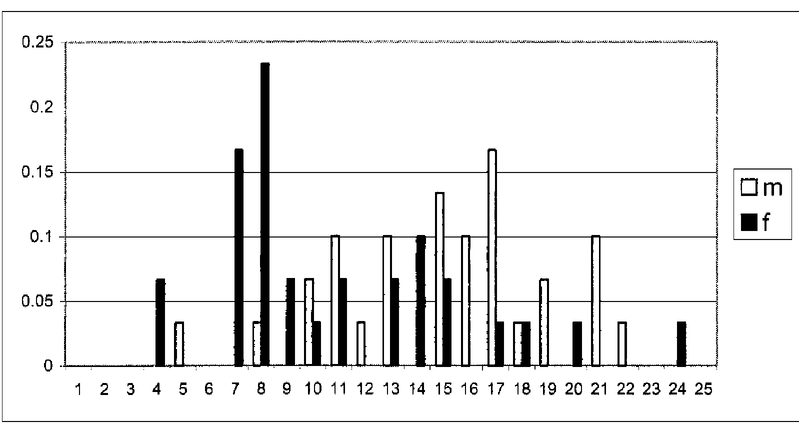

FIGURE II
Number of Mazes Solved under Tournament Condition
The horizontal axis corresponds to the number of mazes solved, and the height of each bar reports the proportion of the 30 male and 30 female participants, respectively, who solved that many mazes.

dom pay closely follows the piece rate scheme. To compare the performance of women and men across treatments and investigate gender differences within treatments, we use the two-sided Mann-Whitney U test, which compares distributions.

III.D. Performance under Noncompetitive Incentive Schemes, the Piece Rate, and the Random Pay

There are no significant gender differences in performance with the piece rate scheme or the random pay scheme, the p-value of the two-sided Mann-Whitney U test is 0.2023 and 0.165, respectively, the somewhat higher performance of men is not significant.[10] There is no significant difference in performance under the two noncompetitive treatments, the random pay and the piece rate for men (p-value 0.6449) or women (p-value 0.6130); the somewhat higher performance under the random pay scheme is not significant.

Participants do not change their performance in case of a random payment or a certain payment of the same expected value. We do not find any evidence that risk aversion influences performance.

III.E. Performance in the Mixed Tournament

There is a significant gender difference in performance in the mixed tournament: the p-value of the two-sided Mann-Whitney U test is 0.0004; the performance of men is significantly higher.

III.F. Performance Differences between Competitive and Noncompetitive Incentive Schemes for Men and Women

Men perform significantly higher in mixed tournaments than under both noncompetitive incentive schemes, the piece rate and the random pay. The p-value of the two-sided Mann-Whitney U test that compares performance of men in mixed tournaments with the piece rate is 0.001, while for the comparison to the random pay treatment it is 0.006.

Women do not significantly differ in their performance in the mixed tournament and the piece rate scheme, the p-value of the

10. Even though men do not significantly outperform women, below we will further investigate the possible implications that men have a somewhat higher performance than women. (In the piece rate treatment the average performance of men is 11.23, and that of women is 9.73. In the random pay treatment the averages are 11.83 and 10.33. All the average performances will be shown in Figure III).

two-sided Mann-Whitney *U* test is 0.62, while it is 0.623 when compared with the random pay. The somewhat higher performance of women in the mixed tournament is not significant.

Our results do not indicate that gender differences toward uncertainty are a driving factor for the gender gap in performance in tournaments. Furthermore, the increase in the gender gap in performance between the noncompetitive and the competitive treatment is driven by an increase of the performance of men and basically no change in the performance of women. The performance of women in mixed tournaments is not characterized by any, let alone a significant, group of women reducing their performance level compared with the noncompetitive incentive schemes. Women do not "give up" when competing against men.

The fact that men significantly change their performance in mixed tournaments compared with a piece rate scheme shows that (at least for men) mazes are an appropriate task to study different incentive schemes. Men do not simply solve mazes for fun, independently of the incentive scheme at hand; solving mazes appears to require real, costly effort.

III.G. Mixed Tournaments Significantly Increase the Gender Gap in Mean Performance Compared with Noncompetitive Incentive Schemes

To measure whether we have a significant treatment effect, i.e., a significant change in the gender differences in performance between the noncompetitive and the competitive treatment, we consider in each treatment the gender differences in average performance. The gender gap in mean performance in the piece rate scheme is 1.5 (mean performance of men is 11.23, that of women is 9.73) compared with 4.2 in mixed tournaments (mean performance of men is 15, that of women is 10.8). We use 1000 iterations of bootstrap to test whether (Gender gap in mean performance in the tournament) − (Gender gap in mean performance in the piece rate treatment) is strictly positive against the null hypothesis that this difference is less than or equal to zero.[11]

11. Specifically, we compute 1000 values for each of the gender gaps in mean performance in the tournament and the piece rate in the following way. For example, to compute the mean performance of men in the piece rate treatment, we randomly select, with replacement, 30 values of male performance out of the 30 values from our experiment. Doing the same for the women in the piece rate, we compute a value for the gender gap in the piece rate. Similarly, we compute a value for the gender gap in the tournament. We calculate the difference between the two, and iterate the process 1000 times.

