Under Pressure: Gender Differences in Output Quality and Quantity under Competition and Time Constraints*

Olga Shurchkov†
Wellesley College

Forthcoming:
The Journal of the European Economic Association

Abstract

Gender gaps in the workplace are widespread. One explanation for gender inequality stems from the effects of the interaction between competition and two pressure sources, namely, task stereotypes and time constraints. This study uses a laboratory experiment to find that the gender gap in performance under competition and preferences for competition can be partly explained by the differential responses of men and women to the above pressures. In particular, while women underperform the men in a high-pressure math-based tournament, women greatly increase their performance levels and their willingness to compete in a low-pressure verbal environment, such that they actually surpass the men. This effect appears largely due to the fact that extra time in a verbal competition improves the quality of women's work, reducing their mistake share. On the other hand, men use this extra time to increase only the quantity of work, which results in a greater relative number of mistakes. A labor market study suggests that the nature of the job and stress levels seem to be correlated with the gender gap in the labor market in a manner consistent with the results of my experiment.

Keywords: Gender Differences, Stereotypes, Competition, Effects of Time Pressure

JEL Classification: C9, J16, J71

*I am grateful to the editor, Stefano DellaVigna, and anonymous referees for their detailed feedback. I would like to thank George-Marios Angeletos, Kristin Butcher, Rachael Croson, Ernst Fehr, Robert Gazzale, Uri Gneezy, Robin McKnight, Gauri Kartini Shastry, Lise Vesterlund, Akila Weerapana, and the participants of the 2008 International ESA Conference, the 24th Congress of the EEA, and the Wellesley College working seminar series for insightful comments. I acknowledge the generous financial support provided by the Shultz Fund (MIT Economics Department) and the Wellesley College Faculty Award. I would also like to thank the Harvard Business School Computer Lab for Experimental Research for providing me with the laboratory space and the access to the subject pool for this study. Finally, I thank the FEEM Award Selection Committee for recognizing this research with a prize at the EEA. All remaining errors are my own.

†Department of Economics, Wellesley College, 106 Central St., Wellesley, MA 02481.
e-mail: oshurchk@wellesley.edu.

1 Introduction

The study of gender differences has a long history in the field of labor economics. Despite a recent policy push toward the equalization of men and women in the workplace and in society, considerable inequality persists, especially when it comes to competition for high profile jobs.[1]

The possible explanations put forth in the literature can be sorted into three categories. The first explanation relies on gender differences in skills and preferences that lead to occupational self-selection (Polachek 1981; Macpherson and Hirsch 1995). The second explanation points to discrimination in the workplace which results in differential treatment of men and women of identical abilities and preferences (Becker 1957; Black and Strahan 2001).

In this paper, I explore the third explanation: men and women of the same ability may perform differently under competition and may exhibit different preferences for competition. In particular, previous literature has established that women perform poorly relative to men of identical skill level once competition is involved (Gneezy, Niederle, and Rustichini (hereafter, GNR) 2003), and that women "shy away from competition" (NV 2007).[2] The first goal of my study is to design an experiment that replicates the results of these seminal papers under the same conditions (a high time pressure mathematical environment that is perceived to favor men[3]). While my task – a numbers-in-numbers puzzle – mimics the mathematical nature of the previously used games, the novel feature is that it admits more than just one correct solution and provides a meaningful measure of relative quality of output in addition to quantity. The results confirm that, in mixed-gender groups, women underperform the men under competition, even though

[1]Bertrand and Hallock (2001) document this fact by gathering data on the five highest-paid executives of a large group of U.S. firms over the period of 1992–1997, where they find that only 2.5 percent of the executives in the sample are women. A similar under-representation of women is found among CEOs at Fortune 500 companies (CNNMoney 2006), tenured faculty at leading research institutions (MIT 1999), and top surgeons in New York City (New York Magazine 2010).

[2]For studies that argue that gender inequality persists due to the innate inability of women to compete, see also Baron-Cohen (2003), Lawrence (2006), and the citations in Barres (2006). Babcock and Laschever (2003) provide yet another possible explanation for gender differences in income and status that stem from the suppressed relative willingness and ability of women to engage in negotiations.

[3]Society generally perceives men to be better than women at following directions and reading maps, while women supposedly tend to follow landmarks when driving (Rahman et al. 2005). When it comes to solving mazes, men are thought to be overwhelmingly superior to women (Pease and Pease 2000, p. 107). Similarly, men are perceived to have higher math abilities than women, while women are perceived to have superior verbal skills. In particular, Pajares and Valiante (2001) note that differences in achievement of middle school students lie in the stereotyped beliefs about gender differences rather than gender itself. Girls report stronger motivation and confidence in writing and receive higher grades in language arts. Boys report stronger performance-approach goals (Pajares and Valiante 2001).

both genders perform equally well in a noncompetitive (piece-rate) treatment (consistent with GNR 2003). Moreover, relative to men, women are considerably less likely to choose to compete in this environment (consistent with NV (hereafter, NV) 2007).

My second goal is to document the effects of changing the environment to be more woman-oriented by relaxing both sources of pressure: the time constraint[4] and the mathematical nature of the task.[5] To this end, I run the experiment with a verbal task – a word-in-a-word puzzle – that still allows me to capture the quality and quantity aspects of performance.[6] In the verbal environment, competition no longer hinders women's performance relative to men; in fact, women significantly outperform the men in the low time pressure tournament. Moreover, in the low pressure verbal setting, women are actually more likely than men to choose competition. To my knowledge, no past study has been able to document such a reversal in preferences for competition.

Finally, I seek to understand the reasons behind the gender differences in performance under competition and preferences for competition under the various conditions of my experiment. Why do women underperform the men in the high time pressure math tournament, and why do women shy away from this kind of game? On the other hand, why do women outperform the men in the low time pressure verbal tournament, and why do women increase their willingness to compete in this type of environment? The explanation comes partly from the previously unexplored quality dimension of the task. In the math tournament, women tend to increase the proportion of mistakes they make under time pressure. In the verbal tournament with extra time, it is the men who tend to make a larger number of mistakes relative to the total possible number of points they can earn. This evidence is consistent with certain theories put forth by evolutionary biologists that suggest that men, as "hunters," tend to have lower attention spans. On the other hand, women, as "gatherers," tend to pay attention to detail and can stay focused on a singular task for a prolonged period of time. For example, men often flick through TV

[4]Time pressure has been recently explored in the context of a field experiment conducted by Paserman (2007).

[5]The math task is perceived to be favorable to men and is therefore associated with a stereotype threat made salient in a mixed-gender setting. The idea of a stereotype threat first appeared in the field of psychology. It describes the fear that certain behavior would confirm an existing stereotype of a group with which one identifies (Steele and Aronson 1995; Spencer, Steele, and Quinn 1999).

[6]Gneezy and Rustichini (2005) use anagram puzzles to see which gender is more likely to self-select into a tournament and obtain results similar to NV (2007). Although Gneezy and Rustichini (2005) use a verbal task, it is fundamentally different from the verbal environment described in my paper. An anagram, like a maze puzzle, has only one correct solution and is more in line with previous literature where quantity is all that matters, since an incorrectly solved puzzle is not penalized. By virtue of the task, the authors' approach also only allows for the investigation of high-time-pressure treatments. My findings validate their results and expand the study to address the previously unexplored questions regarding quality of output and time-pressure effects on competitive outcomes.

channels and do not have the patience to watch commercials, while women are not as averse to sitting through the boring breaks (Sullivan 2001, Pease and Pease 2000).

Since my experiment is highly stylized, the results need to be interpreted and applied to the real world with a degree of caution. Firstly, I emphasize the importance of quality, in addition to quantity, as a measure of performance. Secondly, the choice of experimental tasks in my study reflects the fact that most jobs require not only mathematical skills, but also verbal abilities. Thus, verbal tasks should not be ignored when examining the effects of competition on performance and preferences. Finally, jobs can be ranked by the associated level of stress and time pressure. Examples of high-pressure jobs that put an emphasis on mathematical skills include financial analysts and managers, while actuaries and accountants fall into a lower pressure category. Examples of high-pressure jobs with a verbal emphasis include reporters and news analysts, while low-pressure jobs in the verbal category may include novelists and poets. A simple labor market analysis suggests that the gender gap in earnings is present in the mathematical jobs, and is largest in the jobs that involve intense time pressure. However, the gap is diminished or disappears altogether for jobs with a verbal emphasis. The nature of the job and the stress level seem to be correlated with the gender gap in the labor market in a manner consistent with the results of my experiment.

