ESSAYS ON THE BEHAVIORAL ECONOMICS OF DISCRIMINATION AND STEREOTYPES

A Dissertation

by

MACKENZIE JESSICA ALSTON

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Catherine Eckel
Committee Members,	Jonathan Meer
	Ragan Petrie
	Jane Sell
Head of Department,	Timothy Gronberg

May 2019

Major Subject: Economics

ABSTRACT

While there are people who gain success in life without many complications, there are many others those who struggle and must overcome challenges in order to succeed. In these three chapters, I use experimental methods to study how women and men, black people, and unemployed individuals deal with issues like stereotypes and discrimination that can create barriers between them and their goals.

In the first chapter, I used an online experiment to investigate whether women and men are able to estimate the amount of discrimination that existed when they applied for a stereotypically male task. The findings suggest that both women and men anticipated that hiring managers would be biased against women; however, men, in particular, overestimated how biased the managers would be.

In the second chapter, we examined whether the test scores of black students attending a historically black college are affected by a subtle reminder of the stereotype that black people are not as intelligent as white people. We found that black students who were exposed to this subtle reminder before taking a mock Graduate Record Examination (GRE) performed just as well as black students who were not exposed to the reminder. This result differs from prior study findings, which primarily used black students at predominantly white institutions.

In the third chapter, I used a laboratory experiment to study the resilience of unemployed individuals who had either been fired or randomly laid off. Individuals in the study performed a real effort task for four rounds, and I compared how well they performed on the task before and after they experienced one round without pay. The results show that low-performing individuals' performance is not significantly affected by job loss or the cause of unemployment (low

performance or chance). In summary, these three studies prove that some people are influenced by the barriers they face (e.g., stereotypes and perceived discrimination), but that does not necessarily mean that these barriers will negatively affect their performance or outcomes.

ACKNOWLEDGEMENTS

I would first like to thank my loving and supportive family. My parents, Greg Alston and Myra George, taught me the value of education at a very early age and made sure that both my brother and I received the best education possible. During my years of graduate study, they were always there to offer me advice and encouragement. They did everything from sending me flowers and ice cream to celebrate my accomplishments to answering my midnight phone calls. My brother, Jeremy Alston, also played a significant role and has been one of my top cheerleaders and voices of support. He was always there to lift me up and remind me how talented I am even when I doubted myself. I love him more than words can describe, and I am extremely grateful to have all three of them in my life.

I would also like to thank the members of my committee. This includes Dr. Catherine Eckel, my committee chair, and my other committee members, Dr. Jonathan Meer, Dr. Ragan Petrie, and Dr. Jane Sell. They provided me with not only guidance that shaped me into a better researcher, but they also opened doors for me that allowed me to have unique experiences like attending a conference with Nobel Laureates. I appreciate their continued support and encouragement.

Many other economists have also helped me significantly throughout my academic career. Because they were not members of my dissertation committee, they did not receive official recognition for all that they have done; however, they all played a significant role in my professional success and advancement. Therefore, I wish to acknowledge them here.

Dr. Katie Coffman is one of the first economists who I would like to recognize. She was the first person to give me true training in experimental economics and showed me that

experimental economics was indeed the field I wanted to pursue. I am thankful for her mentoring during the summer of my junior year when I was her research assistant at The Ohio State University, her invitation to help her run an experiment at Harvard University the semester after, and her guidance after I started graduate school. Even though I was never her official student, she freely gave me her time and assistance, and I would not be where I am today without her.

Dr. Mark Hoekstra is another economist who has been indispensable to my academic career. First and foremost, I am grateful for his efforts to recruit me to Texas A&M University. His enthusiasm about me attending Texas A&M was one of the top two reasons why I accepted the university's offer. Since my acceptance, he has given me advice and provided mentoring whenever I needed it. He has also been a fantastic friend and has accepted me completely, even on the days when I dressed up as Iron Man to the office. My time in the department would not have been the same without him.

I call Dr. Bruce Petersen my "econ dad," and I believe that this title is a good indicator of how close our relationship is. While Bruce started as just my undergraduate advisor, he quickly grew to become so much more. As an undergraduate student, I enjoyed stopping by his office to talk about anything from economics to family life. He was always a strong ally and advocate who jumped to help me at every opportunity, even without being asked. We maintained contact after I graduated, and I have always appreciated his presence in my life as a guaranteed cheerleader. I know that he has complete confidence and faith in me, and I have relied on this unconditional love and support many times. He has been a fantastic mentor and friend that I know will be a constant part of my life.

It would also be remiss of me not to mention my "econ sister," Dr. Zoey Wang, who took me under her wing during her final two years at Texas A&M University. Having an older

graduate student in the department to help me navigate graduate school and life in College Station was beyond useful. She helped me apply for workshops, acted as an accountability partner when I was studying for qualifiers, and let me share her hotel room during conferences. She drove me to the grocery store when I didn't have a car, treated me to countless meals, and always checked in to make sure that I was doing okay. Even after she graduated, she's been a major source of support and encouragement for me.

There are many other people who I have to thank, from colleagues like Dr. Marie Mora and Dr. Sarah Jacobson to buddies from my cohort like Piruz Saboury and Karen Yan to friends like Megan Kress and Cheryl Mitchell. I could write a page or more about each of them because their advice, support, and encouragement have been so meaningful to me. For that, I am blessed and will forever be grateful.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Catherine Eckel (chair), Dr. Jonathan Meer, and Dr. Ragan Petrie of the Department of Economics and Dr. Jane Sell of the Department of Sociology.

The second chapter is part of a collaborative project. The research team included Eckel at Texas A&M University, Dr. William (Sandy) Darity at Duke University, Dr. Lawrence McNeil at Prairie View A&M University, and Dr. Rhonda Vonshay Sharpe, president of Women's Institute for Science, Equity, and Race. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University and by the National Science Foundation (SES-1530796). Research funding from Texas A&M University was also provided from the John Van Huyck Fellowship and the M.J. Grove Fellowship in the Department of Economics and the Private Enterprise Research Center. The second chapter was funded by the National Science Foundation (SES-1530796). Additional funding was provided by the American Economic Association (AEA) Mentoring Program, which was supported by the National Science Foundation (SES-1357478 and SES-1730651).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

When people are faced with barriers in their life, they can either surrender to them or overcome them. In my dissertation, I use experiments to examine the resiliency of individuals who are faced with challenges that could prevent them from being successful. In these three studies, I investigate how people react to obstacles like (perceived) discrimination and stereotypes and study how this can affect their decision-making and performance. This allows me to identify which people are capable of prevailing despite these challenges and which people are more likely to struggle.

In the first experiment, I study how men and women react when they apply for a stereotypically-male job. Subjects in this experiment were recruited to act as job candidates and hiring managers. The subjects who took the role of job candidates created profiles that were similar to resumes, and their profiles were shared with the subjects who were acting as hiring managers. These hiring managers picked one person to hire for the stereotypically-male task. However, before the managers made their choice, job candidates had the option to state their preference for which version of the profile the managers would see. One version included their gender, and the other version excluded their gender. I found that male job candidates were willing to pay more for the profiles with gender than the profiles without gender. The opposite was true for female job candidates, suggesting that job candidates believed that managers would discriminate against women and/or in favor of men. Yet job candidates who hid their gender earned a similar amount as they would have if they had revealed their gender. Thus, male job candidates, in particular, appeared to overestimate how biased managers were. This implies that

the occupational segregation that exists in fields like computer programming and engineering could be partially explained by perceptions of discrimination that cause men to apply for stereotypically male jobs at a higher rate than equally talented women.

In the second experiment, my co-authors and I examine how black college students at a historically black university respond to the negative stereotype that black people are not as smart as white people. Students were sorted into two groups: the treatment group and the control group. Those in the treatment group were subtly reminded of the stereotype before taking a mock verbal Graduate Record Examination (GRE), but those in the control group received no such reminders. Previous studies conducted at predominantly white institutions found that the control group answered more questions correctly than the treatment group. This finding is relevant to understanding the black-white achievement gap and suggests that test scores of black students may not be the best measure of their ability. However, we found no evidence that the reminder negatively affected the test scores of black college students at the historically black university. Therefore, future research on students who attend historically black universities and the environment of these institutions may help us discover ways to reduce the black-white achievement gap.

In the last experiment, I study the behavior of individuals before and after they experience a period of unemployment. This experiment consisted of four periods in which subjects participated in a real effort task. In one treatment, subjects became unemployed after the first period if they performed poorly. In the other treatment, subjects became unemployed after the first period by chance. In both treatments, subjects were allowed to work again for pay in the third period. I found that, among low-performing individuals, being fired or being laid off did not

cause their performance to change significantly relative to those who randomly kept their jobs. This suggests that those who experience job loss are not necessarily demotivated.

In summary, these three studies investigate the choices and reactions of people faced with challenges in their lives. In some cases, I find evidence that individuals do let challenges like dealing with potentially biased hiring managers affect their decision-making. But I also find evidence that many individuals are resilient and are not negatively influenced by the negative beliefs of others or by negative events that happened to them in the past.[1]

[1] Please note that the most recent version of all three studies can be found online at www.mackenziealston.com.

CHAPTER II

THE (PERCEIVED) COST OF BEING FEMALE: AN EXPERIMENTAL INVESTIGATION OF STRATEGIC RESPONSES TO DISCRIMINATION

2.1 Introduction

Of the 337 occupations listed in the 2017 United States Current Population Survey, 56% of them have a 50% or higher share of men. But it is not clear from these statistics whether the source of this occupational gender segregation is the result of discrimination against women or women's occupational preferences. Previous research has tried to understand women's demand for female-dominated jobs through differences in ability or preferences (e.g., Paglin and Rufolo, 1990; Niederle and Vesterlund, 2007). However, past literature has also shown that women have been discriminated against when applying for high-paying and male-oriented jobs (e.g., Riach and Rich, 2006; Neumark, Bank, and Van Nort, 1996). This paper explores the role of *anticipated* discrimination as a factor that could explain why there is an underrepresentation of females in certain fields. That is, men and women may make their decision to pursue fields like science, which are male-dominated, based on their perception of whether or not employers are gender biased. If there is a perception that men will be preferred when they apply for these careers, then men may choose to pursue these fields at a higher rate than women, even after controlling for differences in ability and preferences. If people's perceptions are false, then this becomes particularly problematic.

To test whether job applicants can accurately predict how much discrimination exists when applying for stereotypically male jobs, I conducted two experiments using Amazon Mechanical Turk (MTurk), an online job platform hosted by Amazon. I recruited participants from MTurk to

act as workers and hiring managers. The workers completed a stereotypically male task and answered questions that were used to build two resume-like profiles. These profiles included information like their age and past volunteer activity. The profiles also included a noisy signal of the workers' ability, a measure of how well they performed on half of the task. The two profiles were identical, except one did not include the worker's gender.

Each male worker was paired with a female worker who had a similar signal of ability. Their profiles were then shared with a participant from the hiring manager experiment, who had to choose which of the two workers to hire. Workers who were hired were paid $0.40 for each correct answer on the task. If they were not hired, they were only paid $0.10 for each correct answer.

Before the managers made their decision, I elicited the workers' preference for the managers to see the profiles with or without gender. The workers were asked to consider a situation in which their manager was originally supposed to receive review the profiles that included gender. The workers were then asked how much they would be willing to pay so that the managers saw the profiles *without* gender instead. Next, the workers were asked to consider the case where managers were initially supposed to see profiles that excluded gender. They then indicated how much they would be willing to pay so that the managers saw the profiles *with* gender instead.

The workers' maximum willingness to pay to change the profiles was used to measure how much discrimination they believed existed in this study. If the workers believed that the managers would discriminate against women on the basis of their gender, then the workers had an incentive to change which version of the profiles the manager received. Female workers should be willing to pay to remove gender from their profiles if they believed that hiding gender would increase their

chances of being hired. On the other hand, male workers should be interested in using profiles that included gender.

I find that the workers believed that female workers would be disadvantaged during the hiring process due to their gender. In a post-experimental survey, nearly 40 percent of the workers expected men to be chosen more than women when gender was revealed to the managers, and only 8 percent believed that women would be chosen more. When managers did not know the gender of the workers, only 14 percent of workers expected men would be chosen more. More importantly, women were willing to forfeit 10.0 percent of their earnings from being hired to use gender-free profiles. Meanwhile, men were willing to forfeit 9.4 percent of their earnings from being hired to use profiles with gender. In other words, women were willing to pay to make the hiring process gender-free, and men were willing to pay for the opposite. This suggests that both men and women anticipated that women would become targets of discrimination because of their gender.

I determined whether the managers were discriminating based on gender by comparing their hiring decisions when they saw two profiles with gender to when they saw two profiles without gender. Men were hired 66 percent of the time when the managers saw gender on the profiles and 61 percent of the time when the managers did not see gender. This difference is not statistically significant ($p=0.38$; two-tailed t-test). However, there were cases when it appeared as if the managers did let gender influence their decision, especially when they were forced to choose between two workers with similar signals of productivity.

Regardless, workers were both overestimated how much their gender would affect their earnings. The amount that the workers earned from the task when gender was revealed was not significantly different than the amount they earned when their gender was unknown. Therefore, workers should not have been willing to pay to change the profiles, yet I find that workers were

interested in paying for the opportunity. In particular, men appeared to overestimate how much money they should spend to include gender on the profiles. This finding reflects the importance of capturing information about job applicants' perceptions of discrimination and not just the actual levels of discrimination that exists.

If job applicants' perceptions of gender discrimination are not consistent with the true levels of discrimination, then this could result in men applying for stereotypically male jobs at a higher rate than equally qualified women. Consequently, there would be an underrepresentation of females in certain tracks. Finding a way to adjust job applicants' anticipated discrimination to better reflect the true level of discrimination could then affect the gender composition in science, technology, engineering, and math (STEM) and other male-dominated fields. This may be of particular interest for companies, schools, and organizations that wish to increase diversity.[2]

2.2 Gender Discrimination in the Labor Market

Previous literature has tried to answer the question of whether or not men and women are treated differently by hiring managers. A common strategy researchers have used to study discrimination in the labor market is audit and correspondence studies. In these studies, researchers create fictitious resumes that are sent out to real-world companies. These resumes are designed such that they are nearly identical to each other aside from a particular characteristic of the job applicant. In correspondence studies examining gender discrimination, resumes differ by having female names like Ann and male names like Michael. Then researchers compare the percent of female job applicants who received calls back from the companies to the percent of

[2] Adding diversity to these fields could potentially bring several benefits like reducing gender-based differences in earnings (e.g., Kirkeboen et al., 2016; Arcidicano, 2004) and improving group performance (e.g., Woolley et al., 2010; Hoogendorn, Oosterbeck, and Van Praag, 2013).

male job applicants who received callbacks. If the callback rate for one gender is higher than the other, then that is evidence that companies were discriminating. This is indeed what Riach and Rich (2006) found for male-oriented jobs like engineering. Neumark et al. (1996) found that women were less likely to receive callbacks for high-paying jobs like wait staff at high-priced restaurants. Other studies like Goldin and Rouse (2000) have also shown that women can face labor discrimination using quasi-experimental data. Thus, there is evidence that there are cases in which employers favor men over women.

Because audit and correspondence studies focus on the employers' decisions, they are unable to provide insight into the beliefs of the job applicants. Using experiments and survey questions involving job applicants can help fill this gap. Kang et al. (2016) interviewed black and Asian university students who were searching for jobs and internships. They found that 31% of blacks and 40% of Asians admitted to "whitening" their resumes by modifying how their names were written and changing how they described their race- or culture-based extracurricular activities and experiences so that they sounded white or race-neutral. One common reason stated as to why they whitened their resume was to "even the playing field" or "get their foot in the door." This implies that they felt that other racial groups had an advantage over them and suspected that some employers would discriminate against them if their race were revealed.

In their third study, Kang et. al (2016) conducted a correspondence study where they varied the race of the fictitious job applicant as well as whether or not the resumes employers saw were whitened. Completely whitening one's resume increased black applicants' callback rates by 12.5 percentage points and Asian applicants' callback rate by 9.5 percentage points. Thus, this paper suggests that racial minorities may be correct in their belief that employers care about race. However, the study is unable to say whether their predictions about how much

discrimination they will face are accurate. Neither does it discuss gender differences in the behavior of job applicants.

Zafar (2013) used data on sophomores at Northwestern University to understand how men and women choose their majors, which is an important step to determining their career. He found that women believed that they would be treated more poorly than men for careers in natural sciences, math and computer sciences, and engineering. But this was related to their beliefs about discrimination on-the-job instead of during the application process.

To my knowledge, this is the first paper to study whether job applicants anticipate that females will become targets of discrimination and determine whether or not their expectations are correct. By using two experiments involving both workers and hiring managers, I studied the supply and demand side of the same labor market. Consequently, these experiments allowed me to collect information that revealed how much discrimination workers believed existed in this particular labor market as well as who was hired from that same market, a unique feature of this study.

Workers' beliefs about discrimination were captured through their responses to survey questions and their willingness to pay to have (or not have) a gender-blind hiring process. In addition, each hiring decision was recorded, so it was possible to calculate the likelihood that a female was hired if gender was hidden and when gender was revealed. This gave me the ability to determine whether workers were over- or underestimating how biased managers truly were.

Another advantage of this design was that I avoided some sample selection issues. If I asked workers applying for real-world stereotypically male jobs, then I would be restricted to analyzing individuals who had already decided that the benefit of applying for the job outweighed the costs (e.g., the risk of being discriminated against). In this study, the workers

were given limited information about the experiment before agreeing to participate and thus I was able to observe the behavior of workers who might have otherwise perceived the cost of applying to be too high. Therefore, this study will add to the literature on discrimination and gender differences by providing information on whether or not job applicants are able to accurately predict to the level of discrimination that exists in stereotypically male fields.

2.3 About the Mechanical Turk Sample

Subjects in this study were individuals recruited from Amazon Mechanical Turk (MTurk). To be eligible to participate, they were required to be at least eighteen years old, be located in the United States, and have had satisfactory prior job performance on MTurk. Users on MTurk visit the platform to find short-term jobs for pay and can be more diverse than students who participate in laboratory experiments on college campuses. Therefore, MTurk provides a useful platform to study questions related to the labor market. Previous studies like Coffman et al. (2017) have also used it to study gender-based discrimination.

Table 1 describes the characteristics of this study's subjects. On average, subjects were approximately 35 years old. The majority were white, non-Hispanic, earned $1,000 to $3,000 per month, and had a bachelor's degree or higher. Most worked full-time, and roughly 40% were married. Details on the workers' characteristics by gender can be found in Table 18 in Appendix A.

Table 1. Subject Characteristics

	Workers	Managers
Average Age	35.92	35.16
	(10.93)	(10.19)
Percent Male	49.77%	60.09%
Percent White, Non-Hispanic	73.36%	73.71%
Mode Monthly Salary	$1,001 - $2,000	$2,001 - $3,000
Percent with Bachelor's Degree or Higher	57.48%	65.26%
Percent Employed Full-Time	64.02%	75.59%
Percent Employed Part-Time	16.82%	10.80%
Percent Married	39.25%	43.19%
Percent Never Married	53.50%	46.48%

2.4 Experimental Design

2.4.1 Experimental Design for Worker Experiment

This study was divided into two experiments: one for the workers and one for the managers. For the Worker experiment, 430 men and women located in the United States were recruited from Amazon MTurk to participate in a four-part experiment.[3] In Part One of the experiment, Workers answered questions that would help build resume-like profiles. They indicated their gender, age (as a range of values), whether they had volunteered in the past year, whether their former supervisor would consider them to be a good employee, and whether they frequently read the news. They received $0.50 for completing this part, and the full set of questions can be found in Appendix A.

After answering these questions, Workers read instructions for Part Two. They were told that they would have 2.5 minutes to answer ten questions. They would be paid either $0.10 or

[3] Anyone who did not self-identify as male or female was not allowed to participate in the Worker experiment.

$0.40 for each correct answer and $0.00 for any incorrect answers. Workers were told that a third party would determine whether they would be awarded $0.10 or $0.40 per correct answer. To reduce the likelihood that Workers would condition their effort on the task based on their beliefs about their chances of earning the higher piece rate, no other details about earnings were provided until Part Two was finished.

During Part Two, Workers were given 150 seconds to complete the stereotypically male task: a sports trivia quiz consisting of 10 multiple-choice questions.[4] After the quiz was over, subjects were presented with two nearly identical "profiles" (A and B), which were designed to be similar to resumes. Both profiles contained Workers' answers to the questions from Part One (i.e., their age, volunteer activity, former supervisor's impression of them, and upkeep with the news). The profiles also included their Partial Score, which was a noisy signal of their performance and equaled the number of questions they answered correctly from 5 randomly-chosen questions from the sports trivia quiz. The only difference between the profiles was that Profile A contained the Worker's gender and Profile B did not. An example of the profiles is shown in Figure 1.

[4] In a pilot study (n=51), 84 percent of men and 74 percent of women expected that men would answer more questions correctly on the sports trivia quiz than women. Furthermore, 75 percent of men and 70 percent of women believed that others would say that men performed better than women on the sports trivia quiz. Therefore, the sports trivia quiz was considered to be stereotypically-male. The quiz questions are available upon request.

Figure 1. Example of Worker Profile A and Profile B

Profile Version A:

Gender	Female
Age	26 - 34 years old
Volunteered in past year?	No
Considered good employee?	Yes
Reads the news often?	No
Partial Score (out of 5)	3

Profile Version B:

Age	26 - 34 years old
Volunteered in past year?	No
Considered good employee?	Yes
Reads the news often?	No
Partial Score (out of 5)	3

Before Part Three began, the Manager's experiment was explained to Workers. They were told that MTurk participants would be recruited to participate in another study. These subjects, called "Managers," would start their experiment by answering the same set of sports trivia quiz questions as Workers did in Part Two. Managers would then be presented with one profile belonging to a male Worker and one profile belonging to a female Worker. Both Workers would

have similar Partial Scores reported on their profiles (i.e., within one point of each other), and either both profiles would be Profile A's or both would be Profile B's.

The Managers' task was to hire one Worker, i.e. select one of the two profiles they were presented with. The Worker whose profile was chosen would be paid $0.40 for each correct sports trivia quiz answer they had submitted, and the other Worker would be paid $0.10 per correct answer. The Managers would earn $0.20 for each question answered correctly by the Worker they chose, incentivizing them to select the Worker they believed performed better on the quiz.[5]

Workers were told that they would be sorted into 1 of 3 treatment groups: Forced to Show, Forced to Hide, and Optional.[6] Depending on the treatment Workers were sorted into, some Workers had the ability to determine which version of the profiles was shown to Managers. In the Forced to Show group, Workers were required to use Profile A, which included their gender. Workers in the Forced to Hide treatment group were required to use Profile B, which excluded their gender. Lastly, subjects in the Optional group had the autonomy to determine which profiles would be shown to Managers based on their own preferences. Note that if subjects from the Optional group indicated a preference for Profile A (Profile B), then their Managers would see Profile A (Profile B) for both that subject and the subject they were matched with.

Because assignment to treatment occurred after the experiment was over, every subject was asked to indicate their preferences for Profile A and Profile B in case they were chosen for the Optional group. In Part Three, Workers were asked to consider two scenarios. In the first scenario, the default profile was Profile A, meaning that Managers would know the gender of both Workers.

---

[5] To make sure that Workers understood the Manager experiment, they completed a short comprehension quiz. These questions can be found in Appendix A.

[6] In the experiment, these treatments were called Group 1, Group 2, and Group 3, respectively.

Workers were asked if they would be willing to pay various prices to have Managers see version B of the profiles instead. The prices ranged from $0.00 to $1.00 and increased in increments of $0.10. In the second scenario, the default profile was Profile B, meaning that Managers would know that the two profiles belonged to individuals of opposite gender. However, they would not know which profile belonged to the female or the male. Workers were asked if they would be willing to pay $0, $0.10, …, $1.00 to have Managers see Profile A instead.