We reject the null hypothesis at $p = 0.034$. Similarly, the gender gap in mean performance in the random pay treatment is 1.5 (mean performance of men is 11.83, that of women is 10.33), which is significantly smaller than the one of the mixed tournament at $p = 0.047$.

The rest of the study will be devoted to investigate why women do not increase on average their performance in tournaments, and in particular why tournaments have such a significant impact on the gender gap in mean performance.

IV. Do Women Compete?

There are many possible explanations why mixed tournaments result in a significantly higher gender gap in mean performance than the noncompetitive treatments. We will not provide an exhaustive list of explanations, but rather cluster them in several relevant categories.

IV.A. Women Do not Compete against Men

There are several explanations why women do not increase their performance in mixed tournaments, while men do, compared with their performance in a noncompetitive environment.

First, there may be a rational explanation for the significant increase in the gender gap in mean performance in tournaments. If women believe (even if incorrectly) that men are somewhat more skilled in solving mazes and they take the gender of their competitors as a signal of their ability (and maybe even take gender as a signal of their own ability), then a man and a woman face a different situation in the tournament. A man has two male competitors, and three female competitors who are on average less able than male competitors. An equally able woman, however, has two competitors whose ability is drawn from the same distribution as her own, and three male competitors who are on average abler than she is. Therefore, a man and a woman of equal ability have different expectations about their relative ability within the set of their competitors. This may lead to a lower expected performance of the woman than of an equally able man.

The psychology literature offers an additional possible explanation: women may think that they are worse than men at solving mazes in competitive environments, beyond possible actual gender differences. Furthermore, even if women do not internalize a possible stereotype about being less good at tasks like solving mazes, or performing in competitive environments, their performance might suffer in such situations. The idea is that women may experience a "stereotype threat" that provides an additional source of anxiety while performing the task and leads to higher instances of "choking under pressure" [Steele 1997].[12] The effect will be similar to the one where women believe they are less able than men, and where these beliefs are based on actual differences.

These explanations for the performance of women in tournaments hinge on the identity of the competitors. They imply that the performance of women in tournaments is not boosted the same as that of men when competing against men, though women might be capable of being effective in competitive environments and be motivated by competitive incentives.

The results of our experiments so far also allow for another class of explanations, namely that women may not be motivated or able to compete at all, independently of the nature of the competitors.

IV.B. Women Do not Compete at All

These explanations focus only on the ability and behavior of women in tournaments (or even in more general incentive environments) per se.

First, it might be that the costs of effort for women are such that increasing the output would only be possible at very high costs. Hence women effectively cannot solve more mazes. A second explanation is that women might not be sensitive to incentive schemes, at least when the task at hand is solving mazes. Note that in all the experimental treatments discussed so far, the performance of women did not change significantly. Third, women may not like to compete. There are several possible reasons for such preferences. One is that women are not socialized to

12. Steele and Aronson [1995] present a typical experiment of stereotype threat theory. Black and white participants had to take a difficult verbal test. Their two treatments varied the stereotype vulnerability of Black participants by varying whether or not their performance was ostensibly diagnostic of ability, and thus, whether or not they were at risk of fulfilling the racial stereotype about their intellectual ability. As predicted by stereotype threat theory, Blacks underperformed in relation to Whites in the ability-diagnostic condition but not in the nondiagnostic condition, controlling for performance in the Scholastic Aptitude Test. An additional treatment shows that the mere salience of the stereotype, by simply asking participants to indicate their race, can impair the performance of Blacks, even when the test was not diagnostic of ability. Brown and Josephs [1999] have experiments that suggest that positive stereotypes may also be a burden to performance.

compete. A fourth explanation is that the decision to not increase performance (and hence effort) in the tournament as opposed to the random pay treatment can be viewed as a public good. If all participants believe that the ex ante prior for each participant to win is 1/6, then they would all be better off if they ex ante committed to the effort level of the random pay treatment. However, the existence of gender differences in the contribution of public goods remains inconclusive (for an overview see Eckel and Grossman [2000] and Ledyard [1995]).[13]

There remains another possible driving factor for our results.