In an educational setting, the findings suggest that shorter exams (or extra time), especially in hard sciences and math, might reduce the performance gap between men and women and in the long-run increase female participation in the study of these subjects. In order to draw policy implications of the results for the alleviation of the gender gap in the labor market, further empirical investigation needs to be undertaken. This paper suggests that gender differences in quantity and quality of output matter for overall performance. However, the effects of relaxing time constraints on total productivity of both genders remain to be shown.[7]

The rest of the paper is organized as follows. Section 2 describes the treatments, tasks, and procedures. Section 3 presents the basic results. Section 4 provides possible explanations for the findings. Section 5 draws parallels between the experimental findings and labor market observations of the gender gap in earnings and the share of women across various occupation types. Section 6 concludes and discusses potential future research.

[7] Moreover, time pressure cannot and probably should not be fully eliminated from most competitive jobs. However, industries such as investment banking almost pride themselves on their pressure-cooker mentality, with the vice-presidents putting unrealistic deadlines on the associates, and associates piling on the work onto the analysts. (See, for example, Rolfe and Troob, 2000.) This kind of culture is partly self-fulfilling as managers put the pressure on the lower-level employees because they had been treated in a similar fashion. There may be social benefits from dampening this "vicious cycle."

2 Overview of the Experiment

The goal of this study is to ask whether gender differences in performance under competition and in preference for competition persist once we vary the aspects of the work environment and the task at hand. To this end, I conduct a laboratory experiment that involves groups of two men and two women solving mathematical and verbal tasks. Performance differences are captured by two metrics. First, I compare men and women according to their *scores* in a given round. Second, I isolate the quality dimension of performance by observing *mistakes* defined as the number of points lost due to invalid answers. Preferences for competition are elicited through the subjects' *choices* of payment schemes. The degree of confidence is measured by a self-reported performance *rank guess* (out of four group members). A comparison between a piece-rate (non-competitive) payment scheme and a tournament (competitive) payment scheme allows me to capture the effects of competition. Finally, the effects of time pressure are gauged by varying the length of time allowed for completing the tasks. Note that the comparison is within-subjects in terms of the varying degrees of time pressure and competition. However, the comparison is between-subjects when it comes to the verbal and mathematical environments.

From session to session, I vary the order of treatments as to control for learning effects (that may increase performance over time) and tiredness/boredom effects (that may decrease performance over time). The various permutations of treatments are presented in Table 1.

Table 1. Treatment Summary

	Rounds							Total # of
Session	0	1	2	3	4	5	6	Subjects
1 & 3 (V)	P	H, PR	H, T	H, C	L, PR	L, T	L, C	36
2 & 5 (V)	P	H, T	H, PR	H, C	L, T	L, PR	L, C	48
4 (V)	P	L, PR	L, T	L, C	H, PR	H, T	H, C	24
6 (V)	P	L, T	L, PR	L, C	H, T	H, PR	H, C	20
7 & 8 (M)	P	H, PR	H, T	H, C	L, PR	L, T	L, C	44
9 & 10 (M)	P	H, T	H, PR	H, C	L, T	L, PR	L, C	40

Notes: Each cell corresponds to a particular round within a given session. Treatment abbreviations: V – verbal task; M – mathematical task; P – practice; PR – piece-rate; T – tournament; C – choice; H – high time pressure (2 min); and L – low time pressure (10 min).

The following sections provide detailed explanations of the tasks and treatments used in the experiment.

2.1 The Tasks

2.1.1 Verbal Puzzles

For my verbal task, I chose a Word-in-a-Word puzzle where players must form sub-words from the letters of a larger puzzle word.[8] In this task, performance can be maximized by increasing the number of sub-words entered (the quantity dimension) and by minimizing the entry of erroneous sub-words (the quality dimension).

At the beginning of the experiment, each participant was given one minute to solve a practice Word-in-a-Word puzzle.[9] As an example, consider the puzzle word *persuasively.* The table below gives a possible set of entries, as well as the scores for each entered sub-word based on the scoring rules (see Appendix A[10] for details).

Valid Solutions:	**Score:**	**Invalid Solutions:**	**Score:**
persuasive	+7	live (repeated)	0
live	+1	as (too short)	-1
lives	+2	lyver (misspelled)	-2
Total Score:			10 − 3 = **7**

My within-subjects design requires me to change the puzzle from round to round which may prove problematic if the difficulty of the task varies dramatically. If the words were not similar, the results would not be directly comparable across treatments.[11] My main strategy is to choose puzzles with a consistent number of correct sub-words and the maximum number of points. In particular, the range for the number of sub-words is 77-85, and the range for the maximum possible number of points is 132-137. I also check how difficult it is to find the sub-words in any given puzzle. In particular, I count the

[8] All puzzles were computerized using the programming languages Python and Java. The Word-in-a-Word puzzles were based on the games provided by the website www.wordplays.com.

[9] The score in the practice round did not count toward the final payment. For practice, I chose puzzle word *infuscate*. The practice score is used as a baseline measure of ability in solving word puzzles.

[10] Appendices A-D can be found in the online Supplementary Material to this paper.

[11] Previous studies face a similar problem, although with mathematical tasks the difficulty level is easier to control. For example, although GNR (2003) restrict their pool of mazes to a certain level of difficulty, mazes still vary within that difficulty level. Similarly, Ariely et al. (2009) and Gneezy and Rustichini (2005) use anagram puzzles that can vary in difficulty. In my case, because every word is different, it is difficult to find truly identical tasks. Note that the novelty of my experiment comes from the quality dimension of the task that allows subjects to exert effort and creativity by finding multiple solutions within one puzzle. Thus, it is crucial that the same puzzle is always assigned to a given treatment.

number of permutations of letters needed to arrive at any one sub-word and confirm that the puzzle words are relatively close according to this metric.

2.1.2 Math Puzzles

The math puzzle was selected based on the following two criteria. First, I looked for a task that was comparable to the tasks used in the existing literature on competition – in particular, puzzles where subjects add up numbers to gain points (NV, 2007). Second, I wanted the math task to have a quantity and a quality dimension in order to be comparable to my verbal task. The resulting math task is a Numbers-in-Numbers puzzle where players must find combinations of numbers ("solutions") within a given sequence that add up to a specific target number.[12]

At the beginning of the experiment, each participant was given two minutes to solve a practice math puzzle.[13] For example, consider a puzzle sequence is **1034582614** with a target number of **117**. The table below lists a possible set of entries, as well as the scores for each entered sum based on the scoring rules (see Appendix A for details).

Valid Solutions:	**Score:**	**Invalid Solutions:**	**Score:**
103+14	+2	5+112 (5 is a 1-digit number)	-1
54+63	+1	101+16 (1 appears only twice in sequence)	-2
14+43+60	+4	03+114 (0 cannot be used as a placeholder)	-1
Total Score:			$7-4=\mathbf{3}$

While the verbal puzzles were taken directly from a well-known game website, the math puzzles had to be generated by the experimenter. The sequence generation algorithm is explained in Appendix B. In order to preserve consistency across tasks, I picked sequences that are similar along several dimensions. In particular, each sequence has a maximum

[12] Note that the effects of lowering the time pressure may be different in the verbal and in the math tasks. While giving more time to subjects in the verbal task does not necessarily make the task a lot easier, giving more time in the math task may allow the subjects to work out an algorithm that generates solutions more quickly. Because of this caveat, I do not perform the analysis pooling across the verbal and the math treatments. However, my tasks are more comparable along the quality dimension (consider the difference between word-in-a-word puzzles and standard math tasks of adding numbers, for instance). Since exploring the quality aspects of performance is a major focus of the paper, I find my tasks to be appropriate.

[13] The score in the practice round did not count toward the final payment. It is used as a baseline measure of ability in solving math puzzles.

score in the range of 455-461 points, the total number of solutions in the range of 122-123, and the total number of 2-factor solutions in the range of 14-15. Each sequence was randomly assigned to a particular round.

2.2 The Treatments

Each verbal and each math session consists of 7 rounds: 1 practice round (not for pay); 3 high time pressure rounds; and 3 low time pressure rounds. In the latter, the subjects are given considerably more time (10 minutes compared to 2 minutes) to work on puzzles of equal difficulty. Subjects also have the opportunity to finish the low time pressure rounds early. After each of the 6 for-pay rounds, but before the individual scores are revealed, subjects guess their rank out of 4. Each correct guess pays one dollar, but an incorrect guess does not subtract any money from the total.