After Workers indicated their preferences under each of these scenarios, they were told that one of their decisions would be chosen at random. If they were selected to be in the Optional group, then their choice in that randomly-chosen decision would be executed. This procedure is called multiple price list and is a common method used to elicit subjects' true preferences and willingness to pay (WTP).[7]

One decision was selected at random after the experiment was completely over. This chosen decision was whether Workers were willing to pay $0.50 to switch from Profile A to Profile B. The sorting and matching process for determining how Workers were matched with their rivals is explained in Figure 2. If a Worker from the Optional group said he (she) would be willing to pay $0.50 to switch profiles, then $0.50 was subtracted from his (her) final earnings. He (she) was matched with a random female (male) Worker from the Forced to Hide group who had a similar Partial Score. Their Manager would then see Profile B for both Workers. In other words, their Managers would not know which profile belonged to the male Worker and which belonged to the female Worker.

[7] See Kahneman, Knetsch, and Thaler (1990) and Eckel and Grossman (2008) for other examples of studies that use multiple price lists.

If a subject from the Optional Group had said he (she) would *not* be willing to pay $0.50 to switch profiles, then nothing was subtracted from his (her) final earnings. He (she) was matched with a random female (male) from the Forced to Show group, and their Manager would see Profile A for both. In this case, their Manager would clearly be able to identify which profile belonged to the male Worker and which profile belonged to the female Worker.

Figure 2. Sorting Process for Optional Group

	Preference to pay $0.50 for Profile B	**Matched With**	**What Manager Sees**	**Deduction from Earnings**
Female chosen for Optional	Wants gender *hidden*	Male chosen from Forced to Hide	Two profiles *without* gender	$0.50
Female chosen for Optional	Wants gender *shown*	Male chosen from Forced to Show	Two profiles *with* gender	N/A
Male chosen for Optional	Wants gender *hidden*	Female chosen from Forced to Hide	Two profiles *without* gender	$0.50
Male chosen for Optional	Wants gender *shown*	Female chosen from Forced to Show	Two profiles *with* gender	N/A

Approximately 45 percent of Workers were randomly sorted into the Forced to Show group, 45 percent were in the Forced to Hide group, and 10 percent were in the Optional group. After all Workers from the Optional group were paired with an appropriate Worker from the Forced to Show or Forced to Hide group, the rest of the pairing went as follows. The remaining

Workers from the Forced to Show group were paired with a random Worker from the same group, and their Managers saw version A of their profiles. Similarly, all remaining Workers from the Forced to Hide group were paired with a random Worker from the Forced to Hide group, and their Managers saw version B of their profiles.

Before Workers made their decisions about their WTP to change the profiles, they were made aware that Managers would never know anything about Part Three.[8] This meant that Managers would not know the Workers' WTP to change the profiles nor would they know that more than one version of the profiles ever existed. As a result, Workers did not have to worry that the Managers might judge them based on their responses to Part Three.

Once Part Three concluded, Workers completed a post-experimental survey in Part Four. They provided information about their exact age, race, ethnicity, highest level of education, status as an undergraduate student, employment status, monthly salary, and marital status.

They also answered five questions about their beliefs. First, they were asked how many questions they thought their rival (i.e., the Worker with whom they were matched) answered correctly. Workers were reminded of how many questions they had solved correctly themselves and reminded that their rival's Partial Score was similar to theirs before they answered. In the second and third question, Workers estimated how many questions the average male and average female Worker answered correctly. Their answers could range from 0.0 to 10.0 in increments of 0.1.

In addition, Workers were asked two questions that measured whether or not Workers believed that Managers were biased. For the first question, Workers were told that I would

[8] Workers also completed a second comprehension quiz, which can be found in the Appendix.

randomly select two Managers who had seen version A of the profiles. Each Manager had to choose between one male Worker and one female Worker and would see gender on their profiles. There were three possibilities: both of the Managers could choose the male worker, one of the Managers could choose the male worker while the other chose the female worker, or both of the Managers could choose the female worker. Workers guessed which of these three possibilities happened without seeing the profiles of the four selected Workers. This question captured what Workers expected to happen when Managers knew the gender of the Workers.

Similarly, Workers were asked to guess which of the three possibilities was most likely when two other randomly-chosen Managers had seen version B of the profiles. In this case, the Managers would not know the gender of the Workers, so this question captured what Workers expected to happen when the hiring process was gender-blind.

After the experiment was completely over, two pairs of Workers from the Forced to Hide group and two pairs from the Forced to Show group were randomly chosen, and their Managers' decisions were used to determine whether workers received additional earnings for their guesses. Workers were paid $0.10 for each of the five belief-based questions that they answered correctly.

The post-experimental survey ended with four questions about the Workers' beliefs about how much discrimination women face in the world and two questions about their beliefs about how much discrimination they face in Amazon MTurk, specifically. All questions from the post-experimental survey can be found in Appendix A.

The median time spent on the Worker experiment was 15.18 minutes. Workers earned a $2 completion fee and $0.50 for completing Part One. They also had the opportunity to earn bonus payments based on their performance on the sports trivia quiz in Part Two and their guesses in Part Four. On average, Workers who fully completed the experiment earned $4.12.

2.4.2 Experimental Design for Manager Experiment

For the Manager Experiment, 218 MTurk users of any gender were invited to participate in the study. The Manager experiment consisted of three parts. In Part One, Managers were told that MTurk users in a previous study had been asked to complete a sports trivia quiz. Managers were asked to complete the same quiz themselves so that they could become familiar with the task that Workers had completed. They were given 2.5 minutes, just like the Workers, but were not paid for their performance, something they knew ahead of time.

In Part Two, Managers decided whom they wanted to hire. They were shown the profiles of two MTurk users called Worker 1 and Worker 2. If the profiles were version A, the Managers knew which profile belonged to the female and which profile belonged to the male. Otherwise, the Manager only knew that the profiles belonged to two Workers of the opposite sex.[9] In either case, Managers were asked to choose the Worker they wanted to earn $0.40 for each correct answer he/she submitted from the sports trivia quiz. To incentivize Managers to select the Worker with the higher performance, Managers were told that they would earn $0.20 for each correct answer submitted by the Worker they chose.

After Managers chose a Worker, they completed a post-experimental survey similar to the Workers'. They were asked to report their gender, age, race, ethnicity, highest level of education, status as an undergraduate student, employment status, monthly salary, and marital status. They also stated whether they identified as male, female, or another gender.

[9] This was done so that Managers always had the same information about the gender composition of the Workers. There would never be a case where, for example, Managers saw version B of the profiles and believed that they were choosing between two men or two women. Every Manager knew that there was a possibility that they could hire a man and a possibility that they could hire a woman.

Managers answered three questions about their beliefs about the Workers they were assigned to, which can be found in Appendix A. They were asked how many questions they believed Worker 1 and Worker 2 answered correctly out of 10. They were also asked whether they thought Worker 1 was female or male. For Managers who saw version A of the profiles, this question checked to see how well they paid attention to and remembered the profiles. For Managers who saw version B of the profiles, this question indicated whether or not there was any information in the profiles that indirectly revealed the gender of the Worker. For each correct guess, the Manager earned a bonus payment of $0.10. Lastly, the Managers stated their beliefs about discrimination in the world and in MTurk, specifically.

The median time Managers spent on the experiment was 8.13 minutes. Managers earned $2 for completing the experiment. They had the opportunity to increase their earnings depending on the sports trivia quiz score of the Worker they selected and the accuracy of their beliefs. On average, Managers earned $3.32.

2.5 Results[10]

2.5.1 Worker Beliefs About Performance and Actual Performance

At the end of the Worker experiment, Workers were asked to guess how well they believed the average female and male Worker in the experiment performed on the sports trivia quiz. Workers guessed that females answered 4.2 questions correctly, which is significantly less than

[10] The final sample of subjects that were used for the analysis below consist of individuals from MTurk who completed the Worker or Manager experiment and properly submitted the completion code they were given at the end of the experiment. Two subjects from the Worker experiment identified as female on MTurk but as male on the survey. Therefore, data for these Workers were excluded from the analysis related to the Workers. When analyzing the decisions made by the Manager, I excluded these Workers and the Workers and Managers they were paired with. Additionally, if a pair of Worker profiles was viewed by more than one Manager, only the responses from the Manager who completed the experiment first were used. This only occurred two times.

the 5.6 questions they expected men to answer correctly ($p<0.01$; two-tailed t-test). Interestingly, women were more skeptical about women's abilities than men were. Women (n=215) expected women's average quiz score to be 4.0 compared to men (n=213) who predicted women's average quiz score to be 4.4 ($p=0.05$; two-tailed t-test). There were not any significant gender differences regarding men's expected scores ($p=0.28$; two-tailed t-test). In short, both men and women assumed that men would perform better on the task, on average, than women. This confirms that the task was viewed as stereotypically male.

Workers had information about the range of their rival's possible Partial Scores, so I also compared Workers' expectations of their rivals. Men guessed that their female rivals answered 5.4 questions correctly whereas women guessed that their male partners answered 5.7 questions correctly. Thus, Workers expected a larger gender gap between the average male and the average female versus between them and their own rival. Bu men were still predicted to perform better than women ($p=0.09$; two-tailed t-test).

Workers' belief that men outperformed women was correct. On average, men answered 6.3 questions correctly compared to the 5.4 questions women answered correctly ($p<0.01$; two-tailed test). Note that these values are higher than what the Workers predicted and that Workers predicted that the gender gap would be 1.4 instead of 0.9. The distributions of the estimated and true performance by gender are shown in Figures 3 through 6.

Figure 3. Distribution of Estimated Female Worker Performance

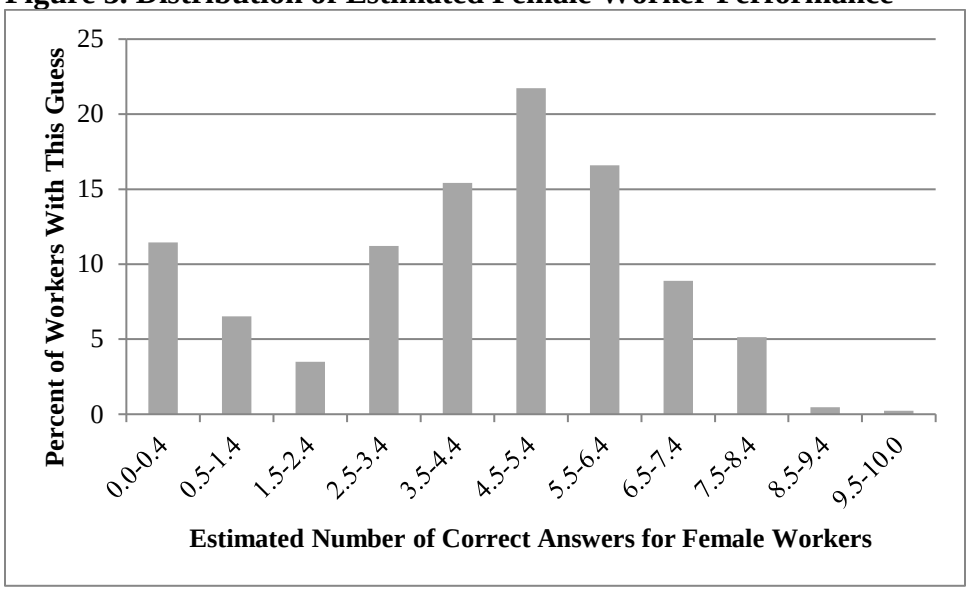


Figure 4. Distribution of Estimated Male Worker Performance

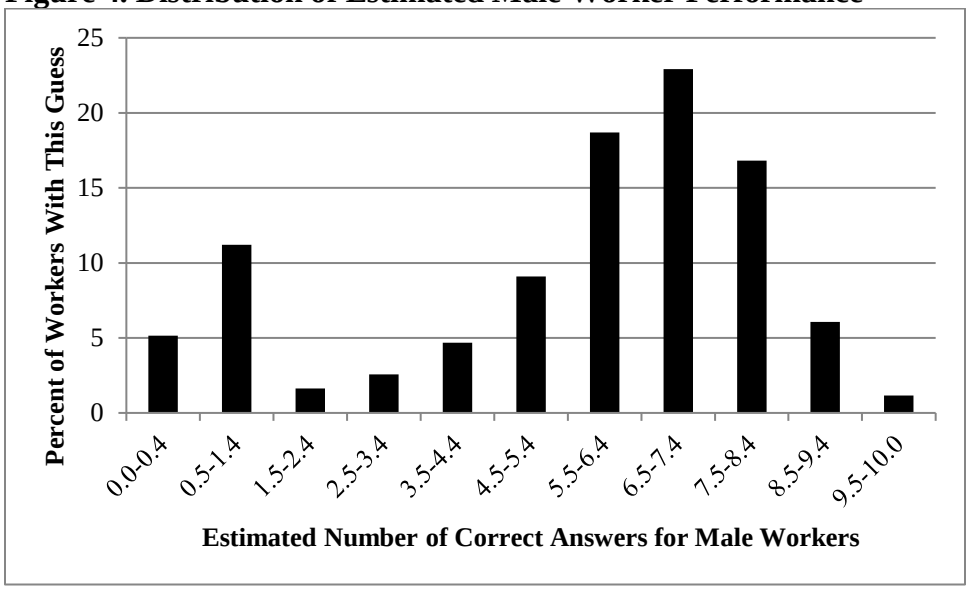


Figure 5. Distribution of True Female Worker Performance

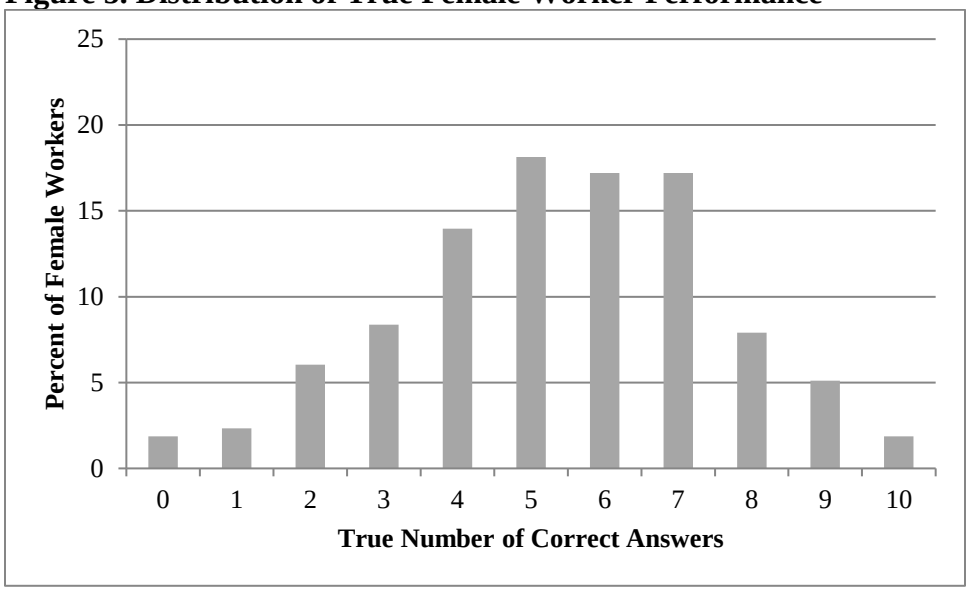
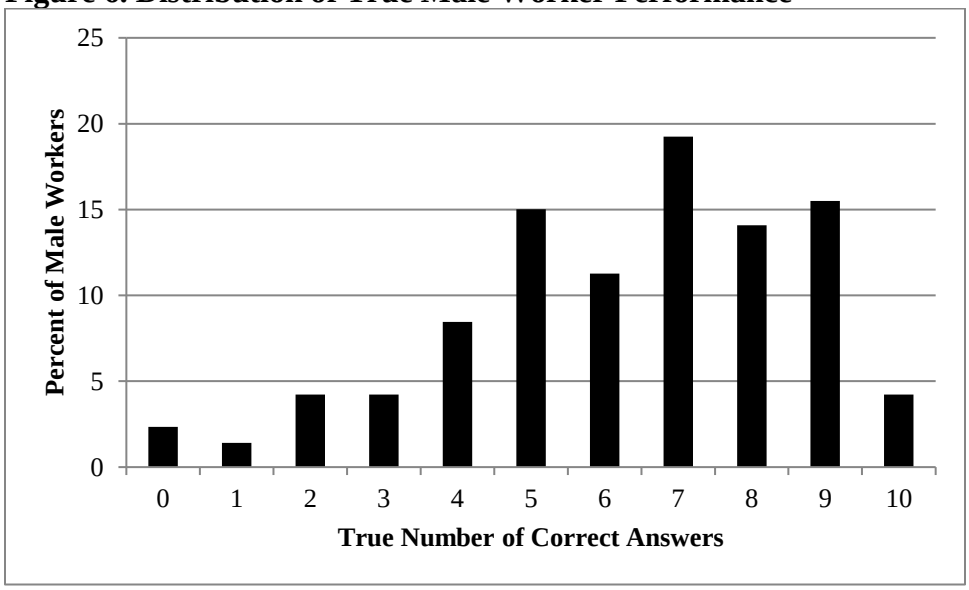


Figure 6. Distribution of True Male Worker Performance



2.5.2 Managers' Hiring Decisions

Managers chose which Worker they wanted to earn $0.40 per correct answer (i.e., who they wanted to hire). In the context of this experiment, I define discrimination as occurring if the frequency in which females were hired is different when Managers saw gender than when they did not see gender. When gender was revealed, Managers hired females 34.2 percent of the time. When gender was not revealed, women were hired 39.4 percent of the time. This difference is not statistically significant (p=0.44; two-tailed t-test). While this appears to suggest that there is no evidence that Managers discriminated against female Workers, it does not account for the fact that Managers saw different combinations of profiles: profiles where the Partial Scores of the two Workers were equal, profiles where the male Worker's Partial Score was higher than the female's, and profiles where the female Worker's Partial Score was higher than the male's.

When Managers were presented with the profiles of two Workers with the same Partial Score, they hired the female Worker 31.8 percent of the time when gender was known. This increased significantly to 56.0 percent when gender was not known (p=.05; Mann-Whitney test). In other words, Managers were far more likely to select the female Worker when they did not know the gender of the person they were hiring. This can be seen in Figure 7.

When Managers knew that the male Worker's Partial Score was higher than the female Worker's Partial Score, Managers hired the female 23.6 percent of the time. The likelihood of a female being hired fell to 20.3 percent when Managers did not know the gender of the Workers. However, this difference is not statistically significant (p=0.67; Mann-Whitney test). In these cases, Managers had the tendency to hire the Worker with the higher Partial Score; gender did not seem to affect their decisions.

In some cases, Managers saw a profile belonging to a female Worker who had a higher Partial Score than the male Worker. When gender was revealed, the female Worker was hired 70.0 percent of the time compared to 86.7 percent of the time when gender was hidden. While this appears to be an economically significant difference, there were only 35 cases in total that fall into this category. Therefore, there are not enough observations to say that the difference is statistically significant (p=0.25; Mann-Whitney test). In sum, there appear to be circumstances when Managers did discriminate. They seemed to pay particular attention to the Workers' gender when they received signals that the Workers' ability was equal. Otherwise, they tended to hire the person whose Partial Score indicated that they were the better performer.

Figure 7. Percent of Female Workers Hired

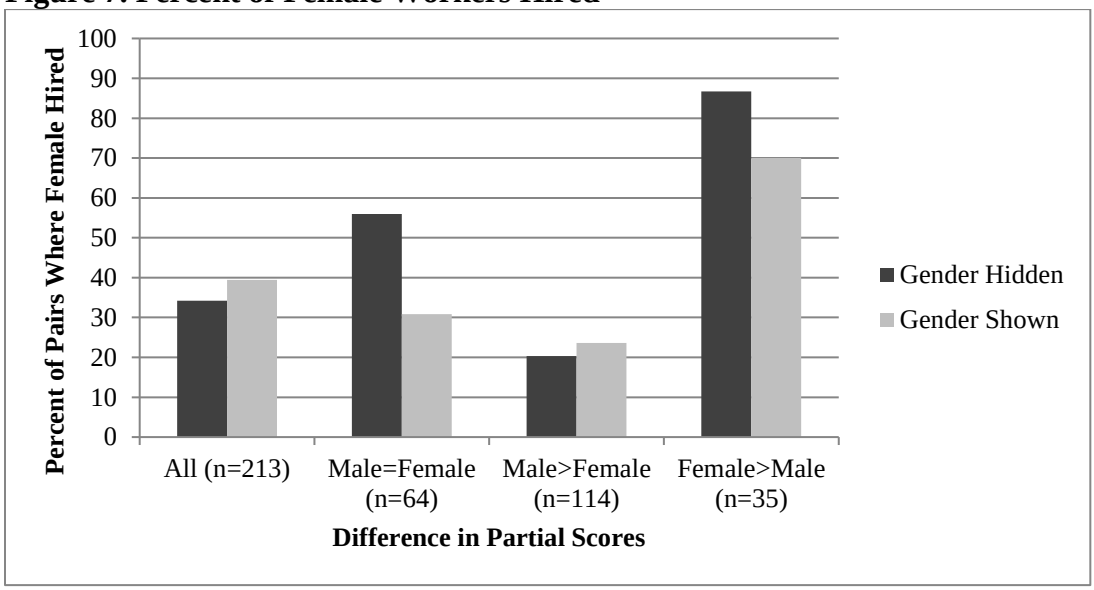


Table 2 reports the estimates of linear probability models where the dependent variable is a dummy variable that equals one if the female Worker was hired. The independent variables include a dummy variable for *Gender Shown*, which equals one if the Manager saw version A of the Profiles. If the coefficient for *Gender Shown* is significant, this means that Managers hired females at different rates depending on whether or not gender was known. The model also includes controls for the every item of the Workers' profiles, except for gender. For example, *Male Volunteers* equals one if the male Worker reported that he volunteered in the past year.

Column 1 of Table 2 reports the estimates for every pairing included in the final sample. Overall, women were 4.5 percentage points less likely to be hired when Profile A was used versus Profile B. This is not statistically significant, suggesting that there was no strong case for discrimination in aggregate. Women's age did play a role in Managers' hiring decisions. Females had a better chance of being hired if they were younger. In addition, their probability of being hired increased by 32.7 percentage points if their former supervisor considered them to be a good employee. Interestingly, women were 18.1 percentage points less likely to be hired if they stated that they frequently read the news.

Table 2. Linear Probability Model Estimates for the Likelihood of Female Being Hired

	(1) All	(2) Male=Female	(3) Male>Female	(4) Female>Male
Gender Shown	-0.0453	-0.152	0.0582	-0.123
	(0.0671)	(0.135)	(0.0778)	(0.169)
Male's Age	0.0379	0.0264	0.0964*	0.00811
	(0.0466)	(0.0859)	(0.0553)	(0.116)
Female's Age	-0.0858*	-0.184**	0.0334	0.0711
	(0.0491)	(0.0906)	(0.0596)	(0.120)
Male Volunteers	0.0387	0.148	0.00306	0.0508
	(0.0676)	(0.134)	(0.0813)	(0.164)
Female Volunteers	0.0351	-0.00121	-0.0209	0.211
	(0.0680)	(0.146)	(0.0742)	(0.157)
Male Good Employee	-0.169	-0.0922	-0.428*	0
	(0.186)	(0.276)	(0.243)	(.)
Female Good Employee	0.327***	0.158	0.196***	0
	(0.0845)	(0.136)	(0.0744)	(.)
Male Reads News	0.109	-0.322	0.283*	-0.0128
	(0.141)	(0.193)	(0.155)	(0.269)
Female Reads News	-0.181**	-0.200	-0.139**	0.0846
	(0.0814)	(0.159)	(0.0652)	(0.237)
Constant	0.333	1.136**	-0.00743	0.183
	(0.318)	(0.454)	(0.356)	(0.588)
N	213	64	114	35

Note: Standard deviations are reported in parentheses. *p<0.01, **p<0.05, ***p<0.01

Column 2 of Table 2 displays the estimates of the same model, except only for pairings where the Partial Scores of the female and male Workers were equivalent. Females were 15.2 percentage points less likely to be hired when Managers knew gender. While this estimate is not statistically significant, it is economically significant. As discussed earlier, women whose Partial Scores were identical to their male rivals seemed to be put at a disadvantage when their gender

was displayed. However, the only statistically significant factor that affected hiring decisions was the age of the female Worker.

