The Effect of Task Choice and Task Assignment on the Gender Earnings Gap: An Experimental Study*

Kai Ou† Xiaofei Pan‡

Abstract

Previous studies explain the gender earnings gap by gender differences in choosing competitive and higher-paying jobs. However, little is known about whether and how women's earnings are affected when they choose more challenging jobs. In this study, we use a novel identification strategy to investigate 1) how the gender earnings gap arises from individuals' self-selecting into different tasks and 2) whether mobilizing women to work on the tasks typically preferred by men increases women's earnings and decreases the earnings gap. Our results show that men who prefer the hard and higher-paid task are more likely to obtain higher earnings regardless of the task they are assigned. In contrast, we find that women obtain higher earnings when they work on a hard and higher-paid task even if their initial take choice is the easy and lower-paid one. Our findings are consistent and robust across task stereotypes. Our results imply that mobilizing women to work on more challenging and rewarding tasks is likely to reduce the gender earnings gap.

JEL-Classification: C91, J16, J24, J31, M52

Keywords: Task Choice Effect, Task Assignment Effect, Gender, Experiment

*The financial support from Bryant University and Florida State University is gratefully acknowledged. We are indebted to comments by Jeffery Carpenter, Yan Chen, Pedro Dal Bó, Laura Gee, Jens Großer, Haifeng Huang, Dorothea Kübler, Svetlana Pevnitskaya, Dmitry Ryvkin, Ulrike Vollstaedt, members at the Eastern Economic Association Conference, Economic Science Association Annual Conference, WEAI Virtual International Conference and seminar audience at Brown University, Southern Connecticut State University, and Florida State University. We thank Philip Solimine, Sarah Warren, and Braeden McNulty for their helpful assistance. All errors remain the authors' own responsibility.

†Assistant Professor, Faculty of the xs/fs Experimental Social Science Cluster, Florida State University. 554 Bellamy Building, Tallahassee, FL 32306-2230. kou@fsu.edu.

‡Assistant Professor, Department of Economics, Bryant University. Faculty Suite F, Room 2420. 1150 Douglas Pike, Smithfield, RI, 02917. xpan@bryant.edu.

1 Introduction

The gender wage gap has been well-documented as a phenomenon that persists across occupations (Azmat and Ferrer 2017; Bertrand and Hallock 2001). In exploring the reasons behind this phenomenon, some studies have found that labor market discrimination or attitudes perpetuating stereotypes towards women help explain the gap (Goldin and Rouse 2000; Neumark 2018). Other studies have shown that women enter negotiations less often and/or fare worse when they do negotiate (Dittrich et al. 2014; Leibbrandt and List 2015). In an experimental setting, studies have found that women are generally less likely to engage in competition than men, which they posit contributes to the gender wage gap (Flory et al. 2015; Niederle and Vesterlund 2007).[1] However, a recent study on the gig economy continues to document a gap even when the above factors are barely present (Cook et al. 2020). Other experimental studies show that gender differences affect how individuals choose between easier or challenging tasks, which may also contribute to gender earnings gap. Buser et al. (2014) find that women are less likely to choose prestigious academic tracks, Bracha and Fershtman (2013) suggest women are likely to spend less time on challenging and rewarding tasks than their male counterparts, and Niederle and Yestrumskas (2008) conclude men choose a challenging task much more frequently than women, even in situations where both genders have equal ability. These studies suggest that, compared to men, women are less likely to self-select into more challenging and higher-paid jobs. If high-performing women are disproportionately left out from higher-paying jobs, then we might expect a large gender earnings gap. Our study contributes to research exploring the factors impacting the gender wage gap through examining: a) the extent there are gender differences in the choice of a difficult task and b) the ways in which a simple intervention such as manipulating task assignment impacts subsequent earnings.

It is possible that even if women change their job choices, they may not be able to reduce the gender earnings gap due to stereotypes based on cultural and social norms

[1] Dohmen and Falk (2011) show that this difference in competition could partially be driven by women's more risk-averse preferences. Thus women become less likely to choose variable payment schemes than a fixed payment one. Niederle and Vesterlund (2011) point out that women's lower confidence in their relative ability may contribute more to their tendency to shy away from competition than risk aversion.

as well as occupation patterns and political institutions. For example, Blau and Kahn (2017) find evidence of both occupation segregation (e.g., greater concentrations of women in administrative support and service positions) and labor market discrimination (e.g., hiring that caters to an employer, co-worker, or customer preferences). In other words, even if a woman prefers a competitive, risk-seeking position, cultural and social factors may make it difficult for women to obtain and succeed in such a position. Indeed, Bowles et al. (2007) find that women who initiate negotiations are penalized more by male evaluators. Moreover, Babcock et al. (2017) suggest that women are more likely to accept assignments with a lower potential for promotion than men, hindering their ability to move up in an organization. In a study of women who do have higher positions and pay, Ong et al. (2019) find that these women are more likely to face a penalty in terms of marriage and children.

Given the discouraging landscape facing women seeking to reduce the wage gap, our study examines the effectiveness of an institutional intervention, such as Affirmative Action, that would make women more willing to opt-in without considering cultural norms or potential social costs. If men earn more because they choose a harder and more rewarding task and women earn less because they choose an easier and less rewarding task, we are interested in whether motivating women to choose the harder task will have a resultant positive impact on their earnings. At a broader level, encouraging women to choose these tasks could contribute to the earnings of high-ability women who are left out in the current system and even improve aggregate social welfare.

While understanding whether the gender earnings gap would be reduced if more women worked on challenging and rewarding tasks would provide critical insight for both scholars and policymakers, there are currently limited studies examining these questions due to at least two empirical challenges. First, to address these questions, we need to manipulate individual job choices while excluding other factors that might be associated with the task selection decision-making process, which is exceedingly difficult to find in observational data or to implement in a natural or field experiment. Second, we need to generate an appropriate counterfactual to observe how women *would have performed had their job choice been different*. In naturally-occurring situations, we are empirically and

ethically unable to know what individuals would do in such a counterfactual world.[2]

Our study addresses these issues by using a novel and unique research design to explore whether and how having men and women work on jobs that they would not have chosen in the first place results in a positive effect on reducing the gender earnings gap. To elicit individuals' preferences on task choices in our experimental study, we asked participants to indicate if they preferred a hard and higher-paid piece-rate task or an easy and lower-paid piece-rate task. To establish the appropriate counterfactual, 55% of participants were assigned their preferred tasks, while the other 45% of participants were assigned the tasks they did not choose.[3] Participants have incentives to sincerely report their preferred tasks because they are more likely to be assigned a task they choose. We establish the counterfactual by investigating the performance and earnings of those participants who are assigned non-preferred tasks.

Specifically, our study focuses on three effects that may contribute to the gender earnings gap: 1) *Task Choice Effect*. We compare earnings within women (men) who chose different tasks yet were assigned the same task to address whether those who self-select into the hard and higher-paid piece-rate task earn more. 2) *Task Assignment Effect*. We compare earnings within women (men) who chose the same task yet were assigned different tasks to investigate whether, controlling for self-selection, working on the hard and higher-paid piece-rate task impacts one's earnings. 3) *Gender Effect*. We compare earnings *between* men and women who *chose the same* and *were assigned the same* task to study whether women earn differently from men, and explore the determinants of the behavioral difference between men and women.

Our research design identifies multiple channels of earning differences that are caused by different effects that have not been documented in the literature. The classical experimental design in studying different preferences across genders normally elicits men and women's decisions between different environments (see Bracha and Fershtman 2013; Gneezy et al. 2009; Niederle and Yestrumskas 2008). While these studies improve our understanding of how men and women behave differently in the labor market, they could

[2]While observational data may provide information on the labor market or where individuals find their jobs, this data typically bundles a number of running variables and uncertainty, making it difficult to learn from these data about the net effect of having people choose a different job.

[3]The mechanism of how the tasks are assigned to participants is common knowledge explained in the experiment.

not address whether earnings difference between genders are caused by their task choices or being assigned to different task environments.[4] Other studies, including Gneezy et al. (2003) and Gneezy and Rustichini (2004), directly compare men's and women's earnings in a competitive task environment and find that a gender earnings gap persists even when the same men and women perform similarly in a non-competitive task. However, they randomly assign subjects to tasks and thus are not able to isolate the extent to which individuals' tasks choices, i.e., the self-selection in the labor market, contribute to the earnings gap between genders. By contrast, our study elicits participant task choices and thus can explore the earnings impact of assigning a task different from the preferred one. To our knowledge, our unique research design is the first one to directly identify whether any earnings gap is a result of task choice effect or is an outcome of task assignment effect. We use the findings from our study to propose that, if the gender earnings gap is an outcome of working on different types of tasks, scholars and policymakers should devise mechanisms and policies to mobilize women and/or men to work on the tasks that will improve gender equality, broadly defined.[5]

Consistent with previous studies, we find that men are more likely than women to choose a more difficult task. We further find that task choices for women are driven by their risk preference and competition preference, while task choices for men are driven by their self-perception of their performance in the tasks. These findings have important implications. If men's task choices are mainly driven by performance considerations, then we would expect a drop in their earnings if low-performing men who chose the easy task were assigned the hard task they did not prefer. On the other hand, since women's task choices are driven by preference considerations – high performance women did not necessarily choose hard and rewarding tasks – they should show an increase in earnings if women who choose the easy task are assigned the hard task.