Within each time pressure treatment, there are three rounds that test the subjects' performance under and preferences for competition.

2.2.1 Noncompetitive Treatments ("Piece-Rate")

The Piece-rate treatment establishes the baseline performance level for the rest of the experiment. Each subject receives X cents for every point earned in this round. No winner is announced, and everyone earns income according to one's own performance.[14]

The puzzle words for the piece-rate verbal treatments are: *carriageway* (high time pressure) and *allopathy* (low time pressure). The puzzle sequences for the piece-rate math treatments are 647029590696014 with target number 165 (high time pressure) and 903538359266169 with target number 182 (low time pressure).

2.2.2 Competitive Treatments ("Tournament")

The Tournament treatment tests whether there is a gender-specific effect of competition on performance. Each subject's total score is compared to the scores of the other three group members. The person with the highest score ("the winner") receives $4X$ cents for

[14] $X = 10$ cents in all verbal sessions and $X = 5$ cents in all math sessions.

every point earned. The other three members of the group receive 0 points. In case of a tie, the winner is determined randomly out of the top performers.[15]

The puzzle words for the competitive verbal treatments are: *ordination* (high time pressure) and *equitable* (low time pressure). The puzzle sequences for the competitive math treatments are 845196336864734 with target number 197 (high time pressure) and 674639419829848 with target number 193 (low time pressure).

2.2.3 Choice Treatments ("Choice")

The Choice treatment elicits the subjects' preferences for competition. Subjects first choose which of the two previous payment schemes they prefer to apply to their subsequent performance. If a subject chooses the piece-rate, she receives X cents for every point she earns in this round. If a subject chooses the tournament, her performance is evaluated relative to the performance of the other three group members in the Tournament round. If the score is higher than the top Tournament score of the other three group members, then the subject receives $4X$ cents for every point earned. The subject receives no earnings in the Choice round if she selects the tournament and fails to get a higher score than the other three group members in the Tournament round.[16]

The puzzle words for the choice verbal treatments are: *memorable* (high time pressure) and *reachably* (low time pressure). The puzzle sequences for the choice math treatments are 497220002195953 with target number 145 (high time pressure) and 436771974115604 with target number 135 (low time pressure).

[15] Note that the expected payoff in this competitive treatment is set to be identical to the expected payoff in the noncompetitive treatment. (The expected payoff is equal to $\frac{1}{4}(\mathrm{Y}_1+\mathrm{Y}_2+\mathrm{Y}_3+\mathrm{Y}_4)$, where Y is the number of points of each group member i.) However, payment is now uncertain. In the pilot experiment (Shurchkov 2009), I disentangle the effects of competition and uncertainty on gender differences in performance by conducting a "random winner" treatment. In this treatment, payment is uncertain, yet independent of the performance of others. Since the results of this treatment do not differ significantly from the competitive treatment, I choose not to repeat this treatment in my main study.

[16] I follow the same scoring procedure as NV (2007) in the choice round. One can think of this as competing against other participants who already performed. One of the advantages is that the performance of a person competing in a tournament is only compared to people who also competed in a tournament. Another advantage is that this scoring procedure keeps the number of competitors at 3 people, regardless of choices. Essentially, in the Choice treatment, participants face an individual decision problem, which depends only on their own ability to beat the Tournament score of others and their preference for tournament relative to the piece-rate payment scheme.

2.3 The Procedure

The experiment was conducted at the Computer Lab for Experimental Research (CLER) at Harvard Business School (HBS) and took place in early 2009. The verbal sessions consisted of a total of 128 people (6 sessions; 27 groups of two men and two women and five groups of all men). The math sessions consisted of a total of 84 people (4 sessions; 21 groups of two men and two women and 3 groups of all women).[17]

Participants were seated in rows and informed that they were grouped with the other people in their row. Even though gender was not emphasized at any point during the study and explicit communication was not allowed, subjects could clearly see the gender composition of their group.

Paper copies of the general instructions were distributed to the participants prior to the beginning of the experiment. Computerized instructions were presented to each participant between rounds explaining the changes to the payment scheme from round to round. Participants were encouraged to ask questions in private if they did not understand these instructions. The subjects had to wait for everyone to finish reading the instructions before they could proceed to the next round.

The time (two or ten minutes, depending on the treatment) ran out automatically. Once the time ran out, the subjects could see their score (in points) and the maximum possible score in a given puzzle. The subjects were *not* given the information about the average performance of their group, their relative ranking, or the genders of those ranking above and below them. Since the program recorded the scores automatically, the subjects did not need to keep track of their winnings from round to round.[18]

[17] CLER recruits subjects via an online registration procedure. Subjects first register for the CLER subject pool. Then, they sign up for studies of their choosing. Most subjects are students at Harvard University (undergraduates and graduates), although students from other Boston-area universities, such as MIT and Boston University, also participate. At any point, a subject can remove him- or herself from the study for any reason.

[18] GNR (2003) leave it to the subjects to record the number of correctly solved mazes, because they use the internet for the experiment. Although the authors monitored the subjects, in order to ensure that they did not lie about their performance, it is possible that some of them were left unwatched at some points during the task. If men tend to lie and overestimate their performance more in the tournament than in the piece-rate treatment, then this would potentially distort the results (see Dreber and Johannesson 2008). That is the reason why I did not use online word puzzles for this experiment, but rather programmed my own version of the game.

At the end of the experiment, each participant filled out a brief questionnaire.[19] The questionnaire asked the subjects demographic questions and inquired about their strategies and beliefs throughout the experiment. In particular, the subjects reported their *age*, their proficiency in the English *language*, and their field of study (*concentration)*. I also measured *gender confidence* by asking each subject to report his or her belief about who would be better in these tasks on average, men or women. (Descriptive statistics are provided in Appendix C, Table C.1.)

At the end, each subject was paid in cash a $10 show-up fee and his or her earnings over the course of the session. Final income was first given in points and then converted to US $ at the different rates according to the payment scheme detailed above. Including the show-up fee, average income was $23.02 (verbal) and $21.68 (math) and maximum income was $61.30 (verbal) and $76.50 (math). The approximate average duration of all sessions was 1 hour 15 minutes.

3 Basic Results: Performance and Preferences

In this section, I examine whether men and women differ in their performance under competition and in their preferences for competition.[20] I begin with a high time pressure environment and ask whether changing the task from mathematical to verbal has any effect on the outcomes. Next, I relax the time constraint and once again focus on the effects of changing the task.

3.1 Math Task under High Time Pressure

A high time pressure mathematical environment contains two sources of pressure for women: the time constraint and the task-stereotype pressure.[21]

[19] Copies of the informed consent forms are available upon request. Full copies of the instructions and the questionnaire can be found in Appendix A.

[20] The results are based on mixed-gender group behavior only since my single-gender sample is too small. See Shurchkov (2009) for a discussion of men-only groups.

[21] The task stereotype pressure is confirmed by the questionnaire responses. Both genders perceive men to be more proficient at the math puzzles: only 30.6 percent of male and female subjects reported thinking that women would be, on average, better than men. (See Appendix C, Table C.1.) Thus, it is

Result 1. *Under high time pressure with a math task: (a) Men and women do not differ in terms of their scores in the piece-rate treatment, but in the tournament, men significantly outperform the women; (b) Men are significantly more likely to self-select into a tournament than women.*

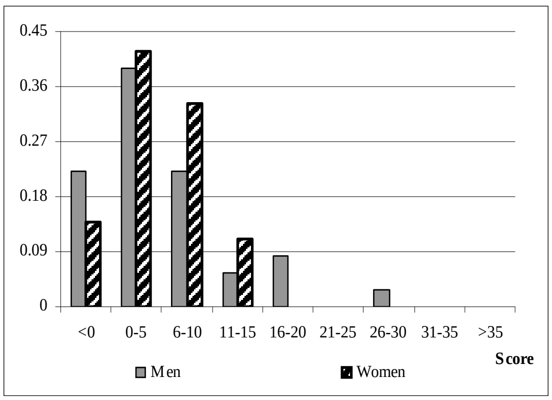

Figure 1a. Distribution of Math Scores by Gender Under Piece-Rate, High Time Pressure

