HARVARD | BUSINESS | SCHOOL

When to Apply?

Katherine B. Coffman
Manuela Collis
Leena Kulkarni

Working Paper 20-062

When to Apply?

Katherine B. Coffman
Harvard Business School

Manuela Collis
Harvard Business School

Leena Kulkarni
Harvard School of Public Health

Working Paper 20-062

This work was funded by the National Science Foundation, grant #1713752.

WHEN TO APPLY?

Katherine B. Coffman*
Manuela Collis
Leena Kulkarni

August 19, 2019

Abstract: Labor market outcomes depend, in part, upon an individual's willingness to put herself forward for different career opportunities. In this paper, we use both laboratory and field experiments to better understand decisions around willingness to apply for higher return, more challenging work, with a focus on gender differences. We find that, in male-typed domains, women are less likely to view themselves as well-qualified for a given opening, both because of differences in forecasts about own ability, but also, and importantly, because of differences in how high men and women believe "the bar" is. We show that these beliefs matter for application decisions. Finally, we provide evidence that a simple intervention, in which required qualifications are stated more precisely and objectively, helps to close the gender gap in willingness to apply, increasing both the average talent and the diversity of the applicant pool.

*Corresponding author: kcoffman@hbs.edu ; 445 Baker Library, Harvard Business School, Boston, MA.
Acknowledgements: We thank the NSF for their generous support of this work.

"Why are you not a full professor - given your eminence?"

[Silence]

"I never applied." - *Donna Strickland, Nobel Laureate in Physics, 2018*

I. Introduction

An important body of work documents the impact of gender bias and discrimination on women's careers (see Riach and Rich 2002 for an overview). Women are less likely to be interviewed for high-status jobs (Fernandez and Mors 2008) and promotions (Ginther and Kahn 2009; Ibarra, Carter, and Silva 2010; Zahidi and Ibarra 2010). Evidence from the laboratory reinforces these findings, with many studies showing that employers in simulated labor markets are more likely to hire men than women for male-typed jobs (Bohnet, van Geen, and Bazerman 2016; Reuben, Sapienza, and Zingales 2014; Coffman, Exley, and Niederle 2018). Once a female worker is hired, she is subject to bias in both formal job evaluation processes (Heilman 2001) and in more informal mentoring (Ibarra, Carter, and Silva (2010)). Firms are devoting significant attention to reducing these biases, with the hope of achieving greater gender diversity throughout the throughout the pipeline.[1]

Of course, when considering the sources of gender gaps in labor market outcomes, discrimination and bias are only one side of the coin. Decisions made by employees themselves also have the potential to have large impacts on gender gaps in outcomes. Candidates decide what types of education to pursue, which jobs to apply to, and when to put themselves up for promotion. Gender differences at these crucial decision nodes could play an important factor. Indeed, social scientists have documented that occupational segregation plays an important role in explaining gender gaps in wages (Altonji and Blank 1999). Choosing an industry, however, is just one of many important choices an employee makes.

In this paper, we study the decisions of candidates about whether and when to apply for different opportunities. We aim to tackle the question of whether there are gender differences in how job-seekers perceive their own qualifications for different job opportunities, and how this impacts their decision about

[1]Salesforce is a prime example. They have spent $6M in an attempt to close the gender wage gap that exists at their firm, auditing their wage data on a regular basis and taking steps to eliminate bias from many stages of their hiring and promotion practices ("2019 Salesforce Equal Pay Update" 2019).

whether or not to apply. These decisions are likely to be key not only at the hiring stage, but also as careers advance, presenting opportunities for promotion.

Outside of controlled experiments, many of these decisions about whether and when to apply may be made in the face of (anticipated) bias, making it hard to isolate the role for self-stereotyping from external bias or fear of bias. We take advantage of controlled frameworks to separate between these stories, allowing us to focus on the role of employee perceptions and decisions, absent employer biases. We ask whether women are as likely as men to see themselves as qualified for challenging, high-paying positions, and whether they apply at similar rates conditional on their degree of qualification.

The past literature on gender differences provides some potential reasons why women may be less likely to apply. Careful laboratory evidence suggests that conditional on having the same ability, women have more pessimistic beliefs about their own ability in male-typed domains compared to men (Niederle and Vesterlund 2007, Coffman 2014; Bordalo et al. 2018, Exley and Kessler 2019). This suggests that if men and women share the same view as to what is required to qualify for a given position, women may be less likely to believe they possess that qualification (holding all else equal). Even conditional on holding the same beliefs, differences in competitive preferences could also drive differences in behavior (Niederle and Vesterlund 2007). It may be that women are less willing to compete for a job opening by applying.

A few clever field experiments have explored some factors that impact job-seekers probability of application, sometimes with differential effects by gender. Consistent with the factors mentioned above, Flory, Leibbrandt, and List (2015) find that an opening that is framed as more "male-typed", more competitive, or with more pay uncertainty deters female applicants more than male applicants. Similarly, Samek (2015) finds that changing from a competitive framing to a non-profit framing can attract more female applicants. Gee (2018) finds that showing job-seekers the number of other applicants increases applications from women more than men, arguably through a reduction in ambiguity. Together, this evidence suggests that decreasing perceived competitiveness and uncertainty about the position may help to reduce gender gaps in application behavior. Female role models can also influence application decisions: Del Carpio and Guadalupe (2018) find that women are more likely to opt into a tech skills training program when presented with an example of a female success story. Our paper will build on this important work by attempting to better understand the decision of whether and when to apply through a series of controlled experiments, including laboratory experiments that can speak to underlying mechanisms.

Our paper tackles the application decision in a few key steps. First, we explore the role for beliefs. We ask how individuals view their own qualification levels, given their talent. We do this across two distinct settings. In the first, we expose individuals to real job advertisements and ask them questions about how

qualified they feel they are for different types of positions. While this gives us data with rich external validity, we cannot incentivize their responses. So, we complement this approach with a laboratory study where we can directly measure the talent of each individual in a simulated labor market environment, and ask them incentivized questions about their believed abilities.

We find significant gender differences for two different types of beliefs. Not only do we see gender differences in individuals' beliefs about their own talent level, but we also see important evidence for gender differences in what individuals believe "the bar" is for a given position. That is, even conditional on having the same belief of their own talent, men and women seem to vary in how qualified they feel for a given opening. Women simply believe the bar is higher. Both of these types of beliefs are predictive of application decisions.

Finally, we turn to policy solutions. The fact that we observe gender differences in beliefs about what it takes to be successful in applying leads us to test a simple intervention: providing more information about "the bar". Given that in many contexts there is uncertainty about exactly what types of qualifications an employer is looking for, heterogeneity in beliefs about what it might take to be hired are perhaps not surprising. We test a simple policy intervention in which we provide more concrete, objective measures of what is needed in order to receive the job, and we ask whether this helps to reduce the gender gap in application rates.

We conducted a series of three experiments to address these questions, including a field experiment with real job applicants. In the first laboratory study, we provide evidence on how female and male students perceive their and their peers' qualifications for real job openings. In the second, we create a controlled laboratory setting online that simulates a labor market, where we exogenously vary whether and how precisely the qualifications for a promotion opportunity are presented to participants. A similar exercise is then repeated as a field experiment on the online labor market platform Upwork with real job applicants.

Across our three studies, we draw three important conclusions. First, on average, women believe themselves to be less qualified for a given job opening than men. However, this gender gap is reduced when the qualifications for the job openings are more objective and clear. Second, men and women also differ in their beliefs about where "the bar" is for a given opening. Even holding fixed how they view themselves, women seem to believe it is more difficult for anyone to be qualified for a given opening as compared to men. Third, clearly stated qualifications help to reduce gender gaps in application rates. In our field experiment, we find that very few talented, qualified women apply for an advanced job in a baseline condition. Once we provide more detailed information on expectations of what is needed to receive the

"Why are you not a full professor - given your eminence?"

[Silence]

"I never applied." - *Donna Strickland, Nobel Laureate in Physics, 2018*

I. Introduction

An important body of work documents the impact of gender bias and discrimination on women's careers (see Riach and Rich 2002 for an overview). Women are less likely to be interviewed for high-status jobs (Fernandez and Mors 2008) and promotions (Ginther and Kahn 2009; Ibarra, Carter, and Silva 2010; Zahidi and Ibarra 2010). Evidence from the laboratory reinforces these findings, with many studies showing that employers in simulated labor markets are more likely to hire men than women for male-typed jobs (Bohnet, van Geen, and Bazerman 2016; Reuben, Sapienza, and Zingales 2014; Coffman, Exley, and Niederle 2018). Once a female worker is hired, she is subject to bias in both formal job evaluation processes (Heilman 2001) and in more informal mentoring (Ibarra, Carter, and Silva (2010)). Firms are devoting significant attention to reducing these biases, with the hope of achieving greater gender diversity throughout the throughout the pipeline.[1]

Of course, when considering the sources of gender gaps in labor market outcomes, discrimination and bias are only one side of the coin. Decisions made by employees themselves also have the potential to have large impacts on gender gaps in outcomes. Candidates decide what types of education to pursue, which jobs to apply to, and when to put themselves up for promotion. Gender differences at these crucial decision nodes could play an important factor. Indeed, social scientists have documented that occupational segregation plays an important role in explaining gender gaps in wages (Altonji and Blank 1999). Choosing an industry, however, is just one of many important choices an employee makes.

In this paper, we study the decisions of candidates about whether and when to apply for different opportunities. We aim to tackle the question of whether there are gender differences in how job-seekers perceive their own qualifications for different job opportunities, and how this impacts their decision about

[1]Salesforce is a prime example. They have spent $6M in an attempt to close the gender wage gap that exists at their firm, auditing their wage data on a regular basis and taking steps to eliminate bias from many stages of their hiring and promotion practices ("2019 Salesforce Equal Pay Update" 2019).

whether or not to apply. These decisions are likely to be key not only at the hiring stage, but also as careers advance, presenting opportunities for promotion.

Outside of controlled experiments, many of these decisions about whether and when to apply may be made in the face of (anticipated) bias, making it hard to isolate the role for self-stereotyping from external bias or fear of bias. We take advantage of controlled frameworks to separate between these stories, allowing us to focus on the role of employee perceptions and decisions, absent employer biases. We ask whether women are as likely as men to see themselves as qualified for challenging, high-paying positions, and whether they apply at similar rates conditional on their degree of qualification.

The past literature on gender differences provides some potential reasons why women may be less likely to apply. Careful laboratory evidence suggests that conditional on having the same ability, women have more pessimistic beliefs about their own ability in male-typed domains compared to men (Niederle and Vesterlund 2007, Coffman 2014; Bordalo et al. 2018, Exley and Kessler 2019). This suggests that if men and women share the same view as to what is required to qualify for a given position, women may be less likely to believe they possess that qualification (holding all else equal). Even conditional on holding the same beliefs, differences in competitive preferences could also drive differences in behavior (Niederle and Vesterlund 2007). It may be that women are less willing to compete for a job opening by applying.

A few clever field experiments have explored some factors that impact job-seekers probability of application, sometimes with differential effects by gender. Consistent with the factors mentioned above, Flory, Leibbrandt, and List (2015) find that an opening that is framed as more "male-typed", more competitive, or with more pay uncertainty deters female applicants more than male applicants. Similarly, Samek (2015) finds that changing from a competitive framing to a non-profit framing can attract more female applicants. Gee (2018) finds that showing job-seekers the number of other applicants increases applications from women more than men, arguably through a reduction in ambiguity. Together, this evidence suggests that decreasing perceived competitiveness and uncertainty about the position may help to reduce gender gaps in application behavior. Female role models can also influence application decisions: Del Carpio and Guadalupe (2018) find that women are more likely to opt into a tech skills training program when presented with an example of a female success story. Our paper will build on this important work by attempting to better understand the decision of whether and when to apply through a series of controlled experiments, including laboratory experiments that can speak to underlying mechanisms.

Our paper tackles the application decision in a few key steps. First, we explore the role for beliefs. We ask how individuals view their own qualification levels, given their talent. We do this across two distinct settings. In the first, we expose individuals to real job advertisements and ask them questions about how

qualified they feel they are for different types of positions. While this gives us data with rich external validity, we cannot incentivize their responses. So, we complement this approach with a laboratory study where we can directly measure the talent of each individual in a simulated labor market environment, and ask them incentivized questions about their believed abilities.

We find significant gender differences for two different types of beliefs. Not only do we see gender differences in individuals' beliefs about their own talent level, but we also see important evidence for gender differences in what individuals believe "the bar" is for a given position. That is, even conditional on having the same belief of their own talent, men and women seem to vary in how qualified they feel for a given opening. Women simply believe the bar is higher. Both of these types of beliefs are predictive of application decisions.

Finally, we turn to policy solutions. The fact that we observe gender differences in beliefs about what it takes to be successful in applying leads us to test a simple intervention: providing more information about "the bar". Given that in many contexts there is uncertainty about exactly what types of qualifications an employer is looking for, heterogeneity in beliefs about what it might take to be hired are perhaps not surprising. We test a simple policy intervention in which we provide more concrete, objective measures of what is needed in order to receive the job, and we ask whether this helps to reduce the gender gap in application rates.

We conducted a series of three experiments to address these questions, including a field experiment with real job applicants. In the first laboratory study, we provide evidence on how female and male students perceive their and their peers' qualifications for real job openings. In the second, we create a controlled laboratory setting online that simulates a labor market, where we exogenously vary whether and how precisely the qualifications for a promotion opportunity are presented to participants. A similar exercise is then repeated as a field experiment on the online labor market platform Upwork with real job applicants.

Across our three studies, we draw three important conclusions. First, on average, women believe themselves to be less qualified for a given job opening than men. However, this gender gap is reduced when the qualifications for the job openings are more objective and clear. Second, men and women also differ in their beliefs about where "the bar" is for a given opening. Even holding fixed how they view themselves, women seem to believe it is more difficult for anyone to be qualified for a given opening as compared to men. Third, clearly stated qualifications help to reduce gender gaps in application rates. In our field experiment, we find that very few talented, qualified women apply for an advanced job in a baseline condition. Once we provide more detailed information on expectations of what is needed to receive the

advanced job, the gender gap is reduced substantially, creating a more gender diverse, and more talented, pool of applicants.

