Performance in competitive Environments: 

Gender differences*

Uri Gneezy, Muriel Niederle and Aldo Rustichini

April 1, 2001

Abstract

In spite of the fact that equal opportunities for men and women have been a priority in many countries, enormous gender differences prevail in most competitive high-ranking positions. We conduct a series of controlled experiments to investigate whether women might react differently than men to competitive incentive schemes commonly used in job evaluation and promotion. We observe no significant gender difference in mean performance when participants are paid proportional to their performance. But in the competitive environment with mixed gender groups we observe a significant gender difference: the mean performance of men has a large and significant, that of women is unchanged. This gap is not due to gender differences in risk aversion. We then run the same test with homogeneous groups, to investigate whether women under-perform only when competing against men. Women do indeed increase their performance and gender differences in mean performance are now insignificant. These results may be due to lower skill of women, or more likely to the fact that women dislike competition, or alternatively that they feel less competent than their male competitors, which depresses their performance in mixed tournaments. Our last experiment provides support for this hypothesis.

Key words: Gender, motivation, competition.

JEL Classification: D81

* Technion, Harvard University and University of Minnesota. We thank Dan Brodkey, Stefano Della Vigna, Al Roth, Gerhard Orosel, Nikita Piankov, Lise Vesterlund.
I. Introduction

High profile jobs remain pervasively male and are a major factor for the gender gap in earnings (Gunderson 1981). 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 work place, which leads to differential treatment of men and women with equal preferences and abilities (Goldin and Rouse 2000, Wennerås and Wold 1997). We propose an alternative explanation, namely that women and men differ in their ability or propensity to perform in competitive environments. Since incentive schemes commonly used in job evaluation and performance are often highly competitive, they may trigger different performances among men and women.

Relatively little attention has been given to gender differences in competitive behavior in the economics and, surprisingly, also in psychology literature.¹ We will present the results of a controlled experiment designed to measure performance precisely, and to exclude any discrimination and any expectation of discrimination.

We have groups of three males and three females perform a real task (solving computerized mazes during 15 minutes). In the benchmark treatment the payoff of participants depends only on their own performance: each one is paid at a fixed piece rate for every maze they solve within the 15 minutes. We find no significant gender difference in performance. To study the effect on performance of a competitive environment we use a tournament scheme: the design is the same, but only the participant who solves the largest number of mazes receives is paid proportionally to the output. The average performance of men significantly increases, while that of women is not affected. As a result, men significantly outperform women on average.

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

¹ For instance a recent review in the American Psychologist on gender differences does not even mention competition, let alone possible gender differences in attitudes toward competition See American Psychologist, March 1995, pages 145 – 171.
uncertainty into payments will affect men and women differently.\footnote{For a review of the economics literature see Eckel, Catherine C. and Philip J. Grossman, (2000), "Differences in the economic decisions of men and women: Experimental evidence", Handbook of Results in Experimental Economics forthcoming. 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.} We introduce uncertainty but no competition in the next treatment. Again only participant is paid (as in the tournament), but this participant is now chosen at random. We do not find gender differences in this treatment.

All the above experiments were conducted in mixed groups of three males and three females. But women might perform differently in single sex than in mixed groups. Such statements are made for example to support single-sex schooling, and are based on the idea that girls, when shielded from the competition with boys, have a higher chance of developing their skills and interests in science. Two main types of reasoning may support this point of view. The first is that if girls are less competitive than boys, the environment in single sex schools for girls might be set to be less competitive than in the mixed sex schools, and hence be more suitable to girls. The second argument is that girls do not dislike competition: they dislike competing against boys. Hence in single sex schools girls will be more competitive, and education more effective, than in mixed schools.

To discern the effect of competitive environments 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 women are not completely unable or unwilling to compete.

The data of our experiment also allow us to consider the impact of different incentive schemes on performance. First, we find that aggregate performance is significantly increased when moving from a piece rate scheme to tournaments. For mixed tournaments, but not for single sex tournaments, this is entirely driven by the boost in performance of males. Furthermore aggregate performance in single sex tournaments is as high as in mixed tournaments. This is also true, when we only compare the performance of the winner of the ten mixed and single sex tournaments.