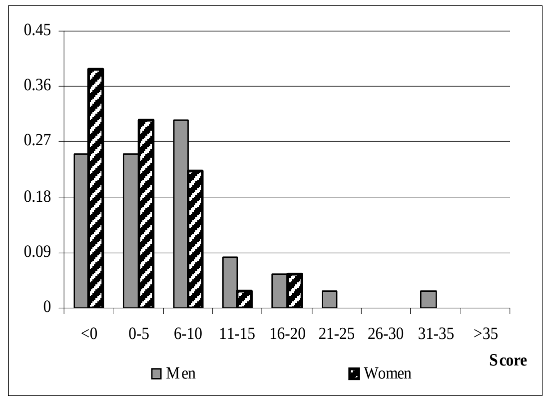

Figure 1b. Distribution of Math Scores by Gender Under Tournament, High Time Pressure

Support for Result 1(a) comes from Figures 1a and 1b which show the distributions of scores achieved by men and women in the piece-rate and in the tournament treatments, respectively. The height of the bars in both figures corresponds to the share of male (grey bars) and female (black bars) participants who achieved the score in a given range. With the mean scores of 5.17 (men) and 5.11 (women), neither gender displays a significant performance advantage in the piece-rate condition (p-value of 0.97).[22] This suggests that the perception that men are inherently better at this math task than women is not supported by the data. Figure 1b shows a significant increase in the negative scores for women in the tournament. With the mean scores of 6.31 and 2.39 for men and women, respectively, women now perform significantly worse than men (p-value of 0.03).[23] Note

fair to say that, in my experiment, there is a stereotype threat against women in the math sessions.

[22] I report simple two-sided t-test p-values. See Table C.1 in Appendix C, for the p-values based on the Mann-Whitney U test that compares distributions. The two tests produce similar results.

[23] Within-subject analysis confirms these basic findings. See Appendix C, Table C.6 for OLS regressions of individual math score as a function of gender, treatment effects and various controls. Note that the results are robust to the inclusion of treatment order conrols and individual subject characterstics. I also confirm that the results are robust to using only the data for the first three rounds of the experiment

that this finding confirms the GNR (2003) result that men significantly outperform the women in a tournament with a stereotypically "male" task.[24]

Result 1(b) derives from the observation that, on average, 44 percent of men and only 19 percent of women self-select into a tournament in the high-pressure math environment and is confirmed by the regressions in Table 2. The difference in means is statistically significant at the 5 percent confidence level (p-value of 0.02). Note that male scores are located in the right tail of the distributions in Figures 1a and 1b.[25] Thus, it is important to condition the probability of entry into the tournament on past performance. Table 2 reports the coefficients from the probit regressions of choice of compensation scheme as a function of the female dummy and various controls. Conditional on past performance (proxied by the score in the preceding tournament round and the difference between the piece-rate and the tournament score), a large and significant gender gap in tournament entry remains (36 percent in specification 1). These results support the NV (2007) findings that women tend to "shy away from competition," at least when faced with a stereotypically "male" task under high time pressure. In order to see whether confidence explains the gender gap in tournament entry, I include the tournament rank guess in specification 2. Participants who are confident (guess a higher rank) are significantly more likely to enter a tournament.[26]

(Appendix D, Table D.2). For more on the within-subject variation, see Figure C.1 in Appendix C.

[24]Several studies (for example, NV (2007), Niederle, Segal and Vesterlund (2008), Booth and Nolen (2009)) are unable to replicate the GNR (2003) result.

[25]I confirm that, out of 1000 simulated sessions with randomly re-arranged mixed-gender groups, on average only 34% of the winners are women, so men win significantly more than 50% of the time (significant at the 1% confidence level). In fact, in only 10 out of 1000 high pressure math sessions, were there more female winners than male winners.

[26]Note that I run specification 2 with rank guesses of 1, 2, and 3 only in order to be able to compare my results with those of NV (2007). Table C.5 in Appendix C provides probit estimates of tournament entry decision for the unrestricted sample of rank guesses across all sessions. The main results are robust to this change. The distribution of math tournament rank guess by gender and time treatment, as well as ordered probit regressions of guessed rank as a function of the female dummy and performance in all treatments are provided in Appendix C, Tables C.2 and C.4, respectively.

Table 2. Probit of Tournament Entry Decision, High Time Pressure Math Sessions

Independent Variables:	Dependent Variable: Choice (Tournament =1)	
	(1)	(2)
Female	-0.36*** (0.00)	-0.38*** (0.01)
Tournament Score	0.03*** (0.00)	0.04** (0.02)
Tournament − Piece-Rate Score	-0.02* (0.06)	-0.03*** (0.01)
Guessed Tournament Rank		-0.36** (0.03)
No. Observations	72	49

Notes: Standard errors clustered at the group level (p-value in parentheses); marginal effects. Other controls include order of tournament and time, age, major (1 = science), native language (1 = English), and reported gender stereotype (1 if women perceived to be better). Significance levels: *10%, **5%,***1%. Guesses of 4 are eliminated in Specification (2).

I have confirmed the previous findings that competition under high time pressure in a mathematical environment hurts women relative to men. Next, I investigate whether the gender gap persists once I relax each of the pressures in isolation.

3.2 Verbal Task under High Time Pressure

The next result concerns the effects of changing the task to be perceived as more "woman-friendly"[27], while maintaining the relatively high time pressure environment.

Result 2. *Under high time pressure with a verbal task: (a) Men and women do not differ significantly in terms of their scores in either the piece-rate or the tournament treatment; (b) Women and men no longer differ in terms of choice of compensation scheme.*

Support for Result 2(a) comes from Figures 2a and 2b which compare the distributions of verbal scores achieved by men and women in the piece-rate and in the tournament treatments, respectively. In the piece-rate treatment, men and women achieve the scores of 12.91 and 14.19, respectively (p-value of 0.30). In the tournament, women (with a mean score of 11.83) actually slightly outperform the men (with a mean score of 9.76), although the difference is only significant at the 10 percent confidence level (p-value of 0.07).

[27]Both genders perceive women to be more proficient at the verbal puzzles: 86.2 percent of female subjects and 60.6 percent of male subjects reported thinking that women would be, on average, better than men (See Appendix C, Table C.1). Thus, in my experiment, there is a stereotype threat against women in the math sessions but against men in the verbal sessions.

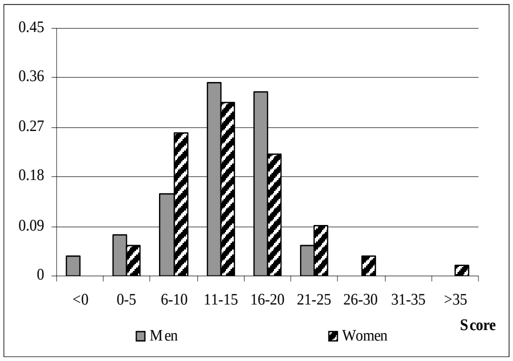

Figure 2a. Distribution of Verbal Scores by Gender Under Piece Rate, High Time Pressure

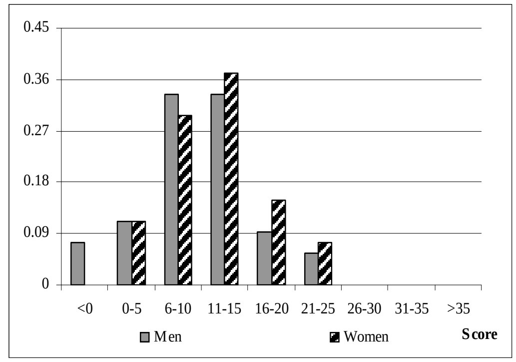

Figure 2b. Distribution of Verbal Scores by Gender Under Tournament, High Time Pressure

Result 2(b) is based on the finding that, on average, 39 percent of men and 30 percent of women choose tournament in this setting and is confirmed by the regressions in Table 3. Although a gap still persists in this high time pressure environment, the difference is not statistically significant (p-value of 0.32).

Note that in this high pressure verbal task, the right tail of the piece-rate score distribution is now dominated by women.[28] Thus, it is again important to condition the decision to enter the tournament on past performance. Probit regressions of the entry decision as a function of the female dummy, conditional on tournament score and various other controls can be found in Table 3.