II. Motivating Evidence from Impressions of Real Job Advertisements

Methods

Our goal in this first study is to collect potential job-seekers' impressions of real job ads. To increase the external validity of the exercise, we constructed a random sample of real job postings, representative of available online postings in the geographic area of our participants.

In April 2018, we performed a search of job advertisements posted to Indeed.com. We searched for full-time job postings in the Boston area that required a Bachelor's degree. We performed two searches: a search of entry-level jobs and a search of mid-level jobs. Within the entry-level search, we downloaded all ads that were returned from our search and randomly selected 20; within the mid-level search, we downloaded all ads that were returned and randomly selected 30. We then read each ad selected. In cases where the ad description did not appear to fulfill our search criteria (for example, not actually being full-time, despite being returned by Indeed.com), we eliminated the ad from our sample. This left us with 38 job ads: 13 entry-level and 25 mid-level. Copies of each ad are available in the Appendix.

In the first part of the study, we collect participant impressions of how well-qualified they feel they are for a given job posting. The structure is as follows. Participants complete four rounds of job ad evaluation. Within each round, participants are given two minutes to view a randomly-selected ad from the set of 38.[2]

We limit participants' time viewing the ads for two reasons. First, we wanted to ensure timely completion of the study, as completion in under 20 minutes was required for the laboratory format we took advantage of. Second, we wanted to limit the extent to which participants simply "looked up answers" when asked questions. We gave them a set period of time to view the ad and form a general impression; then, they proceed to questions about the ad. In this way, they cannot refer back to specific details but instead hopefully give us their holistic impression of the ad.

We ask three questions about perceived qualifications:

1. *On a scale of 1 (Extremely Poorly Qualified) - 10 (Extremely Well Qualified), how well-qualified do you feel you are for this job?*

[2] We create four non-overlapping subsets of the 38 ads. Each bucket contains 3 – 4 entry-level jobs and 6 – 7 mid-level jobs. Within each round of this study, participants view one randomly-selected ad from a given bucket, ensuring that no participant sees the same ad twice in one part of the study.

2. *Thinking of the desired skills, characteristics, and qualifications stated in the advertisement, what percent of those skills, characteristics, and qualifications do you possess?*

3. *More specifically, please list some of the desired skills, characteristics, and qualifications that you do possess, and some of the desired skills, characteristics, and qualifications that you do not possess.*

The first question gets at our core issue: how well-qualified an individual feels for a given position. The second question allows us to start to disentangle different stories for why individuals might vary in how well-qualified they feel they are. Suppose participant *a* feels less well-qualified than participant *b* (as indicated by answers to Question 1). It could be that participant *a* believes she possesses fewer of the stated qualifications than *b*, or it could be that conditional on believing she possesses the same fraction of qualifications, *a* views this as less worthy of a "well-qualified" rating compared to *b.* With the data from question 2, we can test this second explanation; in particular, we can ask whether conditional on providing the same answer to question 2, men and women vary in how well-qualified they believe themselves to be. The third question encourages participants to think of the specific ad they just viewed.

We also ask two decoy questions of our participants:

4. *On a scale of 1 - 10, with 1 being not appealing at all and 10 being extremely appealing, how appealing is this job opening to you?*

5. *Approximately what salary would you expect this job to offer?*

We include these questions so that participants do not become solely focused on qualifications as they read additional job ads in the study. We display the qualification questions first, followed by the decoy questions, in Rounds 1 and 3; we reverse the order in Rounds 2 and 4.

In the second part of this study, we collect additional data from *the same set of* participants about each of these 38 ads. In particular, we recognized that in analyzing the data on perceived qualifications, there were a number of ad characteristics that we would want to collect and use as controls in our analysis. Rather than code these characteristics ourselves, we chose to have participants code these characteristics, ensuring no researcher bias.

The format of the second study is nearly identical to the first. Participants complete four rounds of the study. Within each round, they are given 2 minutes to view one randomly-selected ad from the 38. Note that this randomization operates independently from the randomization in Study 1; thus, participants could be

"Why are you not a full professor - given your eminence?"

[Silence]

"I never applied." - *Donna Strickland, Nobel Laureate in Physics, 2018*

I. Introduction

An important body of work documents the impact of gender bias and discrimination on women's careers (see Riach and Rich 2002 for an overview). Women are less likely to be interviewed for high-status jobs (Fernandez and Mors 2008) and promotions (Ginther and Kahn 2009; Ibarra, Carter, and Silva 2010; Zahidi and Ibarra 2010). Evidence from the laboratory reinforces these findings, with many studies showing that employers in simulated labor markets are more likely to hire men than women for male-typed jobs (Bohnet, van Geen, and Bazerman 2016; Reuben, Sapienza, and Zingales 2014; Coffman, Exley, and Niederle 2018). Once a female worker is hired, she is subject to bias in both formal job evaluation processes (Heilman 2001) and in more informal mentoring (Ibarra, Carter, and Silva (2010)). Firms are devoting significant attention to reducing these biases, with the hope of achieving greater gender diversity throughout the throughout the pipeline.[1]

Of course, when considering the sources of gender gaps in labor market outcomes, discrimination and bias are only one side of the coin. Decisions made by employees themselves also have the potential to have large impacts on gender gaps in outcomes. Candidates decide what types of education to pursue, which jobs to apply to, and when to put themselves up for promotion. Gender differences at these crucial decision nodes could play an important factor. Indeed, social scientists have documented that occupational segregation plays an important role in explaining gender gaps in wages (Altonji and Blank 1999). Choosing an industry, however, is just one of many important choices an employee makes.

In this paper, we study the decisions of candidates about whether and when to apply for different opportunities. We aim to tackle the question of whether there are gender differences in how job-seekers perceive their own qualifications for different job opportunities, and how this impacts their decision about

[1]Salesforce is a prime example. They have spent $6M in an attempt to close the gender wage gap that exists at their firm, auditing their wage data on a regular basis and taking steps to eliminate bias from many stages of their hiring and promotion practices ("2019 Salesforce Equal Pay Update" 2019).

whether or not to apply. These decisions are likely to be key not only at the hiring stage, but also as careers advance, presenting opportunities for promotion.

Outside of controlled experiments, many of these decisions about whether and when to apply may be made in the face of (anticipated) bias, making it hard to isolate the role for self-stereotyping from external bias or fear of bias. We take advantage of controlled frameworks to separate between these stories, allowing us to focus on the role of employee perceptions and decisions, absent employer biases. We ask whether women are as likely as men to see themselves as qualified for challenging, high-paying positions, and whether they apply at similar rates conditional on their degree of qualification.

The past literature on gender differences provides some potential reasons why women may be less likely to apply. Careful laboratory evidence suggests that conditional on having the same ability, women have more pessimistic beliefs about their own ability in male-typed domains compared to men (Niederle and Vesterlund 2007, Coffman 2014; Bordalo et al. 2018, Exley and Kessler 2019). This suggests that if men and women share the same view as to what is required to qualify for a given position, women may be less likely to believe they possess that qualification (holding all else equal). Even conditional on holding the same beliefs, differences in competitive preferences could also drive differences in behavior (Niederle and Vesterlund 2007). It may be that women are less willing to compete for a job opening by applying.

A few clever field experiments have explored some factors that impact job-seekers probability of application, sometimes with differential effects by gender. Consistent with the factors mentioned above, Flory, Leibbrandt, and List (2015) find that an opening that is framed as more "male-typed", more competitive, or with more pay uncertainty deters female applicants more than male applicants. Similarly, Samek (2015) finds that changing from a competitive framing to a non-profit framing can attract more female applicants. Gee (2018) finds that showing job-seekers the number of other applicants increases applications from women more than men, arguably through a reduction in ambiguity. Together, this evidence suggests that decreasing perceived competitiveness and uncertainty about the position may help to reduce gender gaps in application behavior. Female role models can also influence application decisions: Del Carpio and Guadalupe (2018) find that women are more likely to opt into a tech skills training program when presented with an example of a female success story. Our paper will build on this important work by attempting to better understand the decision of whether and when to apply through a series of controlled experiments, including laboratory experiments that can speak to underlying mechanisms.

Our paper tackles the application decision in a few key steps. First, we explore the role for beliefs. We ask how individuals view their own qualification levels, given their talent. We do this across two distinct settings. In the first, we expose individuals to real job advertisements and ask them questions about how

qualified they feel they are for different types of positions. While this gives us data with rich external validity, we cannot incentivize their responses. So, we complement this approach with a laboratory study where we can directly measure the talent of each individual in a simulated labor market environment, and ask them incentivized questions about their believed abilities.

We find significant gender differences for two different types of beliefs. Not only do we see gender differences in individuals' beliefs about their own talent level, but we also see important evidence for gender differences in what individuals believe "the bar" is for a given position. That is, even conditional on having the same belief of their own talent, men and women seem to vary in how qualified they feel for a given opening. Women simply believe the bar is higher. Both of these types of beliefs are predictive of application decisions.

Finally, we turn to policy solutions. The fact that we observe gender differences in beliefs about what it takes to be successful in applying leads us to test a simple intervention: providing more information about "the bar". Given that in many contexts there is uncertainty about exactly what types of qualifications an employer is looking for, heterogeneity in beliefs about what it might take to be hired are perhaps not surprising. We test a simple policy intervention in which we provide more concrete, objective measures of what is needed in order to receive the job, and we ask whether this helps to reduce the gender gap in application rates.

We conducted a series of three experiments to address these questions, including a field experiment with real job applicants. In the first laboratory study, we provide evidence on how female and male students perceive their and their peers' qualifications for real job openings. In the second, we create a controlled laboratory setting online that simulates a labor market, where we exogenously vary whether and how precisely the qualifications for a promotion opportunity are presented to participants. A similar exercise is then repeated as a field experiment on the online labor market platform Upwork with real job applicants.

Across our three studies, we draw three important conclusions. First, on average, women believe themselves to be less qualified for a given job opening than men. However, this gender gap is reduced when the qualifications for the job openings are more objective and clear. Second, men and women also differ in their beliefs about where "the bar" is for a given opening. Even holding fixed how they view themselves, women seem to believe it is more difficult for anyone to be qualified for a given opening as compared to men. Third, clearly stated qualifications help to reduce gender gaps in application rates. In our field experiment, we find that very few talented, qualified women apply for an advanced job in a baseline condition. Once we provide more detailed information on expectations of what is needed to receive the

"Why are you not a full professor - given your eminence?"

[Silence]

"I never applied." - *Donna Strickland, Nobel Laureate in Physics, 2018*

I. Introduction

An important body of work documents the impact of gender bias and discrimination on women's careers (see Riach and Rich 2002 for an overview). Women are less likely to be interviewed for high-status jobs (Fernandez and Mors 2008) and promotions (Ginther and Kahn 2009; Ibarra, Carter, and Silva 2010; Zahidi and Ibarra 2010). Evidence from the laboratory reinforces these findings, with many studies showing that employers in simulated labor markets are more likely to hire men than women for male-typed jobs (Bohnet, van Geen, and Bazerman 2016; Reuben, Sapienza, and Zingales 2014; Coffman, Exley, and Niederle 2018). Once a female worker is hired, she is subject to bias in both formal job evaluation processes (Heilman 2001) and in more informal mentoring (Ibarra, Carter, and Silva (2010)). Firms are devoting significant attention to reducing these biases, with the hope of achieving greater gender diversity throughout the throughout the pipeline.[1]

Of course, when considering the sources of gender gaps in labor market outcomes, discrimination and bias are only one side of the coin. Decisions made by employees themselves also have the potential to have large impacts on gender gaps in outcomes. Candidates decide what types of education to pursue, which jobs to apply to, and when to put themselves up for promotion. Gender differences at these crucial decision nodes could play an important factor. Indeed, social scientists have documented that occupational segregation plays an important role in explaining gender gaps in wages (Altonji and Blank 1999). Choosing an industry, however, is just one of many important choices an employee makes.

In this paper, we study the decisions of candidates about whether and when to apply for different opportunities. We aim to tackle the question of whether there are gender differences in how job-seekers perceive their own qualifications for different job opportunities, and how this impacts their decision about

[1]Salesforce is a prime example. They have spent $6M in an attempt to close the gender wage gap that exists at their firm, auditing their wage data on a regular basis and taking steps to eliminate bias from many stages of their hiring and promotion practices ("2019 Salesforce Equal Pay Update" 2019).

whether or not to apply. These decisions are likely to be key not only at the hiring stage, but also as careers advance, presenting opportunities for promotion.

Outside of controlled experiments, many of these decisions about whether and when to apply may be made in the face of (anticipated) bias, making it hard to isolate the role for self-stereotyping from external bias or fear of bias. We take advantage of controlled frameworks to separate between these stories, allowing us to focus on the role of employee perceptions and decisions, absent employer biases. We ask whether women are as likely as men to see themselves as qualified for challenging, high-paying positions, and whether they apply at similar rates conditional on their degree of qualification.

The past literature on gender differences provides some potential reasons why women may be less likely to apply. Careful laboratory evidence suggests that conditional on having the same ability, women have more pessimistic beliefs about their own ability in male-typed domains compared to men (Niederle and Vesterlund 2007, Coffman 2014; Bordalo et al. 2018, Exley and Kessler 2019). This suggests that if men and women share the same view as to what is required to qualify for a given position, women may be less likely to believe they possess that qualification (holding all else equal). Even conditional on holding the same beliefs, differences in competitive preferences could also drive differences in behavior (Niederle and Vesterlund 2007). It may be that women are less willing to compete for a job opening by applying.

A few clever field experiments have explored some factors that impact job-seekers probability of application, sometimes with differential effects by gender. Consistent with the factors mentioned above, Flory, Leibbrandt, and List (2015) find that an opening that is framed as more "male-typed", more competitive, or with more pay uncertainty deters female applicants more than male applicants. Similarly, Samek (2015) finds that changing from a competitive framing to a non-profit framing can attract more female applicants. Gee (2018) finds that showing job-seekers the number of other applicants increases applications from women more than men, arguably through a reduction in ambiguity. Together, this evidence suggests that decreasing perceived competitiveness and uncertainty about the position may help to reduce gender gaps in application behavior. Female role models can also influence application decisions: Del Carpio and Guadalupe (2018) find that women are more likely to opt into a tech skills training program when presented with an example of a female success story. Our paper will build on this important work by attempting to better understand the decision of whether and when to apply through a series of controlled experiments, including laboratory experiments that can speak to underlying mechanisms.