For a review of the economics literature see Eckel, Catherine C. and Philip J. Grossman, (2000), “Differences in the economic decisions of men and women: Experimental evidence”. Handbook of Results in Experimental Economics forthcoming. 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.
respectively. In the mixed tournaments, around 30 percent of the winner are female. In the single sex tournaments, by design, fifty percent of the winners are female. Of course the choice of homogeneous rather than mixed gender tournament leaves open the question of the hiring policy to be followed on the basis of the results of the tournament. For instance, one might run homogeneous gender tournaments, and then hire the best candidate in the mixed sample, avoiding quotas.

II. Experimental Design

We first establish an environment in which we can meaningfully measure competence and effects of changes in incentive schemes on performance. For this purpose we conduct experiments in which participants have to solve a real task.

Task

Participants were told that the task they have to perform is to solve mazes. The mazes can be found at [http://games.yahoo.com/games/maze.html](http://games.yahoo.com/games/maze.html). 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 were distributed. Only this final part differed between treatments. The different instructions are presented in Appendix 2.

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, as usual, 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, as usual, 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 marked the table correctly, and these records become the data of the experiment.
Subjects
The experiment was conducted at the Technion, Haifa, Israel. Students at Technion get a degree in engineering. Students were recruited with posters on campus that were promising money for participating in an experiment one hour long. 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. 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.³ Participants then received the instructions for the experiment (see Appendix 1). 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.

III. Competitive and Non Competitive Environments
The first experiment aims at establishing a benchmark: the performance of subjects when their reward is independent of the performance of others.

Experiment 1: Piece Rate Payment
Participants were told that they have 15 minutes to solve mazes of difficulty level two, and that they would be paid 2 shekels for every maze they solved. Participants would not know how much other participants earned (i.e. how many mazes they solved.)

³ At the time of the experiment, $1=4.1 Shekels
If males and females are equally skillful in the basic task the distribution of mazes solved by males and females should not be significantly different. There is in fact no significant difference between males and females. The average number of mazes solved for males is 11.23 versus 9.73 for females. The confidence intervals are overlapping, and the p-value of the two-sided Mann-Whitney $U$ test is 0.2023. The somewhat higher average for male is not significant. The results are presented in Figure 1.

<Insert Figure 1>

**Figure 1: Number of mazes solved under proportional payment.** On the horizontal axis the number of mazes solved; on the vertical axis the percentage of participants for the corresponding performance, male and female respectively.

We are going to put on record our main results to outline the progress of the investigation:

1. **There is no significant difference in gender performance in piece rate payment**

We have now established that performance between men and women is not significantly different. Next we want to assess the impact of competition on the performance of males and females.

**Experiment 2: Competitive Pay**

As in Experiment 1 participants have 15 minutes to solve mazes. They are told (in advance) that after those 15 minutes the participant who solved the most mazes will be paid 12 shekels for every maze he or she had solved. In case of a tie, the winners shared the payment equally. The other participants in the group do not receive any payment additional to the show up fee.

To determine the theoretical prediction, consider a group of six participants that are risk neutral and have (common knowledge) ex ante equal ability and effort costs for solving mazes. Then
we expect that the average number of mazes solved in the tournament to be higher than in the case of proportional pay. The reason is that in the tournament the expected marginal gain of solving one more maze is comprised of two terms. The first term is that in the case of winning the participant receives 12 shekels for each additional maze solved. In case the expected probability of winning is 1/6, this first term just equals the incentives for solving mazes in the proportional pay treatment. However in the tournament treatment the participant has an additional incentive to exert effort to solve mazes. By solving more mazes a participant can actually increase the chances of winning. Hence in the case of a group of six risk neutral participants who think they are equally capable of solving mazes, we expect all participants to solve more mazes than in the piece rate treatment (Experiment 1). In general we expect every risk neutral participant who expects an ex ante probability of winning the tournament of at least 1/6 to exert more effort than in the proportional pay treatment and consequently solve on average more mazes. The mean number of mazes solved in the tournament treatment is 12.95 versus 10.48 in the proportional pay treatment. This difference is highly significant, with a \( p \)-value of 0.0073.

2. Performance is significantly higher in tournament than in piece rate payment

Next we investigate whether this average increase in performance is due to all participants uniformly increasing their performance and whether there are any gender differences. The mean for males, 15, is much higher than the mean of the females, 10.8. The \( p \)-value is 0.0004: men have a significantly higher average performance than women. The difference is presented in Figure 2.