Table 3. Probit of Tournament Entry Decision, High Time Pressure Verbal Sessions

Independent Variables:	Dependent Variable: Choice (Tournament =1)	
	(1)	(2)
Female	-0.03 (0.80)	0.05 (0.72)
Tournament Score	0.02* (0.07)	0.02 (0.28)
Tournament – Piece-Rate Score	-0.01* (0.09)	-0.02* (0.08)
Guessed Tournament Rank		-0.17** (0.04)
No. Observations	91	70

Notes: Standard errors clustered at the group level; marginal effects. Other controls include order of tournament and time, age, major (1 = science), native language (1 = English), and reported gender stereotype (1 if women perceived to be better). Significance levels: *10%, **5%,***1%. Guesses of 4 are eliminated in Specification (2).

[28] I confirm that, out of 1000 simulated sessions with randomly re-arranged mixed-gender groups, on average, 59% of the winners are women in the verbal tournament (significantly higher than 50%).

The likelihood of entry into the tournament rises with an increase in rank guess (specification 2), but the female dummy is not significant in either specification.[29] This result does not invalidate the conclusions of the previous literature (NV 2007), but rather adds a novel finding when we consider a different kind of task.

Next, I ask whether the presence of time constraints may differentially affect men and women in terms of their performance in competition and their willingness to compete.

3.3 Math Task under Low Time Pressure

By analogy with the above discussion of the high time pressure results, I begin the analysis of the low time pressure results with the math task perceived to disadvantage women relative to men.

RESULT 3. *Under low time pressure with a math task: (a) Men and women do not differ significantly in terms of their scores in either the piece-rate or the tournament treatment, although having more time significantly improves the performance of both genders; (b) Relative to the high time pressure environment, women are now nearly twice as likely to self-select into competition, while the men maintain a similar likelihood of choosing the tournament.*

Support for Result 3(a) comes from Figures 3a and 3b. The distributions of math scores become more spread out with more subjects at both extremes. In the piece-rate treatment, men and women achieve the mean scores of 29.5 and 19.9, and in the tournament, the scores of 32.3 and 23.1, respectively. However, statistical tests do not show a significant gender difference in either treatment.

[29] The distribution of verbal tournament rank guess by gender and time treatment, as well as ordered probit regressions of guessed rank as a function of the female dummy and performance in all treatments are provided in Appendix C, Tables C.3 and C.4, respectively.

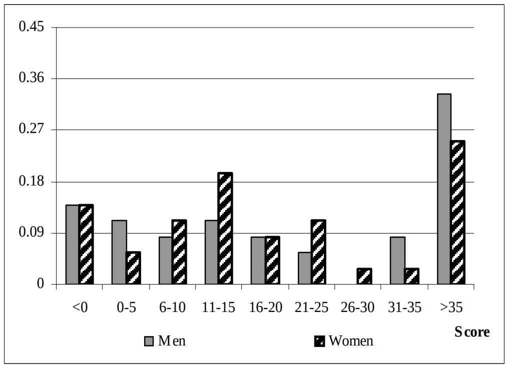

Figure 3a. Distribution of Math Scores by Gender Under Piece Rate, Low Time Pressure

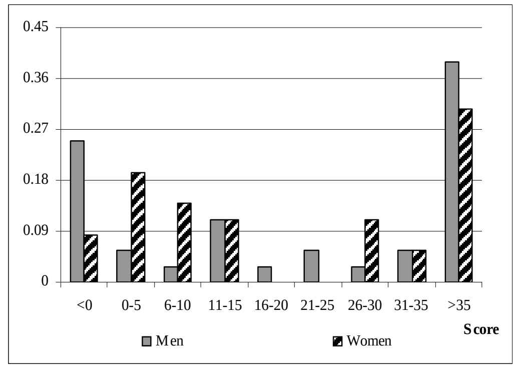

Figure 3b. Distribution of Math Scores by Gender Under Tournament, Low Time Pressure

Support for Result 3(b) comes from Figure 4 which summarizes the average probabilities of tournament entry in the math treatments by gender and time pressure treatment and is confirmed by the regressions in Table 4.

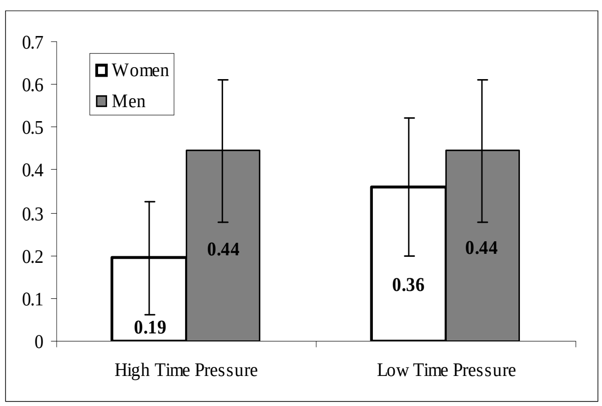

Figure 4. Average Likelihood of Self-Selecting into the Math Tournament (95% C.I.)

The high time pressure panel of Figure 4 illustrates the previously discussed fact that women are significantly less likely to enter competition in a high pressure math environment[30] (consistent with NV 2007). However, relaxing the time pressure results in women

[30] Note that the 95 percent confidence intervals in Figure 4 drawn for the mean of each gender group overlap slightly. However, the difference of 0.25 between men and women is significantly different from zero at the 5 percent confidence level.

nearly catching up to men in terms of their willingness to compete. The increase in the likelihood of women choosing the tournament from 19 percent to 36 percent is significant at the 5 percent confidence level (p-value of 0.03). Conditioning the tournament entry decision on previous performance and other controls confirms this result (the coefficient on the female dummy is not statistically significant in Table 4). Again, this result is different from the previous literature (NV 2007), but it represents a novel finding of what happens when we consider a different kind of environment.

Table 4. Probit of Tournament Entry Decision, Low Time Pressure Math Sessions

Independent Variables:	Dependent Variable: Choice (Tournament =1)	
	(1)	(2)
Female	-0.08 (0.60)	0.05 (0.80)
Tournament Score	0.01^{***} (0.00)	0.01^{*} (0.10)
Tournament − Piece-Rate Score	-0.001 (0.81)	-0.0002 (0.94)
Guessed Tournament Rank		-0.27^{**} (0.02)
No. Observations	72	55

Notes: Standard errors clustered at the group level; marginal effects. Other controls include order of tournament and time, age, major (1 = science), native language (1 = English), and reported gender stereotype (1 if women perceived to be better). Significance levels: *10%, **5%,***1%. Guesses of 4 are eliminated from Specification (2).

3.4 Verbal Task under Low Time Pressure

I have established that relaxing either one of the pressure sources on women helps them achieve levels of performance similar to those of men. The natural next step is to relax both sources.

RESULT 4. *Under low time pressure with a verbal task: (a) Men and women do not differ in terms of their scores in the piece-rate treatment, but in the tournament, women significantly outperform the men; (b) Relative to the high time pressure environment, women are now nearly twice as likely to self-select into competition, while the men maintain a similar likelihood of choosing the tournament.*

Support for Result 4(a) comes from Figures 5a and 5b which compare the distributions of scores achieved by men and women in the low time pressure verbal piece-rate and tournament treatments, respectively. Both men and women increase their scores in the tournament relative to the piece-rate treatment. More importantly, the increase in the performance of women is significantly greater than the increase in the performance of men. As a result, under competition, women achieve a significantly higher mean score of 23.4 relative to the men's 17.8 (p-value of 0.00).[31]

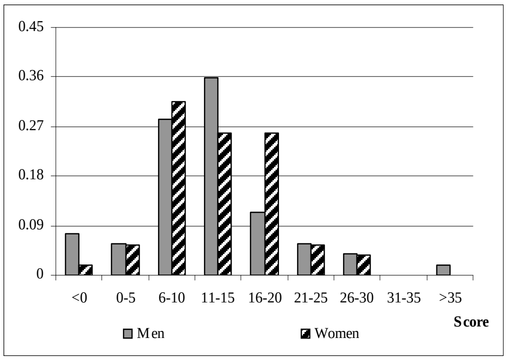

Figure 5a. Distribution of Verbal Scores by Gender Under Piece-Rate, Low Time Pressure

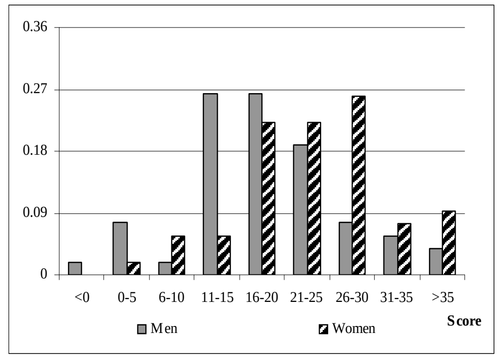

Figure 5b. Distribution of Verbal Scores by Gender Under Tournament, Low Time Pressure

Figure 6 and Table 5 provide evidence in support of Result 4(b).