IV.C. Men Compete too Much

By using a real task, we forgo the advantage of being able to compute the equilibrium outcome in mixed tournaments. We do not know whether the behavior of the women is the "puzzle" that requires an explanation, or the behavior of men. It might be that the men are solving "too many" mazes, because they receive an additional benefit from winning the tournament, or are overconfident about their abilities and hence their chances of actually winning the tournament.[14]

To investigate these issues, we consider single-sex tournaments of six women only and six men only. The reason to have single-sex experiments for each gender, and especially also for men, is to control for the possibility that each gender may only perceive participants of the same sex as "real" competitors. However, behavior in tournaments is affected by the number of competitors. Hence, we do not necessarily want to compare the performance of women in single-sex tournaments with the performance of men in mixed tournaments.

This treatment also tests for the possibility that a significant component of men's performance in tournaments is due to the presence of women (i.e., men only compete when there are women around, and it is easy to come up with an evolutionary story for that).

13. Andreoni and Vesterlund [2001] found gender differences in altruism depending on the costs of providing benefits. Their predictions for our experimental environment would be unclear.

14. For an overview of gender differences in overconfidence, and the vast occurrence of overconfidence, see, e.g., Lichtenstein, Fischhoff, and Phillips [1982].

IV.D. Treatment 4: Single-Sex Tournaments

In the following treatment we have five sessions of groups of six women only and five sessions of groups of six men only. Apart from the gender composition of the group, everything is the same as in the tournament treatment: participants have fifteen minutes to solve mazes. Only the participant who solved the most mazes will be paid 12 shekels for every maze he or she solved. In case of a tie, the winners shared the payment equally. The other participants in the group did not receive any payment additional to the show-up fee. Participants would not know how much other participants earned (i.e., how many mazes they solved and hence the identity of the winner of the tournament).

IV.E. The Performance of Men

The performance of men in single-sex tournaments is not significantly different than in mixed tournaments; the p-value of the two-sided Mann-Whitney U test is 0.5630. The comparison of performance between men in single-sex tournaments as opposed to the piece rate or the random payment is still significant, the p-values are 0.036 and 0.018, respectively.

The performance of men, even though a bit lower in single-sex than in mixed tournaments, is not significantly affected by the gender of their competitors.

IV.F. Are Women Competitive?

To investigate whether women are competitive, we compare the performance of women in single-sex tournaments with that in the noncompetitive environments.

The performance of women in single-sex tournaments is significantly higher than in the noncompetitive treatments. The p-value of the two-sided Mann-Whitney U test when comparing the performance in single-sex tournaments to the piece rate performance is 0.0148, while it is 0.0469 when using the random pay treatment as the noncompetitive incentive scheme. So women do react to tournament incentives and do compete in single-sex groups.

However, the former analysis does not allow us to determine whether women (in single-sex groups) are as capable of competing as men are. To do that, we will compare the performance distributions of men and women in single-sex tournaments. However, we have seen that even in the noncompetitive treatments, men

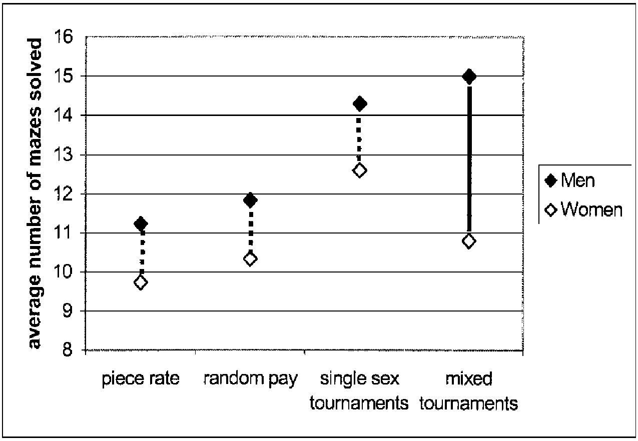

FIGURE III
Averages Performance of the 30 Men and 30 Women in Each of the Treatments

somewhat (though not significantly) outperform women. Therefore, we want to compare gender differences in mean performance across all treatments, and see whether differences in single-sex tournaments resemble differences in noncompetitive environments or differences in the mixed tournament.

There is no significant gender difference in performance in the single-sex tournament treatment; the p-value of the Mann-Whitney U test is 0.135. The somewhat higher performance of men is not significant.

Second, the gender gaps in mean performance are 4.2 for Mixed Tournament, 1.7 for Single-Sex Tournament, and 1.5 for Piece Rate and Random Pay. To confirm that there is a significant reduction in the gender gap in mean performance when moving from mixed to single-sex tournaments, we run 1000 iterations of bootstrap on (Men mixed − Women mixed) − (Men single-sex − Women single sex). We find a p-value of 0.082; hence we have a significant reduction in the gender gap in mean performance when moving from mixed to single-sex tournaments. Furthermore, there is no significant difference when comparing the gender gap in mean performance of the single-sex tournament with the piece rate (the p-value equals 0.459) or the random pay treatment (the p-value equals 0.535).