Our paper tackles the application decision in a few key steps. First, we explore the role for beliefs. We ask how individuals view their own qualification levels, given their talent. We do this across two distinct settings. In the first, we expose individuals to real job advertisements and ask them questions about how

randomly assigned the same ad in both studies, but this was not particularly likely. They are then asked four questions about the ad:

1. *In general do you think the stereotype associated with this job is more female-typed or more male-typed? Use the slider scale below to indicate your answer, where -1 indicates extremely female-typed and 1 indicates extremely male-typed.*

2. *How prestigious would you say this job is? Use the slider scale below to indicate your answer, where 1 indicates not prestigious at all and 7 indicates extremely prestigious.*

3. *Thinking of typical Harvard undergraduates, how well-qualified do you think the average Harvard undergraduate would be for this job? Use the slider scale below to indicate your answer, where 1 indicates not at all qualified and 10 indicates extremely well-qualified.*

4. *Thinking of how the qualifications in the job advertisement were described, how specific, clear, and objective were the stated qualifications? Use the slider scale below to indicate your answer, where 1 indicates not at all clear and 10 indicates extremely specific, clear, and objective.*

We hypothesized that each of these measures could be relevant in predicting participant beliefs about how well-qualified they were. The first gets at the gender-stereotype associated with the job, speaking to the mechanism of Coffman (2014), who finds that individual self-confidence and willingness to put oneself forward is highly dependent on the gender congruency of the domain. If beliefs of own ability are a key driver in predicting beliefs of how well-qualified someone is, then past work, including Coffman (2014) and Bordalo et al. (2018) would suggest that the gender stereotype associated with the job could be an important factor; we would predict that as the maleness of the job posting increased, men would feel relatively more well-qualified (holding true qualification level fixed) while women would feel relatively less well-qualified. The second question allows us to try to separate out differences in the gender stereotype attached to the job from differences in the prestige of the position.

The third question accomplishes two main goals. First, it allows us to better account for variation across ads in how likely it is that any participant in our study feels qualified for that particular ad. Second, it allows us to ask how individuals vary in their beliefs about how well-qualified others are for various openings. Is it the case that individuals who feel poorly qualified believe others would also be poorly qualified for that opening (suggesting it was something about the job), or do they believe others would be well qualified for that openings (suggesting it was more something about themselves).

Finally, the fourth question speaks to the main mechanism we will test in our controlled experiments: whether the specificity and clarity of the stated qualifications matter. The hypothesis is that as qualifications become clearer, holding all else fixed, the gender gap will be reduced: when ads are ambiguous, there is more room for stereotypes and subjective beliefs to shape decisions.

In total 200 participants completed the two studies as part of bundle sessions at the Computer Lab for Experimental Research at Harvard Business School, of which 197 provided information on their gender.[3] The laboratory format we used bundles our study with short studies from other researchers. The placement of our studies within these other studies was varied across session, though note that the two parts of our study are always placed in the same order and appear consecutively.

Results

We provide summary statistics on our participants in Appendix Table B1.

We find that, on average, men view themselves as marginally more well-qualified than women in our sample. Recall that participants rate on a 1-10 scale how well-qualified they feel they are for each of four particular job ads. On average, men rate themselves a 4.65 (2.63 SD) while women rate themselves a 4.22 (2.51 SD). In Table 1, we present a regression that explores the determinants of these ratings, while controlling for ad fixed effects, order of appearance of our study within the session, and demographic controls. In Columns II and III, we include the average rating of the ad in terms of stereotype, prestige, objectivity of stated qualifications, and how well-qualified the typical Harvard undergraduate would be for the ad, as rated by the participants in the second part of the study, as additional controls. Controlling for ad and individual characteristics, we estimate that women rate themselves approximately 0.46 points less well-qualified for a given job opening on average ($p<0.10$).

In Column III, we interact the ad characteristics with the female dummy, to ask whether any of the characteristics of the job opening are differentially important for men and women. We find evidence in support of the main channel we explore in our follow-up experiments: more objectively stated job qualifications, as rated by our participants, help to close the gender gap in perceived qualification level. More objectively stated qualifications directionally decrease the extent to which men feel well-qualified, while this effect is significantly less pronounced for women, serving to decrease the average gender gap in perceived qualification level. None of the other job ad characteristics have predictive power for perceived qualifications or the gender gap in perceived qualifications.

Table 1. The Gender Gap in Perceived Qualification for Real Job Openings

[3] In addition, one participant did not provide data on their age; thus they are excluded from analyses that control for age. We obtain answers to a standard bank of demographic questions that are asked of all participants in bundle sessions by the laboratory.

OLS Predicting How Well-Qualified an Individual Feels for Job Opening			
	I	II	III
Female	-0.46* (0.24)	-0.46* (0.24)	-4.66 (3.01)
Average rating of Male Stereotype for this Ad		0.52 (3.43)	0.81 (3.45)
Average Rating of Prestige for this Ad		-0.59 (0.72)	-0.28 (0.76)
Average Rating of Objectivity of Stated Qualifications for this Ad		-0.38 (0.47)	-0.84 (0.53)
Average Belief of How Well-Qualified Undergrad Would be for this Ad		1.45 (1.26)	1.44 (1.25)
Female x Stereotype			-0.24 (0.85)
Female x Prestige			-0.53 (0.35)
Female x Objectivity			0.78** (0.30)
Female x Avg. Qualified			0.15 (0.26)
Ad Fixed Effects	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes
Order of Study within Session	Yes	Yes	Yes
R-squared	0.20	0.20	0.21
Clusters (Obs.)	196 (784)	196 (784)	196 (784)

Notes: Controls are fixed effects for each ad, fixed effects for each race category, fixed effects for each education category, age, a dummy for fluent in English, and a dummy indicating where our pair of studies fell within the session. * indicates $p<0.10$, ** indicates $p<0.05$, ***indicates $p<0.01$, ****indicates $p<0.001$.

Note that these results are very similar if we instead use as the dependent variable the fraction of qualifications that a participant believes they possess. On average, conditional on ad and individual characteristics, women believe they possess roughly 5pp fewer of the qualifications than men do (male average: 0.50, female average: 0.45, $p<0.05$ see Appendix Table B2). And, conditional on believing they have the same fraction of qualifications, men and women rate themselves as equally well-qualified (Appendix Table B3).

Interestingly, we also see significant gender differences in how individuals rate the average qualification level of a typical Harvard undergraduate for a given ad in Study 2. Men rate the typical Harvard undergraduate as being significantly more well-qualified than women do, providing respective ratings of 6.55 (SD 2.45) and 5.68 (2.57), respectively ($p<0.01$ in regression that clusters at individual level). This suggests that the gender gap in perceived qualification that we observe may also relate to gender differences in beliefs about how challenging it is to be qualified for a given job opening *for anyone*, not just themselves.

While this data provides interesting insights into perceptions of real job postings, it leaves open a number of important issues. In particular, we know very little about our participants in this setting. While we can observe differences in beliefs of how well-qualified they are, we cannot speak to differences in true underlying qualifications. To get at this key issue, we move to the controlled environment of an incentivized online experiment.

III. Evidence from a Controlled Experiment

We use a controlled experiment to better understand what drives participant beliefs about their qualifications, and to link these beliefs to incentivized decisions. In the setting below, we are able to observe the true ability (or qualification level) of each participant. We can then ask, given this ability, what their beliefs look like, about themselves and about their chances of receiving a given job conditional on applying. Importantly, we can also ask participants to make incentivized application decisions, asking whether there are gender differences in willingness to apply for a promotion.

Design

The general structure of the experiment is as follows. In the first part of the experiment, we aim to collect data on participant ability in a diagnostic test. We also elicit participant beliefs about their own performance. Then, we confront participants with a decision. They are asked to decide whether to apply for a promotion – a higher return opportunity – for Round 2 of the experiment. Then, participants complete Round 2. We detail each of these stages below.

Round 1

In Round 1 of the experiment, participants take an ability test that covers general science, arithmetic reasoning, math knowledge, mechanical comprehension, and assembling objects. The questions are drawn from the Armed Services Vocational Aptitude Battery (ASVAB). The questions were chosen for their simple multiple-choice format, and the fact that they cover stereotypically male domains. Participants have

5 minutes to answer up to 30 questions. All 30 questions appear on the same page and can be answered in any order. If Round 1 is chosen for payment, participants receive $0.20 per question answered correctly.

Beliefs of Round 1 Performance

After completing Round 1, each participant is asked to guess their score – that is, how many problems she solved correctly in Round 1 -- and how they rank relative to other MTurkers who are completing the HIT. They receive $0.10 if they guess their score correctly and $0.10 if they guess their bucket of rank correctly (bottom 5%, bottom 20%, bottom 40%, middle 20%, top 40%, top 20%, top 5%). They are then randomized into one of three feedback conditions: receiving either no feedback on Round 1 performance, a signal equal to their true score 60% of the time, or a signal equal to their true score 90% of the time. We then elicit posterior beliefs of Round 1 score from participants in each of the two noisy feedback conditions.

This noisy feedback intervention is the focus of a different paper, detailed in Coffman, Collis, and Kulkarni (2019). We will not focus on it in our analysis, though we will use data on posterior beliefs of Round 1 score and we will account for feedback treatment assignment in the analysis that follows.

Application Decision

Participants are told that they will soon have a chance to participate in a second round of ASVAB problem-solving. Again, they will have 5 minutes to answer up to 30 ASVAB questions, but are told that these questions will be more difficult on average than the questions in Round 1. In this way, Round 1 performance is predictive of Round 2 performance, but there is additional uncertainty due to the increased difficulty. If Round 2 is chosen for payment, they receive $0.20 per problem solved correctly.

Prior to completing Round 2, however, they have to make a decision about whether they want to apply for an "expert-level promotion". Participants are presented with two options:

"Option 1: Accept the novice job. If you choose this option and Round 2 is chosen for payment, you will get a Round 2 completion payment of $2 on top of the $0.20 per problem solved correctly in Round 2."

"Option 2: Apply for the expert-level job. If you choose this option **and** you are chosen to be promoted to the expert-level job, you will get a **promotion bonus** plus an extra $0.20 per problem solved correctly in Round 2, for a total of $0.40 per problem solved correctly. However, if you apply for the expert-level job and you are not promoted, you will only earn the $0.20 per problem solved correctly. You would not earn a Round 2 promotion bonus."

2. *Thinking of the desired skills, characteristics, and qualifications stated in the advertisement, what percent of those skills, characteristics, and qualifications do you possess?*

3. *More specifically, please list some of the desired skills, characteristics, and qualifications that you do possess, and some of the desired skills, characteristics, and qualifications that you do not possess.*

The first question gets at our core issue: how well-qualified an individual feels for a given position. The second question allows us to start to disentangle different stories for why individuals might vary in how well-qualified they feel they are. Suppose participant *a* feels less well-qualified than participant *b* (as indicated by answers to Question 1). It could be that participant *a* believes she possesses fewer of the stated qualifications than *b*, or it could be that conditional on believing she possesses the same fraction of qualifications, *a* views this as less worthy of a "well-qualified" rating compared to *b.* With the data from question 2, we can test this second explanation; in particular, we can ask whether conditional on providing the same answer to question 2, men and women vary in how well-qualified they believe themselves to be. The third question encourages participants to think of the specific ad they just viewed.

We also ask two decoy questions of our participants:

4. *On a scale of 1 - 10, with 1 being not appealing at all and 10 being extremely appealing, how appealing is this job opening to you?*

5. *Approximately what salary would you expect this job to offer?*

We include these questions so that participants do not become solely focused on qualifications as they read additional job ads in the study. We display the qualification questions first, followed by the decoy questions, in Rounds 1 and 3; we reverse the order in Rounds 2 and 4.

In the second part of this study, we collect additional data from *the same set of* participants about each of these 38 ads. In particular, we recognized that in analyzing the data on perceived qualifications, there were a number of ad characteristics that we would want to collect and use as controls in our analysis. Rather than code these characteristics ourselves, we chose to have participants code these characteristics, ensuring no researcher bias.

The format of the second study is nearly identical to the first. Participants complete four rounds of the study. Within each round, they are given 2 minutes to view one randomly-selected ad from the 38. Note that this randomization operates independently from the randomization in Study 1; thus, participants could be

"Why are you not a full professor - given your eminence?"

[Silence]

"I never applied." - *Donna Strickland, Nobel Laureate in Physics, 2018*

I. Introduction

An important body of work documents the impact of gender bias and discrimination on women's careers (see Riach and Rich 2002 for an overview). Women are less likely to be interviewed for high-status jobs (Fernandez and Mors 2008) and promotions (Ginther and Kahn 2009; Ibarra, Carter, and Silva 2010; Zahidi and Ibarra 2010). Evidence from the laboratory reinforces these findings, with many studies showing that employers in simulated labor markets are more likely to hire men than women for male-typed jobs (Bohnet, van Geen, and Bazerman 2016; Reuben, Sapienza, and Zingales 2014; Coffman, Exley, and Niederle 2018). Once a female worker is hired, she is subject to bias in both formal job evaluation processes (Heilman 2001) and in more informal mentoring (Ibarra, Carter, and Silva (2010)). Firms are devoting significant attention to reducing these biases, with the hope of achieving greater gender diversity throughout the throughout the pipeline.[1]

Of course, when considering the sources of gender gaps in labor market outcomes, discrimination and bias are only one side of the coin. Decisions made by employees themselves also have the potential to have large impacts on gender gaps in outcomes. Candidates decide what types of education to pursue, which jobs to apply to, and when to put themselves up for promotion. Gender differences at these crucial decision nodes could play an important factor. Indeed, social scientists have documented that occupational segregation plays an important role in explaining gender gaps in wages (Altonji and Blank 1999). Choosing an industry, however, is just one of many important choices an employee makes.

In this paper, we study the decisions of candidates about whether and when to apply for different opportunities. We aim to tackle the question of whether there are gender differences in how job-seekers perceive their own qualifications for different job opportunities, and how this impacts their decision about

[1]Salesforce is a prime example. They have spent $6M in an attempt to close the gender wage gap that exists at their firm, auditing their wage data on a regular basis and taking steps to eliminate bias from many stages of their hiring and promotion practices ("2019 Salesforce Equal Pay Update" 2019).

whether or not to apply. These decisions are likely to be key not only at the hiring stage, but also as careers advance, presenting opportunities for promotion.