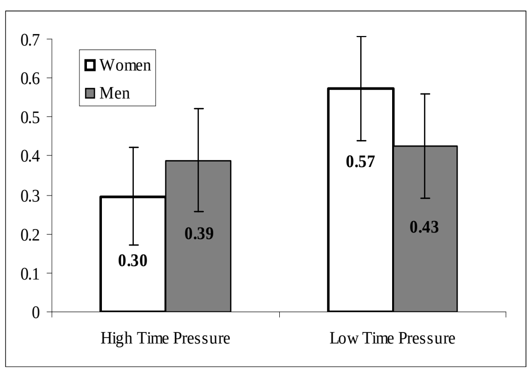

Figure 6. Average Likelihood of Self-Selecting into the Verbal Tournament (95% C.I.)

[31] Again, the within-subject analysis confirms these average results. See Appendix C, Table C.7 for OLS regressions of individual verbal score as a function of gender, treatment effects and various controls. Note that the results are robust to the inclusion of treatment order conrols and individual subject characterstics. I also confirm that the results are robust to using only the data for the first three rounds of the experiment (Appendix D, Table D.3). For more on the within-subject variation in the math and the verbal sessions, see Figure C.1 in Appendix C.

The likelihood that a woman will self-select into the tournament payment scheme nearly doubles with the reduction of time pressure in the verbal environment (significant at the 5 percent confidence level). Note that I am able to not only find environments where men and women no longer differ in terms of their preferences for competition (see the two previous subsections), but also discover a setting (low time pressure verbal task) where women are actually more likely than men to enter a tournament.[32]

Performance in the rounds preceding the choice treatment can play a role. In fact, out of 1000 simulated sessions with randomly re-arranged mixed-gender groups, on average 72 percent of the winners are women in the low time pressure tournament (significantly higher than 50 percent). In order to control for past performance, I run probit regressions of choice of compensation scheme (Table 5). Conditional on past performance, women are actually more likely to enter the tournament than men in this setting (specification 1). Once I control for confidence (rank guess), women remain marginally more likely to choose the tournament than the men.[33]

Table 5. Probit of Tournament Entry Decision, Low Time Pressure Verbal Sessions

Independent Variables:	Dependent Variable: Choice (Tournament =1)	
	(1)	(2)
Female	0.37^{***} (0.01)	0.28^{*} (0.10)
Tournament Score	0.004 (0.67)	-0.003 (0.79)
Tournament − Piece-Rate Score	-0.04^{***} (0.01)	-0.03^{*} (0.10)
Guessed Tournament Rank		-0.26^{*} (0.06)
No. Observations	89	77

Notes: Standard errors clustered at the group level; marginal effects. Other controls include order of tournament and time, age, major (1 = science), native language (1 = English), and reported gender stereotype (1 if women perceived to be better). Significance levels: *10%, **5%, ***1%. Guesses of 4 are eliminated in Specification (2).

[32] Previous studies that use a high pressure mathematical task all find a significant gender gap in tournament entry (NV 2007 and Niederle, Segal and Vesterlund 2008, for example).

[33] Women greatly increase the number of guesses of 1 and 2 under low time pressure (45 guesses) relative to high time pressure (27 guesses) in the verbal tournament (Appendix C, Table C.3). The ordered probit regression of guessed rank as a function of the female dummy, performance and other controls (Appendix C, Table C.4, specification 4) also shows that women are significantly more confident (more likely to report a lower rank guess) than men in the low time pressure verbal tournament.

4 Sources of Gender Differences

4.1 Quality vs. Quantity

Next, I seek to shed light on the origins of the gender differences in performance and preferences for competition. I start by focusing on the quality dimension of my tasks. In particular, I define the "quality-to-quantity ratio" or "mistake share" as the number of points lost due to entering invalid solutions (mistakes) divided by the total possible points (invalid plus valid).

Result 5. *(a) In the math task, women significantly increase the quality of their output in the competitive treatments once the time pressure is reduced; (b) In the verbal task, men significantly decrease the quality of their output in the competitive treatments once the time pressure is reduced, resulting in a large gender gap in mistake share.*

Support for Result 5(a) comes from Figure 7a which reports the average mistake shares in the math task by gender across all treatments. Under high time pressure, the mistake share for women doubles in the tournament relative to piece-rate. The quality reduction for women is significant at the 5 percent confidence level (p-value of 0.03).[34] The tournament mistake share for women falls significantly once the time pressure is reduced (p-value of 0.01). Note that the mistake share for men remains relatively stable over all treatments with a slight increase in the low time pressure tournament. Although the gender differences are not statistically significant in any treatment, the trends suggest that extra time allows women to improve the quality of their work.

Support for Result 5(b) comes from Figure 7b which reports the average mistake shares in the verbal task by gender across all treatments.

[34] This finding is consisitent, for example, with the evidence that women exhibit a decline in performance at high-pressure moments during tennis matches due to an increase in the number of costly errors, while the men's probability of error does not change significantly over the course the match (Paserman 2009).

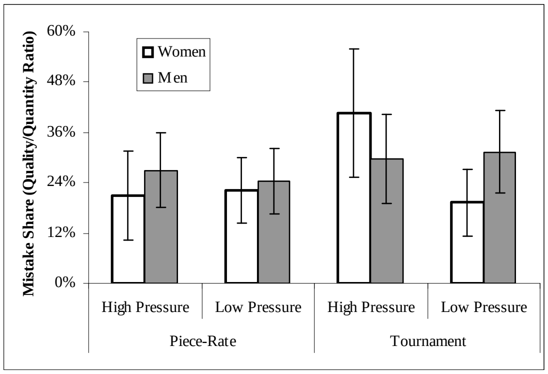

Figure 7a. Average Quality-to-Quantity Ratio by Gender for All Math Treatments (95% C.I.)

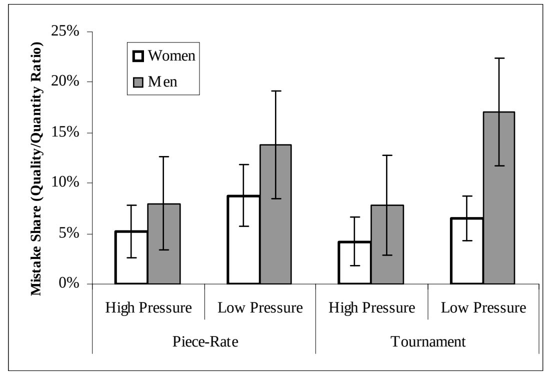

Figure 7b. Average Quality-to-Quantity Ratio by Gender for All Verbal Treatments (95% C.I.)

First, note that the mistake shares of men and women are not significantly different in either of the piece-rate conditions or in the high time pressure tournament. However, the male mistake share is significantly higher under low time pressure than under high time pressure when the two verbal tournaments are compared (p-value of 0.001). As a result, I observe a significant gender gap in mistake share in the low time pressure verbal tournament (p-value of 0.0002) which partly contributes the under-performance of men relative to women in this treatment (see Figure 5b).[35]

Can the gender differences in mistakes be explained by the types of errors people are prone to make? If men are making more mistakes in the low pressure tournament because they are seeking big rewards and entering longer words or longer number combinations, then the explanation may have nothing to do with quality considerations, but rather with a preference for risk-taking behavior. In order to test this alternative theory, I count the average number of letters and the average number of digits making up the mistake entries. I find that women actually make errors on *longer* combinations than the men in the low time pressure math tournament (p-value of 0.04) and on longer words in the high time pressure verbal piece-rate treatment (p-value of 0.02). The gender differences in average

[35] Tables C.5 and C.6 in Appendix C report the OLS regressions of the mistake share as a function of gender, treatment effects and various controls for the math and the verbal sessions, respectively. The results are robust to the inclusion of treatment order and other controls and to only using the first three rounds of data (Web Appendix D, Tables D.2 and D.3).

length of mistake entries are not statistically significant in the rest of the treatments. Note that it may be misleading to only look at mistake entries, so I also check whether men and women differ in terms of the overall average word length for both valid and invalid entries. Simple t-tests show again that there is no significant gender difference for any of the treatments.[36]

A related consideration is that, in word-in-a-word puzzles, there is more than one notion of what constitutes a "mistake." Some mistakes are typos and spelling errors, some do not conform to the rules of the game (too short, proper nouns), while some are words that do not exist in the English language (for example, archaic words, words from a different language, and words that are simply made-up). Since there is no way to be sure that a word actually exists, subjects may face additional uncertainty.[37] Thus, the increase in mistake-making by the men may again indicate a different attitude toward risk rather than a lack of attention to quality. The categorization of mistakes according to the existence criterion is subject to some assumptions (for example, on how to deal with proper nouns). However, no matter what definition I use, I find that the ratio of word-does-not-exist mistakes to the total number of mistakes does not differ significantly by gender. This is true on average, and more importantly, in the low time pressure treatment, where I observe a significant difference in the mistake share for men and women. Thus, I can conclude that the uncertainty in mistake-making is not responsible for the increase in errors made by men.