Figure III represents the average performance of men and women in all the treatments.

In single-sex tournaments, women solve more mazes than in mixed tournaments. However, this difference is not significant. (The p-value of the two-sided Mann-Whitney U test is 0.1025. Using a one-sided test of course makes the results significant at conventional levels.)

Before we investigate our data in more detail, in order to better understand why women do not compete against men, we want to point out that there remains one final possible explanation for our results, which has nothing to do with "competitiveness" of women. There is an alternative explanation for the increase of performance of women in single-sex tournaments compared with all other treatments. Maybe women in single-sex tournaments are only, or mostly, motivated not by the tournament's incentives, but by the fact that there are no male participants present in the experiment. Therefore, we need to test for any possible effect of "absence of male participants."

The new psychology literature on stereotype threat suggests that women in mixed-sex environments may be more alert to their gender, and hence more prone to stereotype threat and reduced performance. However, the one study testing this assertion [Inzlicht and Ben-Zeev 2000] did not simply have men and women perform a task, but made clear that the performance of each participant would be announced publicly, thereby creating a competitive environment. Indeed, the authors find a gender gap in performance on math GRE questions in mixed gender groups, but not in single-sex groups.

IV.G. Treatment 5: Single-Sex Piece Rate Payment

This treatment mimics the "mixed" Piece Rate treatment, but the group is homogeneous: either only six men or six women. Participants were told that they have fifteen minutes to solve mazes. Their reward consists of 2 shekels for every maze they solved. Participants would not know how much other participants earned (i.e., how many mazes they solved). We conducted two sessions for each gender. For women the average performance in the single-sex piece rate is 10, while it is 9.73 in the mixed piece rate, 10.33 in the mixed random pay, and 12.6 in single-sex tournaments. The p-value of the two-sided Mann-Whitney U test when comparing the single-sex piece rate results of women to their mixed piece rate is 0.83. The p-value of the two-sided Mann-Whitney U when comparing the single-sex piece rate results of women to their single-sex tournaments performance is 0.13. For the men, the average performance in single-sex piece rates is 11.08, compared with 11.23 in mixed piece rates and 14.3 in single-sex tournaments. The respective p-values of the two-sided Mann-Whitney U test are 0.77 and 0.05.

Given the average performance of women in the different treatments, we conclude that the increase in performance of women in single-sex tournaments is due to the incentive scheme and not the absence of male participants.

V. The Performance of Men and Women in More Detail

We now show how the average experiences from Figure III translate into individual behavior.

Specifically, we consider for each decile the proportion of women among all the participants whose performance ranks them above this decile.[15] For the single-sex tournament we pool the observations of men and women. The figure shows, for example, that if the top 40 percent of subjects were to be selected based on performance in mixed tournaments, this would result in a female representation of 0.24. On the other hand, if the selection were based on either performance in single-sex tournaments or under piece rate, the representation of females would be higher, around 0.4. Figure IV also suggests that (apart from possibly the highest performance group), women do not receive the same boost in performance as men when moving to a mixed tournament.

The fact that the proportion of women among high performers in single-sex tournaments mimics the proportion under piece rates is another indicator that women, when competing among themselves, receive a boost in performance from competition in single-sex tournaments. Mixed tournaments, however, result in a lower representation of women among all performance deciles but the first and second one.

V.A. *Do Men and Women Increase their Performance in the Same Way?*

We want to present some simple calculations, to estimate whether the performance of women increases in the same way as

15. When deciding about the gender of participants to include among all of those whose performance is exactly on the decile (there is in general more than one participant whose performance ranks them, e.g., tenth), we include men and women proportionally (according to their representation among participants whose performance is exactly the decile performance).

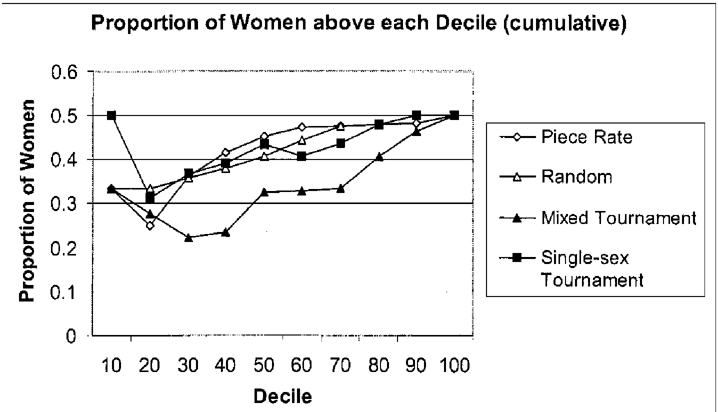

FIGURE IV
For Each Treatment, the Proportion of Women above Each Performance Decile

men's performance when moving from noncompetitive incentive schemes to competitive incentive schemes.