Exploring these conjectures, for men, we find that the earnings of those who choose the hard tasks are significantly higher than those who choose the easy task, regardless of

[4]For instance, Niederle and Vesterlund (2007) find that men and women earn the same in piece rate, yet men are more likely to select a competitive pay scheme (i.e., tournament). However, they do not further compare whether men and women earn differently in a tournament context, which would in any case not distinguish whether earnings are due to participant preference or the compensation scheme.

[5]While we focus on the effect of task choice and task assignment on the gender earnings gap, earnings and jobs are closely associated with other forms of gender inequality.

their task assignment. In contrast, the earnings of women who are assigned the difficult task are significantly higher than those who are assigned the easy task, regardless of their preference for tasks. These results suggest the task choice effect on earnings of men but not that of women. Yet, these results also suggest a task assignment effect on the earnings of women but not on the earnings of men. Importantly, when working on the same assigned task, women who prefer the hard task show no difference in performance from women who prefer the easy task. As a result, when women who prefer the easy task are mobilized to work on the non-preferred hard and rewarding task, they are likely to earn significantly more than they would have otherwise earned in the easy tasks. These are striking findings that have not been documented in previous studies.

Our findings make two important contributions to the literature. First, we identify differential effects of task choice and task assignment on men and women. Our findings suggest that the gender earnings gap could partially be driven by the fact that high ability women may not choose a difficult, better-paying task due to preferences differences between genders. This is different than the pattern of men's task choice, which is based on performance considerations. It further implies that policies and mechanisms to decrease the gender earnings gap should take into consideration that the driving factors that affect the job choices and earnings outcomes for women and men are different. We do not include gender competitions or tournament payment schemes in our research design to minimize the influence of gender differences in engaging in competition (e.g. see a review by Niederle and Vesterlund 2011) on their task selection process and task performance (Gneezy et al. 2003 and Shurchkov 2012). To identify the task choice effect and task assignment effect and to investigate whether mobilizing women to work on the hard and rewarding task is an effective way to reduce the gender earnings gap, our study is designed as a first step toward exploring this important but unanswered question. Our new findings and the identified effects are consistent with the wisdom and results in the broader gender literature, both of those exploring gender differences in competitive environments (e.g., Niederle and Vesterlund 2007) and those in a noncompetitive setting (e.g., Cook et al. 2020), in that the gender differences in task choices should lead to considerable subsequent effects on their earnings.

Second, the results of our study provide scholarly support for workplace mechanisms

that provide greater opportunity for female employees to obtain more challenging and rewarding positions.[6] While this is a normative argument that is not new in and of itself, scholarly evidence in support of this argument is still limited in the political sphere. Several previous studies examine the impact of Affirmative Action (i.e., gender quotas) in a tournament context. In one study, gender quotas are found to encourage women to enter a tournament environment (Balafoutas and Sutter 2012). Other studies find that introducing the quota in the later stage makes the quota more effective (Czibor and Dominguez Martinez 2019; Maggian et al. 2020). These studies further find that quotas encouraging female entrants do not hurt the selection of top performers in the tournament. While these studies examine the effectiveness of gender quotas in encouraging participation, they do not directly address the issue of the gender earnings gap, and thus cannot answer whether (and in what ways) the earning difference can be improved by mobilizing women to work on a different task. Ours is the first study, to the best of our knowledge, that directly tests whether assigning women to a male-preferred difficult task increases their earnings and decreases the gender wage gap.

The paper proceeds as follows. The next section reports our experimental design. We analyze the experimental results and interpret their implications in Section 3. Our discussion and robustness checks are reported in Section 4. We conclude in Section 5.

2 Research Design

In our study, we examine the impact of gender differences on task choices and integrate these findings into an experimental design that explores the relationship between task choice, task assignment, and earnings by gender. We elicit individuals' task choices, then establish a counterfactual to observe how women (and men) *would have performed had their task choice been different.*

To measure task choices, we ask participants whether they prefer to work on an easy and lower-paid piece-rate task or a challenging but higher-paid piece-rate task. These tasks are neutrally framed as Task A and Task B in the experiment. One of the strengths of our research design is that after eliciting the participants' preferred tasks, a computer

[6]The qualitative implications of our results focus on the examination of the mechanism behind policies rather than the policies *per se*.

program randomly assigns tasks such that about 55% of participants are assigned their preferred tasks, while about 45% of the participants are assigned to their non-preferred tasks. Thus, while participants endogenously choose their preferred task, the assignment of the task is *exogenous* and random.

The procedure of our experiment is outlined in Figure 1. First, in Step 1 of the experiment, participants are given sample questions of the two types of tasks they could potentially perform in the experiment. Participants are given detailed instructions on each type of task and they have five minutes to solve each type of sample question — a total of ten minutes for two types of sample questions. Participants are not paid for solving sample questions. The sample questions are used only to inform participants about the characteristics of the two tasks.

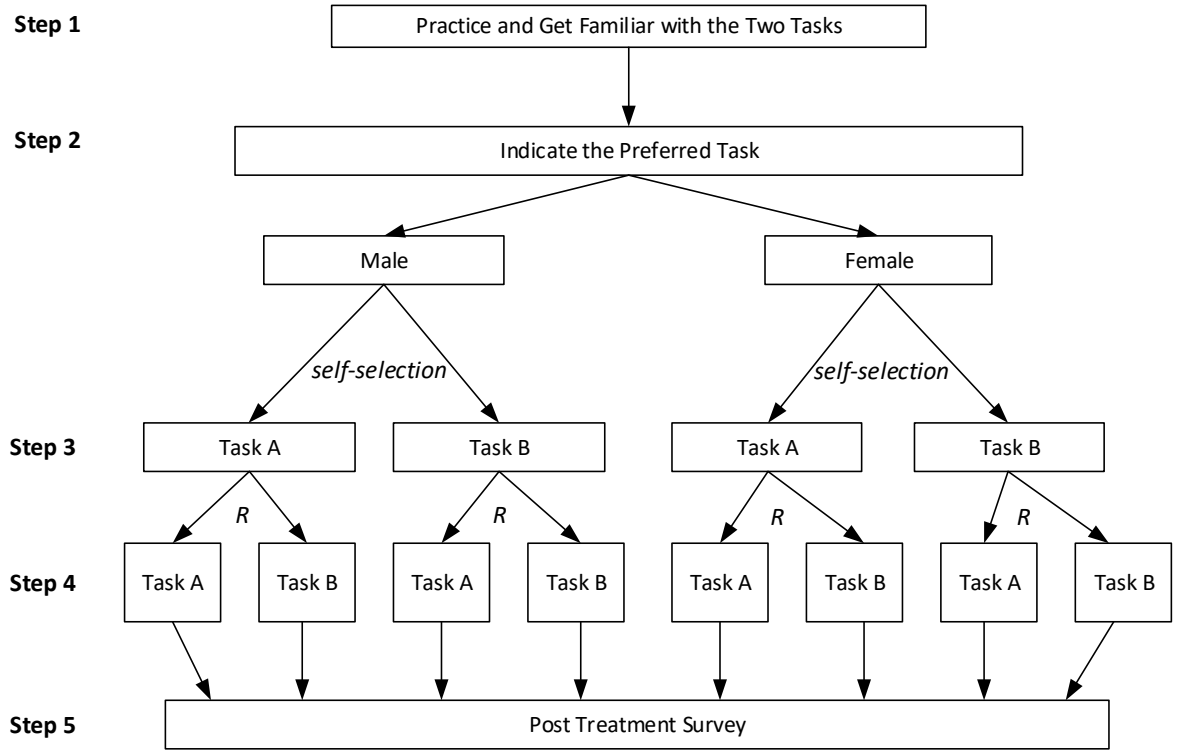
In Step 2 of the experiment, participants are given detailed information on the piece rate for solving each type of task and asked to indicate which task they would prefer to work on. For solving each question correctly in Task A (Task B), a subject is paid a low (high) piece-rate wage.[7] There is no uncertainty about the payment scheme. Conditional on the task assignment, the question difficulty level remains constant throughout the session. Prior to indicating their task choice, participants are informed they may or may not receive their preferred task. It is important to note that, while the assignment of a task is exogenous and random in our experiment, participants have an incentive to truthfully choose their preferred task because they are informed they are more likely to be assigned the task that they prefer to work on.[8]

In Step 3 of the experiment, after participants indicate which task they preferred, the computer program randomly assigned each participant their task. In Step 4, participants

[7]We calculated the piece rate for each task based on individual performance in completing tasks of varying difficulty in the pilot sessions. We chose two tasks that subjects can handle but clearly varied in difficulty levels. Online Appendix B provides details of the pilot sessions. Subjects in our experiments were not informed of our calibration of the wage in the pilot sessions since we wish to limit the role of this calibration information may have played on their task choice and performance. Additional information about how payments were calibrated might elicit choices according to social conformity and/or social comparison, despite our efforts to control those behaviors. Moreover, we also do not know whether this information about how payments were calibrated would interact with preferences and characteristics to affect choices and performance.