Outside of controlled experiments, many of these decisions about whether and when to apply may be made in the face of (anticipated) bias, making it hard to isolate the role for self-stereotyping from external bias or fear of bias. We take advantage of controlled frameworks to separate between these stories, allowing us to focus on the role of employee perceptions and decisions, absent employer biases. We ask whether women are as likely as men to see themselves as qualified for challenging, high-paying positions, and whether they apply at similar rates conditional on their degree of qualification.

The past literature on gender differences provides some potential reasons why women may be less likely to apply. Careful laboratory evidence suggests that conditional on having the same ability, women have more pessimistic beliefs about their own ability in male-typed domains compared to men (Niederle and Vesterlund 2007, Coffman 2014; Bordalo et al. 2018, Exley and Kessler 2019). This suggests that if men and women share the same view as to what is required to qualify for a given position, women may be less likely to believe they possess that qualification (holding all else equal). Even conditional on holding the same beliefs, differences in competitive preferences could also drive differences in behavior (Niederle and Vesterlund 2007). It may be that women are less willing to compete for a job opening by applying.

A few clever field experiments have explored some factors that impact job-seekers probability of application, sometimes with differential effects by gender. Consistent with the factors mentioned above, Flory, Leibbrandt, and List (2015) find that an opening that is framed as more "male-typed", more competitive, or with more pay uncertainty deters female applicants more than male applicants. Similarly, Samek (2015) finds that changing from a competitive framing to a non-profit framing can attract more female applicants. Gee (2018) finds that showing job-seekers the number of other applicants increases applications from women more than men, arguably through a reduction in ambiguity. Together, this evidence suggests that decreasing perceived competitiveness and uncertainty about the position may help to reduce gender gaps in application behavior. Female role models can also influence application decisions: Del Carpio and Guadalupe (2018) find that women are more likely to opt into a tech skills training program when presented with an example of a female success story. Our paper will build on this important work by attempting to better understand the decision of whether and when to apply through a series of controlled experiments, including laboratory experiments that can speak to underlying mechanisms.

Our paper tackles the application decision in a few key steps. First, we explore the role for beliefs. We ask how individuals view their own qualification levels, given their talent. We do this across two distinct settings. In the first, we expose individuals to real job advertisements and ask them questions about how

qualified they feel they are for different types of positions. While this gives us data with rich external validity, we cannot incentivize their responses. So, we complement this approach with a laboratory study where we can directly measure the talent of each individual in a simulated labor market environment, and ask them incentivized questions about their believed abilities.

We find significant gender differences for two different types of beliefs. Not only do we see gender differences in individuals' beliefs about their own talent level, but we also see important evidence for gender differences in what individuals believe "the bar" is for a given position. That is, even conditional on having the same belief of their own talent, men and women seem to vary in how qualified they feel for a given opening. Women simply believe the bar is higher. Both of these types of beliefs are predictive of application decisions.

Finally, we turn to policy solutions. The fact that we observe gender differences in beliefs about what it takes to be successful in applying leads us to test a simple intervention: providing more information about "the bar". Given that in many contexts there is uncertainty about exactly what types of qualifications an employer is looking for, heterogeneity in beliefs about what it might take to be hired are perhaps not surprising. We test a simple policy intervention in which we provide more concrete, objective measures of what is needed in order to receive the job, and we ask whether this helps to reduce the gender gap in application rates.

We conducted a series of three experiments to address these questions, including a field experiment with real job applicants. In the first laboratory study, we provide evidence on how female and male students perceive their and their peers' qualifications for real job openings. In the second, we create a controlled laboratory setting online that simulates a labor market, where we exogenously vary whether and how precisely the qualifications for a promotion opportunity are presented to participants. A similar exercise is then repeated as a field experiment on the online labor market platform Upwork with real job applicants.

Across our three studies, we draw three important conclusions. First, on average, women believe themselves to be less qualified for a given job opening than men. However, this gender gap is reduced when the qualifications for the job openings are more objective and clear. Second, men and women also differ in their beliefs about where "the bar" is for a given opening. Even holding fixed how they view themselves, women seem to believe it is more difficult for anyone to be qualified for a given opening as compared to men. Third, clearly stated qualifications help to reduce gender gaps in application rates. In our field experiment, we find that very few talented, qualified women apply for an advanced job in a baseline condition. Once we provide more detailed information on expectations of what is needed to receive the

"Why are you not a full professor - given your eminence?"

[Silence]

"I never applied." - *Donna Strickland, Nobel Laureate in Physics, 2018*

I. Introduction

An important body of work documents the impact of gender bias and discrimination on women's careers (see Riach and Rich 2002 for an overview). Women are less likely to be interviewed for high-status jobs (Fernandez and Mors 2008) and promotions (Ginther and Kahn 2009; Ibarra, Carter, and Silva 2010; Zahidi and Ibarra 2010). Evidence from the laboratory reinforces these findings, with many studies showing that employers in simulated labor markets are more likely to hire men than women for male-typed jobs (Bohnet, van Geen, and Bazerman 2016; Reuben, Sapienza, and Zingales 2014; Coffman, Exley, and Niederle 2018). Once a female worker is hired, she is subject to bias in both formal job evaluation processes (Heilman 2001) and in more informal mentoring (Ibarra, Carter, and Silva (2010)). Firms are devoting significant attention to reducing these biases, with the hope of achieving greater gender diversity throughout the throughout the pipeline.[1]

Of course, when considering the sources of gender gaps in labor market outcomes, discrimination and bias are only one side of the coin. Decisions made by employees themselves also have the potential to have large impacts on gender gaps in outcomes. Candidates decide what types of education to pursue, which jobs to apply to, and when to put themselves up for promotion. Gender differences at these crucial decision nodes could play an important factor. Indeed, social scientists have documented that occupational segregation plays an important role in explaining gender gaps in wages (Altonji and Blank 1999). Choosing an industry, however, is just one of many important choices an employee makes.

In this paper, we study the decisions of candidates about whether and when to apply for different opportunities. We aim to tackle the question of whether there are gender differences in how job-seekers perceive their own qualifications for different job opportunities, and how this impacts their decision about

[1]Salesforce is a prime example. They have spent $6M in an attempt to close the gender wage gap that exists at their firm, auditing their wage data on a regular basis and taking steps to eliminate bias from many stages of their hiring and promotion practices ("2019 Salesforce Equal Pay Update" 2019).

whether or not to apply. These decisions are likely to be key not only at the hiring stage, but also as careers advance, presenting opportunities for promotion.

Outside of controlled experiments, many of these decisions about whether and when to apply may be made in the face of (anticipated) bias, making it hard to isolate the role for self-stereotyping from external bias or fear of bias. We take advantage of controlled frameworks to separate between these stories, allowing us to focus on the role of employee perceptions and decisions, absent employer biases. We ask whether women are as likely as men to see themselves as qualified for challenging, high-paying positions, and whether they apply at similar rates conditional on their degree of qualification.

The past literature on gender differences provides some potential reasons why women may be less likely to apply. Careful laboratory evidence suggests that conditional on having the same ability, women have more pessimistic beliefs about their own ability in male-typed domains compared to men (Niederle and Vesterlund 2007, Coffman 2014; Bordalo et al. 2018, Exley and Kessler 2019). This suggests that if men and women share the same view as to what is required to qualify for a given position, women may be less likely to believe they possess that qualification (holding all else equal). Even conditional on holding the same beliefs, differences in competitive preferences could also drive differences in behavior (Niederle and Vesterlund 2007). It may be that women are less willing to compete for a job opening by applying.

A few clever field experiments have explored some factors that impact job-seekers probability of application, sometimes with differential effects by gender. Consistent with the factors mentioned above, Flory, Leibbrandt, and List (2015) find that an opening that is framed as more "male-typed", more competitive, or with more pay uncertainty deters female applicants more than male applicants. Similarly, Samek (2015) finds that changing from a competitive framing to a non-profit framing can attract more female applicants. Gee (2018) finds that showing job-seekers the number of other applicants increases applications from women more than men, arguably through a reduction in ambiguity. Together, this evidence suggests that decreasing perceived competitiveness and uncertainty about the position may help to reduce gender gaps in application behavior. Female role models can also influence application decisions: Del Carpio and Guadalupe (2018) find that women are more likely to opt into a tech skills training program when presented with an example of a female success story. Our paper will build on this important work by attempting to better understand the decision of whether and when to apply through a series of controlled experiments, including laboratory experiments that can speak to underlying mechanisms.

Our paper tackles the application decision in a few key steps. First, we explore the role for beliefs. We ask how individuals view their own qualification levels, given their talent. We do this across two distinct settings. In the first, we expose individuals to real job advertisements and ask them questions about how

randomly assigned the same ad in both studies, but this was not particularly likely. They are then asked four questions about the ad:

1. *In general do you think the stereotype associated with this job is more female-typed or more male-typed? Use the slider scale below to indicate your answer, where -1 indicates extremely female-typed and 1 indicates extremely male-typed.*

2. *How prestigious would you say this job is? Use the slider scale below to indicate your answer, where 1 indicates not prestigious at all and 7 indicates extremely prestigious.*

3. *Thinking of typical Harvard undergraduates, how well-qualified do you think the average Harvard undergraduate would be for this job? Use the slider scale below to indicate your answer, where 1 indicates not at all qualified and 10 indicates extremely well-qualified.*

4. *Thinking of how the qualifications in the job advertisement were described, how specific, clear, and objective were the stated qualifications? Use the slider scale below to indicate your answer, where 1 indicates not at all clear and 10 indicates extremely specific, clear, and objective.*

We hypothesized that each of these measures could be relevant in predicting participant beliefs about how well-qualified they were. The first gets at the gender-stereotype associated with the job, speaking to the mechanism of Coffman (2014), who finds that individual self-confidence and willingness to put oneself forward is highly dependent on the gender congruency of the domain. If beliefs of own ability are a key driver in predicting beliefs of how well-qualified someone is, then past work, including Coffman (2014) and Bordalo et al. (2018) would suggest that the gender stereotype associated with the job could be an important factor; we would predict that as the maleness of the job posting increased, men would feel relatively more well-qualified (holding true qualification level fixed) while women would feel relatively less well-qualified. The second question allows us to try to separate out differences in the gender stereotype attached to the job from differences in the prestige of the position.

The third question accomplishes two main goals. First, it allows us to better account for variation across ads in how likely it is that any participant in our study feels qualified for that particular ad. Second, it allows us to ask how individuals vary in their beliefs about how well-qualified others are for various openings. Is it the case that individuals who feel poorly qualified believe others would also be poorly qualified for that opening (suggesting it was something about the job), or do they believe others would be well qualified for that openings (suggesting it was more something about themselves).

Finally, the fourth question speaks to the main mechanism we will test in our controlled experiments: whether the specificity and clarity of the stated qualifications matter. The hypothesis is that as qualifications become clearer, holding all else fixed, the gender gap will be reduced: when ads are ambiguous, there is more room for stereotypes and subjective beliefs to shape decisions.

In total 200 participants completed the two studies as part of bundle sessions at the Computer Lab for Experimental Research at Harvard Business School, of which 197 provided information on their gender.[3] The laboratory format we used bundles our study with short studies from other researchers. The placement of our studies within these other studies was varied across session, though note that the two parts of our study are always placed in the same order and appear consecutively.

Results

We provide summary statistics on our participants in Appendix Table B1.

We find that, on average, men view themselves as marginally more well-qualified than women in our sample. Recall that participants rate on a 1-10 scale how well-qualified they feel they are for each of four particular job ads. On average, men rate themselves a 4.65 (2.63 SD) while women rate themselves a 4.22 (2.51 SD). In Table 1, we present a regression that explores the determinants of these ratings, while controlling for ad fixed effects, order of appearance of our study within the session, and demographic controls. In Columns II and III, we include the average rating of the ad in terms of stereotype, prestige, objectivity of stated qualifications, and how well-qualified the typical Harvard undergraduate would be for the ad, as rated by the participants in the second part of the study, as additional controls. Controlling for ad and individual characteristics, we estimate that women rate themselves approximately 0.46 points less well-qualified for a given job opening on average ($p<0.10$).

In Column III, we interact the ad characteristics with the female dummy, to ask whether any of the characteristics of the job opening are differentially important for men and women. We find evidence in support of the main channel we explore in our follow-up experiments: more objectively stated job qualifications, as rated by our participants, help to close the gender gap in perceived qualification level. More objectively stated qualifications directionally decrease the extent to which men feel well-qualified, while this effect is significantly less pronounced for women, serving to decrease the average gender gap in perceived qualification level. None of the other job ad characteristics have predictive power for perceived qualifications or the gender gap in perceived qualifications.

Table 1. The Gender Gap in Perceived Qualification for Real Job Openings

[3] In addition, one participant did not provide data on their age; thus they are excluded from analyses that control for age. We obtain answers to a standard bank of demographic questions that are asked of all participants in bundle sessions by the laboratory.

OLS Predicting How Well-Qualified an Individual Feels for Job Opening			
	I	II	III
Female	-0.46* (0.24)	-0.46* (0.24)	-4.66 (3.01)
Average rating of Male Stereotype for this Ad		0.52 (3.43)	0.81 (3.45)
Average Rating of Prestige for this Ad		-0.59 (0.72)	-0.28 (0.76)
Average Rating of Objectivity of Stated Qualifications for this Ad		-0.38 (0.47)	-0.84 (0.53)
Average Belief of How Well-Qualified Undergrad Would be for this Ad		1.45 (1.26)	1.44 (1.25)
Female x Stereotype			-0.24 (0.85)
Female x Prestige			-0.53 (0.35)
Female x Objectivity			0.78** (0.30)
Female x Avg. Qualified			0.15 (0.26)
Ad Fixed Effects	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes
Order of Study within Session	Yes	Yes	Yes
R-squared	0.20	0.20	0.21
Clusters (Obs.)	196 (784)	196 (784)	196 (784)

Notes: Controls are fixed effects for each ad, fixed effects for each race category, fixed effects for each education category, age, a dummy for fluent in English, and a dummy indicating where our pair of studies fell within the session. * indicates $p<0.10$, ** indicates $p<0.05$, ***indicates $p<0.01$, ****indicates $p<0.001$.