<Insert Figure 2>

3. Performance of male is significantly higher than that of women in tournaments
The mean performance of men in the tournament is 15, versus 11.2 in the proportional pay treatment. This is a significant difference with a \( p \)-value of 0.001.

4. **Men, on average, react strongly to tournament incentives**

The mean performance of women in the tournament is 10.8, versus 9.73 in the proportional pay treatment. This difference is not significant, with a \( p \)-value of 0.62.

5. **Women, on average, do not react to tournament incentives**

We have a large gender gap in performance in the tournament, but no significant gender gap in the proportional pay treatment. Next we want to show that the gender gap in performance as measured by the difference between the mean male performance and the mean female performance significantly increases when moving from a proportional incentive scheme to a tournament. We use 1000 iterations of bootstrap to test whether (Gender gap in mean performance in the tournament)-(Gender gap in mean performance in the proportional pay treatment) is strictly positive against the null hypothesis that this difference is less than or equal to zero. We can reject the null hypothesis at 0.034. Therefore moving from a proportional pay scheme, where there is no significant gender gap in mean performance, to a tournament scheme, significantly increases the gender gap in mean performance.

Next we explore whether the average absence of a boost in performance of women in the tournament translates to experiences at each performance quintile. In each treatment we first rank the sixty participants by performance (remembering their gender). The first quintile are the top 20 percent participants, these are the twelve highest performing participants in each
We then compute the proportion of women in each performance quintile in each treatment, the results are presented in Figure 3.

Let us first concentrate on top quintile, the 20 percent best participants: It consists of 25 percent women in the proportional pay treatment and 27.8 in the mixed Tournament, i.e. around 3 women out of 12 participants in each treatment. This is significantly lower than fifty percent at \( p=0.04 \) in the proportional pay treatment and \( p=0.06 \) in the mixed tournament using 1000 iterations of bootstrap. Furthermore, in order to be (probabilistically) included in the top quintile in the proportional pay treatment a participant has to solve at least 14 mazes as opposed to 17 in the tournament. If ability plays a role, and it is positively correlated with the number of mazes a participant solves, then it seems that the most able men and women are equally affected by tournament incentives in terms of a boost in performance. However in the second quintile, (20 to 40 percent best participants), there are fewer women in the tournament treatment (19.4 percent) than in the proportional pay treatment (58 percent). So that more men have a relatively high performance, or alternatively there are much fewer women whose performance qualifies them to be in the second best quintile. In order to be in the second best quintile in the tournament treatment a participant has to solve between 15 and 17 mazes as opposed to 11 and 14 in the proportional pay treatment. We do not observe any different gender composition in the third and fourth quintile. Therefore we find an increase in the proportion of women in the last and worst quintile of participants (in terms of their performance in solving mazes.) These last twenty percent of participants in the mixed tournament are comprised of 87.5 percent of women as opposed to 58.3 percent in the proportional pay treatment. The proportion of women in the last and worst quintile in the mixed tournament is significantly higher than in the proportional pay treatment at \( p= 0.047 \).

\footnote{Specifically we include all the participants who solved strictly more mazes than the twentieth best participant, and proportionally add participants from the pool of people that solved as many mazes as the twentieth best participant.}
The cumulative effect of these differences can be seen in Figure 4, in which we consider for each decile the proportion of women among the participants whose performance ranks them higher than this decile. For instance for the forth decile (40), the top forty percent of participants, there are in the proportional pay treatment a proportion 0.417 of women and in the tournament a proportion of 0.236. Using 1000 iterations of bootstrap we find that the proportion of women is significantly lower in the mixed tournaments at 40, 60 and 80 (we only run even even deciles) at p=0.07, p= 0.029 and p=0.047 respectively.

The main result is therefore that the tournament results in a significant increase in the gender gap in mean performance as compared to proportional pay. This gender gap is due to the fact that the average performance of men is significantly increased by tournament incentives, whereas the average performance of women is not significantly affected. On a less aggregate level, the best women, those among the top twenty percent of participants, are as affected by tournament incentives as men. It is among the remaining participants that the men receive a relative higher boost in performance than the women. This leads to a lower representation of women among the second best group (the twentieth to fortiest best participants) and a higher representation of women among the worst quintile. Cumulatively this implies that if one would use the tournament in order to determine the ability of women, then one would only find 23.6 percent among the top forty percent, as opposed to 41.7 in the proportional pay.