[8]We told participants that they were more likely to be assigned their preferred task, but withheld the exact likelihood until after they chose a task. We avoided giving an exact number because misperceptions about probabilities can influence decisions involving probabilities (e.g. Kahneman and Tversky 1979; Albert 2003 and Snowberg and Wolfers 2010).

work individually on their assigned tasks by solving as many questions as they can within 10 minutes. When Step 4 is completed, there is a short post-treatment survey in which we collect demographic information and general risk and competition preferences.



Note: The "*R*" before step 4 stands for *randomization*.

Figure 1: Illustration of Experimental Design

2.1 Identification Strategy

A stylized example, illustrated in Figure 2, captures the heart of our identification strategy.[9] By controlling for self-selection on task choice, we can identify whether differences in earnings are driven by men's and women's task choices versus their actual task assignments. Based on our novel design, we can isolate several distinctive channels that may contribute to the gender earnings gap.

Task Choice Effect: We isolate the different outcomes that are caused by different

[9]The sample size of each subgroup is indicated in the triangles. The sample size of Experiment I (II) is reported outside (inside) of parentheses. We report more information on the treatments and their sample sizes in the following sections.

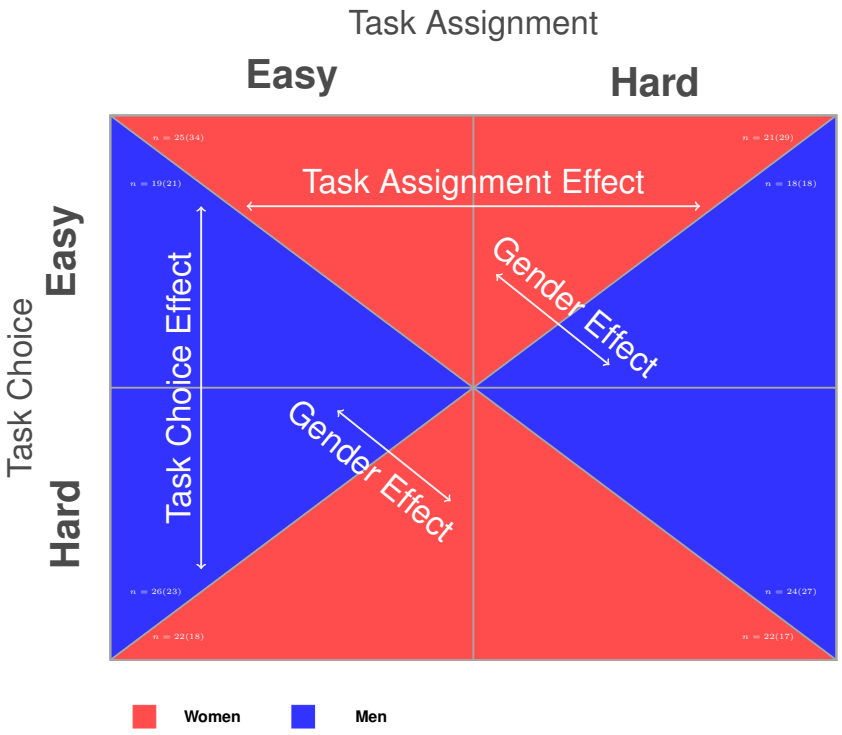


Figure 2: Intuition for Identification Strategy

choices of the easy and hard task. Specifically, we compare earnings within men (women) who are assigned the same task, yet prefer different tasks.

Task Assignment Effect: We identify whether any earnings difference is driven by the task assignment. Specifically, we compare earnings within men (women) who prefer the same task yet are assigned to different tasks.

Gender Effect: We disentangle the earnings difference between men and women that are not correlated with Task Choice Effect or Task Assignment Effect. Specifically, we compare earnings between men and women who prefer the same task and are assigned the same task. We also explore the determinants of the difference of task choice between men and women, which may indirectly contributes to the observed earnings difference.

It is worth clarifying that subjects who chose the same tasks can be randomly assigned to different tasks, comparing the performance of subjects who were assigned their preferred tasks with the performance of subjects who were assigned their non-preferred tasks, we can generate an appropriate counterfactual to observe how individuals *would have performed had their task choice been different*. The randomization strategy allows us to study performance differences at the aggregate level and investigate the average

task choice effect and the task assignment effect. Conditional on the same task choice, we controlled for the hidden characteristics (e.g. including ability) that may explain task choice and earnings differences.

Overall, by comparing the differences within each gender subgroup and between the two gender subgroups, we can separate potential mechanisms driving the gender earnings gap. Through our unique identification strategy, we can test whether a simple policy intervention works for gender equality in earnings. We explore our research questions in a non-competitive setting for two reasons. First, women have been shown to avoid competition more frequently than men do (see Niederle and Vesterlund 2011 for a review). Thus gender differences in willingness to compete would influence task choice, in addition to task difficulty, which confounds the identification of the task choice effect. Second, a competitive setting is inappropriate for exploring the task assignment effect because gender differences in willingness to compete may cause men and women to perform differently. For instance, men are more willing to compete and tend to outperform women in competitive settings (e.g., Gneezy et al. 2003 and Shurchkov 2012). Thus, if we observed men and women performed differently on different tasks, we could not identify whether the competitive setting or the task assignment effect caused the performance differences.

2.2 Experiments and Tasks

Previous literature demonstrates that task stereotypes influence men's and women's decisions, and potentially their earnings (e.g. Dreber et al. 2014 and Gneezy et al. 2003). Mathematical tasks are often considered male-typed tasks whereas verbal tasks are often considered neutral or female-typed. All else equal, while men might over-perform in a male-typed task, women perform as well as men in a female-typed task (e.g. Günther et al. 2010 and Shurchkov 2012). While the results of task stereotypes are often explored in studies with tournaments and competition, little gender differences of performance in a stylized experiment have been found under piece rate (e.g. Niederle and Vesterlund 2007). None of the previous studies have explored the task choice and task assignment effects with different task stereotypes. As a result, we investigate whether those effects influenced the gender earnings gap by conducting Experiment I with math puzzles and Experiment II with verbal puzzles. Through employing these two task stereotypes, we

can identify the extent to which stereotype plays a role in our results.[10]

2.2.1 Experiment I: Math Puzzles

In Experiment I, we use a variation of the number addition game developed by Ariely et al. (2009). In this game, participants are given a set of matrices with 12 numbers in each matrix. See Figure 3 for an example. Participants are then asked to find the numbers in the given matrix that add up to 10. Participants may solve as many matrices as they are able to within the 10-minute time limit. Participants are given one matrix at a time. Once a participant selects the numbers and submits her or his choices, the computer displays the next matrix (drawn from a prepared question set) for the participant. Participants are not informed about whether their answers are correct after they submit each question. In Task A, the objective is to select two numbers that add exactly to 10, while in Task B, the objective is to select three numbers that add exactly to 10. Since each matrix has a unique solution, Task B is relatively more challenging than Task A. At the beginning of the experiment, each participant can use up to 10 minutes for practice, with up to 5 minutes to solve 2 sample questions for each task.[11] For solving each question correctly in Task A (Task B), a subject is paid 0.4 (1.4) US dollars. After participants indicate their preferred task and the computer randomly assigns a task, participants work on the assigned task for 10 minutes. After the first 10-minute period has elapsed, participants are given another 10 minutes to work on the same assigned task but with new questions. Participants are paid based on their performance in one of these two 10-minute tasks; each 10-minute task is equally likely to be chosen.[12]

2.2.2 Experiment II: Verbal Puzzles

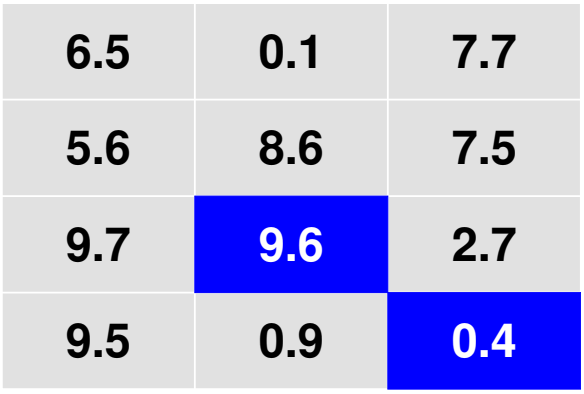
In Experiment II, we use a Word-in-a-Word puzzle where players must form sub-words from the letters of a larger puzzle word.[13] Performance is measured based on the number

[10] We conducted a survey to explore subjects' beliefs of the task stereotypes. The result of this survey is reported in the Online Appendix A4.

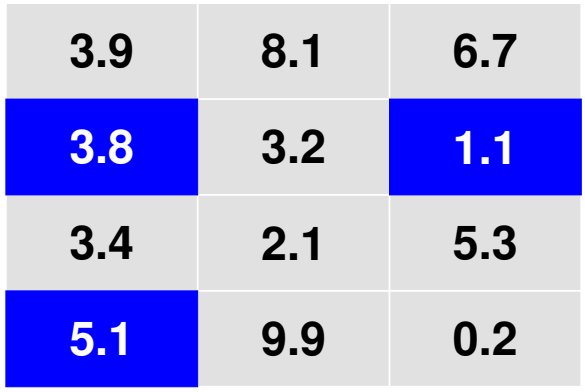
[11] We controlled for the number of practice questions to avoid potential differences in learning between women and men.