Column 3 of Table 2 reports the estimates of the model when the only pairings under consideration are those where the male's Partial Score was higher than the female's. Women were 5.8 percentage points less likely to be hired when gender was displayed on the profiles, but this was not statistically significant. In these situations, the male Worker's age, whether or not the Worker was considered to be a good employee, and the female Worker's habit of reading the news all significantly affected the chances that the female was hired.

Lastly, column 4 of Table 2 considers pairings where the female Worker's Partial Score was higher than the male's. None of the variables had a statistically significant effect on the likelihood of the female being hired, but there were only 35 pairs in this subsample. It should be noted, however, that females were 12.3 percentage points less likely to be hired, which is economically significant.

To summarize, men were hired more, but Managers did not appear to discriminate when one examines all of their hiring decisions collectively. Yet there are indications, though not of statistical significance, that Managers may use gender to influence their hiring decision when they were asked to choose between two Workers with the same Partial Score or two Workers where the female's Partial Score was higher. Gender seemed to play no role when the Partial Score of the male Worker was higher than the female's.

2.5.3 Workers' Beliefs About Discrimination

While the results above found mixed evidence of discrimination, there was clear evidence that many Workers anticipated that Mangers would use gender in their hiring decisions. There were two ways in which I captured this information. The first was by eliciting Workers'

willingness to pay to either include or exclude gender from the profiles using an incentive-compatible elicitation strategy. Except for when it was free to switch profiles, changing one's profile was a costly decision. Therefore, those who indicated a willingness to pay to modify their profiles must be those who believed that there was a benefit from doing so. To determine whether a self-regarding, utility-maximizing Worker should pay to change the profiles, he/she should compare the cost of switching profiles to the expected change in the likelihood of being hired and the resulting change in their earnings. Workers should be willing to make the change so long as the cost to do so is less than the additional earnings they would make from being hired.

Because the task was stereotypically male, I predicted that both men and women would believe that Managers would be biased against women, even when Managers were given a noisy signal of ability. As a result, I expected women to be willing to pay more for Profile B than Profile A because they would want the hiring process to be gender-blind. Men, on the other hand, should expect their chances of being hired to be greater when their gender was revealed to Managers. For this reason, I predicted that male Workers would be willing to pay more for Profile B than Profile A.

When it was free to switch to the gender-free profiles, 82.3 percent of the female Workers were interested compared to 60.1 percent of the male Workers. This difference is significantly different ($p<0.01$; two-tailed t-test). When the price to change the profiles increased to \$0.10, only 45.6 percent of females were willing to pay this amount. However, this percentage is still significantly higher than the 34.3 percent of male Workers who were willing to pay ($p=0.02$; two-tailed t-test). Table 3 displays the percentage of women and men who were willing to pay \$0.00, \$0.10, \$0.20, …, \$1.00 to use version B of the profiles. Women expressed significantly more interest than men when the cost of changing was between \$0.00 and \$0.30 ($p\leq0.10$; two-tailed t-

tests). There is no significant difference between the percent of women and men willing to pay prices ($P) above $0.30. All in all, this implies that women suspected that their gender may be disadvantageous to them. However, most women were not willing to pay more than $0.30 to remove gender from the profiles.

When switching from version B to version A of the profiles was free, 55.9 percent of men were interested in that option. Women found it less appealing, with only 30.7 percent of them expressing a preference for Profile A ($p<0.01$; two-tailed t-test). Once switching became a costly decision, male interest dropped to 32.9 percent and female interest dropped to 13.0 percent. This difference is statistically significant ($p<0.01$; two-tailed t-test). In fact, as can be seen in Table 4, men were always more willing to pay to use version A of the profiles than women ($p<0.01$; two-tailed t-tests). This suggests that men anticipated being favored by Managers when Managers were aware of their gender.

Table 3. Proportion of Workers Willing to Pay to Switch to Version B of the Profiles

Cost ($P) of Switching to Version B	Proportion of Workers Willing to Pay $P		Two-tailed t-test p-value
	Males	Females	
$0.00	0.601	0.823	<0.01
	(0.491)	(0.026)	
$0.10	0.343	0.456	0.02
	(0.476)	(0.499)	
$0.20	0.263	0.335	0.10
	(0.441)	(0.473)	
$0.30	0.160	0.223	0.09
	(0.367)	(0.417)	
$0.40	0.127	0.167	0.236
	(0.333)	(0.374)	
$0.50	0.103	0.112	0.78
	(0.305)	(0.316)	
$0.60	0.066	0.060	0.82
	(0.248)	(0.239)	
$0.70	0.056	0.047	0.65
	(0.231)	(0.221)	
$0.80	0.052	0.042	0.63
	(0.222)	(0.201)	
$0.90	0.038	0.037	0.98
	(0.191)	(0.190)	
$1.00	0.047	0.037	0.62
	(0.212)	(0.190)	

Note: Standard deviations are reported in parentheses.

Table 4. Proportion of Workers Willing to Pay to Switch to Version A of the Profiles

Cost ($P) of Switching to Version A	Proportion of Workers Willing to Pay $P: Males	Proportion of Workers Willing to Pay $P: Females	Two-tailed t-test p-value
$0.00	0.559	0.307	<0.01
	(0.498)	(0.462)	
$0.10	0.329	0.130	<0.01
	(0.471)	(0.337)	
$0.20	0.296	0.098	<0.01
	(0.457)	(0.298)	
$0.30	0.225	0.060	<0.01
	(0.419)	(0.239)	
$0.40	0.202	0.037	<0.01
	(0.402)	(0.190)	
$0.50	0.192	0.023	<0.01
	(0.395)	(0.151)	
$0.60	0.136	0.009	<0.01
	(0.344)	(0.096)	
$0.70	0.099	0.014	<0.01
	(0.299)	(0.118)	
$0.80	0.094	0.019	<0.01
	(0.292)	(0.135)	
$0.90	0.085	0.014	<0.01
	(0.279)	(0.118)	
$1.00	0.085	0.019	<0.01
	(0.279)	(0.136)	

Note: Standard deviations are reported in parentheses.

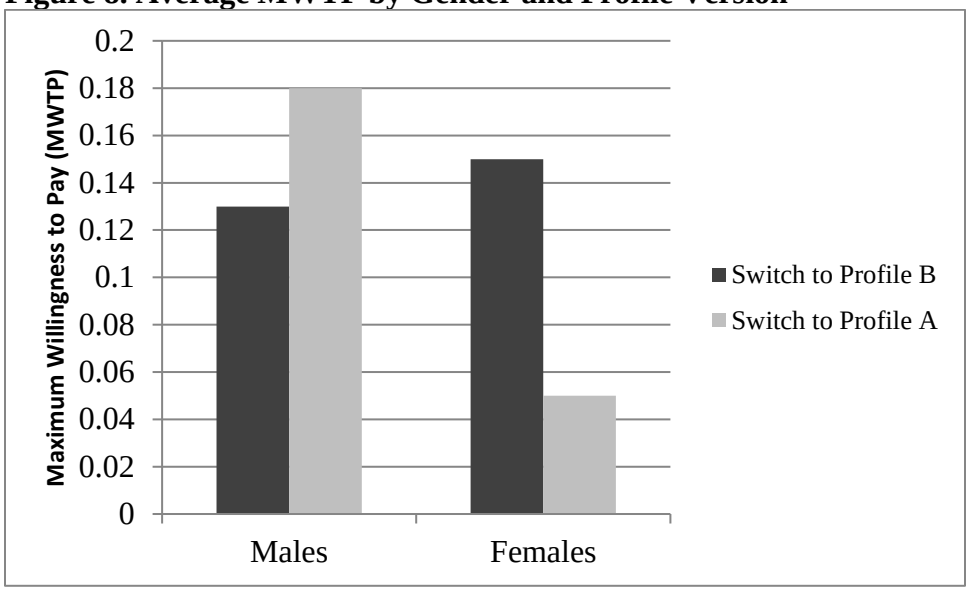
I defined each Worker's maximum willingness to pay (MWTP) to switch profiles as the highest price, $\$\overline{P}$, that he/she was willing to pay such that he/she was unwilling to pay any higher

price.[11] Workers were recorded as having a MWTP of $0 if they said they would never want to switch or if they would pay $0.00 to switch Profiles but nothing higher. Figure 8 displays the average MWTP for male and female Workers to switch from one version of the profile to another. Men's MWTP to use Profile B was $0.13, and they were willing to forfeit 7.5 percent of their earnings from being hired in order to use profiles without gender. Their MWTP to use Profile A was $0.18, and they were willing to forfeit 9.4 percent of their earnings from being hired. The difference in MWTP is significantly different ($p=0.04$; two-tailed t-test) and suggests that men expected to benefit more when Managers knew the gender of the Workers.

Women's MWTP to use Profile B was $0.15, and they were willing to give up 10.0 percent of their earnings from being hired just to gender removed. When it came to their interest in including gender, women's MWTP to use Profile A was only $0.05, and they were willing to spend 4.3 percent of their earnings from being hired to reveal their gender. The fact that they were willing to pay significantly more to use profiles without gender ($p<0.01$; two-tailed t-test) implies that they had a strong interest in making sure that Managers did not know their gender.

[11] Many researchers who use multiple price lists are concerned about how to interpret the action of subjects who indicate a lack of interest in paying $P but a willingness to pay $P+$x ($x>0). (See Andersen et al., 2006 for a discussion on the advantages and disadvantages of multiple price lists.) When this occurred in this experiment, Workers were asked to review their answers but were not forced to make any changes. Therefore, the final answers that Workers submitted were considered as reflecting their true preferences and not as irrational behavior. If $P+$x was the highest price that Workers were willing to pay, then this value was recorded as their MWTP regardless of their WTP (or lack thereof) for prices less than $P+$x.

Figure 8. Average MWTP by Gender and Profile Version



It is also of interest to compare Workers' MWTP for men versus women. While women expressed a higher interest in using Profile B when switching profiles was free, they were reluctant to pay high amounts for that opportunity. Consequently, men and women had a similar MWTP to use Profile B: $0.13 and $0.15, respectively ($p=0.44$; two-tailed t-test). This was not true for MWTP to use version A of the profiles. Men had a greater interest than women in using version A of the profiles even when switching was costly, and this translated to a higher average MWTP to use profiles that included gender for men than for women ($p<0.01$; two-tailed t-test).

In sum, men appeared to believe that they would have an advantage if Managers could identify their profile as the one belonging to the male Worker. This is in contrast to women, who believed that revealing their gender would have negative consequences. This suggests that both

men and women expected women to become targets of discrimination for this stereotypically male task.

The experiment included a second measure of anticipated discrimination. At the end of the experiment, Workers were asked what they expected to happen in randomly chosen cases where two Managers were presented with version A of the profiles. 39.1 percent of women believed that both of the Managers would pick the male Worker. 55.4 percent believed that one Manager would pick the male Worker and one Manager would pick the female Worker. 5.6 percent believed that both of the Managers would pick the female Worker. Thus, over half of the women expected female Workers to be just as appealing to Managers as male Workers when gender was revealed, but there were still women who believed that men would be favored.

Men had similar beliefs. 34.3 percent of them believed that the male Worker would be selected both times, 55.9 percent believed that the male Worker would be selected half of the time, and 9.9 percent believed that the male Worker wouldn't be selected at all. In sum, most male Workers believed that men and women had an equal chance of being hired, but roughly 1 in 3 believed that Managers would prefer men when gender was revealed. Workers were asked to provide the rationale for their guesses, and a selected sample of their responses can be found in Appendix A. Mainly, Workers expected men to be chosen when gender was revealed because they believed that men knew more about sports than women.

Workers also made predictions about what would happen in cases where two randomly chosen Managers did not know the Workers' gender. In this case, 81.4 percent of women and 81.2 percent of men expected one male and one female Worker to be chosen. Workers' explanations for their decision suggest that they expected Managers' choices to be random. However, there were

still some men and women who thought men would always be selected (14.1 percent and 14.9 percent, respectively).

It is important to note that the percentage of Workers who believed that men would always be chosen more than doubled when Workers were asked about Managers who knew the Workers' gender compared to Managers who did not. These responses confirm that many Workers believed that female Workers would be at a disadvantage if Managers could observe gender, but male Workers would be at an advantage. Furthermore, this is consistent with the findings on Workers' MWTP that women preferred to conceal their gender and men preferred to reveal theirs.

2.5.4 Are Workers Overestimating the Cost (Benefit) of Their Gender?

Section 2.5.3 showed that both male and female Workers were unconvinced that Managers would be entirely impartial, and section 2.5.2 revealed that Managers did not always discriminate. In this section, I estimate how much financial advantage or disadvantage Workers received because of their gender and compare this to their MWTP to have gender included or excluded. This allows me to determine whether or not Workers were optimally able to set their MWTP.

Table 5 contains the results of several linear regressions where the dependent variable is *Task Earnings*, which represents the amount of money Workers earned from the sports trivia quiz. This amount depended on their score on the quiz and the hiring decision of the Manager. The regressions also include a dummy variable for all of the items of the Worker's profile. For example, *Reads News* equals one if the Worker indicated that he/she reads the news often. There is also a variable to capture the difference in the Workers' Partial Scores.

Column 1 of Table 5 is a simple regression to see if females earned less than men. Column 2 of Table 5 adds a dummy variable that equals one if Profile A was shown to the Manager, and the third column of Table 5 adds an interaction term between this dummy variable and a dummy

variable for whether or not the Worker was female. If the coefficient on this interaction term is negative and statistically significant, it would suggest that females earned more when they hid their gender from the Managers.

Table 5. OLS Estimates of the Effect of Gender on Worker Earnings in Dollars

	(1)	(2)	(3)
Female	-0.540***	-0.540***	-0.469***
	(0.120)	(0.120)	(0.174)
Gender Shown		-0.0394	0.0243
		(0.106)	(0.163)
Female*Gender Shown			-0.128
			(0.214)
Age	-0.197**	-0.197**	-0.200**
	(0.0817)	(0.0819)	(0.0823)
Volunteered in Past Year	-0.0871	-0.0855	-0.0851
	(0.108)	(0.108)	(0.109)
Good Former Employee	0.511*	0.509*	0.497*
	(0.293)	(0.291)	(0.295)
Reads News	0.393**	0.394**	0.392**
	(0.166)	(0.166)	(0.167)
Diff. in Partial Scores	0.0801***	0.0802***	0.0812***
	(0.0295)	(0.0295)	(0.0297)
Constant	1.668***	1.691***	1.677***
	(0.371)	(0.381)	(0.386)
N	426	426	426

Note: Standard deviations are reported in parentheses. *p<0.10, **p<0.05, and ***p<0.01.

As can be seen in each regression, women earned less than men, even when controlling for differences in age, volunteer history, supervisors' impressions, frequency of reading the news, and Partial Scores. The coefficient on *Female*Gender Shown* in Column 3 is negative but not

statistically significant. However, the results suggest that women earned $0.15 less when their gender was shown versus hidden. Men, on the other hand, earned $0.02 more when their gender was shown versus hidden. So revealing gender had no financial advantage for men, and hiding gender had a small (but statistically insignificant) effect for women.

A self-regarding utility maximizer would set their MWTP such that the MWTP is equal to the financial advantage they would earn from using one version of the profiles over the other. While women's MWTP to switch to version A of the profiles was $0.05, women's MWTP to use version B was $0.15. Based on the coefficients from Table 5 and their statistical significance, both of these values should be $0.00 if women were responding optimally. By the same token, men should not have been willing to pay anything to change their profiles, yet their MWTP to use version A was $0.18. Again, these values should be $0.00 if men were responding optimally.

I find that both women's MWTP to switch to Profile B and men's MWTP to switch to Profile A are significantly different from $0.00 ($p<0.01$; one-sample t-tests). However, women's MWTP to use Profile B exactly matched the difference in earnings for females when gender was shown versus hidden (i.e., the coefficient for *Gender Shown* plus the coefficient for *Female*Gender Shown*). On the other hand, men's MWTP for version A was significantly higher than 0.128 ($p=0.01$; one-sample t-test). In sum, I find that men, in particular, overestimated the cost of being female and the benefit of being male.

2.6 Conclusion

There is an under-representation of women in certain careers like engineers and computer programmers. This could be driven by gender differences in application rates or gender differences in hiring rates. However, an individual's decision to apply to a job may depend on their expected likelihood of being hired. This expectation may further rely on their beliefs about how gender

biased hiring managers are. In this paper, I investigate whether anticipated discrimination can help explain why men are overrepresented in some fields. If men think they will be favored or women think they will be targets of discrimination, then anticipated discrimination could explain why men opt in and women opt out of certain fields.

I conducted two experiments using Amazon Mechanical Turk to test whether men and women can accurately predict how much gender discrimination is present in this experimental labor market. Workers completed a stereotypically male task (a sports trivia quiz) and answered questions to create two resume-like profiles. One version of the profile included gender, and the other did not. Each Worker was matched with another Worker who had similar ability but was of the opposite sex. Workers were then asked how much they would be willing to pay to use one version of the profile over the other if Managers were required to hire only one of the two Workers. Workers' maximum willingness to pay (MWTP) to use the profiles with or without gender were indicators of how much they perceived their gender to be an asset or liability, respectively.

I find that, when changing versions of the profiles was costless, 82.3 percent of women were interested in switching to the profiles without gender relative to 60.1 percent of the men. When Workers were asked about switching to profiles with gender for free, 55.9 percent of men were interested compared to only 30.7 percent of women. When studying Workers' MWTP, I found that men's average MWTP to use profiles with gender was $0.18, which is significantly higher than the $0.13 they were willing to spend to use profiles without gender. Women's average MWTP, on the other hand, was $0.15 to use profiles without gender and $0.05 to use profiles with gender. This difference is statistically significant. All in all, these findings suggest that both men and women assumed that men would have an advantage in the hiring process when gender was revealed. This encouraged men to want to reveal their gender and women to want to conceal it.

Men were indeed hired more than women. Men were hired 65.8 percent of the time when gender was revealed and 60.6 percent of the time when gender was hidden. This difference is not statistically significant, so there was no strong evidence of discrimination when all of the pairs are considered collectively. However, there is suggestive evidence that Managers used gender in their hiring decision under certain circumstances like when the Partial Scores of the two Workers were equivalent. In that case, they were less likely to hire female Workers when gender was revealed than when gender was hidden.

Regardless, men (women) were not significantly better (worse) off financially by having the Managers know their gender. This means that male and female's MWTP to change the profiles should have been zero. Yet women's gender was perceived by both genders to be a disadvantage. Women were willing to forfeit 10.0 percent of their earnings from being hired to remove gender, and men were willing to forfeit 9.4 percent of their earnings to include gender. But men, in particular, seemed to overestimate how much of a role gender played on hiring decisions.

These findings suggest that perceptions about the gender bias of hiring managers matter and may affect men and women's decision to pursue certain careers. If men believe that they have an advantage in stereotypically male fields, they may be overconfident in how much their gender alone will get them ahead if they overestimate how biased hiring managers are. As a result, they may apply to more jobs of this type than what would be optimal if they had correct beliefs about managers' biases. The reverse could happen for women. If they believe that they are disadvantaged, they may apply to fewer jobs of this type than what is optimal because they have overestimated how biased hiring managers will be against them. Further research should investigate ways to help update job candidates' expectations of discrimination, which might affect the gender composition of jobs in stereotypically male fields like STEM.

2.7 References

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics, 9*(4), 383-405.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics, 121*(1-2), 343-375.

Arrow, K. (1973). The theory of discrimination. *Discrimination in labor markets, 3*(10), 3-33.

Becker, G. S. (1957). The economics of discrimination Chicago. *University of Chicago*.

Coffman, K. B., Exley, C. L., & Niederle, M. (2017). When Gender Discrimination Is Not About Gender. Harvard Business School Working Paper.

Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization, 68*(1), 1-17.

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

Hoogendoorn, S., Oosterbeek, H., & Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science, 59*(7), 1514-1528.

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics, 2*(3), 2053168015604648.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98*(6), 1325-1348.

Kang, S. K., DeCelles, K. A., Tilcsik, A., & Jun, S. (2016). Whitened resumes: Race and self-presentation in the labor market. *Administrative Science Quarterly, 61*(3), 469-502.

Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics, 131*(3), 1057-1111.

Krause, A., Rinne, U., & Zimmermann, K. F. (2012). Anonymous job applications in Europe. *IZA Journal of European Labor Studies, 1*(1), 5.

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *Sage Open, 6*(1), 2158244016636433.

Neumark, D., Bank, R. J., & Van Nort, K. D. (1996). Sex discrimination in restaurant hiring: An audit study. *The Quarterly Journal of Economics, 111*(3), 915-941.

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

Paglin, M., & Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics, 8*(1, Part 1), 123-144.

Riach, P. A., & Rich, J. (2006). An experimental investigation of sexual discrimination in hiring in the English labor market. *Advances in Economic Analysis & Policy, 5*(2)

Rinne, U. (2018). Anonymous job applications and hiring discrimination. *IZA World of Labor*.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science (New York, N.Y.), 330*(6004), 686-688.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources, 48*(3), 545-595.

CHAPTER III

THE EFFECT OF STEREOTYPES ON BLACK COLLEGE TEST SCORES AT A HISTORICALLY BLACK UNIVERSITY

3.1 Introduction

According to a 2009 report from the National Center for Education Statistics, 75 percent of white first-time bachelor's students had a cumulative grade point average (GPA) of 3.0 or higher, whereas only 55.3 percent of black students had a similar GPA.[12] Researchers have tried to explain the black-white academic achievement gap through differences in quality of education, parenting, and wealth (e.g., Yeung and Conley, 2008; Hanushek and Rivkin, 2009). In addition, psychologists have explored another factor that could explain why black students are not performing as well as their white peers – stereotype threat. This phenomenon describes the feeling of being at risk of confirming a negative stereotype about one's social group. This feeling can potentially interfere with performance in the domain where the prejudicial belief applies, and it can hold for females and racial groups like Asians (see Shih, Pittinsky, and Ambady, 1999). Our study focuses on the effect that stereotype threat has on black college students.

A widely held prejudice has it that blacks are not as intelligent as whites, or, at minimum, have cognitive deficits relative to whites. This belief has been promoted in works like *The Bell Curve: Intelligence and Class Structure in American Life* (Herrnstein and Murray, 1994), which insinuates that race and IQ are correlated. According to the stereotype threat hypothesis, a black

[12] https://nces.ed.gov/datalab/tableslibrary/viewtable.aspx?tableid=8836

student who is primed or subtly reminded of this prejudice before completing a relevant task will suffer a decline in performance.

It has been proposed that those affected by stereotype threat either surrender to the stereotype or become so focused on the stereotype that they grow distracted from the task at hand. For example, Blascovich et al. (2001) showed that blood pressure rates of black subjects who had been primed to activate stereotype threat were significantly higher than the blood pressure rates of subjects who had not been primed. Regardless of whether individuals consciously change their effort or not, the main hypothesis involving stereotype threat is that those who are primed will perform worse on the task than those who are not. Consequently, stereotype threat could explain why two equally able individuals who are members of different racial groups may earn different grades.