To have an estimate of the increase in performance of men, when moving from the noncompetitive treatments to the mixed tournament, we compare the performance of (different) men across treatments.[16] Consider, for example, a man who solved, say, fifteen mazes in the noncompetitive treatments. This means that among the 60 men in the noncompetitive treatments, he is the twelfth to seventeenth highest performing participant. To estimate his performance in the mixed tournament, we consider the performance of men who are the sixth to ninth highest performers among the 30 men in the mixed tournament. Their performance is 19, 18, 17, and 17 mazes. Therefore, we conclude that a man who solves fifteen mazes in the noncompetitive treatments solves (at least) seventeen in the mixed tournament.

If women receive the same boost in performance as men when moving from a noncompetitive treatment to a mixed tournament, a woman who solved fifteen mazes in the piece rate treatment should solve (at least) seventeen mazes in the mixed tournament.

16. We pool the two noncompetitive treatments, as otherwise, there are some performance values that are missing. For the following analysis, we do not distinguish between a random or a piece rate performance.

We perform this analysis for every number of mazes solved.[17] This yields an "expected mixed tournament" performance of 60 women. This performance is higher than the actual performance of the 30 women in the mixed tournament. The p-value of the two-sided Mann-Whitney U test that compares this "expected mixed tournament" performance with the actual mixed tournament performance is 0.003, which shows that women increase their performance significantly less than men when moving from noncompetitive environments to mixed tournaments.

There could be a good reason why a man and a woman who solve the same number of mazes in the noncompetitive environments perform differently in mixed tournaments: they actually face "different" competitors. A man faces two male and three female competitors, whereas a woman faces three male competitors and two female competitors. We have seen that the performance of women is slightly lower than that of men, so men face an "easier" competition than women.

Consider a man and a woman who solved fifteen mazes in the noncompetitive treatments. For each of them randomly select a group of men and women (among the participants of the noncompetitive treatments) so that the man and the woman are each in a different group of three men and three women. The man who solved fifteen mazes has a probability of 41 percent to be the highest performing participant among his group. The woman, however, has only a chance of 32 percent of being the highest performing participant in her group. Maybe a woman who has a 32 percent chance of being the highest performing participant in a group of three men and three women performing in the noncompetitive treatments (and who solved fifteen mazes) should increase her performance as much as a man who has a 32 percent probability of being the highest performing participant (and solves maybe less than fifteen mazes). It turns out that a man who solved fourteen mazes has a chance of 31 percent of being the highest performer in a mixed gender group. Therefore, one may argue, that the woman who has a 32 percent chance of being the highest performing participant in a group of three men and three women using the noncompetitive treatment (and who solved fifteen mazes), should increase her performance as much as a man

17. In case the equivalent performance of men is missing, we use the next highest performing man as a comparison.

who has a 31 percent chance of being the highest performing participant (and who solved fourteen mazes).

It turns out that for all performance levels a woman who solved x mazes always has a higher chance of being the highest performing participant in a randomly chosen mixed gender group than a man who solved $x - 1$ mazes in the noncompetitive treatments.[18] Therefore, a woman who solved x mazes in the noncompetitive treatments, has a higher probability of being the highest performing participant in her group, and has a higher ability than a man who solved $x - 1$ mazes in the noncompetitive treatment. If a woman and a man who have an equal chance of being the highest performing participant in a mixed gender group under noncompetitive incentive schemes receive the same boost in performance, then a woman who solved x mazes may be expected to increase her performance as much as a man who solved $x - 1$ mazes in the noncompetitive treatment. Using the same strategy as before to determine the boost of men when going from the noncompetitive treatments to the mixed tournaments, we compute an "expected mixed tournament" performance of 60 women. This performance is higher than the actual performance of the 30 women in the mixed tournament, the p-value of the two-sided Mann-Whitney U is 0.04.

Therefore, it seems that the low performance of women in mixed tournaments cannot be solely explained by the hypothesis that a man and a woman of the same ability perform differently in mixed tournaments, solely because the woman faces a tougher competition than the man.

There are several other reasons why women and men of the same ability may behave differently in mixed tournaments. It may be that men and women know nothing about their own ability, but are aware of possible gender differences in the distribution of ability. Using the performances from the piece rate treatment, the random pay treatment and the single-sex tournaments (where we pool the men and the women), a man and a woman in a group of three men and three women faces differences of 0.07 to 0.1 in being the highest performing participants in their group. These differences in probabilities of being the ablest par-

18. However, the woman indeed has a lower chance of being the highest performing participant in a randomly chosen mixed gender group than a man who solved x mazes.

ticipant may warrant the observed differences in behavior of women and men in mixed tournaments.