[12] We can investigate the difference of learning effects between men and women by comparing their performance in the First Try and with it in the Second Try. We explore more about this point in Online Appendices A1 and A2.

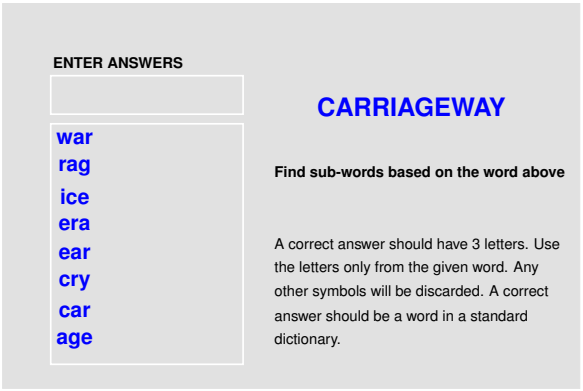
[13] All puzzles were computerized using z-Tree Fischbacher (2007). The framework of our verbal tasks is similar to the verbal task used by Shurchkov (2012). The rules of the verbal tasks are similar to the



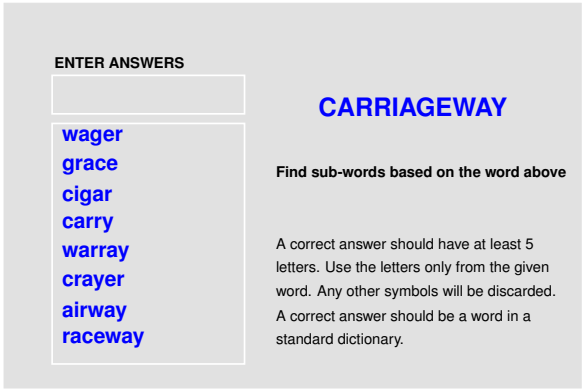
(a) Task A (Math)



(b) Task B (Math)



(c) Task A (Verbal)



(d) Task B (Verbal)

Figure 3: Examples of Task A (Easy) and Task B (Hard)

of sub-words entered. Only the words that can be found in a standard dictionary are counted. Participants' performance is measured based on the total number of correct sub-words they found within the 10-minute time limit. In both Task A and Task B, participants are given the same puzzle words. Task A and Task B differ in terms of the number of letters required in a sub-word.

In Task A, participants are asked to find three-letter sub-words. As an example, consider the puzzle word carriageway. A three-letter sub-word from *carriageway* could be (but are not limited to) "age, car, cry, or war." However, in Task B, they are asked to find the sub-words that consist of at least 5 letters. For example, for the puzzle word *carriageway*, correct sub-words include (but are not limited to) "cigar, grace, raceway, or carriage." For solving each question correctly in Task A (Task B), a subject is paid 0.1 (0.25) US dollars. Screenshots, instructions, and detailed scoring rules of these two verbal tasks can be found in the Online Appendices C2 and C7. Based on the rules of the Word-

games provided by the website www.wordplays.com.

in-a-Word puzzle games and the empirical results we observed in our pilot sessions, Task B is more difficult than Task A. At the beginning of the experiment, each participant was given a total of 10 minutes, with up to 5 minutes to solve a practice puzzle of each type of task. The sub-words solved in the practice are not paid. After participants indicate their preferred task and the computer randomly assigns a task, participants work on the assigned task for 10 minutes. They are paid based on their performance within the time limit. There is no repetition of the 10-minute workload.

In both Experiment I and Experiment II, Task A and Task B are designed in a way such that the essentials are similar, but the levels of difficulty differ. Following the literature (e.g., Fershtman and Gneezy 2011; Kuhnen and Tymula 2012 and Shurchkov 2012), for both experiments, we use participants' performance in practice as a baseline measurement of ability in solving tasks. As stated in Introduction, we wish to make it salient to the participants that Task A is relatively easier (henceforth, the easy task) while Task B is relatively harder (henceforth, the hard task). Moreover, since we used the same design of the easy and hard tasks and the same randomization strategy in both Experiment I (math tasks) and Experiment II (verbal tasks), the findings of the two experiments can be compared qualitatively. Then we can investigate the role of task stereotypes in the interpretation of our findings. As we use a piece-rate payment scheme and there is no uncertainty about the value of each correctly-solved question for each task, we avoid any risk preference impact on participant calculations. It also makes the experiment simple for the participants to understand and enables us to use the participants' earnings to investigate efficiency.

2.3 Subjects and Procedures

The experiments were conducted at the Experimental Social Science Laboratory at Florida State University with subjects who are registered undergraduate students at the university. The experiments were programmed using z-Tree (Fischbacher 2007) and conducted via a computerized network. Subjects were seated at individual computer terminals and not allowed to see other subjects' choices.[14] A total of 508 undergraduate student sub-

[14]Experiment I was conducted in the laboratory. Experiment II and Experiment III were conducted over the internet using unleashed z-Tree (Duch et al. 2020). Experiment III is identical to Experiment I except that participants are always assigned the task that they choose and thus absent of any ran-

jects participated in our study. One hundred and seventy-seven subjects participated in Experiment I, one hundred and eighty-seven subjects participated in Experiment II, thirty-five subjects participated in Experiment III (reported in detail in the Online Appendix A4), sixty-seven subjects participated in our pilot sessions, and forty-two subjects who did not participated in Experiments I, II, or III answered our survey on task stereotypes. No subject participated in more than one session. Average earnings were about \$17 (\$11) in Experiment I (II) in which a \$5 show-up payment was included. Sessions for Experiments I and II included 10 to 25 participants. We report additional information on the experiment instructions and how we conduct the pilot sessions in the Online Appendices B1, B2, and C1.

3 Empirical Analysis

In this section, we present the main results from our experiment regarding the task performance and earnings for the full sample and earnings subsetted by gender. We first report the aggregate results by gender and then report the results of our analysis of the effect of assigning men and women to different tasks. Given the higher difficulty of the hard task, we expect subjects assigned to this task to solve fewer questions correctly. We note that a subject assigned to the easy task may solve enough questions such that the earnings of the two tasks are not significantly different. In this study, subjects' earnings are a linear function of their performance on the assigned task. To facilitate our comparison of the treatment effects, in the following discussion, we use subjects' earnings as our main outcome variable. Unless otherwise noted, we use the non-parametric Mann-Whitney statistical tests to examine the differences between subgroups.

We begin by examining our aggregate results related to male and female participants' task performance and earnings levels. We first examine the Task Choice Effect and investigate whether those who prefer the hard task earn more than those who prefer the easy task.[15] Examining the results of Experiment I (math) by gender, we find that

domization process. Details are reported in the Online Appendix A3. Participants of both Experiment I and Experiment II/III were recruited from the same subject pool of the XS/FS Experimental Social Science Laboratory. We implemented the same standard of experimental economics and used the same protocols to conduct those experiments. Additional details of the experimentation are reported in the Online Appendix C3.

[15]As stated, in our Experiment I, each participant is assigned a task that he or she works on in two

the average earnings of male participants assigned their preferred hard task are \$13.8, which is significantly higher than the \$7.4 earned by male participants who prefer the easy task yet assigned the hard task ($p < 0.001$). Similarly, we find that the average earnings of men who prefer the hard task yet assigned the easy task are \$13.7, which is significantly higher than the \$10.9 earned by male participants assigned their preferred easy task ($p = 0.044$).

We find qualitatively the same pattern on the Task Choice Effect in Experiment II (verbal) for men. The average earnings of male participants assigned their preferred hard task are \$8.3, which is significantly higher than the \$3.3 earned by male participants who prefer the easy task yet assigned the hard task ($p = 0.002$). The average earnings of men who preferred the hard task yet assigned the easy task are significantly higher than the earnings of male participants assigned their preferred easy task (\$7.0 versus \$4.6, $p = 0.012$).

These results suggest that our male participants' preference for the hard task is correlated with a generally better performance in solving the questions in our study, regardless of the task to which they are assigned. Moreover, our data also suggests that the task choice effect on men is consistent across task stereotypes. This leads us to state our first result.