Steele and Aronson (1995) conducted an experiment to test the impact of stereotype threat on the performance of black college students at Stanford University. They found that priming subjects to think about their race made them more aware of the stereotypes associated with a black identity. In addition, they found that subjects who were primed answered fewer questions correctly on a verbal Graduate Record Examination (GRE) than subjects who were not primed. These results were replicated in other studies like Brown and Day (2006) and Davis, Aronson, and Salinas (2006). Furthermore, these findings implied that test scores might misrepresent the true abilities of black students, especially when they have been exposed to a reminder of their race before completing a test.[13] As a result, white students could appear more intelligent even when compared to black students of the same ability.

---

[13] This is the phenomenon of "latent ability." See Walton and Spencer (2009).

The primary goal of our experiment is to determine whether black students from a historically black college respond negatively to stereotype threat on a verbal GRE. Almost all of the studies regarding black stereotype threat (e.g., Steele and Aronson, 1995; Cohen and Garcia, 2005) have been conducted at predominantly white institutions (PWIs). While PWIs are the most common type of institution in the United States, these studies fail to consider how stereotype threat may affect black students at institutions that are predominantly black – that is, historically black colleges and universities (HBCUs).

While a handful of similar studies (e.g., McKay et al., 2002; McKay et al., 2003) have used black students at HBCUs as part of their sample, their focus has not been on the difference between students from these schools and PWIs. Therefore, their analysis pools the black students from each type of institution together without treating them differently. In contrast, our study focuses exclusively on black students at an HBCU.

We ran our experiment in classrooms at an HBCU in Texas. Subjects who participated in the treatment group were primed before completing a set of verbal GRE questions. They were asked to identify their race and ethnicity before the test and read test instructions that emphasized that the verbal GRE was a measure of their intelligence. Subjects in the control group were not asked about their race or ethnicity until after the exam, and their test instructions included no mention of the subjects' verbal ability or intelligence. By comparing the test performance of subjects in the treatment and control groups, we were able to assess the effect of stereotype threat on black students in this setting.

We also varied the level of the threat subjects experienced by using two types of lead experimenters: a white male and a black female. Some sessions were conducted by a white male experimenter, similar to Steele and Aronson (1995). In others, the experimenter was a black

female. Like other studies that have used experimenters of different races (e.g., Marx and Goff, 2005), we predicted that subjects would feel most threatened by a white male experimenter because he would be seen as a critical outsider (given that the student body and faculty were predominantly black). A black female experimenter would be seen as less threatening. Moreover, we anticipated that subjects in the treatment group would have a lower exam score when their experimenter was a white male rather than a black female because those subjects would be exposed to compound threats (i.e., the presence of the white male experimenter plus the primes from the experiment).

Our findings indicate that black students at the HBCU are not affected by stereotype threat regardless of the identity of the experimenter. Overall, black students in the control group answered significantly more questions than black students in the treatment group. Yet we found no significant difference in the number of questions students answered *correctly* in either the white male or the black female experimenter sessions. Finally, we found little evidence to support our prediction that subjects responded differently to the type of experimenter. Having a black female experimenter had a negative effect on the total number of questions answered but no effect on the number of questions answered correctly.

These findings provide preliminary evidence that the test performance of black students at HBCUs is not depressed by reminders about the prejudicial belief that blacks are regarded as less intelligent or less achievement-oriented than whites. This is not to say that we conclude that every black high school student should choose to attend an HBCU. In fact, there is debate over whether attending an HBCU is strictly positive.

For example, Fryer and Greenstone (2010) find that blacks who attended HBCUs suffered from a “wage penalty” relative to blacks who graduated from a PWI, but Price, Spriggs,

and Swinton (2011) find that black students have higher permanent incomes if they attend an HBCU. We are unable to judge the relative net benefits of graduating from an HBCU. However, our experiment does suggest that black students at HBCUs may, at the very least, have the advantage of being less susceptible to stereotype threat. Thus, this may have potential implications for understanding more about the persistence of the black-white achievement gap. Further research on the benefits of attending an HBCU may shed light on ways in which the gap can be narrowed.

3.2 Previous Research on Historically Black Colleges and Universities (HBCUs)

Most HBCUs were established with the primary mission of advancing the education of blacks under conditions of legal segregation in the United States. As of 2016, there were slightly more than 100 HBCUs in the United States, and these schools accounted for 14 percent of bachelor degree's earned by black undergraduates in 2015 – 2016. This is particularly significant given that HBCUs enrolled 9 percent of all black undergraduates in 2015 and 2016.[14] In addition, the National Science Foundation (2015) reports that about 18 percent of science and engineering degrees are awarded to black students by HBCUs. Therefore, HBCUs provide a unique, albeit understudied, environment to investigate stereotype threat.

Although there are mixed findings, many studies have shown that attending an HBCU affords a positive and beneficial experience for students who select such an institution. Nichols and Evans-Bell (2017) find that HBCUs account for higher graduation rates for lower-income black students compared to PWIs. Price et al. (2011) utilized a potential outcomes approach to show that graduating from an HBCU has a positive and significant effect on permanent income.

---

[14] https://nces.ed.gov/fastfacts/display.asp?id=667

They also find that graduating from an HBCU has a positive effect on psychological outcomes like self-image. This confirms the findings of Gurin and Epps (1975) who determined that blacks at HBCUs had positive self-images and high aspirations.

Fries-Britt and Turner (2002) interviewed black students at PWIs and HBCUs to document differences in their experiences in the two environments. Black students at HBCUs reported feeling supported, being a part of the community, finding insulation from discrimination, and having a sense of confidence. On the other hand, black students at the PWIs felt excluded and weighed down by being the "token black" in their classrooms.

The positive impact of HBCUs on student attitudes and perceptions could enhance student performance; for example, studies have shown that self-affirmation improves the test performance of people facing stereotype threat (Mortens et al., 2006). It is not unreasonable to then expect that black students attending HBCUs may be better insulated against stereotype threat than black students at PWIs.

If black students at HBCUs respond differently to stereotype threat than black students at PWIs, it could be the consequence of either selection or environment. Black students who choose to enroll in an HBCU could begin their studies with a lower vulnerability to stereotype threat vis-á-vis black students who enroll at PWIs. However, this is not supported by our other work, which is available upon request. We surveyed incoming black freshmen at two schools in Texas – an HBCU and a PWI – and found that they reported similar levels of stereotype threat vulnerability.

If black students at both types of schools initially are equally susceptible to stereotype threat, it is possible that the environment of an HBCU encourages students to ignore or overcome the effects of stereotype threat. For instance, HBCUs typically have more racially diverse

faculties than PWIs. In 2016, 61 percent of the faculty at our Texas HBCU were black, whereas only 3 percent of the faculty at our Texas PWI identified as black.[15] Black faculty may provide important same-race role models of someone highly educated, whose presence offsets the stereotype that blacks are unintelligent or opposed to academic achievement.

The importance of role models has been supported by previous research showing that students receive higher grades when taught by people that physically resemble them (e.g., Dee, 2004; Egalite, Kisida, and Winters, 2015). There is also evidence in the literature that role models can be influential in overcoming the impact of stereotype threat. For instance, Marx and Roman (2002) showed that when women were primed to think of negative stereotypes, their math scores were higher when there was a competent female experimenter present in the room rather than a male experimenter. Marx and Goff (2005) find a similar result when they primed black subjects and used a black experimenter. Black subjects who had been primed performed just as well as on a verbal test as white subjects when there was a black experimenter. However, they performed worse than white subjects when the experimenter was white.

The content of instruction at HBCUs also might make a difference. The students interviewed in Fries-Britt and Turner (2002) reported appreciating the fact that class lectures highlighted the accomplishments of black scholars and authorities. Correspondingly, McIntyre, Paulson, and Lord (2003) found that women who read about successful women performed better on a hard math test. And Aronson, Fried, and Good (2002) showed that changing black students' mindset about intelligence helped reduce the negative effects of stereotype threat.

---

[15] https://www.pvamu.edu/ir/faculty-data/ and https://accountability.tamu.edu/All-Metrics/Mixed-Metrics/Faculty-Demographics. Note that these statistics on black faculty members do not include international faculty.

Our study answers the question of whether black students at an HBCU are vulnerable to stereotype threat. If any of the factors mentioned above affects black students' vulnerability to stereotype threat, then having more black students attend an HBCU could be an avenue to reduce the black-white achievement gap. These students would be provided with the tools and environment that could decrease the effect of stereotype threat on their performance. Additionally, if specific tools and strategies that successfully mitigate stereotype threat can be identified, it is possible that some of them could be used at PWIs to help black students there.

3.3 Experimental Design and Procedure

Subjects in our experiment were recruited from two sections of an Agricultural Nutrition and Human Ecology course taught by the same professor in Fall 2016, a section of Principles of Microeconomics and a section of Economics and Human Resources taught by a second professor in Winter 2017, and two sections of Principles of Microeconomics taught by the same second professor in Fall 2018. In our white experimenter sessions, there were two researchers: one white male who acted as the lead experimenter and one black female who acted as his assistant. The lead experimenter was the one who read the instructions, directed the experiment, and had a clear leadership role. On the other hand, the assistant handed out and collected materials, helped pay subjects, and maintained a subordinate role. In our black experimenter session, that same assistant took over as the lead experimenter and another black female acted as her assistant.

Once the lead experimenter and assistant arrived at the classroom, the professor briefly introduced them before exiting the room. Because he left the room, students knew that their professor was not observing them. They also were explicitly told that their participation was not related to their class grades. After the professor left, the researchers asked the students to take

their belongings and move into the hallway while one researcher remained in the room to set up the experiment.

The researcher inside the classroom shuffled the test booklets for the control and treatment groups together. Then the researcher randomly placed one numbered card and one test booklet on each of the student desks. In two of the sessions, there were enough desks so that booklets could be placed on every other desk, leaving an empty desk in between. When this was not possible, booklets were sometimes placed on two consecutive desks. At the same time, the researcher outside of the classroom made it clear that students' participation in the study was purely voluntary and passed out numbered cards to any student who chose to participate.

As soon as the room was ready, students were readmitted and asked to find the desk with the card number that matched the one that they had been given in the hallway. This seating arrangement was intended to reduce the probability that students were sitting next to close friends with whom they might exchange information during the experiment.

Each booklet contained five parts and followed much of the same wording used in the materials from Steele and Aronson (1995) and a more recent set of instructions provided by Aronson.[16] In Part One, students in the control group answered six personal questions regarding their age, year in school, major, number of siblings, and mother's and father's highest level of education. Students in the treatment group responded to the same set of questions; however, they were also asked if they identified as Hispanic/Latino and were asked to identify their race. Subjects in the control group answered these two questions after taking the test. The intent was to prime students in the treatment group to think about their race *before* taking the test.

---

[16] The authors would like to thank Joshua Aronson for emailing a copy of his recent instructions.

In Part Two, students received instructions that explained what to expect during the test. Both Part One and Part Two for the control and treatment groups can be found in Appendix B with the accompanying script. Students in the control group read the following instructions:

> The problems you are about to solve are taken from the verbal portion of the GRE (Graduate Record Examination). You will be given twenty-five minutes to answer 18 questions. You will receive $12 in compensation for submitting your answers. Completing this test will allow you to familiarize yourself with the kinds of problems that appear on tests you may encounter in the future. Please try hard to correctly solve as many items as you can to help us in our analysis of the problem solving process.

Students in the treatment group read the following instructions:

> The test you are about to take, the verbal portion of the GRE (Graduate Record Examination), is in large part a measure of your verbal intelligence and verbal reasoning ability. You will be given twenty-five minutes to answer 18 questions. You will receive $12 in compensation for submitting your answers. Completing this test will allow you to familiarize yourself with some of your strengths and weaknesses.

Part Three consisted of 18 multiple-choice questions taken from *The Official GRE Verbal Reasoning Practice Questions* (2014). These questions were identified as easy or medium in difficulty by the authors of the practice book and included a mixture of reading comprehension questions and vocabulary-based questions.[17] For the reading comprehension questions, students

[17] Steele and Aronson (1995) reported using 27 difficult verbal questions from GRE study guides. In a pilot study we conducted, we used 27 verbal questions from GRE practice exams. On average, the control group answered 17.2 questions but only 1.6 were correct (n=5). The treatment group answered 24.5 questions. Of these, 3.67 questions were answered correctly (n=6). We wanted subjects to correctly answer a higher proportion of questions, so we reduced the number of questions and the difficulty level of the questions used. Ultimately, our goal was for subjects in our control group to correctly answer a similar percentage of the test questions as in Steele and Aronson (1995).

read a passage about a topic like the Antarctic and were asked several questions about what they read. Some of the answers could be found directly in the text while others required critical thinking. The vocabulary-based questions asked students to find one or two words that would fit the missing blank in the provided sentences. These questions were chosen to be challenging because Steele and Aronson (1995) argued that black students would find the stereotype more salient when the test was difficult.

Part Four of the booklet asked students to answer a variety of demographic questions. For subjects in the control group, this was the point where information was collected about their race and ethnicity. Other questions that subjects in both groups answered related to their gender, place of birth, cumulative GPA, socio-economic status, family background, high school, and political and religious identity. We also asked them to guess how many questions they believed they answered correctly, how difficult they found the test, how biased they perceived the test to be, and how they believe they performed relative to their peers.

During Part Four, students completed two measures of stereotype threat vulnerability. The first measure was adapted from Steele and Aronson (1995), who reported using eight Likert-scale survey items to measure their subjects' vulnerability to stereotype threat. Five of these statements were included in the published paper, and these were the five that were included in the booklets. An example is, "Some people feel I have less verbal ability because of my race." Students were asked to state how strongly they agreed or disagreed with such statements on a scale of one (strongly disagree) to seven (strongly agree). Their scores on these five statements

---

Steele and Aronson (1995) wanted students to answer approximately 30 percent of the test correctly, and the average score for all students in the control group in our final sample was 28 percent.

were averaged (with two of them reverse-coded first) to determine the first measure of stereotype threat vulnerability. Higher numbers mark higher degrees stereotype threat vulnerability.

The second measure was taken from Aronson, Fried, and Good (2002). Students were asked to rate on a scale of one (strongly disagree) to seven (strongly agree) how much they agreed with two statements. The first statement was, "People make judgments about my racial group based on my race." The second statement was, "People make judgments about my racial group based on my performance." The average of their score for these two statements was taken; higher averages indicate a higher vulnerability to stereotype threat.

To capture how strongly students considered verbal ability to be important to their sense of self, we asked subjects to evaluate four statements on the same seven-point scale described above. Examples of these statements are, "Verbal skills will be important to my career" and "I am a verbally oriented person." Most of these questions were also used in Steele and Aronson (1995).[18]

Part Five was labeled the "Opinion Survey" and included the Multigroup Ethnic Identity Measure (MEIM) created by Phinney (1992). The measure first asked students to write down their race or ethnicity. Next, students evaluated twenty statements on a scale from one (strongly disagree) to four (strongly agree). These statements were designed to capture different elements of an individual's sense of belonging, identity, and behavior as it related to their self-identified race or ethnicity. We used this to measure students' strength of ethnic/racial identity. Higher values for this measure indicate a stronger identification with their race or ethnicity.

---

[18] Part Four also included a short financial literacy quiz taken from *The Wall Street Journal's* online quiz titled "Test Your Financial Literacy" and a measure of locus of control used in Darity, Goldsmith, and Veum (1996).

Following the MEIM, students answered questions about their father's and mother's ethnicity and race. They also indicated the name of the city in which their high school was located. The last questions in the booklet asked for general feedback about the experiment (e.g., stating whether they had a positive experience).

Once the experiment was over, subjects were asked to form a line for payment. After filling out their receipt and turning in their booklet, they were paid $12 in cash and allowed to leave. Occasionally, because the classes had strict start and end times, the experiment took longer than the allotted class time. However, from our observations, most students were still able to complete the survey within sixty minutes. In total, we had 214 subjects.[19] Of these, 193 identified as at least partially black.

3.4 Results

Because we conducted our experiment in classrooms during normal class times, we were not able to control the racial composition of the sessions. Consequently, not all of our subjects were black. See Table 6 for the demographics of the students who agreed to participate in the experiment. In our analysis, we only considered students that identified as black or African-American. If they identified as multiracial, they were included in our analysis so long as one of their self-identified races was black or African-American.

Also, because of the nature of the experiment, some subjects came to class late and thus missed the introduction of the experiment. These subjects were always given control booklets but were excluded from the analysis because they did not have sufficient time to complete the

[19] One subject participated in the study twice. Data from the second time he participated in the study were excluded from all analysis.

booklet. In addition, careful notes were taken of students who were distracted (e.g., using their cell phones during the experiment) and students who skipped ahead. These students were also were excluded from the analysis. This made our final sample size 176: 104 were in the black female experimenter sessions, and 72 were in the white male experimenter sessions.

Table 6. Subject Demographics

	Full Sample (1)	Blacks Only (2)	Final Sample (3)
Percent Blacks	90.19%	100%	100%
Percent Non-Hispanic Whites	1.40%	0%	0%
Percent Hispanic	8.88%	2.59%	2.84%
Percent Female	51.87%	50.78%	51.14%
Average Age	20.55	20.47	20.43
Average Year in School	2.59	2.61	2.60
Average Cumulative GPA	3.02	3.01	3.02
N	214	193	176

Notes: Anyone who indicated that they were at least partially black or African American are counted as black. The final sample includes blacks who came on time, did not skip ahead, and were not distracted. Year in School is 1 for freshmen, 2 for sophomores, 3 for juniors, and 4 for seniors. Three subjects never reported their race. Some other variables (e.g., age and cumulative gpa) were not reported for all subjects.

3.4.1 Results From Black Female Experimenter Sessions

If black students at an HBCU respond the same way to stereotype threat as black students at a PWI, then our results should confirm those of Steele and Aronson (1995). Confirmation

would mean that students in the control group correctly answered more questions than students in the treatment group. Before we analyzed their correct answers, we first examined the number of questions they answered in total.

It is possible that being exposed to the primes caused subjects in the treatment group to answer fewer questions. They may have found it pointless to disprove the stereotype and thus purposely chose to exert less effort, or they may have unconsciously been distracted by the stereotype and unable to answer as many questions. While we could not distinguish between these two possibilities, we were able to determine that subjects in the treatment group answered an average of 17.53 questions. Subjects in the control group answered an average of 17.81 questions, and this difference is not statistically significant (t-test p=0.14).

This result is reported, along with other summary statistics of the treatment effects, in Panel A of Table 7. Figure 9 shows the distribution of questions answered by group. Eighty-nine percent of subjects in the control group answered all of the questions, and seventy-eight percent of subjects in the treatment group answered all eighteen questions. However, the distribution of questions answered for the treatment group is not significantly different from that of the control group (Mann-Whitney p=0.15).

Table 7. Treatment Effect Summary Statistics for Final Sample

Panel A: Black Female Experimenter Sessions

	Control Group	Treatment Group	p-value for two-sample t test
Average Questions Answered	17.81	17.53	0.14
	(0.59)	(1.22)	
Average Correct Answers	4.75	4.80	0.92
	(2.81)	(2.46)	
Average Accuracy (%)	26.86	27.48	0.83
	(16.21)	(13.81)	
N	53	51	

Panel B: White Male Experimenter Sessions

	Control Group	Treatment Group	p-value for two-sample t test
Average Questions Answered	17.84	17.44	0.10
	(0.59)	(1.33)	
Average Correct Answers	5.39	5.21	0.77
	(2.80)	(2.59)	
Average Accuracy (%)	30.24	29.83	0.91
	(15.61)	(14.27)	
N	38	34	

Note: The standard deviations in parenthesis. Questions answered and correct answers have a maximum value of 18. Accuracy has a maximum value of 100.

Figure 9. Distribution of Questions Attempted in Black Female Experimenter Sessions

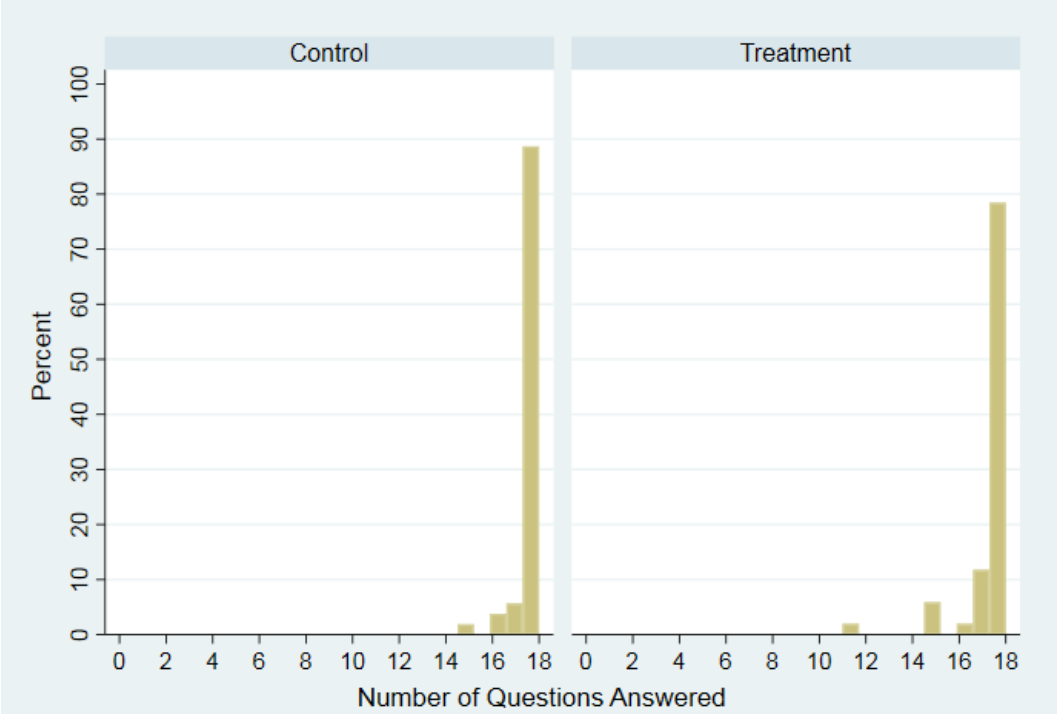


Figure 10 displays the average number of questions subjects in the control and treatment groups answered correctly. The treatment group had an average of 4.80 correct answers whereas the control group had an average of 4.75. This difference is not statistically significant (t-test p=0.92). The distributions of correct answers of these two groups are also similar (Mann-Whitney p=0.90). Because both groups responded to a different number of questions, we compared each group's accuracy, which we calculated as the number of correct answers divided by the number of questions answered. The average control group's accuracy is 26.86%, which is not statistically distinguishable from the treatment group's average of 27.48% (t-test p=0.83). There is no significant difference in the distribution of accuracy, either (Mann-Whitney p=0.68).

As a result, there is no evidence of stereotype threat during the black female experimenter sessions.