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

IV. The Role of Gender Differences in Risk Aversion

One possible explanation for the significant increase in the gender-gap in mean performance when moving from proportional pay to tournaments is gender differences in risk-attitudes. A
tournament schedule differs in two ways from standard proportional pay. First, payment in the
tournament is uncertain, and second, it depends on the performance of the other participants.
There is a large and ongoing literature that examines whether women are more risk-averse than
men. For a review of the economics literature see Eckel and Grossman (2000), for a review of
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. In order to attribute
possible gender differences in tournaments to the competitiveness of the environment, we have
to control whether the introduction of uncertain (random) payments per se might have a
differential impact on the performance of males and females.

**Experiment 3: Random pay**

This treatment was similar to experiment 1, a part from the fact that participants were told
(before solving the mazes) that only one of them would be paid. This participant was chosen at
random and was paid 12 shekels for every maze that he or she had solved. The other
participants were paid only the participation fee.

If the gender gap in tournaments is driven by the uncertainty of the payment results from gender
differences in risk-aversion, then we expect to find that on average males would outperform
females in this experiment. This is so because if we assume that higher effort reduces the time
needed to solve a maze and marginal costs are strictly increasing in effort, then we predict that
participants who are risk-averse should solve fewer mazes than in the proportional pay
treatment. The reason is that the expected payment for each participant is actually 2 shekels for
each maze solved. Hence we expect that participants that are close to risk neutral would solve
as many mazes as in the proportional pay treatment (Experiment 1).

---

5 The reason is that only the expected monetary payment for each maze solved is the same as in the proportional pay
treatment (2 shekels), however the effort costs have to be borne by the participants even when they are not chosen to
be the one to be rewarded. Suppose performance (output) \( y \) is a function of effort \( e \) and ability \( a \). Then a participant in
the proportional pay chooses \( e^* \) such that \( U'(2y(a,e))2y'(a,e) = c'(a,e) \). The same participant chooses the
same optimal effort \( e^* \) in the random pay treatment where \( \frac{1}{6}U'(12y(a,e))12y'(a,e) = c'(a,e) \).
The comparison of males and females performance is presented in Figure 5.

<Insert Figure 5>

**Figure 5: Number of mazes solved under random payment.** On the horizontal axis the number of mazes solved; on the vertical axis the percentage of participants for the corresponding performance, for male and female respectively.

The mean performance of males is 11.83 and of females 10.33, the confidence intervals are overlapping, the \( p \)-value of the Mann-Whitney \( U \) test is 0.165. The mean difference is 1.5 as in the proportional pay treatment. Moreover, when we compare the results of experiment 1 to the results of this experiment, we find that it is not significant among males (\( p \)-value 0.6449) and females (\( p \)-value 0.6130).

6. **There is no significant gender difference in performance with random payment**

From observation 5 we conclude that participants exert the same effort level in the case of a random payment and a certain payment of the same expected value. We do not find any evidence that risk-aversion influences performance. In particular, our results do not indicate that gender differences toward uncertainty are a driving factor for the gender difference in mean performance in tournaments.

**V. Do Women compete?**

There is another possible rational explanation for the significant increase in the gender gap in mean performance in tournaments. Note that in the proportional pay treatment as well as in the random pay treatment we found that on average men solve 1.5 more mazes than women. Even though in our case these differences are not significant, it might be that they would be with a higher participant number. So suppose that men are actually somewhat more able than women. Why would men and women that might find themselves in the same performance quintile in the
proportional pay treatment, react differently to tournaments? If all women believe that men are somewhat more skilled in solving mazes and women take the gender of their competitors as a signal for their ability (and maybe even take gender as a signal for 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 that 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 more able 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, and this may lead to a lower expected equilibrium performance of the woman than of equally able man.

The psychology literature delivers an additional possible explanation: Women may think that they are worse than men in solving mazes in competitive environments, beyond possible actual gender differences. The psychology literature suggests that in some cases women internalize stereotypes about women being less good at solving mazes or performing worse in competitive environments. The effect is of course identical to the one 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 is especially harmed when competing against men, though women might not per se be unable to perform in competitive environments. Hence women might still be effective in competitive environments and be motivated by competitive incentives.