Result 1 (Task Choice Effect) *Male participants' earnings are correlated with their initial task choice. Those who prefer the hard task on average perform better, and thus, earn more than those who prefer the easy task. This result is robust to whether they are assigned the easy or the hard task and whether they work on the math tasks or verbal tasks.*

We further examine the task choice effect among our female participants and find that in Experiment I female participants who work on their preferred hard task earn \$13.3, which is not significantly different from the \$13.2 earned by those who prefer the easy task yet assigned the hard task ($p = 0.92$). We further find that female participants

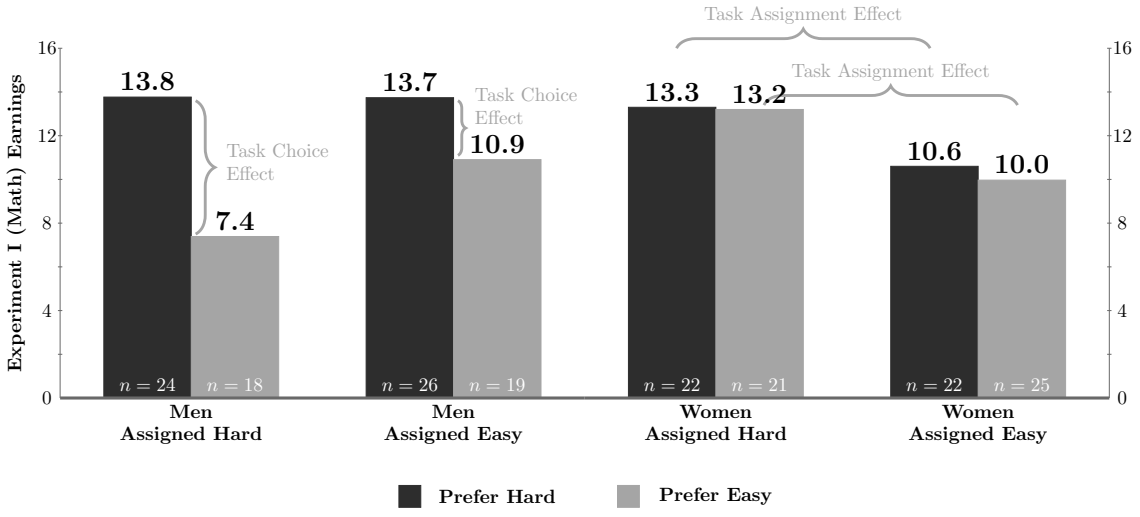
subsequent 10-minute sessions, each with a new set of questions. The results of Experiment I reported in this section are the results for the first 10 minutes. The results of each 10 minutes are qualitatively the same. Whether we pool the data of the 20 minutes together or analyze them separately, the main findings of our study are consistent.

who prefer the hard task yet assigned the easy task earn \$10.6, which is not significantly different from the \$10.0 earned by those assigned their preferred easy task ($p = 0.68$). The results of Experiment II demonstrate a similar pattern among women. That is, we do not find any significant differences between the earnings of those who prefer the hard task and work on the hard task and those who prefer the easy task and work on the hard task (\$8.4 versus \$7.5, $p = 0.73$). We also do not find any evidence that female participants who prefer the hard task yet assigned the easy task earn significantly different from those assigned their preferred easy task (\$4.7 versus \$4.1, $p = 0.15$). In other words, female participants assigned the same task earn statistically the same regardless of their initial task choices. These results suggest that, on average, female participants' task choices are not correlated with their earnings, which is different from the result obtained for our male participants. This leads to our statement of Result 2, followed by our depiction of Results 1 and 2 in Figure 4.

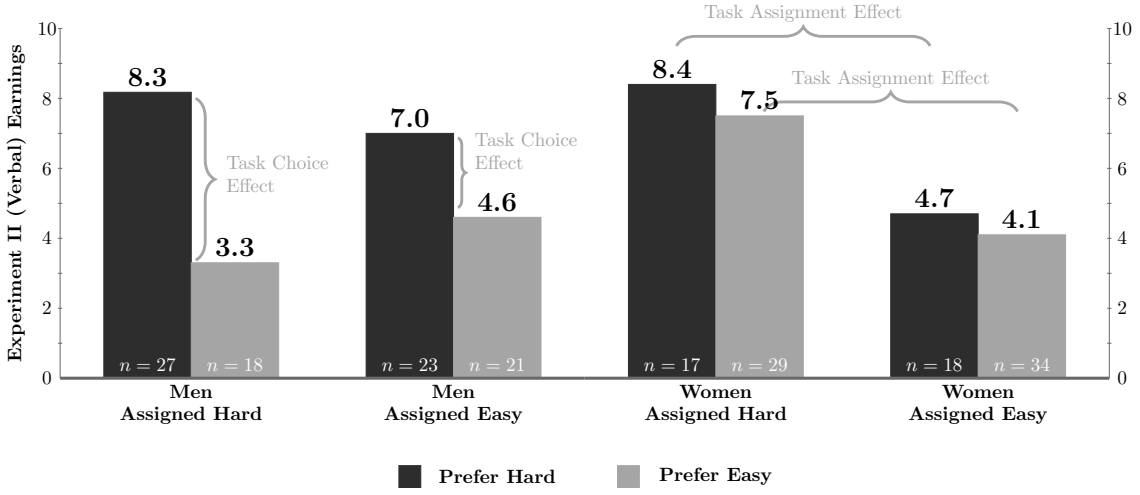
Result 2 (Task Choice Effect) *Female subjects' earnings are not correlated with their initial task choice. When working on the same assigned task, female participants who initially prefer the easy task perform (and thus earn) no differently from female participants who initially prefer the hard task. This result is robust to whether they work on math tasks or verbal tasks.*

We next examine the Task Assignment Effect and investigate whether being assigned a hard task makes a difference in men's and women's earnings in math and verbal tasks. The results of Experiment I show that male participants who prefer the hard task earn about the same when they are assigned the hard as compared to when they are assigned the easy task (\$13.8 versus \$13.7, $p = 0.84$). By contrast, male participants who prefer the easy task earn significantly less when they are assigned the hard as compared to when they are assigned the easy task (\$7.4 versus \$10.9, $p = 0.015$).

Examining the results of Experiment II, we continue to find those male participants who prefer the hard task earn about the same when they are assigned the hard as compared to when they are assigned the easy task (\$8.3 versus \$7.0, $p = 0.39$). Male participants who prefer the easy task earn less when they are assigned the hard as compared to when they are assigned the easy task, but the differences in earnings are not statistically



(a) Experiment I (Math)



(b) Experiment II (Verbal)

Note: The numbers at the top of the bars are the average dollar amount earned by the participants under that bar. The numbers at the bottom of the bar are the sample size of that treatment. The label below each set of the bar shows whether the participants are assigned to the hard or the easy task. The two left (right) sets of bars reflect male (female) earnings.

Figure 4: Earnings for Participants who Prefer the Hard Task (dark grey) versus those who Prefer the Easy Task (light grey)

significant ($3.3 versus $4.6, $p = 0.11$). While the task assignment effect on earnings of male participants who prefer the easy task is significant only when they work on math tasks, the direction that the task assignment intervention has a negative effect on earnings of those who prefer the easy task is the same across Experiment I and Experiment II.

It is worth noting that, in both Experiment I and Experiment II, the easy and hard tasks require the same solving process and differ only in the level of difficulty. Hence, it is not surprising to find those male participants who prefer the hard task earn the same regardless of their task assignment. However, male participants who prefer the easy task are not necessarily capable of solving the hard task and thus earn significantly more when they are assigned the easy task than when they are assigned the hard task. This leads to our statement of our third result.

Result 3 (Task Assignment Effect) *Male participants' earnings, whether they are assigned easy or hard tasks, are not statistically different if they prefer the hard task. Their earnings are lower if they prefer the easy task but are assigned the hard task, although the earnings' differences are only significant in the math tasks but not statistically distinguishable in the verbal tasks.*

We now turn to the task assignment effect for our female participants. The results of Experiment I show that female participants who are assigned the hard task earn more than those who are assigned the easy task, whether they prefer the hard task ($13.3 versus $10.6, $p = 0.057$) or the easy task ($13.2 versus $10.0, $p = 0.041$). The results of Experiment II show a similar pattern. That is, female participants earn more when they are assigned the hard task, whether they prefer the easy task ($7.5 versus $4.1, $p = 0.001$) or the hard task ($8.4 versus $4.7, $p = 0.04$). The results further suggest that female participants are able to adapt when assigned a task, regardless of which task they initially preferred. In particular, it is likely that female participants with high solving ability in our math and/or verbal tasks who may prefer the easy task will perform well if assigned the hard task, leading to an improvement in their earnings. This leads to our statement of our fourth result.

Result 4 (Task Assignment Effect) *When female participants are assigned the hard*

task, their earnings are higher as compared to being assigned the easy task, whether they prefer the easy task or the hard task. This result is robust to whether they work on math tasks or verbal tasks.

These reported aggregate task choice effect and task assignment effect are subject to the influence of participants' ability in solving those tasks. To examine the extent to which our findings are influenced by participants' differences in ability, we conduct OLS regressions and analyze the influence of an individual's preferred versus assigned task on earnings. The results are reported in Table 1. *PreferHard* is a binary indicator that reflects whether a subject prefers the hard task, *AssignedHard* is a binary indicator that reflects whether a subject is assigned the hard task. *TaskAPractice* and *TaskBPractice* are the number of questions solved correctly in the practice phase, which we use as an estimate of an individuals' ability in solving those tasks. *SelfPerception* is individuals' self-reported belief of the percentage of questions they have answered correctly. This information is elicited after participants have finished the assigned task yet before they are informed about their actual performance and final earnings.