Another possibility may be that women underestimate their competence relative to men, so that in principle women may receive the same boost in performance as men, but they do not estimate their ability to be as high as that of an equally able man. In our last experiment we try to measure whether men and women feel differently competent in solving mazes. There is a large psychology literature that shows that all people are overconfident with respect to their ability, but, maybe men even more so than women (e.g., Beyer [1990], Lundeberg, Fox, and Punccohar [1994], and Beyer and Bowden [1997]).

Specifically, we measure whether men and women would make different choices when they can choose the difficulty level at which they will perform the task and in which they will be evaluated. In the next treatment we have subjects choosing the difficulty level of the mazes they will solve. Dweck and Leggert [1988] provide a good overview on the relation between task choice and feelings of competence.

V.B. Treatment 5: Choice of the Level of Difficulty

Participants were told that they have fifteen minutes to solve mazes. At the beginning of the fifteen minutes (after experiencing solving one maze at level 2 as described above), they were asked to fix the level of difficulty for the entire experiment. The payment was a function of the difficulty level: participants who chose level 1 were paid 1 shekel for every maze solved, those who chose level 2 were paid 2 shekels for each maze, and so on. Participants do not learn the choices or earnings of other participants.

Note that participants did not experience the other difficulty levels before setting the level for the entire experiment. Their optimal choice will therefore depend on their estimated ability in solving mazes, on their estimate of the actual difficulty of each level, and on their risk and ambiguity aversion. If men and women do not differ in these respects, then we expect males and females to chose similar levels of difficulty after solving one maze of level 2. Figure V presents the distribution of choices. The mean of the choices is 3.4 for males and 2.6 for females. This difference is statistically significant (the p-value equals 0.0065).

If we assume that all participants think that high levels of ability should lead to optimal choices of higher difficulty, and there is no gender difference in risk or ambiguity aversion, then

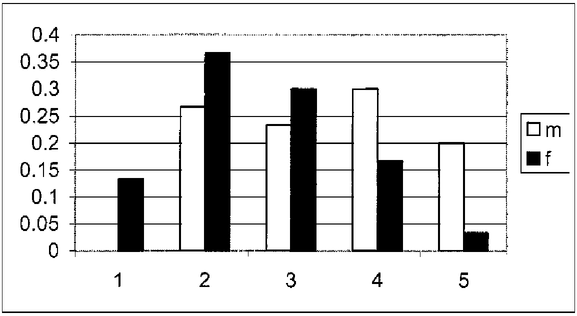

FIGURE V
Choice of Difficulty Level
The horizontal axis corresponds to the difficulty level, and the height of each bar reports the percentage of the 30 male and 30 female participants, respectively, who chose that difficulty level.

it seems clear that men feel more competent than women. The psychology literature has identified that higher choices of difficulty are associated with higher feelings of competence.

VI. THE EFFECT OF TOURNAMENTS

The mean performance of pooled women and men in mixed tournaments is 12.95 as opposed to 13.47 in single-sex tournaments; this difference is not significant (the p-value equals 0.62). Performance in tournaments (mixed and single sex) is significantly higher than under random pay (p-values of 0.032 and 0.0046 for comparison with mixed and single-sex tournaments, respectively), and under a piece rate (p-values of 0.073 and 0.0008 for comparison with mixed and single-sex tournaments, respectively). Hence it is evident that tournament incentives have a strong impact on performance.[19]

The result is related to the finding of Nalbantian and Schotter [1997], who find that the mean performance of their subjects was highest with competitive payment: see the section significantly titled "A Little Competition Goes a Long, Long Way." There are interesting differences, however. They study the effect of group incentive schemes on group performance, while we focus on individual incentives and behavior. Their subjects have noth-

19. The average payment per maze is 2.996 in mixed and 3.044 in single-sex tournaments (as opposed to 2 under the piece rate and the random payment).

ing to prove: the outcome does not depend on any skill or talent of the subject. Even the effort is simply a monetary input, not any real psychological and effort cost. While the setup of Nalbantian and Schotter may seem to provide a more direct control over the decision of the subjects, it also eliminates the psychological aspects (self-confidence) that are at the center of our study.

Tournaments are not only used to provide incentives, but also to determine very high performing participants. Therefore, to compare single-sex and mixed tournaments, we also want to compare the performance of the winning participants in each case.

In the mixed tournament the ten winners were seven men, two women, and one tournament ended in a tie between a man and a woman. The average performance of these ten winners is 19.4. In the ten single-sex tournaments we have of course (by design) five men and five women who win. Their average performance is 20.5. In order to test whether there is a significant difference in performance of the winners, we run 1000 iterations of bootstrap. The average performance of winners in single-sex tournaments is 20.5 which is not significantly different from the average in mixed tournaments 19.26: (p-value is 0.52).