The results of our experiments so far allow however also for another class of explanations, namely that women may not want to or not be able to compete at all, independently of the nature of competitors. These explanations focus only on the ability and behavior of women in tournaments 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 than in the proportional pay treatment, they are less able than men in solving mazes. This is of course also a rational explanation for the gender gap.
A second explanation is that women might not be sensitive to incentive schemes. Note that up to now women performed the same under all incentive schemes. But we are not aware that existing papers on incentive schemes test for gender differences.

Third, even though women might be sensitive to incentive schemes per se, they might not like to compete. There are several possible reasons for such preferences. One is that women are not socialized to compete. Furthermore, 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.

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, equilibrium 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 experiment also tests for the possibility that men only compete and perform in tournaments better when there are women around (and it is easy to come up with an evolutionary story for that.)

**Experiment 4: Single sex groups**

In the following treatment we have five sessions of groups of six women only and five session 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 15 minutes to solve mazes of difficulty level 2. They are told (in advance) that after those 15 minutes the participant that solved the most mazes will be paid 12 Shekels for every maze that she or he had solved. The other participants in the group do not receive any payment additional to the show up fee.
In single sex tournaments men solve on average 14.3 mazes, as compared to 15 in mixed tournaments. The $p$-value is 0.5630 so the difference is not significant. The comparison of performance between men in single-sex tournaments as opposed to random pay or proportional pay is still significant, the $p$-values are 0.036 and 0.018 respectively.

7. **Men performance is the same in mixed and single sex tournaments**

Hence men are not significantly affected by the fact that they do not compete against women any more. They seem to respond to incentives and perform very well under competition. In the single sex tournaments women solve on average 12.6 mazes as opposed to 9.73 in the proportional pay treatment and 10.33 in the random pay treatment. These are significant differences, the $p$-values are 0.0148 and 0.0469 respectively. So women do react to tournament incentives, and it is not true that women are completely unwilling or unable to compete. The mean number of mazes solved by women in mixed tournaments is 10.8 versus 12.6 in single sex tournaments. 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. The different distributions are shown in Figure 6.

<Insert Figure 6>  

**Figure 6:** The performance of women in the mixed versus single-sex tournaments. On the horizontal axis the number of mazes solved; on the vertical axis the percentage of participants for the corresponding performance respectively.

Therefore we conclude that women show an increase in performance when competing against other women as opposed to performing in the proportional pay treatment, the random pay treatment, and competing in mixed tournaments.

7. **Women, on average, react strongly to tournament incentives in single sex groups**
However the fact that women are able to perform in competitive environments does not imply that they are as apt as men in competing. Therefore we compare, in Figure 7, men and women in single-sex tournaments.

<Insert Figure 7>

**Figure 7: Men versus women in single-sex tournaments.** On the horizontal axis the number of mazes solved; on the vertical axis the percentage of participants for the corresponding performance, respectively male and female.

On average men solve 14.3 mazes and women solve 12.6 mazes. The confidence intervals are overlapping, the $p$-value is 0.1346. The difference of 1.7 is not significant.

We have seen a significant gender difference in mean performance in mixed tournaments, where on average men solve 4.2 more mazes than women. The gender gap in mean performance in single-sex tournaments is 1.7. Next we want to confirm that there is a significant reduction in the gender gap in mean performance when moving from mixed to single-sex tournaments. In order to do that we run 1000 iterations of bootstrap on (Men mixed – Women mixed) - (Men single-sex – Women single sex) and find that a proportion 0.082 of the observations is less than or equal to zero. Hence we have a significant reduction in the gender gap in mean performance when moving from mixed to single-sex tournaments.

Furthermore we can compare the gender gap in mean performance in the single-sex tournament of 1.7 to the gender gap in mean performance in the random pay treatment (1.5) and in the proportional pay treatment (also 1.5). We have no significant difference in the gender gap in performance when comparing single-sex tournaments to the random pay treatment (the $p$-value equals 0.535), nor when comparing it to the proportional pay treatment ($p$-value equals 0.459).

Next we investigate how these average experiences translate into individual experiences.

We will again consider for each performance quintile the proportion of women in this quintile. For the single-sex tournaments, we pool the observations of men and women and take quintiles of the overall distribution. The results are presented in Figure 8.
We see that the proportion of women in the first and second quintile (the two best quintiles) in the single-sex tournaments closely follows those in the random pay treatment. Furthermore only in the mixed tournament treatment is the proportion of women in the worst quintile (quintile 5) higher than in all the other treatments.