Figure 10. Mean Correct Answers in Black Female Experimenter Sessions

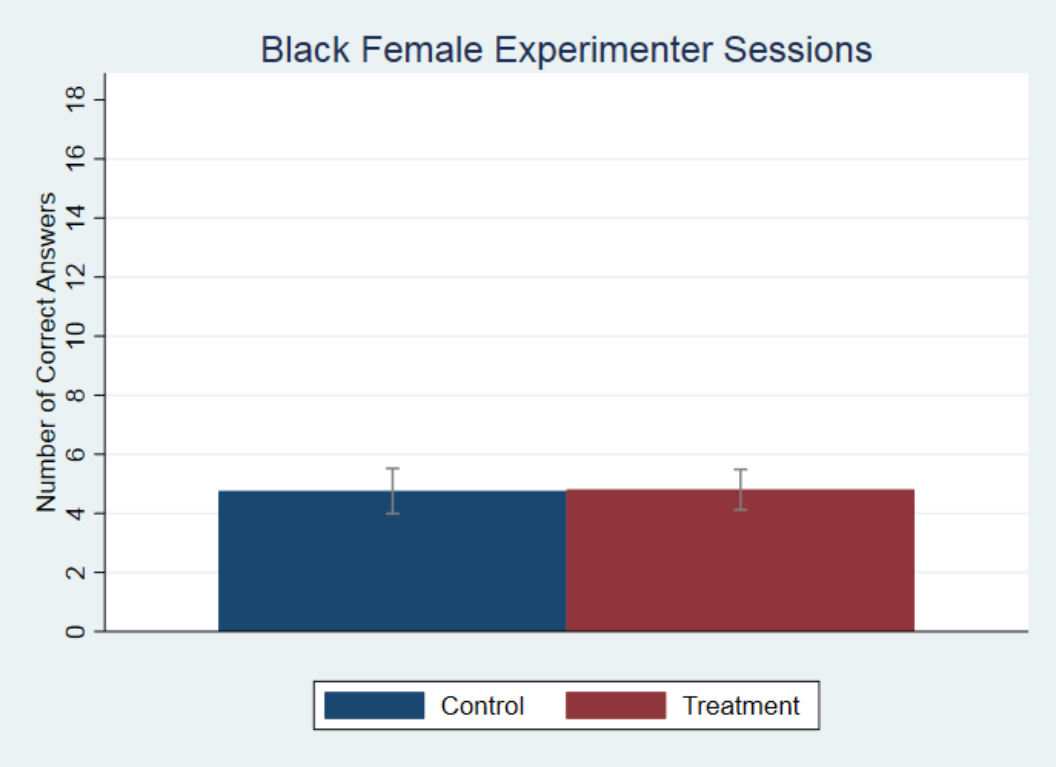


Table 8 shows the results of a Tobit model censored at 18 with the number of questions answered as the dependent variable. Column 1 shows the results for the black female experimenter sessions without any controls. We included a dummy variable called *Treatment*

that equaled one if the subject was in the treatment group. Column 2 shows the same regression with controls for age, gender, GPA, and session effects. In both models, *Treatment* is negative but not statistically significant. When estimating the marginal effects of the second regression, we find that the treatment group would have answered 1.49 questions less than the control group. However, this difference is not statistically significant.

Table 8. Tobit Estimates for Questions Answered

	Black Female Experimenter		White Male Experimenter	
	(1)	(2)	(3)	(4)
Treatment	-1.58	-1.49	-2.55	-2.68
	(1.10)	(1.05)	(1.84)	(1.76)
Age		0.27		-0.25
		(0.23)		(0.28)
Female		-1.55		-0.36
		(0.93)		(1.78)
GPA		-0.35		0.37
		(0.92)		(1.64)
Constant	22.32***	26.00**	24.75***	27.48***
	(1.35)	(6.87)	(2.02)	(7.52)
N	104	95	72	68

Coefficients reported. Robust standard errors in parentheses. Specifications 2 and 4 include session effects. *p<0.10, **p<0.05, ***p<0.01

To estimate the effect of priming on students' tests scores, we ran an ordinary least squares (OLS) model using the number of correct answers and accuracy as the dependent variables in Tables 9 and 10, respectively. The dummy variable *Treatment* is small and only becomes negative once controls are added. More importantly, *Treatment* never reaches statistical significance. This means that there is no evidence of stereotype threat. Those with a higher GPA answered more questions correctly and had higher accuracy. In addition, women were 5.17 percentage points more accurate than men. Overall, we find that priming black subjects to think of negative stereotypes had little effect on their performance on the test when there was a black female experimenter present.

Table 9. OLS Estimates for Correct Answers

	Black Female Experimenter		White Male Experimenter	
	(1)	(2)	(3)	(4)
Treatment	0.05	-0.21	-0.19	0.15
	(0.52)	(0.55)	(0.64)	(0.68)
Age		-0.00		0.19
		(0.07)		(0.18)
Female		0.91		0.28
		(0.55)		(0.74)
GPA		1.59***		-0.44
		(0.58)		(0.64)
Constant	4.76***	0.11	5.40***	2.33
	(0.39)	(2.34)	(0.46)	(3.97)
N	104	95	72	68

Robust standard errors in parentheses. Specifications 2 and 4 include session effects. *p<0.10, **p<0.05, ***p<0.01

Table 10. OLS Estimates for Accuracy

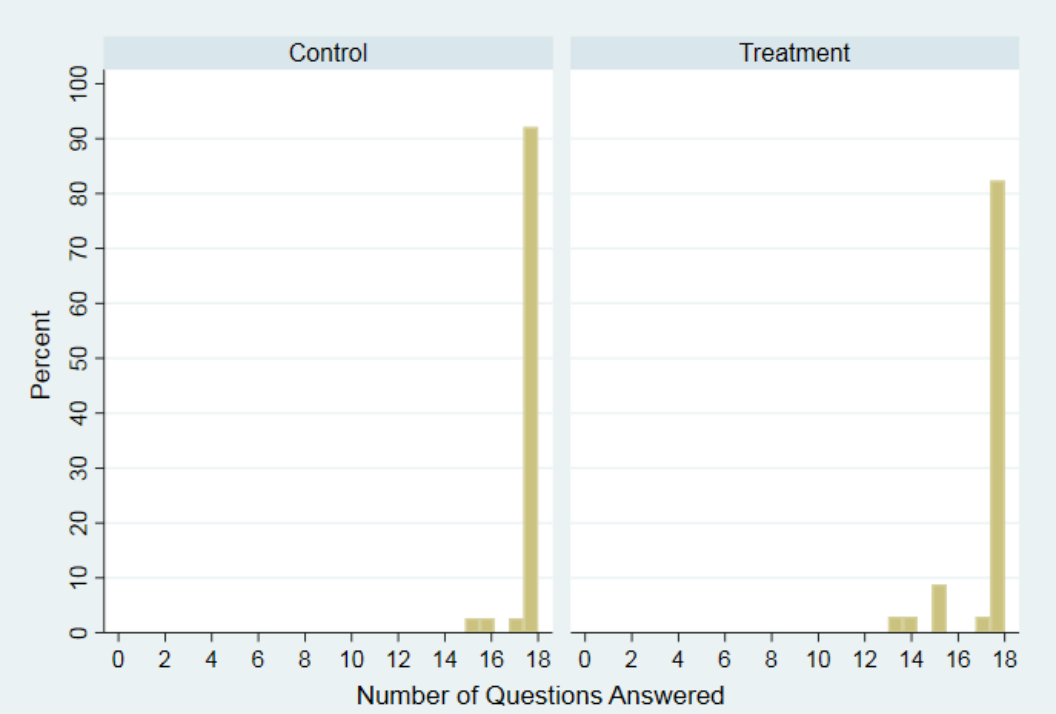
	Black Female Experimenter		White Male Experimenter	
	(1)	(2)	(3)	(4)
Treatment	0.62	-0.99	-0.42	1.43
	(2.95)	(3.08)	(3.52)	(3.72)
Age		-0.10		1.12
		(0.41)		(0.97)
Female		5.27*		1.79
		(3.08)		(4.08)
GPA		9.17***		-2.33
		(3.28)		(3.55)
Constant	26.86***	3.07	30.24***	11.99
	(2.23)	(13.63)	(2.54)	(21.86)
N	104	95	72	68

Robust standard errors in parentheses. Specifications 2 and 4 include session effects. *p<0.10, **p<0.05, ***p<0.01

3.4.2 Results From White Male Experimenter Sessions

For students in the white male experimenter sessions, the average number of questions answered for the control group was 17.84 relative to 17.44 for the treatment group. These distributions are shown in Figure 11. Here we see that 92 percent of subjects in the control group answered every question possible compared to 82 percent of subjects in the treatment group. When comparing the averages and distributions, no significant differences are found (t-test p=0.10 and Mann-Whitney p=0.18, respectively). Therefore, priming subjects seemed to have no influence on how many questions subjects answered when there was a white male experimenter.

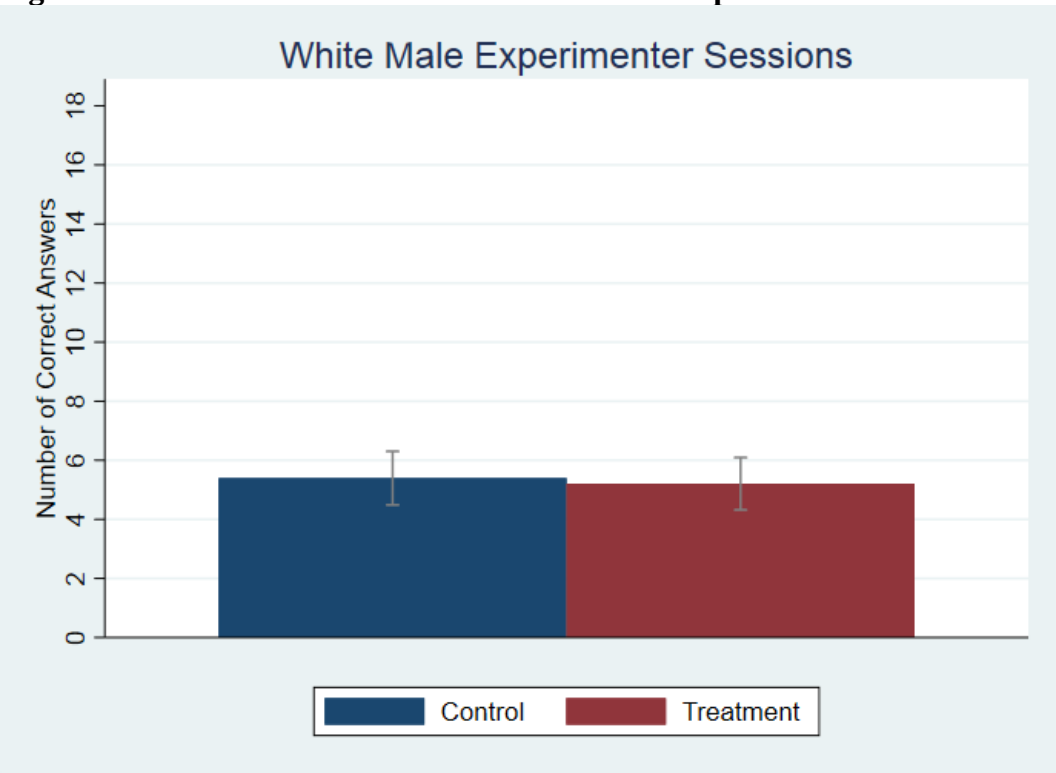
Figure 11. Distribution of Questions Answered in White Male Experimenter Sessions



To understand how subjects' test score may have been affected by the primes, we studied the number of questions subjects answered correctly and their accuracy. Figure 12 displays the mean number of correct answers for each group. The control group correctly answered an average of 5.39 questions. The treatment group's average was 5.21; there is no significant difference in the averages or distributions (t-test p=0.77 and Mann-Whitney p=0.87, respectively). Unsurprisingly, the accuracies of both groups were similar. As reported in Panel B of Table 7, the average accuracy for the control group was 30.24 percent and was 29.83 percent for the treatment group. There is no significant difference in means (t-test p=0.91) or

distributions (Mann-Whitney p=0.94). This suggests that there is no evidence of stereotype threat when there is a white male experimenter.

Figure 12. Mean Correct Answers in White Male Experimenter Sessions



These findings were confirmed in Columns 3 and 4 of Tables 8 through 10. Table 8 reports the results of a Tobit model where the dependent variable is the number of questions answered. Notice that the coefficient for *Treatment* is negative. By examining the marginal

effects, we can say that subjects in the treatment group would have answered 2.68 questions less than subjects in the control group after adding controls, but this difference is not significant. *Treatment* is small and insignificant in Columns 3 and 4 of Table 9, which shows the results of an OLS regression where the dependent variable is the number of correct answers. The same is true in Columns 3 and 4 of Table 10, which reports the results of models where the dependent variable is the subject's accuracy. Neither gender nor GPA had a significant effect on the number of questions answered correctly or their accuracy. This all confirms that there is no evidence to support the belief that stereotype threat affects black HBCU students when a white male experimenter is present.

3.4.3 Results on the Effect of Experimenter Type

While we showed that there was no evidence of stereotype threat in either the black female experimenter sessions or the white male experimenter sessions, it is still possible that students in these two groups responded differently to being primed. Table 11 compares the demographics of black subjects in both sessions. The white male experimenter sessions included older students (t-test $p=0.04$) who were further along in their education (t-test $p<0.01$) than students in the black female experimenter sessions. Students in the white male experimenter sessions also had a lower GPA (t-test $p<0.01$). For this reason, we controlled for age, year in school, and GPA in our regression analysis.

Table 11. Subject Demographics by Experimenter Type (Final Sample)

	Black Female Experimenter (1)	White Male Experimenter (2)	t-test for Difference (3)
Percent Hispanic	1.92%	4.17%	p=0.38
Percent Female	53.85%	47.22%	p=0.39
Average Age	20.04	21.00	p=0.04**
Average Year in School	2.38	2.93	p<0.01***
Average Cumulative GPA	3.14	2.86	p<0.01***
N	104	72	

Notes: p-values report the results of a two-sample t-test with equal variances. Age (n=70), year in school (n=71), and GPA (n=70) were not reported for all subjects in the White Male sessions. GPA (n=95) was not reported for all subjects in the Black Female sessions. *p<0.10, **p<0.05, ***p<0.01

We hypothesized that students with the white male experimenter would experience an added level of threat compared to students with the black female experimenter. If our hypothesis were correct, it would be consistent with findings from studies like that of Marx and Goff (2005) who found that stereotype threat was not as severe when there was a black rather than white experimenter.

To test our hypothesis, we ran four sets of regressions for each dependent variable. For Tables 12 through 14, Column 1 displays the estimates from a regression where we measured the effect of the primes on all of the subjects in our final sample, and Column 2 adds controls. Next, we added a dummy variable in Column 3 called *Black Female* to examine the effect of having a black female experimenter. This dummy variable equaled one if the student had a black female experimenter. Column 4 adds an interaction term called *Treatment*Black Female*, which captures the additional effect of having a black female experimenter for those in the treatment

group. We expected the coefficient on *Black Female* to be positive, which would suggest that students with a white male experimenter performed worse on the exam because they felt threatened by the experimenter. We also predicted that the interaction term would be positive. This would imply that stereotype threat was more detrimental for those who had a white male experimenter rather than a black female experimenter.

Columns 1 through 4 of Table 12 show that the black female experimenter had a negative effect on the number of questions answered, which is the opposite of what we anticipated. However, the difference between the treatment and control groups in terms of the number of questions answered was similar for subjects with both types of experimenters. Therefore, having a black female experimenter had a negative effect on all subjects, independent of whether they were primed.

The results from Table 13 confirm that stereotype threat did not affect students' score since the coefficient on *Treatment* is small and never significant. Furthermore, Columns 3 and 4 show that the presence of a black female experimenter had no significant effect on performance.[20] We also failed to find evidence that students in the treatment group fared any better, relative to the control group, if they had a black female experimenter versus a white male experimenter. We can only say that students with higher GPAs performed better than those with a lower GPA.

[20] We note here that the experimenters came from a PWI. Thus, there is a possibility that even the black female experimenter may have been perceived as "threatening." Our results may have been different if both experimenters had come from an HBCU.

Table 12. Tobit Estimates for Questions Answered (Pooled)

	(1)	(2)	(3)	(4)
Treatment	-1.92*	-1.91**	-1.91**	-2.39*
	(0.96)	(0.93)	(0.93)	(1.46)
Black Female Experimenter			-4.23**	-4.83**
			(1.92)	(2.04)
Treatment * Black Female Experimenter				0.84
				(1.82)
Age		0.43	0.43	0.42
		(0.31)	(0.31)	(0.32)
Female		-0.93	-0.93	-0.86
		(0.85)	(0.85)	(0.88)
Year in School		-1.28*	-1.28*	-1.30**
		(0.65)	(0.65)	(0.64)
GPA		-0.41	-0.41	-0.37
		(0.89)	(0.89)	(0.88)
Constant	23.14***	17.57***	21.85***	22.14
	(1.12)	(6.50)	(6.45)	(6.50)
N	176	163	163	163

Notes: Robust standard errors in parentheses. Specifications 2 through 4 include session effects. *p<0.10, **p<0.05, ***p<0.01

Table 13. OLS Estimates for Correct Answers (Pooled)

	(1)	(2)	(3)	(4)
Treatment	-0.06	-0.19	-0.19	-0.14
	(0.40)	(0.43)	(0.43)	(0.68)
Black Female Experimenter			0.27	0.31
			(0.85)	(0.95)
Treatment * Black Female Experimenter				-0.08
				(0.92)
Age		0.04	0.04	0.04
		(0.07)	(0.07)	(0.07)
Female		0.60	0.60	0.59
		(0.44)	(0.44)	(0.44)
Year in School		0.01	0.01	0.01
		(0.27)	(0.27)	(0.29)
GPA		0.84*	0.84*	0.83*
		(0.47)	(0.47)	(0.47)
Constant	5.02***	1.74	1.47	1.44
	(0.30)	(2.12)	(1.91)	(1.94)
N	176	163	163	163

Notes: Robust standard errors in parentheses. Specifications 2 through 4 include session effects. *p<0.10, **p<0.05, ***p<0.01

Table 14. OLS Estimates for Accuracy (%, Pooled)

	(1)	(2)	(3)	(4)
Treatment	0.15	-0.74	-0.74	-0.19
	(2.26)	(2.41)	(2.41)	(3.76)
Black Female Experimenter			-3.12	3.56
			(4.78)	(5.36)
Treatment * Black Female Experimenter				-0.96
				(5.09)
Age		0.15	0.15	0.15
		(0.39)	(0.39)	(0.40)
Female		3.60	3.60	3.50
		(2.45)	(2.45)	(2.49)
Year in School		0.25	0.25	0.31
		(1.53)	(1.53)	(1.63)
GPA		5.05*	5.05*	5.06*
		(2.64)	(2.63)	(2.65)
Constant	28.27***	11.22	8.10	7.79
	(1.67)	(12.04)	(10.74)	(10.90)
N	176	163	163	163

Notes: Robust standard errors in parentheses. Specifications 2 through 4 include session effects. *p<0.10, **p<0.05, ***p<0.01

The results from Table 14 complement the results discussed above. Mainly, we find that priming black students had no significant effect on their accuracy and neither did the type of experimenter. Furthermore, students in the treatment group were just as accurate, relative to the control group, when they had a white male experimenter as they were when they had a black female experimenter. The only significant predictor of accuracy is GPA. Therefore, these findings fail to support the idea that the test scores of black students at HBCUs are affected by stereotype threat. They also suggest that these students are impartial to the presence of white men when taking difficult tests under threat.

3.5 Conclusion

Stereotype threat is a term used to describe the feeling of being at risk of confirming a negative stereotype about one's social group. Steele and Aronson (1995) used black students at Stanford University to show that the performance of students on a verbal GRE fell when they were reminded, subtly, of the stereotype that blacks are not as smart as whites. Other research has produced similar findings (e.g., Brown and Day, 2016; Davis et al., 2006). Yet the majority of these studies focus on the effect of stereotype threat on black students attending predominantly white institutions.

In our experiment, we examine the role of stereotype threat on the verbal GRE test scores of black students attending a historically black university. There is evidence in economics and other fields that supports the idea that black students at HBCUs may benefit from this type of environment. For example, HBCUs are known for having a diverse faculty, and studies like Egalite et al. (2015) show that young black students have higher reading and math test scores when they have a black teacher. At an HBCU, black professors could serve as a valuable presence for students who are coping with the effects of negative stereotypes about their intelligence or academic motivation. Furthermore, Fries-Britt and Turner (2002) mention that students attending HBCUs are more likely to have the opportunity to learn about the scholarly achievements of other blacks. Research has shown that stereotype threat is reduced when people read about successful people from their identity group (McIntyre et al., 2003), so that is another way in which HBCUs may make a difference. Altogether, these factors could reduce the negative effects of stereotype threat for black students at HBCUs.

We conducted our experiment to test this idea with students at a historically black university in Texas. Students in the treatment group were primed to think about the negative

stereotype about blacks' intelligence before answering questions on a mock verbal GRE. Students in the control group were not primed before taking the exam.

We also varied the level of threat by changing the identity of the lead experimenter. In some sessions, the lead experimenter was a black female (low threat); in others, the lead experimenter was a white male (high threat). According to standard hypotheses about the effects stereotype threat, students in the treatment group should have performed worse than students in the control group. Additionally, we predicted that students with a white male experimenter would answer fewer questions correctly than students with a black female experimenter.

None of these predictions were confirmed in this study. In both the black female experimenter sessions and the white male experimenter sessions, we found little evidence that stereotype threat negatively affects students' test score. Students in the control group correctly answered a similar number of questions as students in the treatment group and had similar accuracy.

We also failed to find strong evidence to support the idea that students' test scores were affected by the identity of the experimenter. Having a black female experimenter had a negative effect on the total questions students answered, but the type of experimenter had no significant effect on the number of questions students answered *correctly*.

Thus, we find preliminary evidence that stereotype threat may not be a problem for black students at HBCUs. If that is the case, then studying the recruitment strategies, teaching methods, on-site resources, and general environment of HBCUs could teach us how to prepare black students to deal with negative stereotypes about their intellect. In turn, when they are confronted with these biases, their performance will not be affected, and their grades will better reflect their latent ability. This may help reduce America's persistent black-white achievement gap.

Therefore, we recommend that future research should investigate differences between HBCUs and PWIs and the students who choose to attend them and study how these differences affect the performance of black college students.

3.6 References

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2), 113-125.

Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*(3), 225-229.

Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on raven's advanced progressive matrices. *Journal of Applied Psychology, 91*(4), 979.

Cohen, G. L., & Garcia, J. (2005). " I am us": negative stereotypes as collective threats. *Journal of Personality and Social Psychology, 89*(4), 566.

Davis III, C., Aronson, J., & Salinas, M. (2006). Shades of threat: Racial identity as a moderator of stereotype threat. *Journal of Black Psychology, 32*(4), 399-417.

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics, 86*(1), 195-210.

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review, 45*, 44-52.

Ehrenberg, R. G., & Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review, 13*(1), 1-17.

Fairlie, R. W., Hoffmann, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review, 104*(8), 2567-2591.

Fries-Britt, S., & Turner, B. (2002). Uneven stories: Successful Black collegians at a Black and a White campus. *The Review of Higher Education, 25*(3), 315-330.

Fryer Jr, R. G., & Greenstone, M. (2010). The changing consequences of attending historically Black colleges and universities. *American Economic Journal: Applied Economics, 2*(1), 116-148.

Goldsmith, A. H., Veum, J. R., & Darity Jr, W. (1996). The psychological impact of unemployment and joblessness. *The Journal of Socio-Economics, 25*(3), 333-358.

Gurin, P., & Epps, E. (1975). Black consciousness, identity, and achievement: A study of students in historically Black colleges.

Hanushek, E. A., & Rivkin, S. G. (2009). Harming the best: How schools affect the black- white achievement gap. *Journal of Policy Analysis and Management, 28*(3), 366-393.

Herrnstein, R. J., & Murray, C. (1994). The bell curve: The reshaping of American life by differences in intelligence. *New York: Free,*

Hoffman, F., & Oreopoulos, P. (2009). *A Professor Like Me: Influence of Professor Gender on University Achievement,*

Lusher, L., Campbell, D., & Carrell, S. (2018). TAs like me: Racial interactions between graduate teaching assistants and undergraduates. *Journal of Public Economics, 159,* 203-224.

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42*(2), 236-243.

Marx, D. M., & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology, 44*(4), 645-657.

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin, 28*(9), 1183-1193.

McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology, 39*(1), 83-90.