Note that these results are very similar if we instead use as the dependent variable the fraction of qualifications that a participant believes they possess. On average, conditional on ad and individual characteristics, women believe they possess roughly 5pp fewer of the qualifications than men do (male average: 0.50, female average: 0.45, $p<0.05$ see Appendix Table B2). And, conditional on believing they have the same fraction of qualifications, men and women rate themselves as equally well-qualified (Appendix Table B3).

Interestingly, we also see significant gender differences in how individuals rate the average qualification level of a typical Harvard undergraduate for a given ad in Study 2. Men rate the typical Harvard undergraduate as being significantly more well-qualified than women do, providing respective ratings of 6.55 (SD 2.45) and 5.68 (2.57), respectively ($p<0.01$ in regression that clusters at individual level). This suggests that the gender gap in perceived qualification that we observe may also relate to gender differences in beliefs about how challenging it is to be qualified for a given job opening *for anyone*, not just themselves.

While this data provides interesting insights into perceptions of real job postings, it leaves open a number of important issues. In particular, we know very little about our participants in this setting. While we can observe differences in beliefs of how well-qualified they are, we cannot speak to differences in true underlying qualifications. To get at this key issue, we move to the controlled environment of an incentivized online experiment.

III. Evidence from a Controlled Experiment

We use a controlled experiment to better understand what drives participant beliefs about their qualifications, and to link these beliefs to incentivized decisions. In the setting below, we are able to observe the true ability (or qualification level) of each participant. We can then ask, given this ability, what their beliefs look like, about themselves and about their chances of receiving a given job conditional on applying. Importantly, we can also ask participants to make incentivized application decisions, asking whether there are gender differences in willingness to apply for a promotion.

Design

The general structure of the experiment is as follows. In the first part of the experiment, we aim to collect data on participant ability in a diagnostic test. We also elicit participant beliefs about their own performance. Then, we confront participants with a decision. They are asked to decide whether to apply for a promotion – a higher return opportunity – for Round 2 of the experiment. Then, participants complete Round 2. We detail each of these stages below.

Round 1

In Round 1 of the experiment, participants take an ability test that covers general science, arithmetic reasoning, math knowledge, mechanical comprehension, and assembling objects. The questions are drawn from the Armed Services Vocational Aptitude Battery (ASVAB). The questions were chosen for their simple multiple-choice format, and the fact that they cover stereotypically male domains. Participants have

5 minutes to answer up to 30 questions. All 30 questions appear on the same page and can be answered in any order. If Round 1 is chosen for payment, participants receive $0.20 per question answered correctly.

Beliefs of Round 1 Performance

After completing Round 1, each participant is asked to guess their score – that is, how many problems she solved correctly in Round 1 -- and how they rank relative to other MTurkers who are completing the HIT. They receive $0.10 if they guess their score correctly and $0.10 if they guess their bucket of rank correctly (bottom 5%, bottom 20%, bottom 40%, middle 20%, top 40%, top 20%, top 5%). They are then randomized into one of three feedback conditions: receiving either no feedback on Round 1 performance, a signal equal to their true score 60% of the time, or a signal equal to their true score 90% of the time. We then elicit posterior beliefs of Round 1 score from participants in each of the two noisy feedback conditions.

This noisy feedback intervention is the focus of a different paper, detailed in Coffman, Collis, and Kulkarni (2019). We will not focus on it in our analysis, though we will use data on posterior beliefs of Round 1 score and we will account for feedback treatment assignment in the analysis that follows.

Application Decision

Participants are told that they will soon have a chance to participate in a second round of ASVAB problem-solving. Again, they will have 5 minutes to answer up to 30 ASVAB questions, but are told that these questions will be more difficult on average than the questions in Round 1. In this way, Round 1 performance is predictive of Round 2 performance, but there is additional uncertainty due to the increased difficulty. If Round 2 is chosen for payment, they receive $0.20 per problem solved correctly.

Prior to completing Round 2, however, they have to make a decision about whether they want to apply for an "expert-level promotion". Participants are presented with two options:

"Option 1: Accept the novice job. If you choose this option and Round 2 is chosen for payment, you will get a Round 2 completion payment of $2 on top of the $0.20 per problem solved correctly in Round 2."

"Option 2: Apply for the expert-level job. If you choose this option **and** you are chosen to be promoted to the expert-level job, you will get a **promotion bonus** plus an extra $0.20 per problem solved correctly in Round 2, for a total of $0.40 per problem solved correctly. However, if you apply for the expert-level job and you are not promoted, you will only earn the $0.20 per problem solved correctly. You would not earn a Round 2 promotion bonus."

randomly assigned the same ad in both studies, but this was not particularly likely. They are then asked four questions about the ad:

1. *In general do you think the stereotype associated with this job is more female-typed or more male-typed? Use the slider scale below to indicate your answer, where -1 indicates extremely female-typed and 1 indicates extremely male-typed.*

2. *How prestigious would you say this job is? Use the slider scale below to indicate your answer, where 1 indicates not prestigious at all and 7 indicates extremely prestigious.*

3. *Thinking of typical Harvard undergraduates, how well-qualified do you think the average Harvard undergraduate would be for this job? Use the slider scale below to indicate your answer, where 1 indicates not at all qualified and 10 indicates extremely well-qualified.*

4. *Thinking of how the qualifications in the job advertisement were described, how specific, clear, and objective were the stated qualifications? Use the slider scale below to indicate your answer, where 1 indicates not at all clear and 10 indicates extremely specific, clear, and objective.*

We hypothesized that each of these measures could be relevant in predicting participant beliefs about how well-qualified they were. The first gets at the gender-stereotype associated with the job, speaking to the mechanism of Coffman (2014), who finds that individual self-confidence and willingness to put oneself forward is highly dependent on the gender congruency of the domain. If beliefs of own ability are a key driver in predicting beliefs of how well-qualified someone is, then past work, including Coffman (2014) and Bordalo et al. (2018) would suggest that the gender stereotype associated with the job could be an important factor; we would predict that as the maleness of the job posting increased, men would feel relatively more well-qualified (holding true qualification level fixed) while women would feel relatively less well-qualified. The second question allows us to try to separate out differences in the gender stereotype attached to the job from differences in the prestige of the position.

The third question accomplishes two main goals. First, it allows us to better account for variation across ads in how likely it is that any participant in our study feels qualified for that particular ad. Second, it allows us to ask how individuals vary in their beliefs about how well-qualified others are for various openings. Is it the case that individuals who feel poorly qualified believe others would also be poorly qualified for that opening (suggesting it was something about the job), or do they believe others would be well qualified for that openings (suggesting it was more something about themselves).

Finally, the fourth question speaks to the main mechanism we will test in our controlled experiments: whether the specificity and clarity of the stated qualifications matter. The hypothesis is that as qualifications become clearer, holding all else fixed, the gender gap will be reduced: when ads are ambiguous, there is more room for stereotypes and subjective beliefs to shape decisions.

In total 200 participants completed the two studies as part of bundle sessions at the Computer Lab for Experimental Research at Harvard Business School, of which 197 provided information on their gender.[3] The laboratory format we used bundles our study with short studies from other researchers. The placement of our studies within these other studies was varied across session, though note that the two parts of our study are always placed in the same order and appear consecutively.

Results

We provide summary statistics on our participants in Appendix Table B1.

We find that, on average, men view themselves as marginally more well-qualified than women in our sample. Recall that participants rate on a 1-10 scale how well-qualified they feel they are for each of four particular job ads. On average, men rate themselves a 4.65 (2.63 SD) while women rate themselves a 4.22 (2.51 SD). In Table 1, we present a regression that explores the determinants of these ratings, while controlling for ad fixed effects, order of appearance of our study within the session, and demographic controls. In Columns II and III, we include the average rating of the ad in terms of stereotype, prestige, objectivity of stated qualifications, and how well-qualified the typical Harvard undergraduate would be for the ad, as rated by the participants in the second part of the study, as additional controls. Controlling for ad and individual characteristics, we estimate that women rate themselves approximately 0.46 points less well-qualified for a given job opening on average ($p<0.10$).

In Column III, we interact the ad characteristics with the female dummy, to ask whether any of the characteristics of the job opening are differentially important for men and women. We find evidence in support of the main channel we explore in our follow-up experiments: more objectively stated job qualifications, as rated by our participants, help to close the gender gap in perceived qualification level. More objectively stated qualifications directionally decrease the extent to which men feel well-qualified, while this effect is significantly less pronounced for women, serving to decrease the average gender gap in perceived qualification level. None of the other job ad characteristics have predictive power for perceived qualifications or the gender gap in perceived qualifications.

Table 1. The Gender Gap in Perceived Qualification for Real Job Openings

[3] In addition, one participant did not provide data on their age; thus they are excluded from analyses that control for age. We obtain answers to a standard bank of demographic questions that are asked of all participants in bundle sessions by the laboratory.

OLS Predicting How Well-Qualified an Individual Feels for Job Opening			
	I	II	III
Female	-0.46* (0.24)	-0.46* (0.24)	-4.66 (3.01)
Average rating of Male Stereotype for this Ad		0.52 (3.43)	0.81 (3.45)
Average Rating of Prestige for this Ad		-0.59 (0.72)	-0.28 (0.76)
Average Rating of Objectivity of Stated Qualifications for this Ad		-0.38 (0.47)	-0.84 (0.53)
Average Belief of How Well-Qualified Undergrad Would be for this Ad		1.45 (1.26)	1.44 (1.25)
Female x Stereotype			-0.24 (0.85)
Female x Prestige			-0.53 (0.35)
Female x Objectivity			0.78** (0.30)
Female x Avg. Qualified			0.15 (0.26)
Ad Fixed Effects	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes
Order of Study within Session	Yes	Yes	Yes
R-squared	0.20	0.20	0.21
Clusters (Obs.)	196 (784)	196 (784)	196 (784)

Notes: Controls are fixed effects for each ad, fixed effects for each race category, fixed effects for each education category, age, a dummy for fluent in English, and a dummy indicating where our pair of studies fell within the session. * indicates $p<0.10$, ** indicates $p<0.05$, ***indicates $p<0.01$, ****indicates $p<0.001$.

Note that these results are very similar if we instead use as the dependent variable the fraction of qualifications that a participant believes they possess. On average, conditional on ad and individual characteristics, women believe they possess roughly 5pp fewer of the qualifications than men do (male average: 0.50, female average: 0.45, $p<0.05$ see Appendix Table B2). And, conditional on believing they have the same fraction of qualifications, men and women rate themselves as equally well-qualified (Appendix Table B3).

Interestingly, we also see significant gender differences in how individuals rate the average qualification level of a typical Harvard undergraduate for a given ad in Study 2. Men rate the typical Harvard undergraduate as being significantly more well-qualified than women do, providing respective ratings of 6.55 (SD 2.45) and 5.68 (2.57), respectively ($p<0.01$ in regression that clusters at individual level). This suggests that the gender gap in perceived qualification that we observe may also relate to gender differences in beliefs about how challenging it is to be qualified for a given job opening *for anyone*, not just themselves.

While this data provides interesting insights into perceptions of real job postings, it leaves open a number of important issues. In particular, we know very little about our participants in this setting. While we can observe differences in beliefs of how well-qualified they are, we cannot speak to differences in true underlying qualifications. To get at this key issue, we move to the controlled environment of an incentivized online experiment.

III. Evidence from a Controlled Experiment

We use a controlled experiment to better understand what drives participant beliefs about their qualifications, and to link these beliefs to incentivized decisions. In the setting below, we are able to observe the true ability (or qualification level) of each participant. We can then ask, given this ability, what their beliefs look like, about themselves and about their chances of receiving a given job conditional on applying. Importantly, we can also ask participants to make incentivized application decisions, asking whether there are gender differences in willingness to apply for a promotion.

Design

The general structure of the experiment is as follows. In the first part of the experiment, we aim to collect data on participant ability in a diagnostic test. We also elicit participant beliefs about their own performance. Then, we confront participants with a decision. They are asked to decide whether to apply for a promotion – a higher return opportunity – for Round 2 of the experiment. Then, participants complete Round 2. We detail each of these stages below.

Round 1

In Round 1 of the experiment, participants take an ability test that covers general science, arithmetic reasoning, math knowledge, mechanical comprehension, and assembling objects. The questions are drawn from the Armed Services Vocational Aptitude Battery (ASVAB). The questions were chosen for their simple multiple-choice format, and the fact that they cover stereotypically male domains. Participants have

5 minutes to answer up to 30 questions. All 30 questions appear on the same page and can be answered in any order. If Round 1 is chosen for payment, participants receive $0.20 per question answered correctly.

Beliefs of Round 1 Performance

After completing Round 1, each participant is asked to guess their score – that is, how many problems she solved correctly in Round 1 -- and how they rank relative to other MTurkers who are completing the HIT. They receive $0.10 if they guess their score correctly and $0.10 if they guess their bucket of rank correctly (bottom 5%, bottom 20%, bottom 40%, middle 20%, top 40%, top 20%, top 5%). They are then randomized into one of three feedback conditions: receiving either no feedback on Round 1 performance, a signal equal to their true score 60% of the time, or a signal equal to their true score 90% of the time. We then elicit posterior beliefs of Round 1 score from participants in each of the two noisy feedback conditions.

This noisy feedback intervention is the focus of a different paper, detailed in Coffman, Collis, and Kulkarni (2019). We will not focus on it in our analysis, though we will use data on posterior beliefs of Round 1 score and we will account for feedback treatment assignment in the analysis that follows.

Application Decision

Participants are told that they will soon have a chance to participate in a second round of ASVAB problem-solving. Again, they will have 5 minutes to answer up to 30 ASVAB questions, but are told that these questions will be more difficult on average than the questions in Round 1. In this way, Round 1 performance is predictive of Round 2 performance, but there is additional uncertainty due to the increased difficulty. If Round 2 is chosen for payment, they receive $0.20 per problem solved correctly.

Prior to completing Round 2, however, they have to make a decision about whether they want to apply for an "expert-level promotion". Participants are presented with two options:

"Option 1: Accept the novice job. If you choose this option and Round 2 is chosen for payment, you will get a Round 2 completion payment of $2 on top of the $0.20 per problem solved correctly in Round 2."

"Option 2: Apply for the expert-level job. If you choose this option **and** you are chosen to be promoted to the expert-level job, you will get a **promotion bonus** plus an extra $0.20 per problem solved correctly in Round 2, for a total of $0.40 per problem solved correctly. However, if you apply for the expert-level job and you are not promoted, you will only earn the $0.20 per problem solved correctly. You would not earn a Round 2 promotion bonus."

