Under Pressure: Gender Di¤erences in Output Quality and Quantity under Competition and Time Constraints

Olga Shurchkov[y]

October 2010

Abstract

Gender gaps in the workplace are widespread. One explanation for gender inequality stems from the e¤ects of the interaction between competition and two pressure sources, namely, task stereotypes and time constraints. This study uses a laboratory experiment to …nd that the gender gap in performance under competition and preferences for

1 Introduction

The study of gender di/erences has a long history in the eld of labor economics. Despite

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 e/ects of changing the environment to be more woman-oriented by relaxing both sources of pressure: the time constraint[4] and the math-

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,

2 Overview of the Experiment

The goal of this study is to ask whether gender di/erences 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 di/erences 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 de ned as the number of points lost due to invalid

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

number of permutations of letters needed to arrive at any one sub-word and con rm 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 taskr]1Jrottabe1 11.955 Tf -310.90.163[(.)]TJ.21 .

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

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

At the end of the experiment, each participant lled out a brief questionnaire.[19] The

Result 1.

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 e/ects. 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). Signi cance levels: 10%, 5%, 1%. Guesses of 4 are eliminated in Speci cation (2).

I have con rmed the previous ndings 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 e/ects of changing the task to be perceived as more woman-friendly [27]

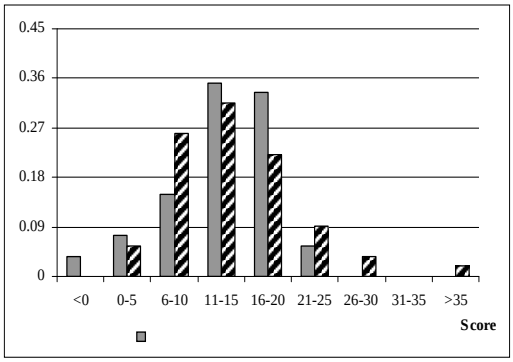

Figure 2a. Distribution of Verbal Scores by Gender Under Piece Rate, High Time Pressure

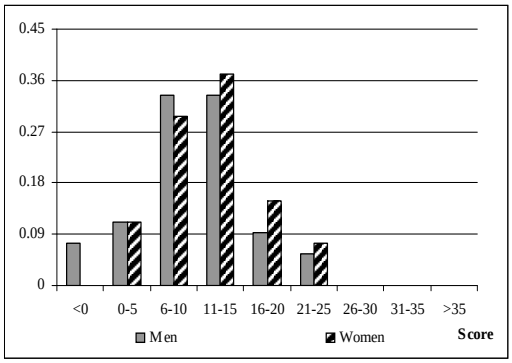

Figure 2b. Distribution of Verbal Scores by Gender Under Tournament, High Time Pressure

Result 2(b) is based on the nding that, on average, 39 percent of men and 30 percent of women choose tournament in this setting and is con rmed by the regressions in Table 3. Although a gap still persists in this high time pressure environment, the di/erence is not statistically signi cant (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		-0.3Tnd in Table 3..(3)39Td[((0.)00.08

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 treat-

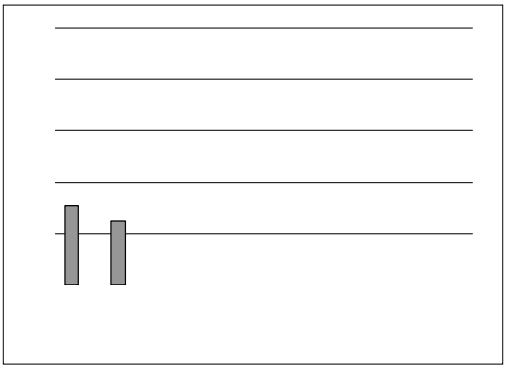

Figure 3a. Distribution of Math Scores by Gender Under Piece Rate, Low Time Pressure

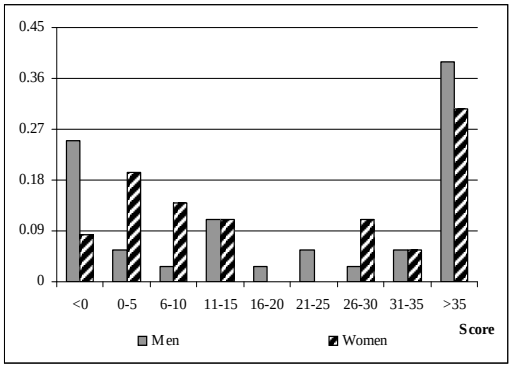

Figure 3b. Distribution of Math Scores by Gender Under Tournament, Low Time Pressure

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 con rmed by the regressions in Table 4.