The next section explores other potential sources of gender differences in the math and verbal tournaments.

4.2 Quitting and Confidence

Quitting or giving up may be an indication of frustration and a defeatist attitude that can result in lower performance. The follow-up questionnaire directly asks the subjects about their effort in the game and whether they gave up at any point during the experiment.

[36] A variation on this analysis using the ratio of the number of long mistake words (5+ letters) to the total number of mistake words produces similar results.

[37] Note that this issue is not relevant for the math game.

While the questionnaire responses should be taken with caution, some noteworthy patterns emerge.

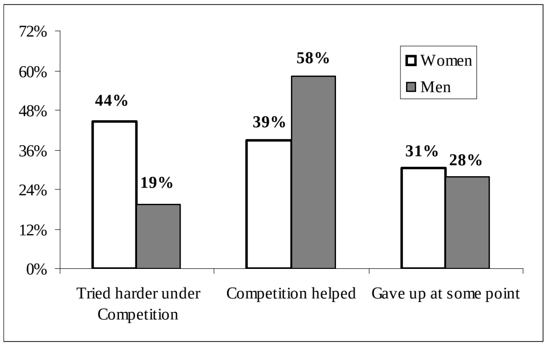

Figure 8a. Questionnaire Responses by Gender for All Math Treatments

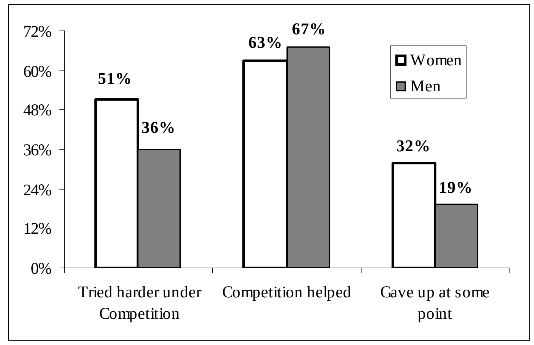

Figure 8b. Questionnaire Responses by Gender for All Verbal Treatments

In both types of sessions (math and verbal), a larger share of women reported that they tried harder under competition than under piece-rate (see Figures 8a and 8b). However, in the numbers game, men were significantly more likely to view competition as helpful (see Figure 8a). This difference disappeared in the verbal game with both genders finding competition equally helpful (see Figure 8b). Finally, both in math and in verbal sessions, a slightly larger share of women reported giving up at some point, although the gender difference is not significant (see Figures 8a and 8b).

Self-reports of giving up may not be trustworthy since men and women might have different likelihoods of telling the truth and of recalling what actually happened. Thus, I use a more reliable metric: actual quitting behavior in the low time pressure rounds.[38] In order to check whether there is a gender gap in quitting behavior, I run probit regressions of quitting as a function of the female and the competition dummies, rank guess, and other controls. The results are reported in Table 6.

In the math tournament, a woman is 24 percent more likely to quit the game than a man in the same treatment (column 1), which is consistent with the notion that women may be less confident in an environment that is perceived to give men the advantage. By

[38] Recall that a subject may withdraw from a 10-minute round at any point by clicking a "Finish" button.

contrast, quitting behavior in the verbal task shows no significant gender difference under either compensation scheme. For both genders, the probability of quitting the word game falls significantly under competition relative to the piece-rate treatment.

Table 6. Probit of Quitting Decision, Low Time Pressure, Math and Verbal Sessions

Independent Variables:	Dependent Variable: Quit (Quit =1)			
	Math Task		Verbal Task	
	(1)	(2)	(3)	(4)
Female	-0.15**	-0.15*	0.01	-0.03
	(0.03)	(0.07)	(0.96)	(0.77)
Competition (Tournament = 1)	-0.03	-0.08	-0.21**	-0.22***
	(0.72)	(0.36)	(0.01)	(0.01)
Female×Competition	0.24***	0.23**	-0.08	0.003
	(0.00)	(0.02)	(0.39)	(0.97)
Guessed Rank (out of 4)		0.05		0.08
		(0.34)		(0.20)
No. Observations	144	111	181	154

Notes: Robust standard errors clustered at the group level; p-values in parentheses; marginal effects. Other controls include: order of tournament and time, age, major (1 = science), native language (1 = English), and reported gender stereotype (1 if women perceived to be better). Guesses of 4 are eliminated in Specifications (2) and (4).
Significance levels: * 10%, ** 5%, *** 1%.

The willingness of women to stay in the game on par with the men may be a contributing factor to the observed performance improvement in the verbal environment.

5 Labor Market Evidence of Variation in Gender Differences across Job Types

The experimental results suggest that gender differences in performance under competition and preferences for competition vary across different types of environments. This section investigates whether similar patterns emerge in the real labor market. In particular, the experimental finding that men display a greater affinity for competition and perform better in high-pressure mathematical environments would suggest that we should expect to see a higher share of men and a relatively larger gender gap in earnings in those settings. On the other hand, the finding that women are more likely to self-select into competition and exhibit higher performance in relatively lower-pressure verbal environments would suggest that we should see a higher share of women and a much smaller gender gap in earnings for those types of jobs.

In order to address these issues, I conduct a simple labor market study that uses individual-level data from IPUMS CPS for years 2003-2009. The relevant data include individual real earnings, gender, occupation, and other demographic variables. I categorize occupations into high-pressure/math, high-pressure/verbal, low-pressure/math, and low-pressure/verbal based on the pressure and stress classifications from CareerCast.com. Clearly, very few jobs are purely mathematical or verbal. An example of the former would be a mathematician, while an example of the latter would be a writer. Most other jobs entail some aspects of both skills. For the purposes of this simple analysis, I restrict the sample to consist of data on individuals in occupations that are exemplars of each category (full list of included occupations is given in Table 7).

Table 7. Examples of Jobs in Each Category

	High Pressure	**Less Pressure**
Math	Financial managers; Financial analysts; Securities, commodities, and financial services sales agents; Physicians and surgeons	Accountants and auditors; Actuaries; Mathematicians; Statisticians
Verbal	Announcers; News analysts, reporters, and correspondents; Advertising and promotions managers; Public relations managers	Writers and authors; Librarians; Archivists, curators, and museum workers

Table 8 provides the main findings from the labor market study. Panel A presents the results of OLS regressions of real earnings as a function of the female dummy and various controls, including demographic variables and year fixed effects, clustering standard errors on the regional level for each subset of occupations from Table 7. I observe the largest gender earnings gap for the high pressure math jobs (Table 8, specification 1). The gender gap is reduced but remains significant in specifications 2 and 3 of Table 8. Finally, the gender gap disappears for relatively less stressful jobs of verbal nature. The differences in the earning gaps across the various occupation categories are consistent with my experiments findings.[39]

[39] Note that the data do not show a full reversal of the gender earnings gap in specification 4 of Table 8. However, the fact that women do not earn more than men under this scenario in the real world might

Table 8. Determinants of the Gender Earnings Gap and the Share of Women by Occupation Type

	(1)	(2)	(3)	(4)
	High Pressure Math Jobs	High Pressure Verbal Jobs	Less Pressure Math Jobs	Less Pressure Verbal Jobs
	Panel A: Gender Earnings Gap (Dependent Variable: Real Earnings; Sample: Within Job Category)			
Female	-17.37***	-7.57**	-7.63***	-6.79
	(2.09)	(2.59)	(1.52)	(3.79)
R^2	0.02	0.07	0.03	0.04
Observations	10984	1199	8343	2012
	Panel B: Share of Women (Dependent Variables: Job Categories; Sample: Full)			
	-0.19***	-0.02***	0.13***	0.08***
	(0.01)	(0.006)	(0.02)	(0.008)
R^2	0.06	0.02	0.03	0.05
Observations	22538	22538	22538	22538
	Panel C: Average Share of Women in Each Job Category			
Female Share	43%	46%	62%	71%

Notes: Robust standard errors clustered at the region level in parentheses. All specifications include: age, age^2, education, race, marital status, the number of children, and year fixed effects (2003-2009). Panel A reports coefficients from OLS regressions of real earnings for four occupation sub-samples. Panel B reports coefficients from OLS regressions of job type dummies for the entire sample pooled across the occupation categories. Significance levels: * 10%, ** 5%, *** 1%.