McKay, P. F., Doverspike, D., Bowen- Hilton, D., & Martin, Q. D. (2002). Stereotype Threat Effects on the Raven Advanced Progressive Matrices Scores of African Americans 1. *Journal of Applied Social Psychology, 32*(4), 767-787.

McKay, P. F., Doverspike, D., Bowen-Hilton, D., & McKay, Q. D. (2003). The effects of demographic variables and stereotype threat on Black/White differences in cognitive ability test performance. *Journal of Business and Psychology, 18*(1), 1-14.

National Science Foundation. (2015). *Report to Congress on Advancing Historically Black Colleges and Universities.* Retrieved from https://www.nsf.gov/ehr/Pubs/HBCU_Report_Oct_2015.pdf

Nichols, A. H., & Evans-Bell, D. (2017). A look at Black student success: Identifying top-and bottom-performing institutions. The Education Trust. Retrieved from http://edtru.st/2mD7pWL

Official GRE verbal reasoning practice questions with practice for the analytical writing measure. Volume 1: (2014). New York : McGraw-Hill, 2014.

Phinney, J. S. (1992). The multigroup ethnic identity measure: A new scale for use with diverse groups. *Journal of Adolescent Research, 7*(2), 156-176.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*(1), 80-83.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology, 77*(6), 1213.

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science, 20*(9), 1132-1139.

Yeung, W. J., & Conley, D. (2008). Black–white achievement gap and family wealth. *Child Development, 79*(2), 303-324.

CHAPTER IV

FIRED UP?: HOW PAST UNEMPLOYMENT AFFECTS WORKER MOTIVATION AND PRODUCTIVITY

4.1 Introduction

According to the Bureau of Labor Statistics, 6.5 million Americans faced unemployment in January 2019. Beyond the obvious consequence of lost personal earnings, past research has shown that being unemployed has far-reaching, negative effects in sometimes surprising ways. For example, Oreopoulos, Page, and Stevens (2008) showed that children of displaced fathers suffered from lower annual earnings once they became adults compared to children whose fathers did not experience an employment shock. Clark, Knabe, and Rätzel (2010) listed some non-pecuniary costs of unemployment and included a decrease in social status, time spent with non-family members, and sense of purpose. Furthermore, they find that individuals in secure jobs have the highest life satisfaction scores, but individuals who are unemployed and have bad job prospects have the lowest life satisfaction scores. In fact, Clark and Oswald (1994) explored the relationship between happiness and unemployment and discovered that individuals perceive unemployment as adding more disutility than a divorce would. And Baum, Fleming, and Reddy (1986) found that individuals who had been unemployed for a longer period of time showed more signs of stress than employed individuals. Relatedly, Eliason and Storrie (2009) reported that displaced workers are more likely to commit suicide and die due to alcohol-related reasons shortly after losing their job. So it seems clear that unemployment can have strong, detrimental consequences for those who have lost their jobs.

When individuals face so many hardships post-unemployment, it begs the question of (if and) how the unemployed cope with all of these hardships. In this study, I am particularly interested in how past unemployment changes individuals' productivity at their next place of employment and what drives this performance change. People who have been fired because of their poor performance may subsequently feel demotivated after receiving this negative feedback about their performance or ability. Losing their job may hurt their self-esteem, outlook on life, and other factors that would negatively impact their performance at their next job. On the other hand, past unemployment could motivate workers to try harder if they believe that improving their performance in their next job could help them retain that job. Additionally, it is possible that the cause of unemployment (e.g., firing versus layoff) affects people's reactions.

To determine whether individuals' performance rises or falls after a period of unemployment, I designed an experiment where I simulate two forms of job loss in a laboratory setting. Undergraduate subjects participated in four rounds of a real effort task. In one treatment, subjects who performed poorly in the first round were told so and informed that they would not be able to work in the next round for payment. In essence, they were fired for one round. These subjects had the knowledge that their performance was relatively weak, and they faced a period of unemployment as a result. Both of these factors could have affected their performance in the third round when they were rehired.

In another treatment, subjects' ability to work for payment in the second round was randomly determined and did not depend on their performance in the first round. This means that those who lost their job or, as I will refer to it from now on, were laid off knew that their performance played no role in their employment status. Regardless, they still had to suffer through a period of unemployment like the fired subjects. As such, I investigated whether not

working for pay in the second round could affect subjects who had been laid off the same way that it did for subjects who had been fired.

Overall, job loss had little effect on subjects. Fired subjects performed worse in the third round than subjects who were not fired, but there was a performance gap between them in the two previous rounds as well. More interestingly, the performance of (low-performing) fired subjects in the third round was similar to low-performing subjects who were laid off but would have been fired had they been in the other treatment. In other words, low-performing subjects responded to being laid off the same way as they did to being fired. Furthermore, there was no significant difference in performance between low performing subjects who were not laid off and (low-performing) subjects who had been fired. Neither was there a significant difference in performance between low-performing subjects who were and were not laid off. This suggests that subjects were unaffected from losing their job. Similar statements can be made about performance growth, or the change in performance pre- and post-unemployment. In general, low-performing subjects who faced job loss did not appear to be encouraged or discouraged relative to their similarly-able counterparts who kept their jobs. This implies that employers who are considering hiring someone who has recently lost their job should focus more on the quality of the candidate versus the reason for their unemployment because low-performing individuals are going to have similar productivity if they have been fired or laid off.

This study also examines the effect that unemployment has on self-esteem and locus of control, two psychological traits that several studies have connected with unemployment (e.g., Darity and Goldsmith, 1996; Goldsmith, Veum, and Darty, 1997; Kanferg, Wanberg, and Kantrowitz, 2001). In this experiment, I find that losing one's job did not affect one's locus of control. However, fired subjects tended to have more of an external locus of control than subjects

who were not fired, even prior to the start of the experiment. Regarding self-esteem, subjects experienced an increase in self-esteem over time, but only laid off subjects seemed to be affected by job loss.

The rest of the paper is organized as follows. Section two of the paper covers past studies that have explored the relationship between unemployment, job search, and effort. There is also a brief discussion of how feedback on performance can affect behavior. Section three of the paper outlines the design of the experiment. Section four describes the results on how unemployment affects performance and how being fired for cause may have a differential effect. Lastly, section five summarizes the experiment and its findings.

4.2 Literature Review

When trying to understand the relationship between past unemployment and future employment, many researchers have focused on job search behavior. For instance, Clark et al. (2010) showed that job search behavior was more likely among unemployed individuals who had a greater drop in life satisfaction. These people also made greater wage concessions when searching for their next job. Bennett et al. (1995) surveyed employees before and after they were laid off and found that people who found the layoff to be fair and people who were upset were less likely to participate in job search activities. Kanfer and Hulin (1985) studied individuals four weeks after they had been laid off. They found that layoff victims who had successfully found work submitted a similar number of job applications as layoff victims who were unsuccessful in their search. The main difference they saw was in the applicants' confidence in their ability to be successful at the job search process (e.g., in their ability to complete job applications). Hence, emotional responses and attitudes about one's capabilities have been shown to play a role in how actively unemployed try to find new jobs.

The previously mentioned studies show a connection between unemployment and future employment as it relates to job search behavior. Another way that unemployment could affect future employment is through changes in the worker's performance. This study explores that relationship. There have been other studies that have investigated the connection between unemployment and performance. Some papers were able to get precise measures of performance by utilizing real effort tasks. For example, Baum et al. (1986) recruited employed and unemployed individuals and asked them to find simple objects hidden within larger geometric shapes. They found that accuracy on this task was higher among the employed. This then suggests that unemployed individuals perform worse than employed individuals.

Many other studies that investigate the relationship between unemployment and performance used self-reported measures of effort and productivity. For instance, Brandes et al. (2008) surveyed managerial and administrative employees who had survived a previous wave of layoffs at their companies and asked them to report their level of job insecurity and work intensity, which was used as a measure of effort. They discovered a negative correlation between job insecurity and effort. These studies seem to suggest that unemployed individuals and people who fear future unemployment might have lower performance than those in stable jobs.

Not all studies reached the same conclusion. Probst et al. (2007) conducted an experiment where they threatened to layoff subjects in the treatment group after the first work period. Subjects in the treatment group increased their performance in the second work period compared to subjects in the control group who were not threatened. They also surveyed employees to measure their job insecurity and counterproductive work behavior. They identified a negative relationship between these two variables. This then suggests that unemployed people and people who feel at risk of becoming unemployed may be more productive than their steadily employed

peers. Consequently, there is no clear consensus in the literature as to whether past unemployment increases or decreases performance after an individual has become reemployed.

Furthermore, since these studies either restrict their sample to subjects who have been laid off or pooled all types of unemployed individuals together, they do not study differences between being fired and being laid off. One difference between the two is the type of feedback they receive right after they lose their jobs. More specifically, people who have been fired receive negative feedback about their performance that people who have been laid off do not. There are several studies that examined the influence of feedback on performance. Elliot and Harackiewicz (1994) conducted an experiment where one group of subjects played pinball and were given information about whether or not they had achieved a stated goal and whether they had surpassed the 65th percentile of pinball scores. These subjects spent more of their break period playing pinball than subjects who were only told whether or not they achieved the goal. This implies that feedback on one's relative performance can encourage people to exert more effort on tasks. Azmat and Iriberri (2010) took advantage of a natural experiment where high school students were told whether they performed better or worse than the class average. They found that this information increased grades by 5%. So past studies have shown that information on relative performance can help improve future performance. Therefore, fired individual may behave differently than individuals who have been laid off.

This study also aims to get a better understanding of the psychological effects of job loss. Darity and Goldsmith (1996) proposed the idea that unemployment can lower self-esteem and make individuals feel like their life is outside of their control (i.e., have an external locus of control), both of which hurts their productivity. Self-esteem may be of particular concern for people who have been fired for cause because they are being told that their performance was not

up to par. People who have lost their job may feel a loss of control over their lives, especially victims of layoffs who had no control over their employment status. In fact, Kanfer et al. (2001) discovered correlations between locus of control, self-esteem, job search behavior, and duration of unemployment. Therefore, people who suffer from job loss may also experience changes in their psychological traits like their self-esteem and locus of control, which may then impact their success at finding a job and their performance thereafter.

In this study, I used an experiment to study the impact of past unemployment on both performance and psychological traits. By using a real effort task, I captured a clean measure of performance and did not have to rely on self-reported measures of productivity. Measures of subjects' self-esteem and locus of control were taken before and after job loss, so I was also able to determine if job loss affected certain psychological traits. In addition, I compared subjects who were fired because of poor performance to subjects who lost their job due to bad luck. Consequently, I was able to investigate whether the cause of job loss influenced how individuals responded to being unemployed.

4.3 Experimental Design

One hundred and fifty-six subjects from Texas A&M University were recruited in November 2016 and February 2018 to participate in this experiment using ORSEE (Greiner, 2015). Table 15 shows the characteristics of these subjects. Fifty-eight of them were assigned to the Performance Based Unemployment Treatment (PBUT), and ninety-eight subjects were assigned to the Random Unemployment Treatment (RUT). Two days before coming to the laboratory, subjects received an email asking them to complete an online, pre-experimental survey. They were informed that this survey needed to be completed before they came to the laboratory. In compensation, they would earn an additional $2, which would be paid to them at

the end of their session. Of the 35 pre-experimental surveys sent to each section, 30, 31, 34, 32, 33, and 28 surveys were completed for the first, second, third, fourth, fifth, and sixth session, respectively.

Table 15. Background Characteristics of Subjects by Treatment Group (Participants Only)

	Performance	Random	p-value
Age	20.05	20.43	0.09*
	(1.19)	(1.38)	
Fraction Female	0.47	0.55	0.30
	(0.50)	(0.50)	
Fraction White	0.67	0.60	0.38
	(0.47)	(0.49)	
Fraction Black	0.16	0.15	0.97
	(0.37)	(0.36)	
Fraction Hispanic	0.24	0.26	0.85
	(0.43)	(0.44)	
Initial Self-Esteem	4.60	5.02	0.07*
	(1.59)	(1.24)	
Initial Locus of Control	2.81	2.89	0.68
	(1.10)	(1.15)	
GPA	3.18	3.18	0.96
	(0.48)	(0.52)	
Income	1.95	1.71	0.03**
	(0.66)	(0.61)	
Financial Aid Recipient	0.59	0.62	0.66
	(0.50)	(0.49)	
Observations	58	98	

Notes: *p<0.1, **p<0.05, ***p<0.01. Numbers in parentheses are standard deviations. Self-esteem is measured on a scale of 0 to 30. Higher numbers indicate higher levels of self-esteem. Locus of control is measured on a scale of 1 to 4. Higher numbers indicate an internal locus of control. Income is measured on a scale from 1 to 3. Higher numbers indicate higher financial constraint.

The pre-experimental survey asked for their email address, age, sex, race/ethnicity, major, and whether or not they attended high school in the United States. In between these demographic questions, subjects completed two personality trait measures. The Rosenberg Self-Esteem Scale (Rosenberg, 1965) measures an individual's self-worth. Subjects were presented with ten statements (e.g., I feel like I have a number of good qualities.) that they were asked to strongly agree, agree, disagree, or strongly disagree with. The second personality trait measure captured subjects' locus of control (i.e., whether they believed that outside forces or fate controlled their lives). The Abbreviated Rotter was taken from Goldsmith et al. (1997) and consists of four pairs of statements. For each pair (e.g., What happens to me is my own doing. Sometimes I feel that I don't have enough control over the direction my life is taking.), subjects had to identify which statement they felt described them better. These two personality trait measures are provided in Appendix C.

When subjects arrived at the laboratory, they were told that they would be participating in an experiment consisting of several rounds and that they would earn tokens at an exchange rate of 100 tokens to $1.50. Then they were given the instructions for Round One, in which subjects performed a real effort task for four minutes in exchange for 300 tokens. Before the round officially began, each subject had to successfully solve two practice problems. The encryption task came from Benndorf, Rau, and Sölch (2014) who designed a task to minimize learning over time. For each problem, subjects saw a table that assigned a random, three-digit number to each letter in the alphabet. They were also given a random, three-letter "word" to encode. They used the table to find the three numbers associated with each letter and then proceeded to the next problem when their submission was correct. The next problem presented them with a new word

to encode and a new table to help them do so. Screenshots of the practice problems and an example of a problem from Round One are located in Appendix C.

After Round One was over, subjects received feedback on how many problems they solved correctly and how many tokens they earned. In addition, they were presented with Employment Message One. Each subject in the RUT was randomly assigned a digit between zero and one. Those whose digit fell below .5 became "unemployed" or "laid off." More specifically, Employment Message One told them that they had a 50% chance of being allowed to participate in the next round for payment, and it was randomly determined that they would not be participating for pay. Instead, they were told that the next round could be used for practice. On the other hand, subjects whose digit was .5 or higher remained "employed" and were given a similar explanation for how employment status was determined. The only difference was that they were told that they would be allowed to participate in Round Two for another 300 tokens.

Subjects in the PBUT were ranked based on their performance in Round One relative to others in their session. This was explained to subjects in Employment Message One. Those subjects who fell in the top half (i.e., the number of problems they solved in Round One was greater than the session's median) were told so and informed that they would be allowed to participate in Round Two for 300 tokens. In other words, they remained "employed." Those who fell in the bottom half (i.e., the number of problems they solved in Round One was less than the session's median) were informed that they could practice in Round Two but would not receive any payment for their effort. This was because they were "unemployed" or "fired." The precise wording used in Employment Message One for RUT and PBUT subjects can be seen in the screenshots in Appendix C.

Before beginning Round Two, all subjects completed the same Rosenberg Self-Esteem Scale (Rosenberg, 1965) and Abbreviated Rotter (Goldsmith et. al, 1997) from the pre-experimental survey. This allowed me to capture their self-esteem and locus of control after they had worked for one period and then possibly lost their job. These measures were then compared to their responses from the pre-experimental survey. After they finished answering the questions, subjects were given a short recap of the instructions for the encryption task and a reminder of how long the round would last and how many tokens they would earn for their time (i.e., 300 tokens if they were employed and 0 tokens otherwise). Round Two then proceeded in the same fashion as Round One. However, the set of words and tables were all new.

At the end of Round Two, subjects learned how many problems they had solved in that round and were reminded of how many tokens they had earned. Below this information was a recap of the instructions for the encryption task. Subjects were told that they would earn 300 tokens in Round Three compensation for their time. This was true for all subjects irrespective of their treatment group or previous employment status. As a result, the subjects' behavior in Round Three showed what happened to subjects once they had been reemployed.

After Round Three ended, subjects got the same feedback as they did in previous rounds (i.e., number of correct answers and tokens earned). In addition, they received Employment Message Two, which can be found in Appendix C. Subjects in the RUT were assigned a new random digit between zero and one, which determined their employment status in Round Four. Subjects in the PBUT were given a new ranking that depended on how others in their session performed in Round Three, and only those in the top half were paid for their efforts in Round Four. Then Round Four proceeded just as Round Two had.

Once Round Four was finished, subjects were told that the experiment was completed. They were told how many tokens they had earned in total and how much they had earned in dollar amount, including the $2 payment from the online, pre-experimental survey. Before subjects were paid, they completed a questionnaire that asked for their cumulative grade point average (GPA), financial security, previous job experience, rating for how much they enjoyed the task, their thoughts during Round Three, and how fair they thought the experiment was. The exact questions are listed in Appendix C.

Each session lasted approximately one hour. At the end of the session, subjects were paid in private with the maximum payment being $20.00 and the minimum being $11.00. The average payment for PBUT subjects was $15.89 while the average payment for subjects in the RUT was slightly but insignificantly lower at $15.50 (p=0.51).

4.4 Results

In the PBUT, 26 subjects were fired after Round One, and 32 kept their jobs. In the RUT, 49 subjects were laid off after Round One, and 49 were not. While subjects had the potential of losing their jobs after Round Three, for the rest of the analysis, "(not) fired subjects" will only refer to subjects from the PBUT who earned zero (three hundred) tokens in Round Two. Similarly, "(not) laid off subjects" will only refer to subjects from the RUT who earned zero (three hundred) tokens in Round Two.

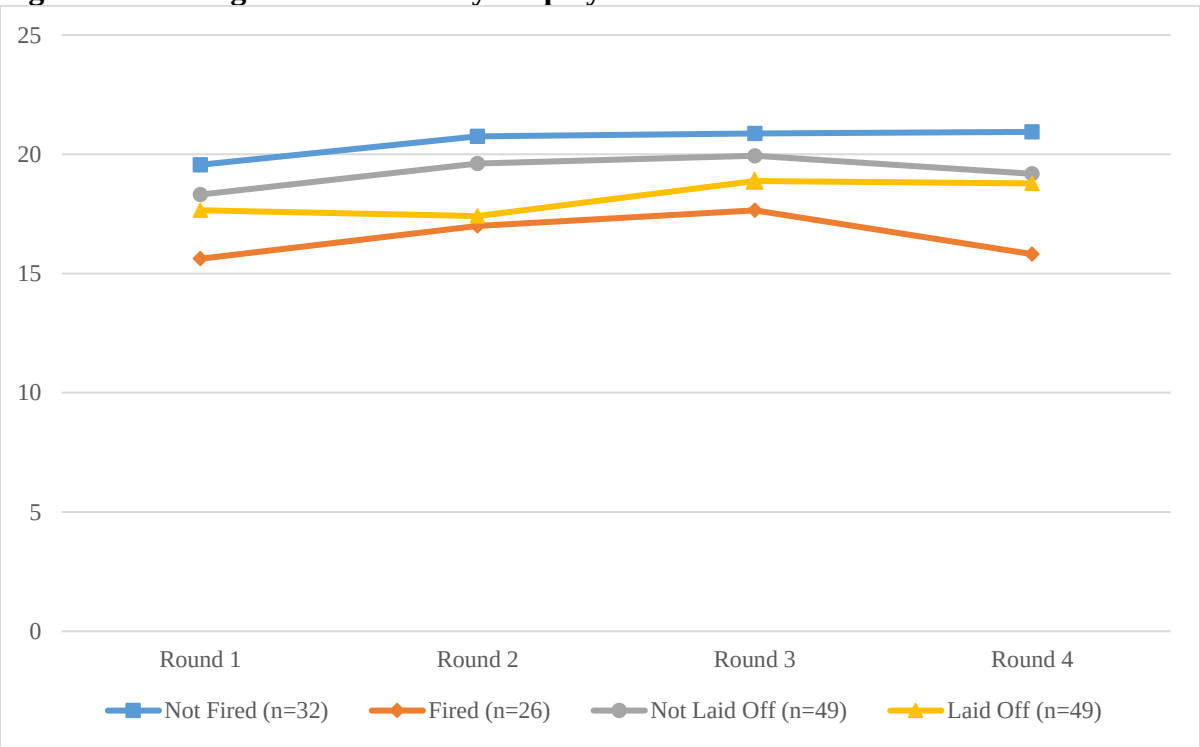
4.4.1 Initial Round One Performance

Figure 13 shows the average performance of not fired, fired, not laid off, and laid off subjects by round. Because subjects' employment status in Round Two was determined by Round One performance for subjects in the PBUT, it was expected that not fired subjects would solve more problems correctly in Round One compared to fired subjects. Indeed, those who were

not fired solved 19.6 problems, on average, compared to the 15.6 problems solved by subjects who were fired. This difference is statistically significant ($p<0.01$).

On the other hand, employment status in Round Two was determined randomly for subjects in the RUT, so there should not have been a significant difference in the average Round One performance of not laid off subjects compared to laid off subjects. Subjects who were not laid off after Round One solved 18.3 problems, on average, which was not significantly different than the 17.7 problems solved by laid off subjects ($p=0.31$).

Figure 13. Average Performance by Employment Status and Round



4.4.2 Round Two Performance

Subjects in the PBUT all received information about their relative standing as soon as Round One was over, but only not fired subjects spent Round Two working for payment. In Round Two, not fired subjects solve 20.8 problems, on average, while fired subjects solve 17.0 problems. This difference is statistically significant ($p<0.01$). So fired subjects performed worse than not fired subjects in Round Two.

Laid off subjects performed significantly worse than not laid off subjects in Round Two ($p<0.01$). Laid off subjects solved 17.4 problems, which was less than the 19.6 problems solved by their employed counterparts. This suggests that subjects in the RUT responded to losing their job after Round One. They could be responding to the notification that they were "laid off," or they could be responding to the lack of financial incentives.

Fired subjects could also be responding to the notification that they lost their job or the lack of incentives. However, it is hard to tell if either of these explanations is true from comparing the performance of fired subjects to not fired subjects because they have different abilities. For this reason, it is important to compare fired subjects to other low-performing subjects. There were 43 subjects in the RUT who performed below their session's median in Round One. In other words, these are subjects who would have been fired had they been in the PBUT. There were 25 low-performing laid off subjects, and they solved 15.7 problems. According to a Mann-Whitney test, fired subjects performed better than low-performing laid off subjects ($p=0.03$). Given that there was no statistically significant difference between their performance in Round One ($p=0.32$), this implies that low-performing subjects solved more problems if they are fired than if they are laid off.

There were 18 low-performing not laid off subjects, and they solved 16.9 problems. This is not significantly different from the number of problems solved by fired subjects according to a Mann-Whitney test ($p=0.81$). However, a Mann-Whitney test suggests that low-performing not laid off subjects did solve significantly more problems than low-performing laid off subjects ($p=0.05$). The fact that low-performing laid off subjects performed worse than fired subjects and not laid off subjects in Round Two is interesting. It could be that they did not like losing their jobs and recognized that practice would not affect the probability that they retained their jobs in the future. Consequently, they exerted less effort on the task than their peers.

4.4.3 Effect of Past Unemployment on Performance

In Round Three, each subject participated for payment again. However, some subjects had spent the previous round unemployed, and this may have had an effect on how well they performed in Round Three. On average, not fired subjects solved 20.9 problems, which is significantly higher than the 17.7 problems fired subjects solved ($p<0.01$). Not laid off subjects solved 19.9 problems, and laid off subjects solved 18.9 problems. This difference was not significant ($p=0.11$).