VII. Discussion and Conclusion

We conducted controlled experiments to test the hypothesis that men and women react differently to competitive incentive schemes when competing against one another. When participants are paid a piece rate, there is no significant gender difference in performance. Our main finding is that using tournament incentives in mixed gender groups resulted in a significant increase relative to the benchmark in performance of male participants, but not of female participants. As a result, we observe a significant gender gap in tournaments. We show that this difference is not due to the uncertainty of the payment in tournaments, through gender differences in risk aversion. To understand a possible reason for this gender gap, we investigate whether women do not show an increase in performance in competitive environments per se, or only when competing against men. We ran single-sex tournaments where women (and men) still have to perform in a competitive environment, though now compete only against their own gender. This results in an increase in mean performance of women and a decrease in the gender gap in mean performance.

Several reasons may account for those results. Note that in all treatments on average men slightly outperform women. This can lead to gender differences in optimal performance in mixed tournaments. The reason is that men face only two male competitors, whereas women face three male competitors. There might be an additional effect that women feel less competent than men (beyond what would be warranted by the slight male advantage). Our last experiment provides support for the conjecture that women feel less competent than men in their ability to solve mazes.

Whether the poor performance of women relative to men in mixed gender group tournaments is a rational response or a psychological phenomenon, our results show that single-sex tournaments elicit more comparable performance from women and men. Furthermore, in our experiments, running single-sex as opposed to mixed tournaments came at no cost. The performance of the winners of single-sex tournaments is no lower than of the winners of mixed-sex tournaments (even though the winners of the single-sex tournaments are constrained to be always half women and half men).

Gneezy and Rustichini [2002] found gender differences in competition with young children (nine to ten years of age). In their experiment, the children first run separately, in which case the speed of girls is the same as that of boys. Then children are matched in pairs according to their performance (which results in some mixed and some same sex pairs), and run a second time. Competition has a positive effect on performance. This effect is stronger on boys than it is on girls, and the gender composition of the competing pair is important. An important difference in the results is that girls are motivated when competing against boys, but when two girls are matched, performance does not improve. However, the setup differs in several ways. First, there is continuous feedback on (relative) performance; second, competing children are matched according to their original performance (and this is common knowledge); third, the subjects are small children (with only an intrinsic motivation); and finally the task is physical rather than mental. Both the continuous feedback, and the fact that the children know that they are in ability matched pairs, would eliminate the reasons we provide for the lower performance of women in mixed tournaments.[20]

While the reasons for the attrition rate of women in science and engineering remain unclear, there exists mounting evidence that the women's low feelings of confidence and competence play a key role. A recent report titled "Women's Experiences in College Engineering" [Goodman, Cunningham, and Lachapelle 2002] shows that women are not dropping out of engineering programs because of poor performance. Many women who left mentioned negative aspects of their school's climate such as competition, lack of support, and discouraging faculty and peers. Positive perceptions of self-confidence were highly associated with staying in the program, and increased with the existence of mentor programs, opportunities for networking with practicing female engineers and clubs like the society for Women Engineers. In general, confidence in one's abilities and optimism have been shown to be strongly related to academic performance [Chemers, Hu, and Garcia 2001]. Furthermore, in male-dominated graduate programs, female students show lower feelings of competence than male students show [Ulku-Steiner, Kurtz-Costes, and Kinlaw 2000].

Our experiments also allow for an analysis of the impact of different incentive schemes when participants are required to provide real effort. We observe a significant increase in mean performance when moving from a noncompetitive scheme, such as a piece rate (or the random payment), to competitive schemes, such as the mixed and single-sex tournaments. Furthermore, single-sex tournaments are as effective as mixed tournaments in eliciting performance of all participants and in terms of eliciting high performance of the winners of the tournaments.

THE UNIVERSITY OF CHICAGO, GRADUATE SCHOOL OF BUSINESS, AND TECHNION
STANFORD UNIVERSITY
UNIVERSITY OF MINNESOTA

REFERENCES

Andreoni, James, and Lise Vesterlund, "Which Is the Fair Sex? Gender Differences in Altruism," *Quarterly Journal of Economics,* CXVI (2001), 293–312.

20. A continuous feedback on performance also allows for subjects to monitor each other's performance and collude, and not exert higher effort, simply because the setup is competitive.