Figure 9 presents the cumulative effects in which we consider for each decile the proportion of women among the participants whose performance are at least as good as those of participants in the respective decile. For the single-sex tournament we pool the observations of men and women. The figure shows that for the forth decile, 40, among the forty percent best participants, there are in the mixed tournament treatment $x$ percent women, as opposed to $y$ in the single-sex tournament. XXXX

VI. Feelings of competence

We have seen that one reason for the behavior of women in mixed and single-sex tournaments is that a men and a women in the mixed tournament face different competitors and hence have a different probability of being the most able participant in their group. To estimate these probability differences, we consider the performance of participants in the other treatments. Let us for instance assume that the performance in the piece rate treatment is a perfect measure of ability. Consider a men and a women that are both in the second best performance quintile and solve the same number of mazes, hence of the same ability. We can then calculate the chances for the men to be the most able participant in the group that comprises two other men and three women. We repeat this exercise for women. This reveals that for a man and a women that are both in the second best performance or ability quintile and have equal ability, the probability difference of being the most able participant in the tournament is between 1.3 to 5.1 percent.
This difference is of course due to the fact that men have a somewhat higher performance and
women face more male competitors than men. We can repeat this exercise with the
performance of participants in the random pay and single sex tournament and find similar gender
differences of winning the tournament. To provide an idea about the magnitude of those
differences, note that it is in general true that a woman in the second quintile that solves x mazes
has a higher chance of having the highest performance in a mixed gender group than a man that
solved x-1 mazes. A man and a woman whose performance is in the second quintile face a
difference of winning the tournament of 0.006 to 0.033 in the different treatments. If a
participant doesn’t know anything about the level of her performance, only the gender, than they
face differences of winning the simulated tournaments with the performance measures of the
experiment of 0.07 to 0.1. Furthermore these differences in probabilities of being the most able
participant may warrant different behavior of women in mixed and single sex tournaments of.
We do want to however explore whether another reason, one from the psychology literature,
would be an additional driving force for explaining our results. We want to measure whether
men and women feel equally competent or confident in their ability of solving mazes. Specifically
we measure whether men and women would make different choices when they can chose the
difficulty level in which they can perform the task and will be evaluated. In the next experiment
we have subjects choosing the difficulty level in which they will solve mazes. In section x we will
discuss the relevant psychology literature and applications.
In this section we wish to measure whether men and women feel equally competent or confident
in their ability of solving mazes. Specifically we measure whether men and women would make
different choices when they can choose the difficulty level in which they can perform the task
and will be evaluated. In the next experiment we have participants choosing the difficulty level in
which they will solve mazes.

**Experiment 5: Choice of the Level of Difficulty**

Participants were told that they have 15 minutes to solve mazes. At the beginning of the 15
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. 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 two. Figure 10 presents the distribution of choices.

<Insert Figure 10>

The mean of the choices is 3.4 for male and 2.6 for female. 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 it seems clear that men feel more competent than women. The psychology literature (described below), has identified that higher choices of difficulty are associated with higher feelings of competence. We put this last result on record:

9. Men choose higher difficulty levels than women in solving mazes

VII. The effect of tournaments
The mean performance 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 with random pay ($p$-value of 0.032 and 0.0046 for comparison with mixed and single sex tournaments respectively), and than with proportional pay ($p$-value 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
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.

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 nothing to prove: the outcome does not depend on any skill or talent of the subject. Even the effort is simply a monetary input, not a real psychological 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 from the scene the psychological aspects (self-confidence) that are at the center of our study.

In the mixed tournament the ten winners were 7 men, 2 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) 5 men and 5 women that win. Their average performance is 20.5. In order to test whether there is a significant difference in performance of the winners, we run 1,000 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 0.52).

Running single-sex as opposed to mixed tournaments is like a strong form of affirmative action. In our case in the single-sex tournaments we effectively sets quotas. The percentage of female winners by design reflects the representation of women in the contestant pool, namely 0.5. In our experiment, implementing quotas in the form of running single sex versus mixed tournaments comes at no cost in performance in any aspect: neither in terms of overall performance, nor in terms of performance of the winner.