Table 16 reports the estimates of several ordinary least squared regressions where the dependent variable is how many problems the subject solved. The first specification only includes a dummy variable called Lost Job that equals one if the subject was fired or laid off. The second specification adds controls for age, whether the subject was white, gender, GPA, and financial security. The third specification includes a dummy variable, Performance Treatment, which equals one for all subjects in the PBUT. Then the fourth specification adds an interaction term between Lost Job and Performance Treatment. This equals one for subjects who were fired.

Table 16. Effect of Unemployment on Round Three Performance

Dependent Variable: Round 3 Performance

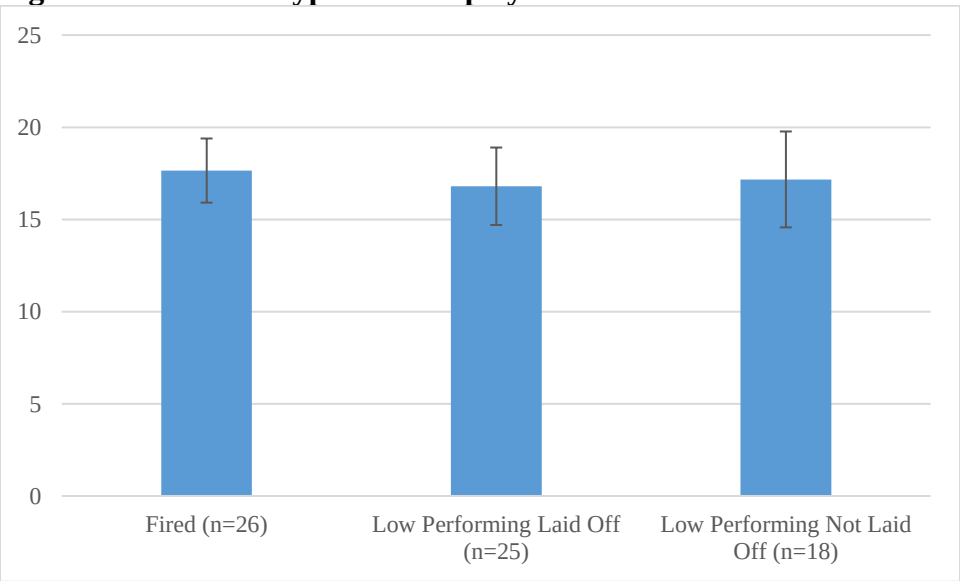
	(1)	(2)	(3)	(4)
Lost Job	-1.86***	-1.78***	-1.78***	-0.86
	(0.46)	(0.48)	(0.48)	(0.67)
Performance Treatment			-0.05	1.16*
			(0.43)	(0.61)
Lost Job * Performance Treatment				-2.56***
				(0.87)
Constant	20.31***	20.31***	20.38***	20.89***
	(0.31)	(4.27)	(4.22)	(4.19)
Controls	N	Y	Y	Y
Observations	156	156	156	156

Notes: Robust standard errors are reported in parentheses. *p<0.1, **p<0.05, ***p<0.01. Controls include age, a dummy for being white, a dummy for being female, GPA, and income. Income is measured on a scale from 1 to 3. Higher numbers indicate higher financial constraint.

Based on the results displayed in columns 1 through 3 of Table 16, it seems like subjects who lost their job performed worse in Round Three than subjects who kept their job. However, column 4 reveals that this is driven by subjects in the PBUT. Overall, subjects in the PBUT answered 1.2 more questions in Round 3 compared to subjects in the RUT. However, subjects who were fired answered 2.6 fewer questions than subjects who randomly kept their job. This seems to suggest that being laid off has no effect on Round Three performance. But it is hard to make any substantive claims about being fired because those who were fired likely had lower ability than those who were not. To make proper comparisons, those who were fired need to be compared to those with similar ability as was done in section 4.4.2.

Figure 14 shows the average Round Three performance of the low performing subjects. Fired subjects solved 17.7 problems, low-performing laid off subjects solved 16.8 problems, and low-performing not laid off subjects solved 17.2 problems. The results of a Mann-Whitney test suggest that there is no significant difference in performance between fired and low-performing laid off subjects (p=0.17), fired and low-performing not laid off subjects (p=0.50), or low ability laid off subjects and low-performing not laid off subjects (p=0.72). This implies that neither job loss nor the reason for job loss has an effect on performance for low-performing subjects.

Figure 14. Effect of Type of Unemployment on Round Three Performance



To address the fact that ability was heterogeneous, I also studied the difference between subjects' Round Three and Round One performance. Not fired subjects improved by 1.3 problems whereas fired subjects improved by 2.0 problems. This difference is marginally significant (p=0.08). On the other hand, not laid off subjects improved by 1.2 problems, which was similar to laid off subjects who improved by 1.6 problems (p=0.31). This suggests that exogenous unemployment has no effect on performance growth, but performance-based unemployment might.

Table 17 reports estimates of several ordinary least squared regressions where the dependent variable is the difference between Round Three and Round One performance. The independent and control variables are the same as those in Table 16. If subjects became motivated after losing their job, Lost Job should be positive. If subjects became discouraged, it should be negative. The results indicate that losing one's job did not have a significant effect on one's relative Round Three performance. Neither did being a subject in the PBUT or specifically being fired. In other words, subjects who lost their job had a similar change in performance as subjects who kept their job, irrespective of the cause of unemployment.

Table 17. Effect of Unemployment on Relative Round Three Performance

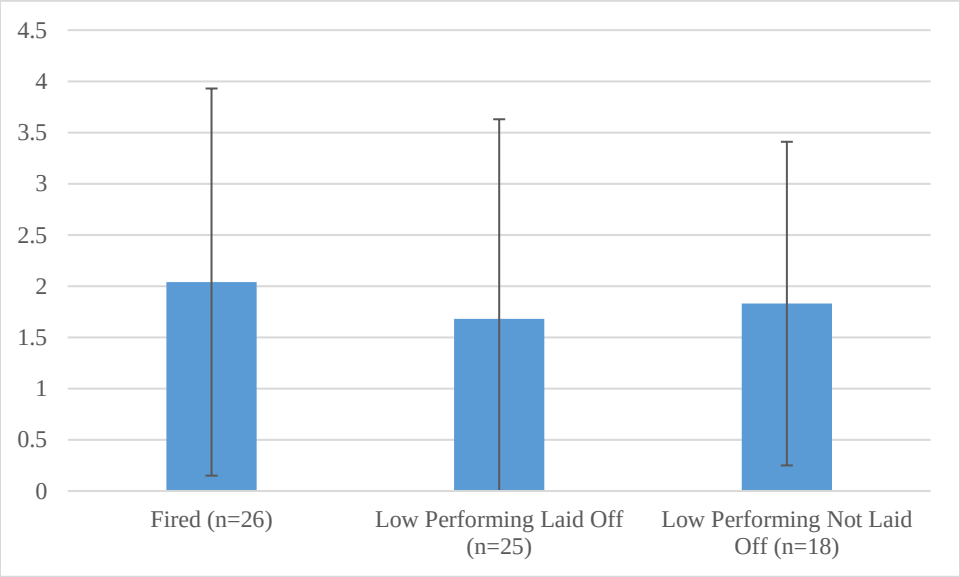
Dependent Variable: Round 3 – Round 1 Performance

	(1)	(2)	(3)	(4)
Lost Job	0.00	-0.14	-0.12	-0.44
	(0.30)	(0.32)	(0.32)	(0.40)
Performance Treatment			0.32	-0.10
			(0.30)	(0.35)
Lost Job * Performance Treatment				0.88
				(0.55)
Constant	1.51***	2.89***	2.38***	2.21
	(0.16)	(2.92)	(2.93)	(2.89)
Controls	N	Y	Y	Y
Observations	156	156	156	156

Notes: Robust standard errors are reported in parentheses. *p<0.1, **p<0.05, ***p<0.01. Controls include age, a dummy for being white, a dummy for being female, GPA, and income. Income is measured on a scale from 1 to 3. Higher numbers indicate higher financial constraint. No controls were significant except for GPA in (4).

Figure 15 displays the average relative Round Three performance for fired, low-performing laid off, and low-performing not laid off subjects. These averages are 2.0, 1.7, and 1.8, respectively. Based on Mann-Whitney tests, there is no statistically significant difference between fired and low performing laid off subjects (p=0.68), fired and low performing not laid off subjects (p=0.85), or low performing laid off and low performing not laid off subjects (p=0.76). This suggests that performance growth for low performing subjects did not depend on whether or not the subject experienced unemployment nor did it depend on the type of unemployment they faced.

Figure 15. Effect of Type of Unemployment on Relative Round Three Performance



4.4.4 Self-Esteem, Locus of Control, and Performance

Measures of subjects' self-esteem and locus of control were collected before subjects came to the laboratory and immediately after they received Employment Message One. Studies have found positive correlations between these two personality traits (e.g., Judge et al., 2002), but the two traits could affect unemployed subjects in different ways. Because self-esteem is a measure of self-worth, I predicted that subjects who were fired would experience a decrease in self-esteem after being told that their performance as not high enough. Since locus of control was designed to test how strongly individuals felt like they had control over what happened to them in their lives, I anticipated that the idea of fate or luck dictating life's events would resonate with subjects who lost their job with a particular effect on those who had been laid off.

On average, subjects who were fired after Round One initially reported 19.1 out of 30 on the Rosenberg Scale where a higher number indicates a higher level of self-esteem. Subjects who were not fired reported 18.3 out of 30, and this difference is not statistically significant ($p=0.58$). When they took the Rosenberg Scale a second time, both types of subjects had higher self-esteem ($p<0.01$ for fired subjects and not fired subjects). Fired subjects reported a self-esteem of 21.6, and not fired subjects reported 21.4. Again, there is no significant difference between the two types in the second measure of self-esteem ($p=0.86$), and the change in self-esteem was similar for both ($p=0.54$). Thus, being fired did not seem to influence subjects' self-esteem.

Subjects who were laid off initially reported 20.3 out of 30 on the Rosenberg Scale while subjects who were not laid off reported 20.2 out of 30. This difference was not statistically significant ($p=0.57$). Similar to the PBUT subjects, laid off and not laid off subjects had a higher level of self-esteem when they completed the Rosenberg Scale again ($p<0.01$ for both). Laid off subjects reported 23.8 and not laid off subjects reported 22.9, which was not significantly different ($p=0.38$). However, laid off subjects had a marginally larger change in self-esteem than not laid off subject ($p=0.08$). This would suggest that being laid off led to a slight increase in subjects' self-esteem.

Subjects also answered questions to measure their locus of control at two different times. Initially, fired subjects reported 2.5 out of 4 on Abbreviated Rotter, where a higher number indicates a higher sense of internal locus of control. Not fired subjects had a significantly higher score of 3.1 out of 4 ($p=0.04$). After Round One, fired subjects reported 2.5, and not fired subjects reported 3.1. This difference is statistically significant ($p=0.07$), but both types experienced a similar change in locus of control ($p=0.97$). Therefore, the PBUT did not appear to

have an effect on subjects' locus of control; however, those who kept their job after Round One tended to be individuals who held a stronger belief that they had control over their own lives. When subjects in the RUT first completed the Abbreviated Rotter, laid off subjects reported 2.9 out of 4, and not laid off subject reported 2.8 out of 4. This difference is not statistically significant (p=0.64). Laid off subjects reported 3.2, and not laid off subjects reported a similar number of 2.9 (p=0.22). The change in locus of control for those who were laid off was similar to the change experienced by those who weren't laid off (p=0.26). As such, there is no evidence to confirm the prediction that laid off subjects would adopt an external locus of control after experiencing job loss.

4.5 Conclusion

The main purpose of this paper was to study how past unemployment affected individuals' performance when they became reemployed. Additionally, the paper sought to determine whether job loss was connected to changes in two psychological traits, self-esteem and locus of control. To study these questions, I designed a laboratory experiment in which subjects participated in four rounds of work. After the first round, subjects in the PBUT who answered fewer problems than the group's median were told that they were fired and were not allowed to work for pay in the next round. Subjects in the RUT were told that they were not allowed to work for pay in the second round if they had bad luck. However, every subject was allowed to work for pay again in the third round. Comparing the change in performance between the first round and the third round allowed me to see if subjects' productivity changed after they experienced a period of unemployment.

The results indicated that losing one's job does not harm subjects' performance. Subjects who were fired solved fewer problems in Round Three than subjects who were not fired. But

there was a performance gap between these two groups of subjects starting in Round One. Fired subjects performed just as well in Round Three as low-performing subjects who randomly kept their job and low-performing subjects who randomly lost their job. This implies that being fired did not discourage or demotivate subjects, and that low-performing subjects responded to being fired no differently than to being laid off.

Unemployment did not seem to have many effects on self-esteem or locus of control, either. Self-esteem was higher for subjects when they were in the middle of the experiment compared to before the experiment. Subjects who were laid off experienced a slightly bigger increase in self-esteem than subjects who were not laid off, but subjects who were fired experienced a similar change in self-esteem as those who were not. Locus of control was relatively stable over time, but subjects who were fired always had more of an external locus of control than subjects who were not fired. So, while there might be a correlation between locus of control and employment status in the PBUT, being fired did not seem to cause individuals to have a larger or smaller shift in their locus of control compared to if they had not been fired. But there was evidence of an effect of being laid off on self-esteem.

The findings of this study may be particularly informative for employers who are looking to hire new employees. They may prefer to hire a job candidate who has been laid off over a job candidate who has been fired because they assume that lay off victims are better performers than those who have been fired. In general, this may be a rational assumption to make. However, the distinction between being fired and being laid off in terms of performance is minor if the job candidates are low-performing. So if two job candidates are similar in terms of indicators of their past productivity, then employers may want to reconsider how much weight they place on the cause of unemployment.

4.6 References

Azmat, G., & Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, *94*(7), 435-452.

Baum, A., Fleming, R., & Reddy, D. M. (1986). Unemployment stress: Loss of control, reactance and learned helplessness. *Social Science & Medicine*, *22*(5), 509-516.

Benndorf, V., Rau, H. A., & Sölch, C. (2014). Minimizing learning behavior in experiments with repeated real-effort tasks. *Available at SSRN 2503029*.

Bennett, N., Martin, C. L., Bies, R. J., & Brockner, J. (1995). Coping with a layoff: A longitudinal study of victims. *Journal of Management*, *21*(6), 1025-1040.

Brandes, P., Castro, S. L., James, M. S., Martinez, A. D., Matherly, T. A., Ferris, G. R., & Hochwarter, W. A. (2007). The interactive effects of job insecurity and organizational cynicism on work effort following a layoff. *Journal of Leadership & Organizational Studies*.

Clark, A. E., & Oswald, A. J. (1994). Unhappiness and unemployment. *The Economic Journal*, *104*(424), 648-659.

Clark, A., Knabe, A., & Rätzel, S. (2010). Boon or bane? Others' unemployment, well-being and job insecurity. *Labour Economics*, *17*(1), 52-61.

Darity, W., & Goldsmith, A. H. (1996). Social psychology, unemployment and macroeconomics. *The Journal of Economic Perspectives*, *10*(1), 121-140.

Eliason, M., & Storrie, D. (2009). Does Job Loss Shorten Life? *The Journal of Human Resources*, (2). 277-302.

Elliot, A. J., & Harackiewicz, J. M. (1994). Goal setting, achievement orientation, and intrinsic motivation: A mediational analysis. *Journal of personality and social psychology*, *66*(5), 968-980.

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

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114-125.

Goldsmith, A. H., Veum, J. R., & Darity, W. (1997). Unemployment, joblessness, psychological well-being and self-esteem: Theory and evidence. *The Journal of Socio-Economics*, *26*(2), 133-158.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct?. *Journal of Personality and Social Psychology*, *83*(3), 693-710.

Kanfer, R., & Hulin, C. L. (1985). Individual differences in successful job searches following lay-off. *Personnel psychology*, *38*(4), 835-847.

Kanfer, R., Wanberg, C. R., & Kantrowitz, T. M. (2001). Job search and employment: A personality–motivational analysis and meta-analytic review. *Journal of Applied psychology*, *86*(5), 837.

Oreopoulos, P., Page, M., & Stevens, A. H. (2008). The Intergenerational Effects of Worker Displacement. *Journal Of Labor Economics*, *26*(3), 455-483.

Probst, Tahira M., et al. "Productivity, counterproductivity and creativity: The ups and downs of job insecurity." *Journal of Occupational and Organizational Psychology* 80.3 (2007): 479-497.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton university press.

Stephens, M. (2004). Job Loss Expectations, Realizations, and Household Consumption Behavior. *The Review of Economics and Statistics*, (1). 253-269.

CHAPTER V

CONCLUSIONS

In this dissertation, I used experiments to study the preferences and decisions of individuals who were faced with challenges like the threat of being the target of discrimination or confirming stereotypes. The purpose was to try to better understand how people respond to these difficulties and how these difficulties may affect their performance and success. As such, I investigated men and women applying for a stereotypically male job, black college students taking an intelligence test, and individuals returning from a period of unemployment.

In my first study, I focused on gender discrimination in the labor market. There is an under-representation of women in careers like engineers, but this could be the result of women applying to these jobs less than men or women being hired less than men. This study investigated whether anticipated discrimination can explain why men dominate certain careers. Individuals from an online job market platform were recruited to act as job candidates who created resume-like profiles and applied for a stereotypically male job. Before their profiles were sent to hiring managers (other individuals from the platform), the job candidates had the opportunity to pay to modify their profiles to either include or exclude gender. Managers then decided who they wanted to hire. I found that men were willing to pay more to include gender on the profiles rather than exclude it. Women, on the other hand, were willing to pay more to exclude gender. This suggests that both men and women believed that women were likely to be discriminated against because of their gender. However, I find men and women earned approximately the same amount regardless of whether or not their gender was known. Thus, people tended to overestimate how much discrimination women faced in this experiment, men especially. So the occupational

segregation that exists in fields like computer programming and engineering could be partially explained by perceptions of discrimination that cause men to apply for stereotypically male jobs at a higher rate than equally talented women.

Stereotypes are also the main topic of my second study where my co-authors and I conducted an experiment to determine whether negative stereotypes impair the performance of black students on the Graduate Record Examination (GRE). Some researchers have tried to explain part of the black-white achievement gap through a phenomenon called stereotype threat, the fear of confirming a negative stereotype about one's social group. They almost always used black students from predominantly white institutions (PWIs) as subjects in their experiments and confirmed the harmful effects of stereotypes on test performance. Our goal was to determine if black students at historically black colleges and universities (HBCUs) are more resilient against stereotypes. We recruited black students from an HBCU and sorted them into two groups: treatment and control. Those in the treatment group were subtly reminded of the stereotype that black people are not as intelligent as white people before taking a mock GRE. Those in the control group received no such reminder but answered the same GRE questions. We failed to replicate previous studies' findings that the control group outperforms the treatment group. Instead, we found that the test scores of both groups were similar. This suggests that we may discover a way to reduce the black-white achievement gap if we learn more about the type of students who choose to attend HBCUs and if we study the environment of these campuses.

In my third study, I designed a lab experiment to study whether past unemployment has a positive or negative effect on workers' performance once they become employed again. Subjects performed a real effort task for one round. Afterward, some subjects were told that they were fired because of poor performance and others were told that they had been randomly laid off.

Consequently, these subjects were unemployed and did not receive any payment in the following round. However, every subject was allowed to work for pay again in the third round. I found that subjects who were fired performed just as well as those of similar ability who were laid off or who were randomly allowed to stay employed. These results suggest that, despite all of the negative consequences involved with job loss, being fired may not be entirely demotivating.

APPENDIX A

APPENDIX FOR CHAPTER II

Additional Table

Table 18. Worker Characteristics by Gender

	Males	Females	Two-tailed t-test p-value
Exact Age	34.35 (10.51)	37.47 (11.14)	p<0.01
Percent who volunteered in the past year	44.60%	52.0%	p=0.12
Percent whose supervisors would consider them a good employee	95.77%	98.60%	p=0.08
Percent who read the news often	93.43%	84.65%	p<0.01
Partial Score	3.13 (1.25)	2.76 (1.30)	p<0.01
N	213	215	

Note: Standard deviations are written in parentheses.

Selected Worker Comments

Selected comments from male Workers who predicted that males would be chosen twice when version A profiles were used:

- I think people assume males stereotypically know more about sports. I think the managers will want to earn their bonus for correct answers and assume males are better at answering sports questions.
- I think the manager would feel the male would know more about sports.
- Think the manager might be biased

Selected comments from female Workers who predicted that males would be chosen twice when version A profiles were used:

- With money being on the line if they choose the person who got more right, I think that managers will go with the male workers based on the stereotype of men knowing more about sports.
- I think people will trust males more to answer sports questions correctly.
- men are stereotyped as more knowledgeable about sports

Selected comments from male Workers who predicted that males would be chosen once when version A profiles were used:

- because both will have a similar score

- I don't think gender is an issue here.
- to be as fair as possible

Selected comments from female Workers who predicted that males would be chosen once when version A profiles were used:

- I would hope it's kept fair and based just on scores.
- Hope that sexism isn't as bad as i think it is.
- Some men have faith in women's ability to know a little bit about sports.

Selected comments from male Workers who predicted that males would be chosen zero times when version A profiles were used:

- I think people are more sensitive to not being sexist now
- I think people tend to be more sympathetic towards women.
- I would pick females if I were them.

Selected comments from female Workers who predicted that males would be chosen zero when version A profiles were used:

- The questions are all sports related so I believe it is more fair to reward the female.
- this is just a random assumption that they picked 2 girls for this.
- Women are better

Selected comments from male Workers who predicted that males would be chosen twice when version B profiles were used:

- Since males are more likely to get a higher score, they would be picked due to their high score
- Because they will have a higher score.
- the males partial score is more likely to be higher than the females so the manager would probably pick that.

Selected comments from female Workers who predicted that males would be chosen twice when version B profiles were used:

- Because I think it's more likely for the males to have a slightly higher partial score.
- Their scores will be higher so they will choose them.
- I think the males probably did better on the questions so they have a higher partial score possibly

Selected comments from male Workers who predicted that males would be chosen once when version B profiles were used:

- It will be up to chance which gender is chosen.
- I think it's random chance if the manager does not have access to the gender info.
- Random since gender is removed

Selected comments from female Workers who predicted that males would be chosen once when version B profiles were used:

- Without the gender bias, there is a 50/50 chance on who they will choose. It all comes down to which worker had the most correct answers, and I'm willing to bet that my fellow female workers didn't do too shabby.
- It would probably be a 50-50 chance.
- I think with genders removed it will more likely end up a split because the managers won't see the genders.

Selected comments from male Workers who predicted that males would be chosen zero times when version B profiles were used:

- female get a more sympathetic ear and males are supposed to just accept what they are given.
- Females tend to volunteer to causes more than males. The managers will look at volunteering as more compassionate.
- The manager will probably choose those who donate to charity which I believe to be more likely women.

Selected comments from female Workers who predicted that males would be chosen zero when version B profiles were used:

- Just a toss up -- I feel lucky and that 2 females will win.
- Better scores will be picked.
- I trust in there judgments

Part One Instructions

Please answer the questions below. Your responses will be used to create a Profile.

You will be <u>given a bonus payment of $0.50 for completing Part 1</u>.

What is your gender?

a. Male
b. Female
c. Other (Please explain.) _____

What is your age?

a. 18 – 25 years old
b. 26 – 34 years old
c. 35 years old or older

Have you volunteered in the past 12 months?

a. Yes
b. No

Would your former supervisor say that you were a good employee?

a. Yes

b. No
c. N/A (Never had a supervisor.)

Do you read the news (online or from a newspaper) often?
a. Yes
b. No

Comprehension Quiz #1

To make sure you understand what happens in Study #2, please answer the following questions.