Participants complete a price list, informing us for different sizes of the promotion bonus ranging from $0 to $6 whether they prefer to apply for the expert-level job. Note that we build in an opportunity cost of applying for the promotion: a worker who chooses to apply for the expert-level job forgoes the $2 completion payment given for novice job applicants. Thus, a worker who applies for but does not receive the expert-level job earns less than a worker who simply accepts the novice job. This creates the incentive to only apply for the expert-level job if the worker believes she has sufficient probability of receiving it.

We wanted to structure the promotion opportunity in a way that mimicked some of the uncertainty about probability of receiving the promotion that would be present outside of the laboratory. In particular, we wanted the probability of receiving the promotion to both be tied to individual performance in Round 1 (mimicking the role of resume or prior experience), but also reliant on the discretion of the "employer".

To achieve this, we recruited 10 other MTurk workers in a separate study to make contingent decisions about which workers they would choose to promote. In particular, we explained the experimental setup described above to these MTurk participants and asked them to make a series of hiring decisions. We told them that they would be randomly paired with a worker who applied for the expert-level promotion in the subsequent study. If they chose to hire that worker, they would receive $0.25 for each problem solved correctly by that worker in Round 2. If not, they would receive a $1.25 fixed payment. These employers made a series of decisions about what workers they would be willing to hire. In particular, they are asked for a series of possible Round 1 scores (i.e. 3 problems solved correctly, 4 problems solved correctly, etc.), whether they would want to hire that worker. For instance, if the worker paired with them ended up having a Round 1 score of "7", would they want to hire that worker for the promotion?

We use these employer decisions to make promotion decisions for all workers who apply to the expert-level promotion. Workers who apply are divided evenly and randomly among the 10 employers and the employers' contingent decisions are used to determine whether they are promoted or not.[4] That is, suppose a worker has a Round 1 score of "7" and applies for the expert-level promotion. She would be randomly paired with and employer and we would look at whether that employer was willing to hire a worker with a Round 1 score of "7". If the employer was willing, she would be hired for the promotion. If the employer was not willing, she would not be hired. Both workers and employers have complete information on this process.

[4] And, one of the matched workers for each employer is randomly selected to determine the employer's payoffs.

Treatment Intervention

The use of a set of employers also allows us to identify and communicate a qualification for the promotion. We vary whether this qualification is communicated across two treatments.

In our control treatment, "No qualifications", we provide workers with a short information section prior to their application decision entitled, "Should I apply?". Within this section, we reinforce information on the potential returns to promotion – reminding them of the details of Round 2 and in particular that problems will be harder on average than in Round 1 – and on the chances of promotion – reminding them of the incentives the employers faced.

Participants in our "Qualifications" treatment receive the same language, but with one additional sentence that communicates the qualification: "While we can make no guarantees regarding your particular application, most employers indicated that they required a Round 1 score greater than 10 in order to be willing to promote a worker." In this way, we introduce a qualification in the form a rough threshold on Round 1 score that provides some additional guidance to workers as to what it might take to be promoted.

Workers then make their application decisions, indicating for each possible promotion bonus whether they would like to apply for the expert-level promotion or not. They are informed that if Round 2 is chosen for payment, one bonus level will be randomly chosen and their decision for that promotion bonus level will be implemented. However, they are not informed of the outcome of this randomization or whether they have been promoted until after their completion of the study.

Beliefs about Promotion

After completing their application decisions, we ask participants questions aimed at two potential mechanisms for explaining promotion decisions. The first question asks participants how many problems they expect to solve correctly in Round 2, allowing us to calculate what their believed returns to promotion are (conditional on being promoted). They receive $0.10 if they guess this score exactly correctly. The second question is unincentivized, and asks participants what they believe the probability is that someone

with their Round 1 score would be promoted (conditional on applying). This gets at their believed probability of promotion.

Round 2

Participants then complete the Round 2 problems, with 5 minutes to solve up to 30 more challenging ASVAB questions. If Round 2 is chosen for payment, they are compensated in accordance with their performance, decisions, and the decisions of the employer they are randomly paired with if they chose to apply for promotion.

Following Round 2, they answer brief demographic questions about themselves: gender, education level, race, and whether they attended high school in the United States. They then complete a series of risk preference questions. Finally, they answer two questions about their decisions on MTurk in general, indicating whether they are reluctant to have their payments *on MTurk specifically* depend on chance or on the decisions of other MTurkers. This allows us to speak to whether their application decisions in our study might be distorted by an MTurk-specific skepticism about having payments be less transparent.

Results

The experiment was conducted in May 2018 with 1,502 workers. Table 2 provides summary statistics on the workers in our study. We control for the demographic variables collected in the analysis that follows. Men outperform women on average in Round 1: 10.96 versus 9.65 problem solved correctly ($p<0.001$). Men on average rank in the 54^{th} percentile, while women rank in the 46^{th} percentile on average ($p<0.001$).

Conditional on Round 1 performance and demographic characteristics, women believe they performed significantly worse on average than men. A woman believes she scored 0.7 points worse than a man with the same score ($p<0.001$), and believes she places 7.2pp worse in the distribution of performers ($p<0.001$) (see Appendix Table B4, Columns I and II). This average gender gap in believed performance conditional on true score remains unchanged after the provision of noisy feedback, with women continuing to estimate scores that are 0.7 points worse than similarly able men ($p<0.001$) (see Appendix Table B4, Column III). These results are consistent with Coffman (2014), Bordalo et al. (2018) and others, who find that women's beliefs about own ability are more pessimistic than men's in male-typed fields, even conditional on true ability.

Table 2. Summary Statistics on Workers

	Men	**Women**	**P-value**
White	0.80	0.81	0.65
Black	0.06	0.09	0.08
Asian	0.10	0.06	0.01
Attended HS in US	0.98	0.96	0.05
HS Only	0.11	0.085	0.06
Some College/Assoc.	0.36	0.37	0.86
Bachelors	0.39	0.40	0.76
Advanced Degree	0.14	0.15	0.36
Rd. 1 Score	10.96	9.65	<0.001
Rd. 2 Score	8.44	7.14	<0.001
Qualifications Treatment	0.49	0.50	0.78
60% Feedback	0.33	0.31	0.35
90% Feedback	0.32	0.35	0.20
N	798	704	

Notes: p-values from binary variables are from two-tailed test of proportions. Comparisons of Round 1 and Round 2 scores use two-tailed t-tests.

Women and men enter the application stage with different beliefs about their capabilities, even conditional on having the same performance. Therefore, if beliefs about performance factor into application decisions (through shaping beliefs about probability of receiving the promotion), we have reason to expect a gender gap in application rates. Furthermore, this belief gap carries through into beliefs about Round 2 performance. When we ask participants to forecast their Round 2 performance, women estimate that they will solve approximately 0.6 problems fewer than men, conditional on having the same Round 1 performance ($p<0.001$, see Appendix Table B4, Column IV). Because of this, believed returns to promotion also vary by gender, another reason why we might expect a gender gap in application rates.

Application Decisions

How do these beliefs translate into participants' application decisions? We start by exploring our control treatment, where participants receive very little guidance on where "the bar" is for promotion. We find that women's beliefs of their probability of being promoted (conditional on applying) are significantly lower than men's. Women believe they have a 39% chance of being promoted on average, while men believe they have a 48% chance of being promoted. Even once we condition on Round 1 performance, women believe they are significantly less likely to be promoted (Table 3, Column I, 7pp, $p<0.001$).

Table 3. Gender Differences in Believed Probability of Promotion

OLS Predicting Believed Probability of Promotion *No Qualifications Treatment*					
	I	II	III	IV	V
Female	-7.17**** (1.64)	-5.54**** (1.59)	-3.14** (1.41)	-7.24**** (1.63)	-2.55* (1.39)
Round 1 Score	1.78**** (0.18)	0.34 (0.27)	0.55**** (0.17)	1.78**** (0.18)	-0.18 (0.24)
Prior Belief of Absolute Ability		0.89** (0.39)			0.20 (0.34)
Posterior Belief of Absolute Ability		0.72 (0.45)			0.53 (0.39)
Belief of Round 2 Score		0.87*** (0.34)			0.69** (0.29)
Prior Belief of Relative Ability (percentile rank)			50.7**** (2.96)		47.1**** (3.01)
Took Common Risk Gamble				-2.84* (1.59)	-2.56* (1.33)
Feedback Treatment Controls	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes
R-squared	0.19	0.26	0.42	0.20	0.44
N	759	759	759	759	759

Notes: Demographic controls are fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US. Feedback treatment controls are dummies for each feedback treatment (no signal, 60% signal, 90% signal). * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.

Next, we include beliefs as explanatory variables. First, we control just for beliefs about own absolute ability (Column II). While these have predictive power, they do not significantly reduce the gender gap in believed probability of promotion, which remains 6pp. If we include instead beliefs of relative ability, we see more explanatory power, suggesting beliefs of relative ability capture something important about this decision (Column III). This is worth noting, given that this isn't a competitive environment. It may be that beliefs of relative ability are better measures of feeling "good enough" than beliefs about absolute ability. Finally, we include a risk preference measure, including a dummy for whether participants chose a 50-50 gamble over $4.80 and 0 over a $2 fixed payment for sure[5]; we find that more risk-averse individuals if

[5] This is a risk preference question all participants are asked. It is a rough approximation of the average risk in the promotion decision. The $4.80 is approximately the average expected return to promotion at the median promotion bonus level (a $3 bonus, plus an extra $0.20 for each of the Round 2 problems solved (which we estimated at an average of 9, although in practice it was closer to 8)). Participants also answer a series of personalized risk preference questions, that feed in the information they provide on their expected number of Round 2 problems solved and believed probability of promotion to create objective risk preference questions that parallel the risk we estimate they face in

anything view themselves as more likely to be promoted, but that this is orthogonal to the gender gap. If we include all of these variables, we estimate that women continue to believe they are approximately 3pp less likely to be promoted (p=0.06).

This residual gap could perhaps be suggestive of the view that men and women have different beliefs of where the bar is, with women believing employers are likely to have higher standards for promoting someone.

Next we consider our key behavioral outcome: willingness to apply for promotion at different wages. Figure 1 presents a simple bar graph showing the fraction of men and women who apply at each wage, as well as the fraction of men and women who never apply. Note that conditional on applying, the average minimum wage at which men and women apply is nearly identical: 264 cents for men and 260 cents for women. But, significantly more women than men choose to never apply (22% of women and 16% of men, p=0.04). From this point forward, we will code the decision to never apply as a minimum willingness to accept of 650 cents, 50 cents more than the maximum promotion bonus we offered. With this coding, the average min. wage required to induce a man to apply is 325 cents, while for women it is 344 cents (p=0.21).

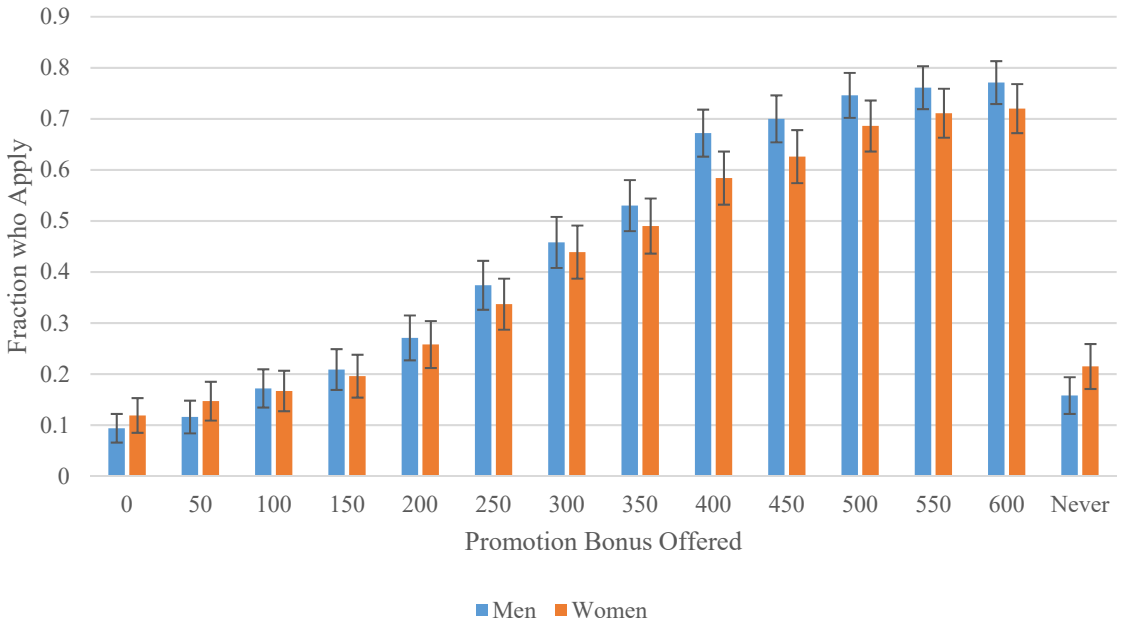

Figure 1. Willingness to Apply for Promotion by Gender and Promotion Bonus Size

deciding whether to apply for promotion. Given that these are a function of participant-specific inputs, they are not as useful as common controls in a regression, and given the lack of predictive power of risk preference for the decision, we choose not to analyze them further.

Table 4 replicates Table 3 above, showing that when we switch from believed probability of probability to minimum promotion bonus at which someone was willing to apply for promotion, there is no significant gender gap.[6] This suggests that other factors are predicting the behavioral decision to apply at different promotion bonus levels that were not relevant when we simply asked for believed probability of being promoted.

What could these factors be? Given our setting, the role for preferences for promotion seems limited. All workers must complete the same Round 2 problems independent of their decisions, so a preference for more challenging work cannot play a role. Additionally, decisions are quite anonymous, so social norms or image concerns surrounding applying for promotion seem unlikely to play a large role. One factor that does seem to matter is risk preferences. While risk preferences do not predict beliefs of receiving the promotion, they do predict application decisions. More risk tolerant participants being much more willing to apply for promotion at lower promotion bonus levels. Including risk preference as a control (see Column IV) does not change the estimated gender gap.