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 signi cant at the 5 percent con dence level (p-value of 0.03). Conditioning the tournament entry decision on previous performance and other controls con rms this result (the coe¢ cient on the female dummy is not statistically signi cant in Table 4). Again, this result is di/erent from the previous literature (NV 2007), but it represents a novel nding of what happens when we consider a di/erent 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 e/ects. 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). Signi cance levels: 10%, 5%, 1%. Guesses of 4 are eliminated from Speci cation (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 di¤er in terms of their scores in the piece-rate treatment, but in the tournament, women signi…cantly outperform the men; (b) Relative to the high time pressure environ-

performance of women is signi cantly greater than the increase in the performance of men. As a result, under competition, women achieve a signi cantly higher mean score of 23.4 relative to the men s 17.8 (p-value of 0.00).[31]

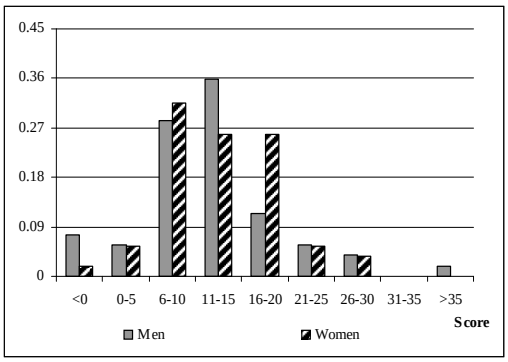

Figure 5a. Distribution of Verbal Scores by Gender Under Piece-Rate, Low Time Pressure

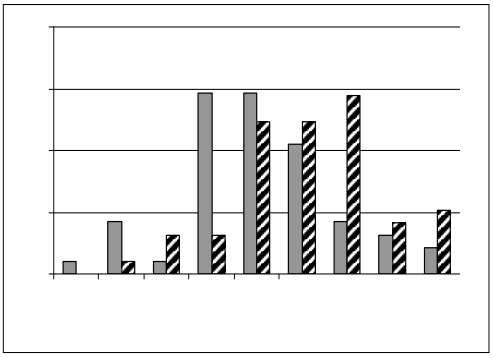

Figure 5b. Distribution of Verbal Scores by Gender Under Tournament, Low Time Pressure

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

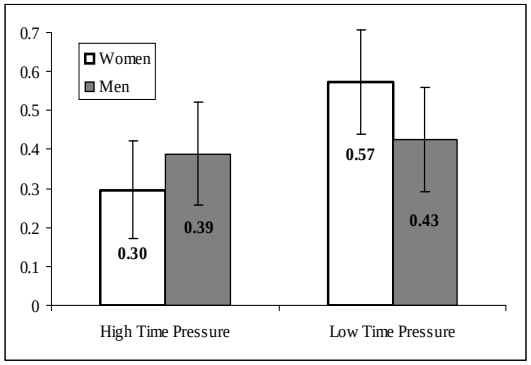

Figure 6. Average Likelihood of Self-Selecting into the Verbal Tournament (95% C.I.)

[31] Again, the within-subject analysis con rms these average results. See Appendix C, Table C.7 for OLS regressions of individual verbal score as a function of gender, treatment e/ects and various controls. Note that the results are robust to the inclusion of treatment order conrols and individual subject characterstics. I also con rm that the results are robust to using only the data for the rst 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 (signi cant at the 5 percent con dence level). Note that I am able to not only nd environments where men and women no longer di/er 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 (signi cantly 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 (speci cation 1). Once I control for con dence (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 e/ects. 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). Signi cance levels: 10%, 5%, 1%. Guesses of 4 are eliminated in Speci cation (2).

[32] Previous studies that use a high pressure mathematical task all nd a signi cant 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, speci cation 4) also shows that women are signi cantly more con dent (more likely to report a lower rank guess) than men in the low time pressure verbal tournament.

4 Sources of Gender Di/erences

4.1 Quality vs. Quantity

Next, I seek to shed light on the origins of the gender di/erences in performance and preferences for competition. I start by focusing on the quality dimension of my tasks. In particular, I de ne 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 signi…cantly increase the quality of their output in the competitive treatments once the time pressure is reduced; (b) In the verbal task, men signi…cantly 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

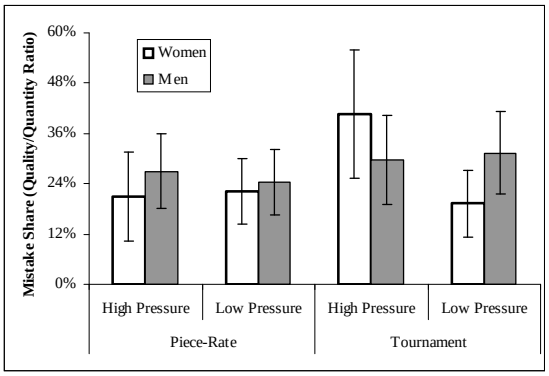

Figure 7a. Average Quality-to-Quantity Ratio by Gender for All Math Treatments (95% C.I.)