1. True or False: The Manager is an adult MTurk worker located in the United States.
a. True ***Correct Answer
b. False

2. True or False: The Manager will see your Profile and the Profile of a [male/female] MTurk worker who participated in this study.
a. True ***Correct Answer
b. False

3. True or False: If the Manager decides that Worker 2 will be paid $0.40 for each correct answer from Part 2, then you will be paid $0.10 for each correct answer.
a. True ***Correct Answer
b. False

4. True or False: The Manager gets a bonus payment based on how many questions the Worker they chose answered correctly.
a. True ***Correct Answer
b. False

5. True or False: The Manager will know your Total Score and your Partial Score.
a. True
b. False ***Correct Answer

Comprehension Quiz #2

Please answer the questions below.

Remember, Version A Profiles include:
- Gender
- Age
- Volunteered in past year?
- Considered good employee?
- Reads the news often?

- Partial Score (out of 5)

And Version B Profiles include:

- Age
- Volunteered in past year?
- Considered good employee?
- Reads the news often?
- Partial Score (out of 5)

1. True or False: The Manager knows that there are 2 versions of the Profiles (Version A and Version B).
   a. True
   b. False ***Correct Answer
2. True or False: The Manager will know how much money you were willing to pay to avoid using the Version A Profiles (gender included).
   a. True
   b. False ***Correct Answer
3. True or False: It's possible for the Manager to see your Version A Profile (gender included) and Worker 2's Version B Profile (gender removed).
   a. True
   b. False ***Correct Answer
4. True or False: If the Manager sees the Version B Profiles (gender removed), then they will know with certainty that your Profile is the one that belongs to a female.
   a. True
   b. False ***Correct Answer

Post-Experimental Worker Survey (Part Four)

Part Four Instructions

Please follow the instructions for the questions below.

Please answer the 5 questions below to the best of your ability. <u>You will earn $0.10 for each correct guess</u>.

1. How many questions do you think Worker 2 answered correctly?
Remember: You answered [NUMBER] questions correctly, and you will be partnered with a [male/female] from this study who had a similar Partial Score as you.

2. We will randomly pick 2 pairs of Workers whose <u>Version A Profiles (gender included) will be shown to the Managers</u>. This will give us 2 Managers, 2 male Workers, and 2 female Workers. Each Manager will have to choose between the male and female Worker and will know their gender.

Which Workers do you think will be chosen to be paid $0.40? [***order randomized***]

a. 2 male Workers
b. 2 female Workers
c. 1 male Worker and 1 female Worker

Why did you choose this option?

3. We will randomly pick 2 pairs of Workers whose Version B Profiles (gender removed) will be shown to the Managers. This will give us 2 Managers, 2 male Workers, and 2 female Workers. Each Manager will have to choose between the male and female Worker but won't know which Worker is male and which one is female.

Which Workers do you think will be chosen to be paid $0.40? [***order randomized***]

d. 2 male Workers
e. 2 female Workers
f. 1 male Worker and 1 female Worker

Why did you choose this option?

4. How many questions do you think male Workers in this study answered correctly on average?

5. How many questions do you think female Workers in this study answered correctly on average?

Please tell us more about yourself.

1. What is your age?

2. How do you identify? (Check all that apply.)

- ❑ White
- ❑ Hispanic, Latino, or Spanish origin
- ❑ Black or African American
- ❑ Asian (ex. Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, etc.)
- ❑ American Indian or Alaska Native
- ❑ Middle Eastern or North African (ex. Lebanese, Iranian, Egyptian, Syrian, Moroccan, Algerian, etc.)
- ❑ Native Hawaiian or Other Pacific Islander
- ❑ Some other race, ethnicity, or origin (Please explain.) _____

3. What is the highest level of school you have completed or the highest degree you have received?

a. Some high school
b. High school graduate or other equivalent (high school diploma, GED, etc.)
c. Some college but no degree
d. Bachelor's degree (ex. BA, AB, BS)
e. Master's degree (ex. MA, MS, MBA)
f. Associate's degree
g. Professional school degree (ex. MD or JD)

h. Doctorate degree (ex. PhD)
i. Other (Please explain.) _____

4. Are you currently an undergraduate student?
a. Yes
b. No

5. What is your current employment status (excluding your MTurk activity)?
a. Unemployed
b. Employed part-time
c. Employed full time
d. Other (Please explain.) _____

6. What is your monthly salary? (If you can't remember or don't know the exact amount, please give your best guess.)
[***dropdown list***]

7. What is your current marital status?[21]
a. Now married
b. Widowed
c. Divorced
d. Separated
e. Never married

How strongly do you agree with the statements below?

1. On average, women are more educated than men.[22]
a. Strongly Disagree
b. Disagree
c. Neither Disagree or Agree
d. Agree
e. Strongly Agree

2. On average, men face more discrimination than women when applying for jobs.
a. Strongly Disagree
b. Disagree
c. Neither Disagree or Agree
d. Agree
e. Strongly Agree

3. On average, women are more likely to be paid less than men for doing the same job.

[21] The wording for this question came from the U.S. Census.

[22] Questions 1 through 4 are based on questions used in Pew Research Center surveys.

a. Strongly Disagree
b. Disagree
c. Neither Disagree or Agree
d. Agree
e. Strongly Agree

4. Discrimination in the workplace has decreased over the past 10 years.
a. Strongly Disagree
b. Disagree
c. Neither Disagree or Agree
d. Agree
e. Strongly Agree

Lastly, please answer these two questions.

1. When you first joined Mechanical Turk, which statement would you have agreed with the most?
a. HITs on Mechanical Turk do not favor one gender over the other.
b. Male workers have an advantage on Mechanical Turk.
c. Female workers have an advantage on Mechanical Turk.

2. Prior to completing this particular HIT, which statement would you have agreed with the most?
a. HITs on Mechanical Turk do not favor one gender over the other.
b. Male workers have an advantage on Mechanical Turk.
c. Female workers have an advantage on Mechanical Turk.

Post-Experimental Manager Survey (Part Three)

Part 3 Instructions

Please follow the instructions for the questions below.

Please answer the 3 questions below to the best of your ability. You will earn $0.10 for each correct guess/answer.

1. How many questions do you think Worker 1 answered correctly (out of 10)?
Why did you guess this number?

2. How many questions do you think Worker 2 answered correctly (out of 10)?
Why did you guess this number?

3. Was Worker 1 female or male? (If you don't remember or don't know, take your best guess.)
a. Female
b. Male

Note: The rest of the questions that Managers answer are nearly identical to those from the Post-Experimental Worker Survey, starting with “What is your age?”

APPENDIX B

APPENDIX FOR CHAPTER III

Script

Thank you for participating in today's experiment!
Your seat is labeled with a number that corresponds with the number card you were given before. Please find your seat.

WAIT FOR THEM TO GET SEAT

Thank you for participating in today's experiment!
Before we get started, please turn off your phones and put all of your belongings on the ground beside you. [In a moment, we will pass out booklets./As you may have noticed, there is a booklet on your desk.] It is really important for the experiment that you remain quiet and keep your eyes on your booklet for the duration of the experiment.

Your responses will have no effect on your course grade; however, you will have the opportunity to earn money during this experiment, which I will explain in a moment.

To start, please open your booklet.

WAIT FOR THEM TO OPEN THEIR BOOKLET

There is a blank line on the first page. Please write your subject ID number there. This is the same number that you were given earlier and that is on your desk.

WAIT FOR THEM TO WRITE THEIR NUMBERS

As with many assessments, your packet may differ from the packet that others receive. There is no need to be concerned about this. Simply follow the instructions written in your packet.

You may now turn the page.

WAIT FOR THEM TO TURN THE PAGE TO PART ONE: QUESTIONNAIRE

Okay, you will now be asked to fill out this questionnaire. Once you have finished the questionnaire, please wait until further instruction.

WAIT FOR THEM TO FILL OUT PART ONE: QUESTIONNAIRE

You may now turn the page to read the test instructions. We will give you three minutes to do so. Please read all of the instructions carefully, and then I will let you know when it is time to turn the page.

WAIT THREE MINUTES FOR THEM TO READ PART TWO: INSTRUCTIONS

You may now turn the page. You will have twenty-five minutes to answer the questions. We will keep track of the time remaining on the board.

START TIME, WRITE TIME REMAINING "25 MINUTES" ON BOARD, AND UPDATE TIME EVERY 5 MINUTES

Okay, time is up for the test. Please turn to Part Four. Once you have finished Part Four, please continue on to Part Five. You may begin to answer the questions now.

WAIT FOR THEM TO FILL OUT PART FOUR: QUESTIONNAIRE AND PART FIVE: OPINION SURVEY

All right. Thank you again for participating in today's experiment. We will call you up one by one to receive payment by your subject ID number. Please wait patiently at your desk until your number has been called. When we call your number, please bring your test booklet and subject ID card with you.

If people are still working on it and class time is ending soon:
Alright everyone! It is [time]. If you have finished answering the questions, feel free to come up to the front to turn in your booklet. Please bring your subject ID card with you, too. Otherwise, if you're still working on the questions, feel free to take your time.

If anyone asks about their score:
Hang around after we pay everyone, and we can discuss this more.
Hand graded test but don't let them keep it

Control Part One

1. What is your age? _____
2. What year are you in school?
    a. Freshman
    b. Sophomore
    c. Junior
    d. Senior
    e. Graduate student
3. What is your major? _____
4. How many siblings do you have? _____

5. What is your mother's highest level of education?
    a. No high school completed
    b. Some high school, no diploma
    c. High school graduate
    d. Some college, no degree
    e. Vocational degree
    f. Associate's degree
    g. Bachelor's degree
    h. Master's degree
    i. Doctoral degree
    j. I don't know.
6. What is your father's highest level of education?
    a. No high school completed
    b. Some high school, no diploma
    c. High school graduate
    d. Some college, no degree
    e. Vocational degree
    f. Associate's degree
    g. Bachelor's degree
    h. Master's degree
    i. Doctoral degree
    j. I don't know.

Control Part Two

The problems you are about to solve are taken from the verbal portion of the GRE (Graduate Record Examination). You will be given twenty-five minutes to answer eighteen questions. You will receive $12 in compensation for submitting your answers. Completing this test will allow you to familiarize yourself with the kinds of problems that appear on tests you may encounter in the future.

Please try hard to correctly solve as many items as you can to help us in our analysis of the problem solving process.

Treatment Part One

1. What is your age? _____
2. What year are you in school?
    a. Freshman
    b. Sophomore
    c. Junior
    d. Senior
    e. Graduate student
3. What is your major? _____
4. How many siblings do you have? _____

5. What is your mother's highest level of education?
    a. No high school completed
    b. Some high school, no diploma
    c. High school graduate
    d. Some college, no degree
    e. Vocational degree
    f. Associate's degree
    g. Bachelor's degree
    h. Master's degree
    i. Doctoral degree
    j. I don't know.
6. What is your father's highest level of education?
    a. No high school completed
    b. Some high school, no diploma
    c. High school graduate
    d. Some college, no degree
    e. Vocational degree
    f. Associate's degree
    g. Bachelor's degree
    h. Master's degree
    i. Doctoral degree
    j. I don't know.
7. Do you identify as Hispanic/Latino?
    a. Yes
    b. No
8. How do you identify? **Circle all that apply.**
    a. White/Caucasian
    b. Black/African American
    c. East Asian
    d. Middle Eastern
    e. South Asian
    f. Native American
    g. Pacific Islander
    h. Other (Please specify.) _____

Treatment Part Two

The test you are about to take, the verbal portion of the GRE (Graduate Record Examination), is in large part a measure of your verbal intelligence and verbal reasoning ability. You will be given twenty-five minutes to answer eighteen questions. You will receive $12 in compensation for submitting your answers. Completing this test will allow you to familiarize yourself with some of your strengths and weaknesses.

It is absolutely vital that you try to do your best on this test. Please try hard to correctly solve as many items as you can to help us in our analysis of your verbal ability.

APPENDIX C

APPENDIX FOR CHAPTER IV

Rosenberg Self-Esteem Scale (Rosenberg, 1965)

For each statement below, please select whether you strongly agree, agree, disagree, or strongly disagree.

1. I feel that I'm a person of worth, at least on an equal plane with others.
2. I feel that I have a number of good qualities.
3. All in all, I am inclined to feel that I am a failure.
4. I am able to do things as well as most other people.
5. I feel I do not have much of which to be proud.
6. I take a positive attitude toward myself.
7. On the whole, I am satisfied with myself.
8. I wish I could have more respect for myself.
9. I certainly feel useless at times.
10. At times I think I am no good at all.
11.

Abbreviated Rotter (Goldsmith, Veum, and Darity, 1997)

Which statement do you agree with more?

a. What happens to me is my own doing.
b. Sometimes I feel that I don't have enough control over the direction my life is taking.

Which statement do you agree with more?

a. When I make plans, I am almost certain that I can make them work.
b. It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune anyhow.

Which statement do you agree with more?

a. In my case, getting what I want has little or nothing to do with luck.
b. Many times we might just as well decide what to do by flipping a coin.

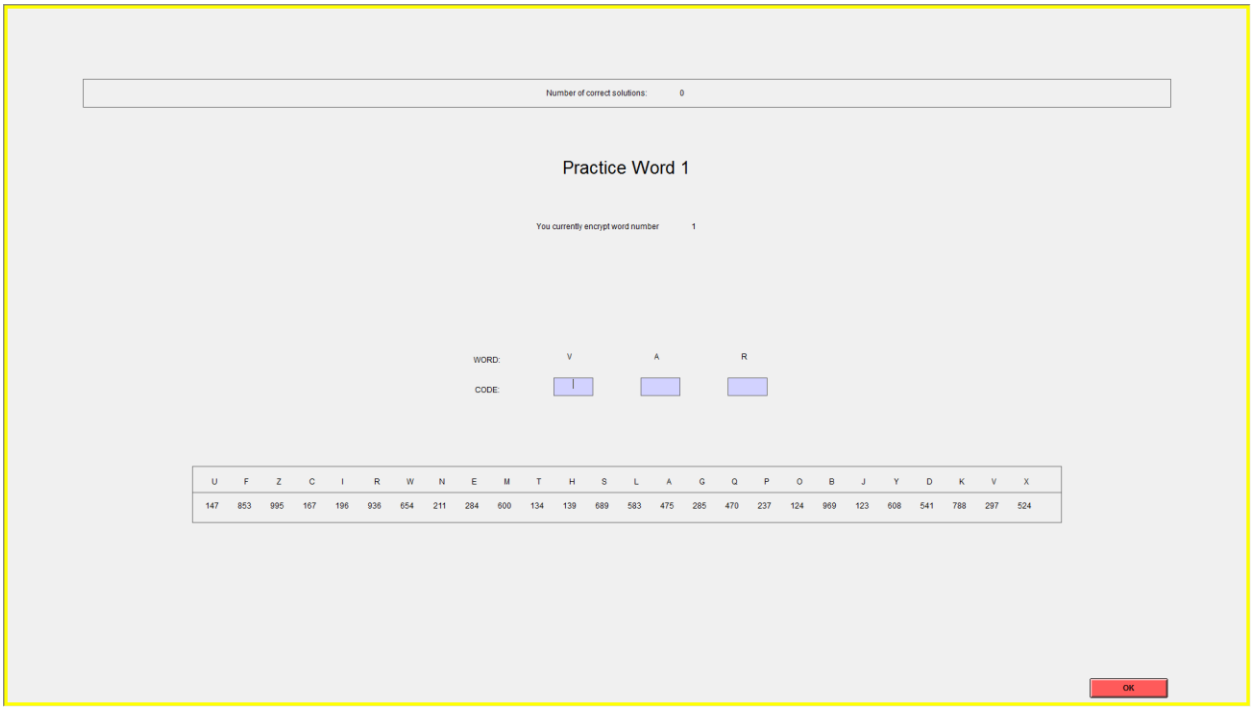
Which statement do you agree with more?

a. It is impossible for me to believe that chance or luck plays an important role in my life.

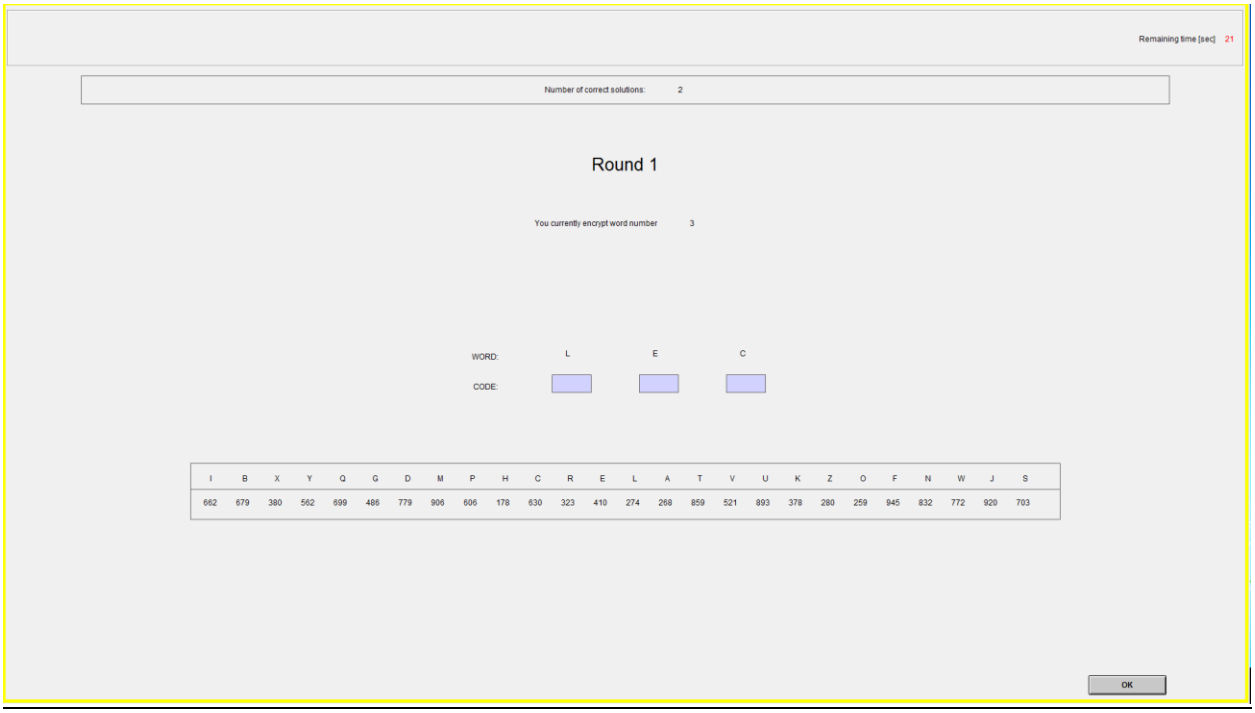
b. Many times I feel that I have little influence over the things that happen to me.

c.

Screenshot of Practice Problem



Screenshot of Round One Problem



Employment Message One: PBUT – Fired

You have completed **Round 1**. You correctly answered X problems and earned 300 tokens.

The scores in **Round 1** for every participant were recorded, and each participant was ranked in terms of how many problems they correctly solved. For example, the participant who solved the most problems was ranked #1. In terms of ranking, you fell in the bottom half. Therefore, you will not be allowed to participate in **Round 2** for 300 tokens. Instead, you will get the opportunity to practice solving encryption problems for 0 tokens.

Before the practice session begins, you will answer a few questions.

Employment Message One: PBUT – Not Fired

You have completed **Round 1**. You correctly answered X problems and earned 300 tokens.

The scores in **Round 1** for every participant were recorded, and each participant was ranked in terms of how many problems they correctly solved. For example, the participant who solved the most problems was ranked #1. In terms of ranking, you fell in the top half. Therefore, you will be allowed to participate in **Round 2** for 300 tokens.

Before the encryption task begins, you will answer a few questions.

Employment Message One: RUT – Laid Off

You have completed **Round 1**. You correctly answered X problems and earned 300 tokens.

For every participant, the computer has randomly determined whether he/she will be allowed to participate in **Round 2** for payment. Each participant had a 50% chance of being allowed to participate for payment. In your case, it has been randomly determined that you will not be allowed to participate in **Round 2** for 300 tokens. Instead, you will get the opportunity to practice solving encryption problems for <u>0 tokens</u>.

Before the practice session begins, you will answer a few questions.

Employment Message One: RUT – Not Laid Off

You have completed **Round 1**. You correctly answered X problems and earned 300 tokens.

For every participant, the computer has randomly determined whether he/she will be allowed to participate in **Round 2** for payment. Each participant had a 50% chance of being allowed to participate for payment. In your case, it has been randomly determined that you will be allowed to participate in **Round 2** for 300 tokens.

Before the encryption task begins, you will answer a few questions.

Employment Message Two: PBUT – Fired

You have completed **Round 3**. You correctly answered X problems and earned 300 tokens.

The scores in **Round 3** for every participant were recorded, and each participant was ranked in terms of how many problems they correctly solved. For example, the participant who solved the most problems was ranked #1. In terms of ranking, you fell in the bottom half. Therefore, you will not be allowed to participate in **Round 4** for 300 tokens. Instead, you will get the opportunity to practice solving encryption problems for 0 tokens.

Before the practice session begins, you will answer a few questions.

Employment Message Two: PBUT – Not Fired

You have completed **Round 3**. You correctly answered X problems and earned 300 tokens. The scores in **Round 3** for every participant were recorded, and each participant was ranked in terms of how many problems they correctly solved. For example, the participant who solved the most problems was ranked #1. In terms of ranking, you fell in the top half. Therefore, you will be allowed to participate in **Round 4** for 300 tokens.

Before the encryption task begins, you will answer a few questions.

Employment Message Two: RUT – Laid Off

You have completed **Round 3**. You correctly answered X problems and earned 300 tokens.

For every participant, the computer has randomly determined whether he/she will be allowed to participate in **Round 3** for payment. Each participant had a 50% chance of being allowed to participate for payment. In your case, it has been randomly determined that you will not be allowed to participate in **Round 4** for 300 tokens. Instead, you will get the opportunity to practice solving encryption problems for <u>0 tokens</u>.

Before the practice session begins, you will answer a few questions.

Employment Message Two: RUT – Not Laid Off

You have completed **Round 3**. You correctly answered X problems and earned 300 tokens.

For every participant, the computer has randomly determined whether he/she will be allowed to participate in **Round 3** for payment. Each participant had a 50% chance of being allowed to

participate for payment. In your case, it has been randomly determined that you will be allowed to participate in **Round 4** for 300 tokens.

Before the encryption task begins, you will answer a few questions.

Post-experiment Questionnaire

1. What is your cumulative GPA? If you do not remember the exact number, please give your best estimate.
2. How difficult is it for you to live on your income right now?
    a. Not at all difficult
    b. Somewhat difficult
    c. Very difficult
3. Do you receive any form of financial aid?
    a. Yes
    b. No
4. Do you have a job nor or have you held a job in the past 6 months?
    a. Yes
    b. No
5. If you answered "Yes" to the previous question, how many hours a week do/did you work on average?
6. How would you rate the encryption task?
    a. Did not like it
    b. It was ok
    c. Liked it
    d. Really liked it
    e. Loved it
7. While you were working on Round 2, were you expecting a fourth round?
    a. Yes
    b. No
8. As you were participating in Round 3, what were you thinking about?

9. As you were parting in Round 3, how concerned were you about your chances of being allowed to participate in the encryption task (for tokens) again?
    a. Not concerned at all
    b. Somewhat concerned
    c. Very concerned
10. How fair did you find this experiment?
    a. Extremely unfair
    b. Unfair
    c. Somewhat unfair
    d. Fair
    e. Extremely unfair
11. Please describe your experience with this experiment.
12. What do you think this experiment was about?
13. Feel free to leave any additional comments about the experiment below.