The results reported in Table 1 are consistent with the task choice effect and task assignment effect, suggesting that our main findings are robust to the check of participants' practice performance in solving these tasks. That is, for female participants, their earnings are affected by whether they are assigned the hard task regardless of which task they initially prefer. For instance, in a 10-minute session, depending on whether working on math or verbal tasks, women assigned the hard task earn about \$3.3 to \$3.4 more than those assigned the easy task. We further observe that whether we control for the number of solved questions in Task A and B in the practice phase, the task assignment effect has a consistently significant impact on female participants' earnings. However, whether they prefer the easy task or the hard task has little correlation with their earnings. This contrasts with the findings from our male participants, who show a positive task choice effect on their earnings. Depending on whether working on math or verbal tasks, men who prefer the hard task earn \$2.3 to \$2.4 more than those who prefer the easy task. Moreover, we find a positive correlation between the number of questions solved correctly in the practice stage and higher earnings beyond that related to *PreferHard* or *AssignedHard*, suggesting that participants' earnings are correlated with their baseline

Table 1: Analysis of the Influence of Preferred and Assigned Task on Earnings

Variables	Panel A: Female							
	Math				Verbal			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PreferHard			0.298				0.058	
			(1.132)				(0.991)	
AssignedHard				3.426***				3.293***
				(0.931)				(0.721)
SelfPerception	0.054***		0.038	0.073†	0.031		0.044**	0.027
	(0.020)		(0.036)	(0.037)	(0.020)		(0.021)	(0.019)
TaskAPractice		0.688	0.409	0.134		0.053	0.088	0.099
		(1.150)	(1.458)	(1.277)		(0.069)	(0.069)	(0.067)
TaskBPractice		1.572**	1.153	0.579		0.260†	0.223	0.205†
		(0.646)	(0.932)	(0.724)		(0.141)	(0.139)	(0.115)
Control Varibles	No	No	Yes	Yes	No	No	Yes	Yes
Observations	90	90	90	90	98	98	98	98
R-squared	0.058	0.077	0.329	0.447	0.019	0.097	0.226	0.376
Variables	Panel B: Male							
	Math				Verbal			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PreferHard			2.335**				2.437**	
			(1.103)				(0.946)	
AssignedHard				-1.456				0.206
				(1.054)				(0.827)
SelfPerception	0.070**		0.039†	0.030	0.107***		0.022	0.075***
	(0.029)		(0.021)	(0.026)	(0.020)		(0.028)	(0.024)
TaskAPractice		5.622***	3.439**	3.902**		0.057	0.069	0.052
		(0.529)	(1.603)	(1.711)		(0.091)	(0.091)	(0.096)
TaskBPractice		2.588***	0.597	1.550**		0.312***	0.219**	0.258**
		(0.710)	(0.804)	(0.774)		(0.095)	(0.109)	(0.113)
Control Varibles	No	No	Yes	Yes	No	No	Yes	Yes
Observations	87	87	87	87	89	89	89	89
R-squared	0.054	0.117	0.555	0.536	0.187	0.192	0.389	0.350

Note: OLS specification. Dependent variable: Earnings in US dollars.
Panel A reports the results of female subjects and Panel B reports the results of male subjects. *PreferHard* and *AssignedHard* are dummy variables that represent the subjects' choices and the assigned task. Control variables include Ethnicity, AdvantagedBackground (dummy variable= 1 in Experiment I (math) if the subject's major involves calculations; dummy variable= 1 in Experiment II (verbal) if the subject is a native speaker), session variables (i.e. session effects caused by gender compositions detailed in the Online Appendix C4), and cognitive reflection test scores (based on Frederick 2005 detailed in the Online Appendix C6).
Robust standard errors in parentheses.
Level of significance: *** $p < 0.01$, ** $p < 0.05$, † $p < 0.1$.

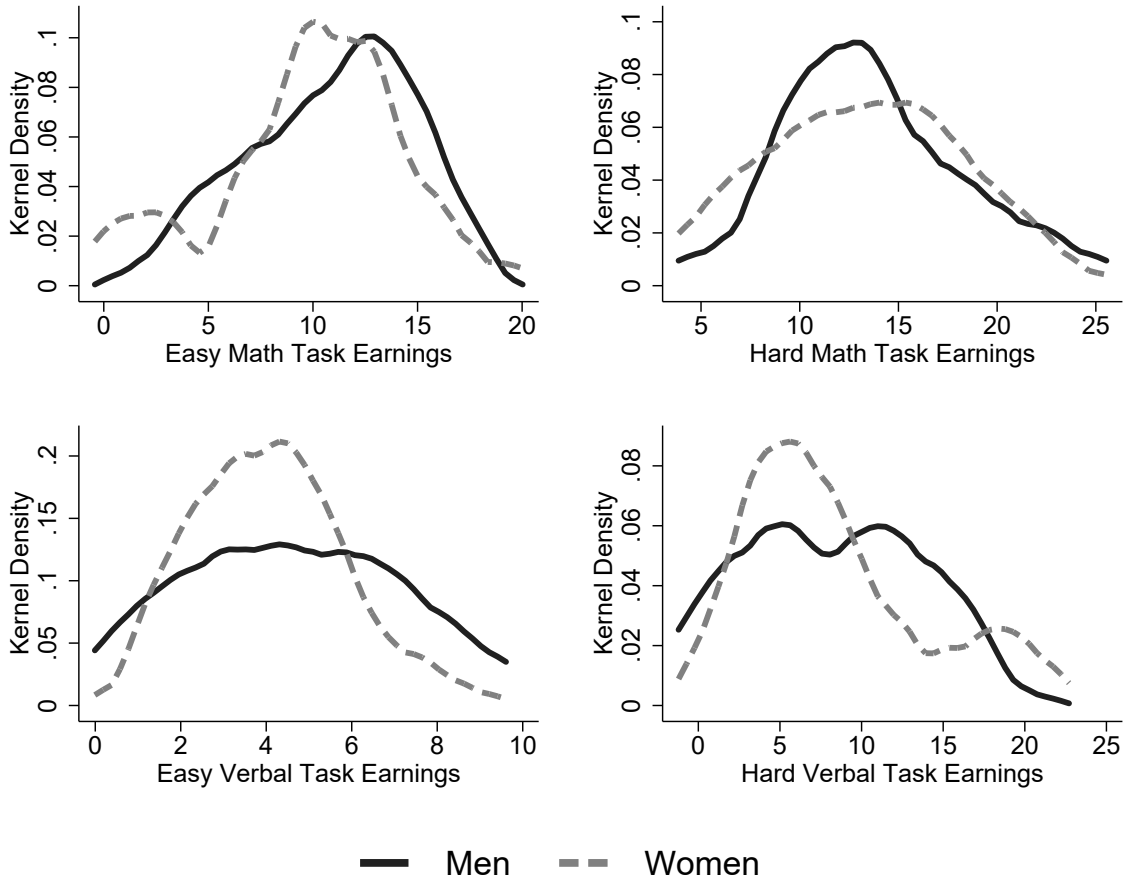
ability in solving questions, which is most evident in their performance in solving the hard questions (Task B) in our experiments. Although the correlation is not always statistically significant (i.e. p-value$<$ 0.05), the pattern of the positive correlations is the

same across task stereotypes and genders. We also find a positive correlation between self-perception of performance and earnings, but it is not always statistically significant when being combined with other driving factors.

To further explore our findings that women earn more on the hard task regardless of the task they initially preferred, whereas men who prefer the easy task earn less if assigned to the hard task, we consider whether our female and male subjects differ in their ability to solve the task. To examine the Gender Effect, we first compare the earnings of women and men who are assigned their preferred hard task and find that their average earnings are statistically indistinguishable both in Experiment I (\$13.8 versus \$13.3, $p = 0.88$) and in Experiment II (\$8.3 versus \$8.4, $p = 0.99$). Similarly, when we compare the earnings of women and men assigned their preferred easy task, we find that their earnings are statistically indistinguishable both in Experiment I (\$10.9 versus \$10, $p = 0.45$) and in Experiment II (\$4.6 versus \$4.1, $p = 0.42$). In Figure 5, we report the Kernel density distribution of men's and women's earnings. An F-test comparing the standard deviations across women and men and a Kolmogorov-Smirnov (K-S) test comparing the general distribution of earnings each show no significant difference in the distribution of earnings between women and men at a conventional significance level. Overall, our results are consistent with the findings of Niederle and Vesterlund (2007) and Dreber et al. (2014) that, in general, men and women exhibit similar abilities in solving simple experimental tasks in non-competitive environment. We now state our fifth result.

Result 5 (Gender Effect) *Men and women who are assigned their preferred tasks earn the same amount of earnings in both math and verbal tasks.*

Although there is no gender difference in performance, we find clear evidence that men are more likely to choose the hard and higher-paid task than women. That is, among the 364 participants who participated in Experiment I or Experiment II, 58% of men choose the hard task, but only 43% of women choose the hard task (Pearson chi-square test, $p = 0.005$). When we further break down the choices by task stereotypes, we find mixed results. In Experiment I (math), 57% of men choose the hard task while 49% of women choose the easy task (Pearson chi-square test, $p = 0.253$). In Experiment II (verbal), 56% of men choose the hard task while 36% of women choose the easy task (Pearson chi-



Note: The horizontal line shows earnings by gender while the vertical line shows the Kernel density. The panel on the left (right) shows the earnings density distribution for men and women who are assigned their preferred easy (hard) task.

Figure 5: Distribution of Earnings by Gender and Task Assignment

square test, $p = 0.005$). Despite the insignificance of gender difference in task choice in Experiment I, the direction of the choice difference between men and women is consistent with that in Experiment II. Our findings are consistent with the observations reported in Niederle and Yestrumskas (2008) and Bracha and Fershtman (2013) who find men are more likely to choose (or spend longer time on) the more challenging task. This leads to our sixth result.