---

6 The average payment per maze in 2,996 mixed and 3,044 in single sex tournaments.
VIII. Related Psychology Literature

Stereotypes

One explanation for the observation that women perform worse in the mixed tournaments relative to the single sex tournament is that women may internalize a stereotype about women being less able in solving mazes or performing in competitive environments. Many papers are devoted to documenting the ongoing prevalence and pervasiveness of stereotypes. For a recent overview see e.g. Fiske 1999. MacraeC.N., Stangor and Hewstone ,96, Eds, “Stereotypes and stereotyping”, New York, Guilford. Dovidio and Gaertner ,’86, “Prejudice, discrimination and racism“, San Diego, CA, Academic Press.

One consequence of the pervasiveness of these stereotypes is that they can be easily, even automatically, accessed, even by people who do not consciously endorse or agree with them (cf. Greenwald and Banaji). Here we are specifically concerned that these stereotypes and prejudices may influence stigmatized individuals. Classic discussions of stigmatization (Allport 1954) conceptualize the experience of being stigmatized as involving some internalization of the stigmatizing images and stereotypes of one’s group, an internalization that, in turn, can alter, even damage the individual personality. However the self-esteem of stigmatized individuals does not differ from other people’s self-esteem. Stigmatized individuals have various mechanisms in order to protect their self-esteem. Specifically they tend to psychologically disengage and disidentify from domains in which they are threatened by stereotypes, i.e. expecting and experiencing poor performance and failure. This disidentification from domains in which one’s group is disadvantaged or threatened by negative stereotypes has the potential to lead to systematic group differences in aspirations, skills and achievements, even when individual capabilities do not warrant these differences (Crocker and Major 1989).

However participants in our experiment are Technion undergraduate students, who receive a degree in engineering. Therefore we expected the least to find an internalization of a negative
stereotype. However the fact that these female students chose lower levels of difficulty in the last experiment (Choice of difficulty level) is an indicator of such an internalization.

**Stereotype Threat Theory**

A new theory in the psychology literature, put forward by Steele (Steele 1990 and 1997), hinges not on the fact that women believe that they are less able, hence internalizing any possible stereotype. Rather women only feel a higher pressure in a tournament where they compete against men, which will harm their performance through “chocking”. We give a more thorough comparison to the psychology literature later.

There are however many stigmatized individuals who seem to resist the stereotype and are identified with domains in which they are susceptible to prejudice, in which they may be threatened and harmed by the stereotype. These could be for instance the women in our participant pool who do get a degree in engineering. Claude Steele (1990, 1997) posits that even those stigmatized individuals are harmed by prejudices. They are vulnerable to what Claude Steele named “stereotype threat”, the threat that arises when one is in a situation of doing something for which a negative stereotype about one’s group applies. It is the threat that a negative stereotype about a group to which one belongs becomes self-relevant, usually as a plausible interpretation for something one is doing, for an experience one is having, or for a situation one is in, that has relevance to one’s self-definition. It means that anything one does or any of one’s features that conform to it make the stereotype more plausible as a self characterization in the eyes of others and also in one’s own eyes. Furthermore, it has been shown that test anxiety and choking disrupt performance through a variety of mediating mechanisms: interfering anxiety, reticence to respond, distracting thoughts, self-consciousness and so on. The stereotype threat simply presents an additional degree of self-threat, one that is not borne by people not stereotyped this way. Therefore stigmatized individuals perform less well in situations that pose stereotype threat. Stereotype threat theory consists of two necessary parts. First the individual has to identify with the task at hand, care about her or his own
performance respectively. Second the situation has to pose a threat to the individual of conforming to a stereotype, as described above.

The seminal paper by Steele and Aronson (1995) presents a typical experiment of its kind. Black participants had to take a difficult verbal test. The 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 under-performed in relation to Whites in the ability-diagnostic condition but not in the non-diagnostic 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 also positive stereotypes may be a burden to performance. Specifically they argue that in mathematics where the stereotype is that men are good and women are bad, men might be concerned that they might not excel, whereas women are more concerned that they might perform poorly. In as far as excess concerns are disruptive, women should performance less well on a mathematics test that is described as indicative of weak abilities than on the same mathematics test described as indicative of exceptional abilities, and men should show the reverse pattern. These are indeed the experimental results. Furthermore giving participants an external handicap, i.e. an external excuse for failure – the computer crashes, the participants can not take any practice problems prior to the test – improves their performance only in the condition in which they experienced excess concern before.