Bertrand, Marianne, and Kevin F. Hallock, "The Gender Gap in Top Corporate Jobs," *Industrial and Labor Relations Review,* LV (2001), 3–21.
Black, Sandra, and Philip E. Strahan, "The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry," *American Economic Review,* XCI (2001), 814–831.
Blau, Francine, and Lawrence M. Kahn, "Gender Differences in Pay," *Journal of Economic Perspectives,* XIV (2000), 75–99.
Beyer, Sylvia, "Gender Differences in the Accuracy of Self-Evaluations of Performance," *Journal of Personality and Social Psychology,* LIX (1990), 960–970.
Beyer, Sylvia, and Edward M. Bowden, "Gender Differences in Self-Perceptions: Convergent Evidence from Three Measures of Accuracy and Bias," *Personality and Social Psychology Bulletin,* XXIII (1997), 157–172.
Brown, Ryan P., and Robert A. Josephs, "A Burden of Proof: Stereotype Relevance and Gender Differences in Math Performance," *Journal of Personality and Social Psychology,* LXXVI (1999), 246–257.
Byrnes, James P., David C. Miller, and William D. Schafer, "Gender Differences in Risk Taking: A Meta-Analysis," *Psychological Bulletin,* LXXV (1999), 367–383.
Chemers, Martin M., Li-tze Hu, and Ben F. Garcia, "Academic Self-Efficacy and First-Year College Student Performance and Adjustment," *Journal of Educational Psychology,* XLIII (2001), 55–64.
Daly, Martin, and Margo Wilson, *Homicide* (New York, NY: Aldine de Gruyter, 1988).
Dweck, Carol S., and Ellen L. Leggett, "A Social-Cognitive Approach to Motivation and Personality," *Psychological Review,* XLV (1988), 256–273.
Eckel, Catherine C., and Philip J. Grossman, "Differences in the Economic Decisions of Men and Women: Experimental Evidence," mimeo, 2000.
Gneezy, Uri, and Aldo Rustichini, "Gender and Competition at a Young Age," mimeo, 2002.
Goldin, Claudia, and Cecilia Rouse, "Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians," *American Economic Review,* XL (2000), 715–742.
Goodman, Irene F., Christine M. Cunningham, and Cathy Lachapelle, "The Women's Experience in College Engineering," Goodman Research Group, Inc., 2002.
Harwarth, Irene, Mindi Maline, and Elizabeth DeBra, "Women's Colleges in the United States: History, Issues, and Challenges," Office of Educational Research and Improvement at the U. S. Department of Education report, 1997.
Inzlicht, Michael, and Talia Ben-Zeev, "A Threatening Intellectual Environment: Why Females Are Susceptible to Experiencing Problem-Solving Deficits in the Presence of Males," *Psychological Science,* XI (2000), 365–371.
Ledyard, John O., "Public Goods: A Survey of Experimental Research," in *Handbook of Experimental Economics,* John H. Kagel and Alvin E. Roth, eds. (Princeton, NJ: Princeton University Press, 1995).
Lichtenstein, Sarah, Baruch Fischhoff, and Lawrence Phillips, "Calibration and Probabilities: The State of the Art to 1980," in *Judgment Under Uncertainty: Heuristics and Biases,* Daniel Kahneman, Paul Slovic, and Amos Tversky, eds. (Cambridge University Press: 1982).
Lundeberg, Mary A., Paul W. Fox, and Judith Punccohar, "Highly Confident but Wrong: Gender Differences and Similarities in Confidence Judgments," *Journal of Educational Psychology,* LXXXVI (1994), 114–121.
Nalbantian, H. R., and A. Schotter, "Productivity under Group Incentives: An Experimental Study," *American Economic Review,* LXXXVII (1997), 314–341.
Polachek, Solomon W., "Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure," *Review of Economics and Statistics,* LXIII (1981), 60–69.
Rabin, Matthew, "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica,* LXVIII (2000), 1281–1292.
Schuss, Deborah Gastfreund, "Bucking the Trend. Wellesley Students Flock to Economics," *Boston Sunday Globe,* February 4, 2001.
Solnick, Sara J., "Changes in Women's Majors from Entrance to Graduation at Women's and Coeducational Colleges," *Industrial and Labor Relations Review,* XLVIII (1995), 505–514.

Steele, Claude M., "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance," *American Psychologist,* LII (1997), 613–629.

Steele, Claude M., and Joshua Aronson, "Stereotype Vulnerability and the Intellectual Test Performance of African Americans," *Journal of Personality and Social Psychology,* LXIX (1995), 797–811.

Ulku-Steiner, Beril, Beth Kurtz-Costes, and C. Ryan Kinlaw, "Doctoral Student Experiences in Gender-Balanced and Male-Dominated Graduate Programs," *Journal of Educational Psychology,* XLII (2000), 296–307.

Wennerås, Christine, and Agnes Wold, "Nepotism and Sexism in Peer-Review," *Nature,* CCCLXXXVII (1997), 341–343.