Panels B and C of Table 8 concern the differences in the share of women across difference occupation types. First, note that I find some evidence of sorting of men and women into job types. According to the simple averages for each of the subsets of occupations, women represent only 43 percent of the sample of workers in high-pressure mathematical jobs, while women represent 71 percent of workers in the relatively less high-pressure verbal jobs (Panel C). Panel B summarizes OLS regressions of the job category (an indicator of whether the individual holds an occupation within a given job category) as a function of the female dummy and various controls, including demographic variables and year fixed effects, pooling the data across all four job categories and clustering standard errors on the regional level. I find that a woman is almost 20 percent less likely to work at a high-pressure math job than a man of similar characteristics. Women are also slightly less likely to work at high-pressure verbal jobs, although the magnitude of the

be explained by other confounding factors (parenthood, discrimination, etc.).

correlation is greatly reduced relative to the mathematical environment. On the other hand, women are more likely to hold occupations that involve relatively less pressure (both mathematical and verbal).

Thus, I find that my experimental results are consistent with what we see in the real labor market. Women favor occupations that involve relatively less on-the-job pressure and that are of verbal nature, and the gender gap in earnings is must less pronounced in those settings relative to the high-pressure mathematical environments.

6 Conclusion

This paper uses a controlled experiment to explore the sources of gender differences in performance under competition and preferences for competition. First, I confirm the previous finding that women underperform the men in competitive settings whenever two types of pressure are present: task stereotype (math task) and time constraints. Furthermore, women are significantly less likely to compete under these conditions.

Relaxing either one of the pressure sources benefits women, such that there is no longer a significant gender difference in performance and preference for competition.

Once both sources of pressure on women are removed, I find that women actually surpass the men in competition and therefore earn a higher average payoff. Women also choose competition significantly more often in this setting.

Next, I explore the possible explanations for the gender differences I find. An important contributing factor turns out to be the differential response of the two genders to competition under reduced time pressure. While women seem to use the extra time to increase the quality of their work (reducing the number of mistakes per total points), men use the time to increase the quantity, producing a higher volume of work, but also increasing the share of mistakes.

The results so far suggest several directions for future research. First, changing the puzzle order, while keeping the payment treatment order fixed, would provide greater control over the task difficulty from treatment to treatment. Note, however, that the sheer number of permutations necessary to achieve this higher degree of control may be prohibitively large. Second, single-gender sessions can be conducted in order to see whether group composition generates strong enough stereotype threat to affect performance. Preliminary evidence suggests that this type of stereotype threat has an asymmetric impact on men and women. While women may perform better in math tournaments with female-only groups (GNR 2003), men do not seem to perform significantly better in verbal tournaments with male-only groups (Shurchkov 2009).

The evidence documented in this paper suggests that the effect of competition on gender-specific outcomes depends greatly on the environment at hand. This evidence seems to be consistent with the observations of gender gaps in the real labor market. The results yield certain policy implications. In the workplace, women and men face competition not only in terms of their ability to perform jobs of mathematical nature, but also in terms of their verbal abilities, such as writing reports, creating presentations, and talking to clients. Competition benefits men in the former, yet benefits women in the latter, especially when given sufficient time to complete the task. In addition, the quality of work is at least as important as the quantity and the rate of output. My novel tasks that involve multiple solutions add the dimension of quality which explains part of the gender difference in both math and verbal environments. In an educational setting, shorter exams (or extra time), especially in hard sciences and math, may close the performance gap between men and women and in the long-run increase female participation in the study of these subjects.

References

[1] Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar (2009). "Large Stakes and Big Mistakes." *Review of Economic Studies*, 76, 451–469.

[2] Babcock, Linda, and Sarah Laschever (2003). *Women Don't Ask: Negotiation and the Gender Divide*. Princeton University Press.

[3] Barres, Ben (2006). "Does Gender Matter?" *Nature*, 442(13), 133-136.

[4] Baron-Cohen, Simon (2003). *The Essential Difference. Men, Women, and the Extreme Male Brain*. Allen Lane, p. 256.

[5] Becker, Gary (1957). *The Economics of Discrimination*. The University of Chicago Press.

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

[7] Rahman, Qazi, Davinia Andersson, and Ernest Govier (2005). "A Specific Sexual Orientation-Related Difference in Navigation Strategy." *Behavioral Neuroscience*, 119(1), 311-316.

[8] Black, Sandra, and Philip Strahan (2001). "The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry." *American Economic Review*, 91(4), 814-831.

[9] Booth, Alison, and Patrick Nolen (2009). "Gender Differences in Risk Behaviour: Does Nurture Matter?" IZA Discussion Papers No. 4026.

[10] CNNMoney.com (2006). "Women CEOs for FORTUNE 500 companies." http://money.cnn.com/magazines/fortune/fortune500/womenceos/

[11] Dreber, Anna, and Magnus Johannesson (2008). "Gender Differences in Deception." *Economics Letters*, 99, 197-99.

[12] Gneezy, Uri, Muriel Niederle and Aldo Rustichini (2003) "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics*, 118(3), 1049-74.

[13] Gneezy, Uri, and Aldo Rustichini (2004) "Gender and Competition at a Young Age." *American Economic Review*, 94(2), 377-81.

[14] Gneezy, Uri, and Aldo Rustichini (2005) "Executives versus Teachers: Gender, Competition and Self Selection." Working Paper, University of California, San Diego.

[15] Gneezy, Uri, Kenneth Leonard, and John List (2008). "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society." NBER Working Paper No. 13727.

[16] Lawrence, Peter (2006). "Men, Women, and the Ghosts in Science." *PLOS Biology*, 4(1), 13-15.

[17] Macpherson, David, and Barry Hirsch (1995). "Wages and Gender Composition: Why Do Women's Job Pay Less?" *Journal of Labor Economics*, 13(3), 426-471.

[18] Massachusetts Institute of Technology (1999). *A Study on the Status of Women Faculty in Science at MIT*.

[19] New York Magazine (2010). "The Top 1,119 Physicians." http://nymag.com/bestdoctors/

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

[21] Niederle, Muriel, Carmit Segal, and Lise Vesterlund (2008). "How Costly is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness." NBER Working Paper No. 13923.

[22] Pajares, Frank, and Giovanni Valiante (2001). "Gender Differences in Writing Motivation and Achievement of Middle School Students: A Function of Gender Orientation?" *Contemporary Educational Psychology*, 26(3), 366-381.

[23] Paserman, M. Daniele (2007). "Gender Differences in Performance in Competitive Environments: Evidence from Professional Tennis Players." CEPR Discussion Paper No. 6335.

[24] Pease, Allan, and Barbara Pease (2000). *Why Men Don't Listen and Women Can't Read Maps: How We're Different and What to Do About It*. Welcome Rain.

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

[26] Rolfe, John, and Peter Troob (2000). *Monkey Business: Swinging Through the Wall Street Jungle*. Warner Books.

[27] Shurchkov, Olga (2009). "Gender Differences in Output Quality and Quantity under Competition and Time Constraints: Evidence from a Pilot Study." FEEM Working Paper No. 356.

[28] Spencer, Steven, Claude Steele, and Diane Quinn (1999). "Stereotype Threat and Women's Math Performance." *Journal of Experimental Social Psychology*, 35(1), 4-28.

[29] Steele, Claude, and Joshua Aronson (1995). "Stereotype Threat and the Intellectual Test Performance of African-Americans." *Journal of Personality and Social Psychology*, 69, 797-811.