Result 6 (Gender Effect) *In general, men are more likely to choose the challenging and higher-paid tasks than women, even if men and women have no difference in their performance in such tasks.*

Based on Results 5 and 6, we see that men and women perform about the same

in both easy and hard tasks regardless of in Experiment I or Experiment II, but men choose hard and higher-paid tasks more frequently than women. It is likely that some women, whose performance are similar to men shy away from more challenging tasks and are thus under-represented in such tasks given their ability. All else equal, the under-representation is likely to result in gender differences in economic outcomes and potentially lead to a significant gender earnings gap. Importantly, our results imply that women's earnings are shaped by the task assignment rather than their task choice. We find no evidence showing that female participants' task choices are correlated with their potential performance and consequent earnings. By contrast, we find that men's earnings are shaped by their task choice. From a policy perspective, our findings suggest that *enabling women to take on more challenging tasks can improve their earnings, but motivating men to do so may have little or even a negative impact on theirs depending on their task preference.*

Our results also lend insight to gender quota policies by providing scholarly data that supports the idea that increasing and reserving more positions and opportunities for women in challenging and rewarding jobs may improve the earnings of women without hurting the interests of men. From a workplace standpoint, the effectiveness of a quota system (e.g. Affirmative Action) is controversial (Balafoutas et al. 2016). Debates have resulted in some states banning affirmative action policies. The main criticism is the concern that such policies may lead to reverse discrimination (e.g. enhance women's welfare at the cost of men). However, our results suggest that such policies may not necessarily hurt the benefits of men at the aggregate level, since the men in our study who preferred the harder tasks earned the same payment whether they are assigned the hard task or the easy task. It is essential to notice that reserving opportunities for competent women should not affect competent men who prefer the hard task since the task assignment effect has little influence on their earnings. Given the observation that the task assignment intervention seems to negatively affect the earnings of those men assigned the hard task but prefer the easy task, analyzing the differential effects of task choice and task assignment on men and women could be beneficial to evaluate policies such as Affirmative Action. Based on the results of our study, task assignment intervention is likely to effectively reduce the gender earnings gap between men and

women. To facilitate our solutions to addressing the research questions proposed in the Introduction, there are no gender competitions in our setting. Our results may not explain the gender earnings gap in labor markets in which there is serious competition between genders. While there are a number of differences between our laboratory experimental setting and the workplace context, our results suggest that policies aimed at reducing the gender earnings gap may apply different strategies toward men and women.

4 Discussion

The results of our experiments show evidence of a task choice effect on men's earnings and task assignment effect on women's earnings. We find that women under-represented in harder and higher-paying jobs can result in a gender earnings gap, even in an environment absent of competition. Moreover, across different task stereotypes, women are likely to earn significantly more if they work on the harder and higher-paid task that most women in our study did not choose in the first place. In this section, we address some additional results from our experiments and examine a few explanations that may contribute to the differential task choice and task assignment effects on men and women. First, we examine whether factors including risk preferences, willingness to engage in competition, and self-perception affect men's and women's task choice decisions differently. Second, we exploit the participants' performance in the practice to investigate whether the choice differences between men and women cause differences in earnings in the experiment.

Before continuing our discussion, we note that in the Online Appendix A3, we report the results of a robustness check, Experiment III, that we conducted to explore whether being assigned the preferred task in a randomization procedure in Experiment I may lead participants to perform differently. Experiment III is identical to Experiment I except that participants are always assigned the task that they choose and thus absent of any randomization process. By comparing the performance results of those assigned their preferred tasks in Experiment I with the performance results of participants who always work on the task they choose in Experiment III, we find little evidence that being assigned to their preferred task plays a role in performance. Moreover, we find that our main results are robust to session effects regardless of the stereotype of the task and we

report those additional results in the Online Appendix C4.

4.1 Determinants of Men's and Women's Task Choices

We start our discussion by examining the conjecture that men and women may make task choice decisions based on different considerations. A number of studies have found that men and women differ in their risk preferences, willingness to engage in competition, and level of confidence (see a review by Croson and Gneezy 2009). Given that men and women have similar performance in solving the easy and hard tasks, but men are significantly more likely to choose the hard and rewarding tasks than women, it is natural to suspect that women's choices are driven by something different from that of men's. To explore this conjecture, we develop a Logit choice model and use the data from our unincentivized post-experiment survey to measure risk preference, willingness to engage in competition and self-perception of performance (as a proxy of confidence).

The results of the Logit regressions are reported in Table 2. Female is a binary indicator to represent a subject's gender, *TaskAPractice* and *TaskBPractice* is the number of questions solved correctly in the practice phase. *RiskSeeking* and *CompetitionPreferences* are continuous variables on a scale from 1 to 10, with 1 indicating the least and 10 the most. *Self-Perception* is individuals' self-reported belief of the percentage of questions they have answered correctly.

Based on the results reported in Table 2, we find that *TaskBPractice* has a positive and significant influence on the choice of the hard task among men and women, which suggests that individuals' baseline ability in solving the hard tasks are positively correlated with their choice of the hard task. The more hard questions they solve in the practice phase, the more likely they are to choose the hard task. This is true for both men and women in both math and verbal tasks.

Examining our findings by gender, we find that risk preferences have no significant effect on men's task choice, but a significantly positive effect on women's choices of the hard task. Moreover, we find *SelfPerception* influences only men's choice of a hard task, but not that of women. If we regard this self-perception as a proxy of confidence or consideration of ability, then these results suggest that men who perceived themselves good at solving the hard task are more likely to choose the hard task, but women who

Table 2: Predictive Analysis of Choosing the Hard Task (Logit Model, Marginal Effect)

Variables	Panel A: Female							
	Math				Verbal			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RiskSeeking	0.115***			0.245***	0.066***			0.066**
	(0.044)			(0.064)	(0.026)			(0.032)
CompetionPreference	-0.088**			-0.197***	0.022			0.018
	(0.041)			(0.070)	(0.025)			(0.034)
SelfPerception		0.004		-0.001		0.003		0.005
		(0.003)		(0.008)		(0.003)		(0.004)
TaskAPractice			-0.176	-0.504†			-0.006	-0.010
			(0.125)	(0.233)			(0.008)	(0.010)
TaskBPractice			0.333***	0.577**			0.043***	0.037**
			(0.090)	(0.224)			(0.015)	(0.015)
Control Varibles	No	No	No	Yes	No	No	No	Yes
Observations	90	90	90	90	98	98	98	98
R-squared	0.120	0.017	0.109	0.503	0.081	0.008	0.069	0.234
Variables	Panel B: Male							
	Math				Verbal			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RiskSeeking	0.030			0.035	0.062†			0.098
	(0.033)			(0.058)	(0.034)			(0.061)
CompetionPreference	0.015			0.008	0.003			-0.004
	(0.030)			(0.059)	(0.030)			(0.035)
SelfPerception		0.012**		0.016**		0.048***		0.063**
		(0.005)		(0.006)		(0.016)		(0.031)
TaskAPractice							-0.010	-0.018
							(0.010)	(0.018)
TaskBPractice			0.455***	0.702***			0.060***	0.038**
			(0.126)	(0.232)			(0.018)	(0.016)
Control Varibles	No	No	No	Yes	No	No	No	Yes
Observations	87	87	87	87	89	89	89	89
R-squared	0.017	0.067	0.143	0.400	0.043	0.502	0.118	0.615

Note: Logit specification. Dependent variable: Choosing the hard task. The coefficients report the marginal effects of the Logit regression at the mean.
Panel A reports the results of female subjects and Panel B reports the results of male subjects. In Panel B Columns 3 and 4, because almost all male participants who failed to answer all Task A practice questions chose the easy task, *TaskAPractice* is omitted in the regressions. *RiskSeeking* and *CompetitionPreferences* are continuous variables on a scale from 1 to 10, with 1 indicating the least and 10 the most. *Self-Perception* is a continuous variable with values between 0 and 100. Control variables include Ethnicity, AdvantagedBackground (dummy variable= 1 in Experiment I (math) if the subject's major involves calculations; dummy variable= 1 in Experiment II (verbal) if the subject is a native speaker), session variables (i.e. session effects caused by gender compositions detailed in the Online Appendix C4), and cognitive reflection test scores (based on Frederick 2005 detailed in the Online Appendix C6).
Robust standard errors are in parentheses.
Level of significance: *** $p < 0.01$, ** $p < 0.05$, † $p < 0.1$.

perceived themselves good at solving the hard task are not necessarily more likely to choose the hard task. Importantly, our research design contains no element of risk (e.g. as that in a lottery payment scheme) or competition (e.g. as that in a tournament), which means women's task choices are influenced by preferences that are independent of the nature of the tasks. Our findings imply that both women's and men's performance in solving the hard task influences their decisions to choose the hard task, but women's choices of hard tasks are further influenced by their risk preferences and sometimes their willingness to engage in competition. In contrast, neither men's risk preference nor their competition preference have any significant influence on their choice of the hard tasks. Instead, men's self-perception of performance is positively correlated with their choice of a hard task. Hence, we find suggestive evidence that men's choice of the hard tasks are likely based on considerations of whether they can perform well in the tasks, whereas women's choices are not driven by such considerations.

4.2 Women Learn Faster than Men

The main results of our study suggest that what task women work on has a significant influence on their earnings. That is, women earn more when they are assigned to a hard yet rewarding task. This leads us to expect that women may be more adaptive to a new task environment. To examine this conjecture, in this section, we investigate whether men and women differ in their learning.