We collected data on a few other arguably less interesting possibilities. For instance, 29% of men and 25% of women reported that they would **never** choose to have their payment on MTurk depend upon the decisions of someone else if they had a choice, suggesting they would be highly reluctant to apply for promotion at any price, for reasons independent of our study. When asked on a 1-7 scale how reluctant they would be to have their payment depend upon chance or the decisions of others, the average response is similar for men and women (4.4 and 4.5, respectively). Both of these measures are indeed predictive of application decisions, although their inclusion does not change the estimated gender gap.

In sum, when there are no clearly stated qualifications for promotion, women believe they are significantly less likely to be promoted than men are, conditional on applying. This gender gap is partially explained by beliefs of ability, and in particular relative ability. When we look at willingness to apply, we estimate that, conditional on ability, women are directionally less willing to apply (they demand larger promotion bonuses in order to be willing to apply), but the gender gap is not significant. It may be that risk preferences and other context-specific factors play a large role in shaping application decisions, overshadowing the role for believed probability of promotion.

[6] The two measures are correlated: in the No Qualifications Treatment, minimum bonus at which a participant applied for promotion and believed probability of promotion have a negative correlation of 0.32 on average. Note that the correlation is stronger for women than men (0.38 and 0.25, respectively). Similarly, if we add believed probability of promotion to any of the specifications of Table 4, it is a highly significant predictor of decisions.

Table 4. Willingness to Apply for Promotion

OLS Predicting Minimum Acceptable Promotion Bonus *No Qualifications Treatment*						
	I	II	III	IV	V	VI
Female	9.57 (15.4)	2.30 (15.4)	4.07 (15.6)	8.00 (15.2)	-2.82 (15.4)	-2.63 (15.4)
R1 Score	-9.51**** (1.71)	-3.13 (2.63)	-7.84**** (1.89)	-9.53**** (1.69)	-2.12 (2.62)	-2.02 (2.61)
Prior Belief of Absolute Ability		0.77 (3.77)			2.53 (3.75)	3.31 (3.74)
Posterior Belief of Absolute Ability		-5.49 (4.40)			-6.79 (4.34)	-7.70* (4.35)
Belief of Round 2 Score		-6.06* (3.28)			-5.97* (3.23)	-5.56* (3.23)
Prior Belief of Relative Ability			-69.0** (32.9)		-46.8 (33.4)	-40.4 (33.5)
Took Common Risk Gamble				-69.0**** (14.8)	-71.7**** (14.7)	-66.2**** (14.7)
Reluctance to have payoff depend on others						11.1** (4.55)
Never want payoff dependent on others						20.5 (18.0)
Feedback Treatment Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.06	0.08	0.07	0.09	0.11	0.13
N	759	759	759	759	759	758

Notes: Demographic controls are fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US. Feedback treatment controls are dummies for each feedback treatment (no signal, 60% signal, 90% signal). * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.

Do Qualifications Help?

Our evidence from real job ads suggested that more clearly stated qualifications might help to reduce the gender gap in willingness to apply. We implement this in our experiment and test it, randomly assigning half of our participants to receive additional information on where "the bar" is. We look at how this change impacts believed probability of promotion and willingness to apply for promotion.

We find that, on average, these stated qualifications significantly reduce men's beliefs about their chances of being promoted (by approximately 3.5pp). For women, however, stated qualifications directionally increase their beliefs of being promoted (by approximately 1.2pp) (see Table 5, Column I). Note that this is true independent of whether we control for individual beliefs of own ability and risk preferences, suggesting that the mechanism is indeed operating through communicating where "the bar" is (see Column II). Thus, overall, stated qualifications significantly reduce the gender gap in believed probability of promotion.

Table 5. The Impact of Clearly Stated Qualifications on Believed Probability of Promotion

OLS Predicting Believed Probability of Promotion						
	All Participants		Unqualified Participants (Round 1 score < 10)	Unqualified Participants (Believed Round 1 score < 10)	Qualified Participants (Round 1 score ≥10)	Qualified Participants (Believed Round 1 score ≥10)
	I	II	III	IV	V	VI
Female	-6.96**** (1.62)	-2.30* (1.38)	-2.14 (2.80)	-2.18 (1.71)	-2.49 (1.92)	-3.64 (2.34)
Qualification Treatment	-3.49** (1.56)	-3.53*** (1.32)	-4.40** (2.04)	-5.53*** (1.76)	-2.06 (1.71)	-0.22 (1.97)
Female x Qualification Treatment	4.65** (2.28)	3.85** (1.92)	2.14 (2.78)	2.93 (2.42)	5.40** (2.62)	6.12* (3.20)
Round 1 Score	2.00**** (0.13)	-0.13 (0.17)	-0.54 (0.42)	-0.07 (0.23)	-0.29 (0.25)	-0.29 (0.26)
Belief and Risk Preference Controls	No	Yes	Yes	Yes	Yes	Yes
Feedback Treatment Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.20	0.26	0.31	0.32	0.42	0.36
N	1502	1502	706	920	796	582

Notes: Demographic controls are fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US. Feedback treatment controls are dummies for each feedback treatment (no signal, 60% signal, 90% signal). * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.

Of course, we might expect heterogeneous effects depending upon whether a participant actually possesses the desired qualification (or believes she possesses the desired qualification). So, in Columns III – VI, we

split the sample across unqualified and qualified by participants. We do this both by actual score (i.e. truly qualified or not, Columns III and V), and believed Round 1 score (i.e. believes themselves to be qualified or not, Columns IV and VI). In each of these specifications, we control for beliefs of absolute and relative ability, and risk preferences of our participants.

When we split the sample into those who are unqualified according to the stated qualification and those who are qualified, we see an interesting pattern. As we might expect, the qualifications treatment reduces the believed probability of being promoted significantly among the unqualified group, but still directionally more so from men than women (see Columns III and IV). Among qualified participants, the reduction in the gender gap achieved by the stated qualifications is significant. Men's beliefs are directionally lower when there are clearly stated qualifications, while qualified women's beliefs increase significantly, eliminating the gender gap completely.

But, once again, when we turn to the behavioral measure of the minimum promotion bonus at which a participant was willing to apply for promotion, these results do not hold. Table 6 replicates Table 5, but using the behavioral dependent variable. Here, we see in general no significant impact of the stated qualifications treatment, and no significant gender differences in responsiveness on average. The estimated effects are much noisier relative to Table 5, and even the directional patterns are not clear.

What can we make of these results? We have a number of key findings: given the same Round 1 performance, women have more pessimistic beliefs of their own performance in Round 1, both in terms of absolute and relative performance. These gender differences in beliefs also manifest in expectations of the future: women believe they will solve fewer problems in Round 2 than men do, reducing the perceived returns to being promoted.

When no clearly stated qualifications are given for promotion, we find that women believe they have a significantly lower chance of being promoted than men. This is true even conditional on measured performance and measured beliefs about performance. Adding more objective information on the qualifications required for promotion help to reduce this gap. In particular, clearly stated qualifications reduce men's beliefs about their probability of being promoted, particularly unqualified men, while directionally boosting women's perceived chances, particularly qualified women. In this way, qualifications are effective at reducing observed gender gaps in believed chances of success.

However, this does not translate into significant differences in application behavior. Application decisions in our experiment, while correlated with believed probability of promotion, are also predicted by other factors. While some of these other factors might be externally relevant, such as risk preferences, others seem much less likely to be so, such as worries about having others' determine their payoffs on MTurk. To

address this issue, we move to the field, where we can provide a third test of our hypothesis in a real job search setting.

Table 6. The Impact of Clearly Stated Qualifications on Willingness to Apply

OLS Predicting Minimum Acceptable Promotion Bonus						
	All Participants		Unqualified Participants (Round 1 score < 10)	Unqualified Participants (Believed Round 1 score < 10)	Qualified Participants (Round 1 score ≥10)	Qualified Participants (Believed Round 1 score ≥10)
	I	II	III	IV	V	VI
Female	8.76 (15.1)	-6.20 (14.9)	-5.13 (23.7)	1.79 (20.3)	-10.8 (18.8)	-22.8 (21.5)
Qualification Treatment	-6.36 (14.6)	-7.13 (14.2)	-5.46 (24.4)	8.47 (20.7)	-9.90 (16.7)	-30.5* (18.2)
Female x Qualification Treatment	6.25 (21.2)	9.36 (20.7)	12.9 (33.6)	-12.6 (28.5)	7.52 (25.6)	50.4* (29.4)
Round 1 Score	-9.93**** (1.21)	-3.86** (1.84)	0.25 (5.05)	-4.48* (2.68)	-3.27 (2.44)	-3.04 (2.40)
Belief and Risk Preference Controls	No	Yes	Yes	Yes	Yes	Yes
Feedback Treatment Controls	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.064	0.12	0.085	0.095	0.12	0.11
N	1502	1498	703	917	795	581

Notes: Demographic controls are fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US. Feedback treatment controls are dummies for each feedback treatment (no signal, 60% signal, 90% signal). * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.

IV. Using Qualifications in the Field

In this section, we report the results of a field experiment testing the idea that more clearly stated qualifications can help to attract more talented female applicants. We ran a field experiment on an online employment platform called "Upwork." Upwork (previously Elance-oDesk) is the largest global freelancing website (Upwork n.d.). In a nutshell, Upwork facilitates match-making between freelance workers and potential employers. To implement our study, we act as employers on Upwork, posting job advertisements, inviting a pool of workers to view and apply to our ads, and then tracking application rates. We make job offers to the most qualified workers that apply to each ad, and provide them the opportunity

to complete the job for the advertised pay. Freelancers are unaware of their participation in an experiment at the time that they make the decision of whether or not to apply to the job opening.

We start by providing a bit of institutional context on Upwork. Freelancers who register with Upwork can advertise their services by creating a profile. This profile is publicly available and can be searched for and viewed on the Upwork website. A profile can include the following information: the freelancer's first name and last initial, photo, state of residence, hourly rate, self-reported education, self-reported skills, self-reported work experience, number of jobs completed on Upwork, hours worked on Upwork, reviews from previous Upwork employers, and availability status.

In addition, freelancers have the opportunity to take standardized tests of their skills and aptitudes in different domains. Upwork offers hundreds of so called "Skills Tests" for free with the topics of those tests ranging from Adobe to XML ("Skills Tests" n.d.). Upworkers are encouraged to take as many tests as they would like, especially those which are linked to their advertised job category, and have the option to retake a test after 180 days. For each test, Upwork provides information on the number of freelancers who are qualified for that test and the number of freelancers who have already taken this test. These tests essentially serve as verified evaluations of capabilities, and freelancers have the option of displaying the results of these tests on their profile.

We take advantage of these skills tests in the design and implementation of our experiment. We started by identifying all available Upworkers that are residents of the United States and have displayed on their profile either the results of the Management Skills Test (Wave 1 of experiment) or the Analytical Skills Test (Wave 2 of experiment). This gives us a pool of workers that have completed a test of interest. We then compiled the profile information for each of the Upworkers in this pool, creating a dataset with a wealth of information about each worker. We attempt to capture all commonly available profile features, including posted hourly rate, state, hours worked on Upwork, jobs completed on Upwork, indicator of currently available, measure of current availability (more than 30 hours/wk, less than 30, as needed), education level (indicators for profile listed a college degree, an MBA, or another graduate degree), a set of indicators for listing skills in different job categories,[7] and the total number of tests they have chosen to display on their profile. On top of that, we enter into the dataset an indicator of our best guess of freelancer gender, using both the first name and photograph when available to make a determination.[8]

[7]Upwork assigns each job to one of the following categories: Web/Mobile/Software Development, IT & Networking, Data Science & Analytics, Engineering & Architecture, Design & Creative, Writing, Translation, Legal, Administrative Support, Customer Service, Sales & Marketing, and Accounting & Consulting. Note that each of these categories has up to 83,000 sub-categories. We captured the freelancer's capabilities at the category level.

[8] Gender determinations were done by one member of the research team, prior to treatment assignment, and double-checked in the final dataset by another member of the research team.

The dataset also contains the freelancer's score on the test of interest (either Management Skills or Analytical Skills) on a normalized 1 – 5 scale. This is a score computed by Upwork, but workers have discretion over whether to display their score. Only workers who choose to display their score appear in our dataset. We also record the number of minutes it took the freelancer to complete the test, made available by Upwork alongside the worker score.

Having created this dataset, we reach out to each worker in our dataset via Upwork, inviting them to apply to our positions. Every invitation contains the following information. Freelancers are informed of two jobs. One job is an "intermediate level job" while the other job represents the more challenging but also better compensated job, the "expert level job". Both types of jobs (intermediate and expert) require writing essay-style answers to two questions and are advertised to take one hour. We offer pay of $70 for the intermediate level job and of $150 for the expert level job.

All participants are presented with both options and are invited to apply to either of the two jobs (the worker can choose either job to apply to, but are told that they can apply to no more than one). All freelancers receive generic information on desired characteristics of a successful applicant for the expert job that reads as follows: "We are looking for candidates with [management expertise / experience in analytical thinking], as demonstrated through education, past work experience, and test scores. Successful applicants will also have strong writing and communication skills."

Each worker is randomly assigned to one of three treatments that vary in the additional level of information that is provided about the desired qualifications for the expert level job. In our *control* treatment, workers are provided with no additional information on the desired qualifications. In our *normative* treatment, freelancers are provided with a prescriptive statement about whether to apply for the expert level job. The job description states that "we invite applicants with a [Management / Analytical] Skills test score of [3.75 / 4.05] to apply for the expert-level job." In our *positive* treatment, freelancers are provided with a descriptive statement about the desired qualifications. The job description states that "we expect that most successful applicants to the expert-level job will have a [Management / Analytical] Skills test score above [3.75 / 4.05]".

A few features of our design are worth noting. First, we reach out to workers rather than simply post the jobs in order to boost response rates, increasing the extent to which our ads are visible to workers and ensuring random assignment to treatment. Each freelancer in our dataset is able to view and apply to a job for exactly one of the three treatments. Second, we chose this design with two jobs because we worried that by directly contacting workers and inviting them to apply, we might already be "de-biasing workers" – our invitation alone might suggest to workers that indeed they are qualified for our opening. To remedy this, we use two jobs, an intermediate level job and an expert level job, and use the decision to apply to the

expert level job, rather than the intermediate level job, as our outcome of interest. In this way, even if we are signaling to workers that they are likely a good fit for one of our positions because of our invitation, it is still the case that they face a less obvious decision about whether to apply to the expert level or intermediate level job. Finally, we had to make a discretionary decision about what the right test score qualification was for our study. We choose to use scores within each test sample that are challenging to achieve (just under 25% of our participants have a test score at or above the stated qualification), but still allow for a reasonable sample size of participants who are "qualified" according to our test score qualification.