In order to apply stereotype threat theory, we can either assume that technion undergraduate women care a lot about their ability of solving mazes. This seems rather implausible. Also it would not explain why women are more threatened by the stereotype when competing against men as opposed to being paid by piece rate. Further, we can assume that they care about the fact that they are able to perform well in competitive environments, the (possible) stereotype
being that women are less competitive, perform less well in competitive environments. The mixed tournament is of course a competitive situation, and then stereotype threat theory posits that the performance of women should suffer. Implicitly the assumption would be that women do not feel threatened by such a stereotype when only competing against women. However, one could see the tournament certainly as a situation indicative of exceptional abilities in solving mazes or exceptional abilities of performing in competitive environments. When there is a stereotype about women being less able in these domains, there is one that men are more able in these domains. Then, according to the findings of Brown and Josephs (1999) that also positive stereotypes may be a burden to performance, we might expect that men should be more concerned about performing well, and this additional stress should dampen their performance in mixed tournaments, without any specific negative impact on the performance of women.

Our experiments differ in many ways from experiments in the psychological literature. First we introduce incentives for our participants to perform the task. Second, we do not manipulate the presentation and description of the task in order to generate different behavior. Rather we only change the incentive scheme, leaving the description of the task constant. Therefore we can study whether the competitiveness of the environment has an effect on performance per se. The comparison of mixed to single-sex tournaments is a new design that lets us differentiate between the impact of the incentive scheme per se and the impact of the nature of the competitors.

IX. Discussion and Conclusion

We conduct controlled experiments to test the hypothesis that men and women react differently to competitive incentive schemes. We found that when participants are paid proportional to their performance there is no significant gender difference in performance. Using tournament incentives, however, results in a significant increase relative to the benchmark in performance in male participants, but not in 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. Several reasons may account for those results. First let us note that in all treatments on average men (un-significantly) slightly outperform women. This can lead to gender differences in optimal performance in mixed tournaments. The reason is that men only face 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 of solving mazes.

The results of our experiments 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 non competitive scheme, such as proportional pay (or the random pay treatment), 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. Naturally single sex tournaments present a strong form of affirmative action. The proportion of women among the winner of the tournaments reflects the gender composition of the participant pool (50 percent), whereas in mixed tournaments, only around twenty percent of the winners are females.
X. References


Goldin, Claudia, BOOK


XI. Appendix

Data

In the following two tables we record the raw data, organized according to the gender of the participants. It is important to remember that the participants were different in different treatments. Also participants in the treatment choice, for instance, were of both genders (half of them were male and the other half female).

Table 1.a: Data for male participants

<table>
<thead>
<tr>
<th>choice</th>
<th>random</th>
<th>tournament</th>
<th>proportional</th>
<th>single-sex</th>
<th>participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>22</td>
<td>22</td>
<td>19</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>21</td>
<td>18</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>21</td>
<td>18</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>21</td>
<td>16</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>19</td>
<td>16</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>19</td>
<td>16</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>18</td>
<td>15</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>17</td>
<td>15</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>17</td>
<td>15</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>17</td>
<td>14</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>17</td>
<td>13</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>17</td>
<td>12</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>16</td>
<td>11</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>16</td>
<td>11</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>16</td>
<td>11</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>15</td>
<td>10</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>15</td>
<td>9</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>15</td>
<td>9</td>
<td>13</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>15</td>
<td>9</td>
<td>13</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>13</td>
<td>8</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>13</td>
<td>8</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>13</td>
<td>8</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>12</td>
<td>8</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>11</td>
<td>8</td>
<td>11</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>30</td>
</tr>
</tbody>
</table>
Table 1.b: Data, for female participants

<table>
<thead>
<tr>
<th>choice</th>
<th>random</th>
<th>tournament</th>
<th>proportional</th>
<th>single-sex</th>
<th>participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>23</td>
<td>24</td>
<td>18</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>20</td>
<td>17</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>18</td>
<td>15</td>
<td>23</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>17</td>
<td>14</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>15</td>
<td>14</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>15</td>
<td>12</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>14</td>
<td>12</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>14</td>
<td>12</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>14</td>
<td>11</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>13</td>
<td>11</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>9</td>
<td>9</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>26</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>28</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>29</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>30</td>
</tr>
</tbody>
</table>