In Experiment I, each subject participates in two 10-minute task sessions. After a subject completes the first 10-minute session, the computer records the subject's performance and earnings without informing subjects of their outcomes. The subject is then given the *same* task (with different questions) and another 10 minutes in which to work on the task. The performance and earnings in one of these two 10-minute blocks is randomly and equally likely chosen by the computer to determine how much pay a subject will receive from the task. For simplicity of exposition, we refer to the first 10 minutes as the *First Try* and the second 10 minutes as the *Second Try*.

We find that the average earnings in the First (Second) Try are 11.8 (12.9) dollars, suggesting that the average earnings are significantly higher in the Second Try ($p < 0.001$). Examining the results by gender, we find women (78%) tend to exhibit stronger

learning effects than men (64%, $p = 0.0496$). We find that 73% (55%) of men assigned the easy (hard) task receive more earnings in the Second Try (Pearson chi-square test, $p = 0.071$). However, for women, we find that 87% (67%) of women assigned the easy (hard) task receive more earnings in their Second Try (Pearson chi-square test, $p = 0.024$). When we break down the analysis by task choice and task assignment, the results are more nuanced. We find that when participants work on the preferred task, there are no performance (and thus) outcome differences between men and women. However, when they work on the non-preferred task, women are more likely to exhibit stronger learning effects and receive significantly higher earnings in their Second Try. We report detailed results in the Online Appendix A2. These results provide initial evidence of the gender learning differences when individuals need to adapt to an unchosen task. However, as a counterfactual comparison, these results should be interpreted with some caution. In particular, fast learners are not necessarily more adaptive in every possible scenario. While we take women's faster learning than men as evidence that they are more adaptive, we acknowledge that our measurement may not be the most accurate method. Nevertheless, our results can provide directions for future research on similar topics.

5 Concluding Remarks

In this study, we demonstrate that when women are assigned to a more challenging task, they experience an increase in earnings, regardless of their task choice or the task stereotypes. In contrast, we find that men assigned their non-preferred hard task experience a reduction in earnings. From these results, we conclude that assigning women more challenging tasks in the workplace could serve as an effective intervention to improve women's earnings and thus reduce the gender earnings gap. Furthermore, we find that on top of ability considerations, women's task choices are further driven by their preferences while men's task choices are further driven by their perceptions of their performance in the experimental tasks.

The findings of our study have strong policy implications. Our study involves tasks with an increasing level of difficulty, such as might be found in the workplace when advancing across one's career, and can especially provide insight into policies related to

cabinet or top management posts. It is also in line with previous research on the effect of the gender quota system on promoting high-performing women to engage in a tournament (Niederle et al. 2013). The evidence from our study can be used to achieve a better understanding of the gender wage gap. By documenting the earnings gains for women from challenging jobs even if those jobs are not initially preferred, it may encourage women to become more risk-seeking and more confident in challenging environments, which will ultimately decrease the gender gap. To mitigate the social and political obstacles these women may encounter in doing so, the creation of policies and environments that encourage female workers to choose challenging jobs could assist in reducing the gender earnings gap.

As with any experiment, a level of caution is needed in an interpretation of our results. First, following the studies on gender (see a recent review by Niederle 2016), we focus on the difference of earnings between men and women and examine whether mobilizing women to the challenging and rewarding tasks may function as a simple and effective intervention to reduce the earnings gap. Our investigation may be different from a utilitarian benchmark which would internalize the possible social costs (including the penalty in terms of marriage and children as demonstrated by Ong et al. 2019) into a utilitarian welfare function.[16] Also, our research is designed to isolate differential effects that may contribute to the gender wage gap. We aim to investigate whether creating environments and encouraging women to engage in challenging and rewarding tasks will increase gender equity and reduce the gender wage gap. Even within our experimental setting, participants have the option to choose whether to continue when they learn they have been assigned to a non-preferred task. In the workplace setting, while challenging and rewarding tasks should be made available and even encouraged, employers should not interpret our results as an indication that such tasks should be mandated. Meanwhile, as a first step to investigate whether mobilizing women to work on a challenging and rewarding task reduces gender earnings gap, competition is absent for the sake of a clean identification of effects. Future research may build upon our results by including competition into the compensation scheme.

[16] Barbulescu and Bidwell (2013) show that compared to men, women prefer jobs with a better anticipated work-life balance, and women have a lower identification with (or receiving offer from) stereotypically masculine jobs.

In addition, the tasks used in our experiments vary in their levels of difficulty and are conducted in a controlled laboratory experiment. In future research, it is important to explore whether the intervention of task assignment has the same effect when a broader range of participants other than student populations are included. This corresponds with the message that women should be encouraged to be more open to their job choices and be mobilized to move from one industry to another. The primary policy implication of this study is that mobilizing women to take on challenging and rewarding tasks can improve gender equality and reduce the gender earnings gap. These results from the laboratory suggest the need for a field intervention to test this relationship in a less controlled setting.

References

Albert, J. H. (2003). College students' conceptions of probability. *The American Statistician*, 57(1):37–45.

Ariely, D., Gneezy, U., Loewenstein, G., and Mazar, N. (2009). Large stakes and big mistakes. *Review of Economic Studies*, 76(2):451–469.

Azmat, G. and Ferrer, R. (2017). Gender gaps in performance: Evidence from young lawyers. *Journal of Political Economy*, 125(5):1306–1355.

Babcock, L., Recalde, M. P., Vesterlund, L., and Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, 107(3):714–47.

Balafoutas, L., Davis, B. J., and Sutter, M. (2016). Affirmative action or just discrimination? a study on the endogenous emergence of quotas. *Journal of Economic Behavior & Organization*, 127:87–98.

Balafoutas, L. and Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science*, 335(6068):579–582.

Barbulescu, R. and Bidwell, M. (2013). Do women choose different jobs from men? mechanisms of application segregation in the market for managerial workers. *Organization Science*, 24(3):737–756.

Bertrand, M. and Hallock, K. F. (2001). The gender gap in top corporate jobs. *Industrial and Labor Relations Review*, 55(1):3–21.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Bowles, H. R., Babcock, L., and Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, 103(1):84–103.

Bracha, A. and Fershtman, C. (2013). Competitive incentives: working harder or working smarter? *Management Science*, 59(4):771–781.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447.

Cook, C., Diamond, R., Hall, J., List, J. A., and Oyer, P. (2020). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *The Review of Economic Studies*.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.

Czibor, E. and Dominguez Martinez, S. (2019). Never too late: Gender quotas in the final round of a multistage tournament. *Journal of Law, Economics, and Organization*, 35(2):319–363.

Dittrich, M., Knabe, A., and Leipold, K. (2014). Gender differences in experimental wage negotiations. *Economic Inquiry*, 52(2):862–873.

Dohmen, T. and Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, 101(2):556–90.

Dreber, A., von Essen, E., and Ranehill, E. (2014). Gender and competition in adolescence: task matters. *Experimental Economics*, 17(1):154–172.

Duch, M. L., Grossmann, M. R., and Lauer, T. (2020). z-tree unleashed: A novel client-integrating architecture for conducting z-tree experiments over the internet. *Journal of Behavioral and Experimental Finance*, 28:100400.

Fershtman, C. and Gneezy, U. (2011). The tradeoff between performance and quitting in high power tournaments. *Journal of the European Economic Association*, 9(2):318–336.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Flory, J. A., Leibbrandt, A., and List, J. A. (2015). Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *Review of Economic Studies*, 82(1):122–155.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42.

Gneezy, U., Leonard, K. L., and List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5):1637–1664.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074.

Gneezy, U. and Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, 94(2):377–381.

Goldin, C. and Rouse, C. (2000). Orchestrating impartiality: The impact of" blind" auditions on female musicians. *American Economic Review*, 90(4):715–741.

Günther, C., Ekinci, N. A., Schwieren, C., and Strobel, M. (2010). Women can't jump?—an experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization*, 75(3):395–401.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–91.

Kuhnen, C. M. and Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. *Management Science*, 58(1):94–113.

Leibbrandt, A. and List, J. A. (2015). Do women avoid salary negotiations? evidence from a large-scale natural field experiment. *Management Science*, 61(9):2016–2024.

Maggian, V., Montinari, N., and Nicolò, A. (2020). Do quotas help women to climb the career ladder? a laboratory experiment. *European Economic Review*, 123:103390.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3):799–866.

Niederle, M. (2016). Gender. In Kagel, J. and Roth, A. E., editors, *Handbook of Experimental Economics*, chapter 8, pages 481–553. Princeton University Press, Princeton.

Niederle, M., Segal, C., and Vesterlund, L. (2013). How costly is diversity? affirmative action in light of gender differences in competitiveness. *Management Science*, 59(1):1–16.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, 3(1):601–630.

Niederle, M. and Yestrumskas, A. H. (2008). Gender differences in seeking challenges: The role of institutions. Technical report, National Bureau of Economic Research.

Ong, D., Yang, Y. A., and Zhang, J. (2019). Hard to get: The scarcity of women and the competition for high-income men in urban china. *Journal of Development Economics*, page 102434.

Shurchkov, O. (2012). Under pressure: gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, 10(5):1189–1213.

Snowberg, E. and Wolfers, J. (2010). Explaining the favorite–long shot bias: Is it risk-love or misperceptions? *Journal of Political Economy*, 118(4):723–746.