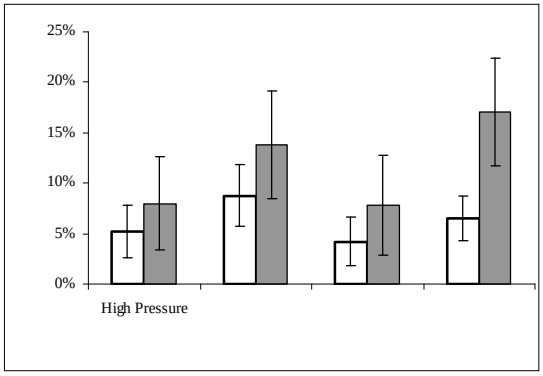

Figure 7b. Average Quality-to-Quantity Ratio by Gender for All Verbal Treatments (95% C.I.)

First, note that the mistake shares of men and women are not signi cantly di/erent in either of the piece-rate conditions or in the high time pressure tournament. However, the male mistake share is signi cantly 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 signi cant 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 di/erences 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 nd 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 di/erences 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 e/ects 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 rst three rounds of data (Web Appendix D, Tables D.2 and D.3).

length of mistake entries are not statistically signi cant 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 di/er in terms of the overall average word length for both valid and invalid entries. Simple t-tests show again that there is no signi cant gender di/erence 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 di/erent 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 di/erent 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 de nition I use, I nd that the ratio of word-does-not-exist mistakes to the total number of mistakes does not di/er signi cantly by gender. This is true on average, and more importantly, in the low time pressure treatment, where I observe a signi cant di/erence 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 di/erences in the math and verbal tournaments.

4.2 Quitting and Con dence

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 e/ort in the game and whether they gave up at any point during the experiment.

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

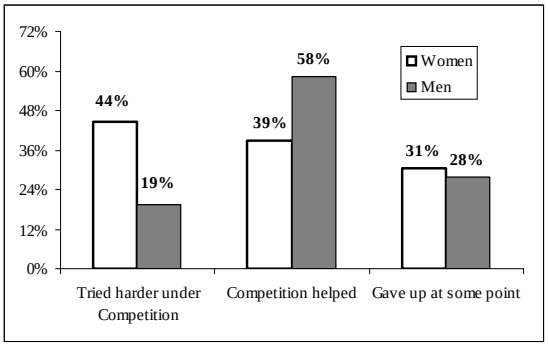

Figure 8a. Questionnaire Responses by Gender for All Math Treatments

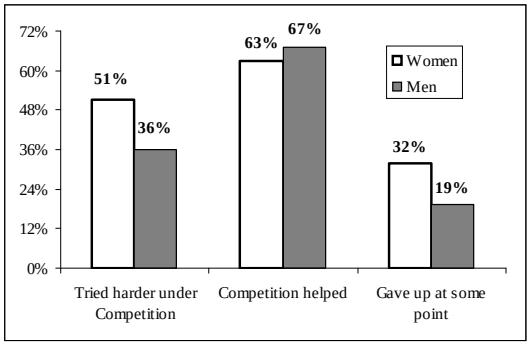

Figure 8b. Questionnaire Responses by Gender for All Verbal Treatments

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 signi cantly more likely to view competition as helpful (see Figure 8a). This di/erence disappeared in the verbal game with both genders nding 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 di/erence is not signi cant (see Figures 8a and 8b).

Self-reports of giving up may not be trustworthy since men and women might have di/erent 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 con dent 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 signi cant gender di/erence under either compensation scheme. For both genders, the probability of quitting the word game falls signi cantly under competition relative to the piece-rate treatment.

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

ments 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 classi cations 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 nancial 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 ndings 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 xed e/ects, 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, speci cation 1). The gender gap is reduced but remains signi cant in speci cations 2 and 3 of Table 8. Finally, the gender gap disappears for relatively less stressful jobs of verbal nature. The di/erences in the earning gaps across the various occupation categories are consistent with my experiments ndings.[39]

[39] Note that the data do not show a full reversal of the gender earnings gap in speci cation 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;				

correlation is greatly reduced relative to the mathematical environment. On the other

hibitively 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,

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 Dont Ask: Negotiation and the Gender Divide. Princeton University Press.

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

[4] Baron-Cohen, S1.geD(6)11().408Td[([)6(410(5)11(1)10( )261(4)10(6)113)16(o)-16(e)9Td[(R)66(

[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? A¢ rmative Action in Light of Gender Di/erences in Competitiveness. NBER Working Paper No. 13923.

[22] Pajares, Frank, and Giovanni Valiante (2001). Gender Di/erences 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 Di/erences in Performance in Competitive