Freelancers who were interested in our positions were able to contact us to apply through the Upwork website. We then made hiring decisions using a pre-determined algorithm. We computed a "hiring score" ranging from 0 to 100 for each worker that was a function of the desired qualifications communicated to them within the job advertisement, assigning a weighted score based upon their experience (100 points if they completed any job on Upwork, 0 points if they have no Upwork experience, weight: 10%), education (as indicated by degrees held, 0 points for no stated education, 60 points for completed College education, 80 points for a Masters degree, 90 points for an MBA degree, 100 points for an MBA and another graduate degree, weight: 20%), and test score on the test of interest (their skills test score converted into a 100 point scale, weight: 70%). We made job offers to the two workers with the best hiring scores for each posting (two intermediate offers and two expert offers within each treatment, for each wave, for a total of 24 offers).[9] Note that only workers who applied to the expert-level job were eligible for the expert-level job; we selected the best two hiring scores within the set of workers who applied to each particular posting.

Results

We start by providing some descriptive statistics of the freelancers in our study. Men and women vary in many dimensions in our sample: men post greater hourly rates (in line with work by Dubey et al. (2017) and Foong et al. (2018) and are more likely to advertise skills in Web Development, IT, Data Science, Engineering, Design, and Accounting. Women, on the other hand, have more experience on Upwork and are more likely to advertise Writing skills, Administrative Support skills, and Customer Service skills. We note, however, that we should likely interpret these skills advertisements

[9] Freelancers who receive job offers are simultaneously told of the study and offered the opportunity to withdraw their data. We had no freelancers request removal from the study; 20 of the 24 workers we made offers to accepted the job and completed it for pay.

carefully; as Murciano-Goroff (2018) documents, women are less likely to post skills on resumes in the tech domain, even given the same level of experience and skill.

Table 7. Summary statistics for Upwork Freelancers in our Dataset

	Men	**Women**	
Requested Hourly Rate	44.0	30.8	p<0.001
Hours Worked on Upwork	321	559	p=0.07
Jobs Worked on Upwork	13.6	15.6	p=0.47
Total Tests Displayed	6.2	7.3	p=0.002
Available Less than 30hrs/wk	0.18	0.22	p=0.13
Available More than 30hrs/wk	0.43	0.41	p=0.41
Available as Needed	0.37	0.35	p=0.46
College Degree	0.73	0.73	p=0.76
MBA Degree	0.14	0.08	p=0.001
Other Graduate Degree	0.20	0.21	p=0.59
Web/Mobile/Software Development	0.20	0.08	p<0.001
IT & Networking	0.08	0.005	p<0.001
Data Science & Analytics	0.18	0.11	p=0.001
Engineering & Architecture	0.04	0.01	p<0.001
Design & Creative	0.19	0.15	p=0.09
Writing	0.32	0.45	p<0.001
Translation	0.05	0.06	p=0.41
Legal	0.05	0.05	p=0.84
Administrative Support	0.25	0.49	p<0.001
Customer Service	0.04	0.11	p<0.001
Sales & Marketing	0.15	0.16	p=0.80
Accounting & Consulting	0.22	0.13	p<0.001
Analytical Skills Score	3.73	3.58	p<0.001
Time Taken on Analytical Test (minutes)	50.4	47.9	p=0.07
Management Skills Score	3.55	3.42	p<0.001
Time Taken on Management Test (minutes)	19.7	20.7	p=0.09
Proportion Qualified by Test Score	0.29	0.19	p<0.001
Proportion in Analytical Skills Dataset	0.42	0.45	p=0.39
N	545	558	

Men outperform women on average in both qualification tests – the Management Skills test and the Analytical Skills test. And, a greater fraction of men than women are qualified for our expert level job according to their test score (i.e. have a test score greater than or equal to the stated test score threshold).

Overall, 20% of men and 19% of women in our sample apply to one of our job postings.[10] This aggregate rate is relatively constant across the three treatments, with 19% of men and women applying in the control, 21% of men and 19% of women applying in the Positive treatment, and 20% of women and 17% of women applying in the Normative treatment.

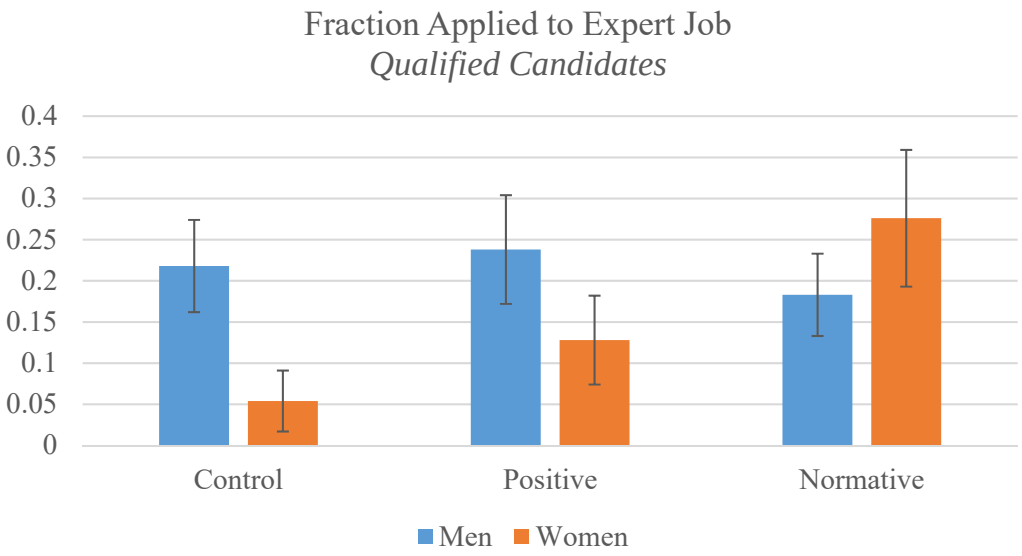

Figure 2. Fraction of Qualified Applicants that Apply to Expert Job

However, the key question is how application rates to the expert job vary by treatment and by qualification level. We first consider the rates of application to the expert job among qualified applicants – those applicants who have a test score at least as high as the stated threshold. Figure 2 demonstrates the results by gender and treatment. We see that the fraction of qualified men who apply to the expert level job is quite steady across treatment, ranging from 18 – 24%. The application decisions of qualified women, however, vary more widely across treatment, with the lowest rate of application in the control – 5% -- and the highest rate of application in the normative treatment – 28%. In terms of these raw means, we see suggestive

[10] Of the 212 participants who apply to our position, 15 do not apply to strictly one job. For all participants who either (i) fail to specify which of the two jobs they wish to apply to, or (ii) explicitly apply to both jobs, our research team contacted them after their initial application and asked them to clarify which job they were choosing to apply to. The 15 participants who we code as "applying to both jobs" either did not reply to this request or explicitly refused to clarify. Our analysis codes these workers as having applied to the expert job and as having applied to the intermediate job.

evidence that qualified women respond positively to more clearly stated qualifications in a way that qualified men do not.

How do more clearly stated qualifications impact unqualified applicants? Application rates to the expert level job are quite low across for this sub-population across all treatments, for both men and women. Again, men's application rates are quite flat across the treatments. The rate at which women apply to the expert-level job when they are unqualified is highest in the control and lowest in the normative treatment.

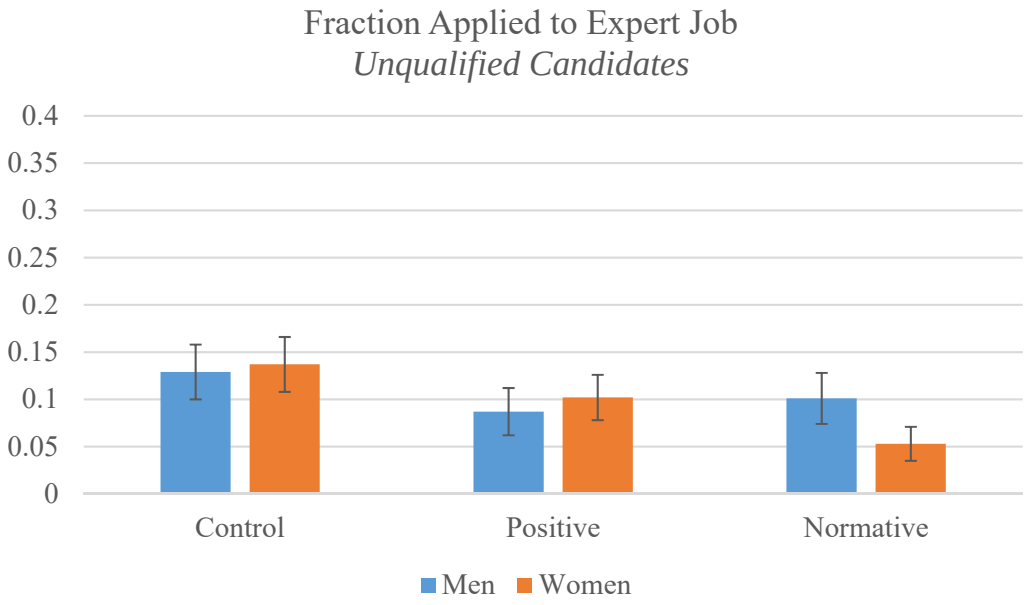

Figure 3. Fraction of Unqualified Applicants that Apply to Expert Job

In Table 7, we use regression analysis to explore these questions formally. In Columns I – III, we focus on qualified applicants (those with test scores greater than or equal to the advertised threshold). We predict the decision to apply to the expert level job from treatment assignment, using the control treatment as our baseline. We control for all profile information included in our summary statistics table, excluding the self-reported skills.[11] Consistent with Figure 3, we estimate that our treatments have no impact on qualified men's decisions. Qualified men are equally likely to apply to the expert level job independent of how clearly stated the desired qualifications are. Women's decisions, however, do vary by treatment. Qualified women are 12pp more likely to apply in our positive treatment relative to the control ($p=0.16$), and 23pp more likely to apply in our normative treatment relative to the control ($p<0.05$).

[11] Results are unchanged if we exclude these dummies.

In Columns IV – VI, we analyze the decisions of unqualified applicants. Again, we estimate that men's decisions are not impacted by our treatments. For women, we estimate that, relative to the control treatment, the normative treatment deters applications from unqualified women by approximately 9pp ($p<0.05$). However, in an interacted model, we cannot reject that the deterrence effect is of a similar size for men and women.

Table 7. Application Rates to Expert Level Job

OLS Predicting Decision to Apply to Expert Job						
	I	II	III	IV	V	VI
	Qualified Men	Qualified Women	All Qualified	Unqualified Men	Unqualified Women	All Unqualified
Positive Treatment	-0.006 (0.088)	0.12 (0.084)	0.022 (0.080)	-0.055 (0.039)	-0.039 (0.0345)	-0.051 (0.038)
Normative Treatment	-0.010 (0.082)	0.23** (0.094)	0.0003 (0.074)	-0.036 (0.039)	-0.086** (0.035)	-0.040 (0.037)
Female			-0.20** (0.086)			0.008 (0.038)
Female x Positive			0.11 (0.12)			0.016 (0.051)
Female x Normative			0.28** (0.12)			-0.043 (0.051)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.16	0.32	0.14	0.11	0.082	0.072
N	157	105	262	388	453	841

Notes: Qualified applicants are those with a test score greater than or equal to the advertised threshold. Controls are posted hourly rate, hours worked, jobs worked, total tests posted, normalized test score, time taken to complete the test, college degree dummy, MBA dummy, other graduate degree dummy, dummies for each category of availability (less than 30 hrs/wk, more than 30 hrs/wk, as needed), dummies for each self-reported skill, and a dummy for being in the second wave of experiment.

Figure 4 allows us to consider how the distribution of applications varies across treatments for qualified applicants. In particular, we consider the total number of applications received for each posting from qualified applicants, both to the intermediate level job and the expert level job. For men, we see the total number of applicants is nearly identical across treatment, and within each treatment, most qualified men apply to the expert level job. For women, introducing qualifications increases the total number of applicants. And, in the normative treatment, a larger share of qualified women who apply choose to apply to the expert level job rather than the intermediate level job. More clearly stated qualifications seem to help sort women, but not men.

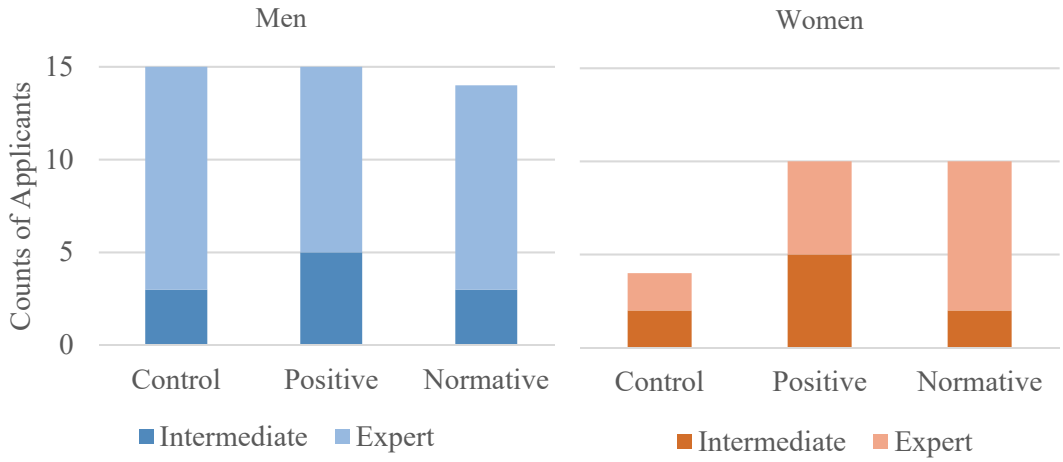

Figure 4. Distribution of Applications among Qualified Candidates

Because of the treatment effects for qualified female applicants, the qualifications serve to increase the total number of qualified applicants to our expert job. The share of women in the pool of qualified candidates also increases. This can be seen in Figure 5, which simply counts the total number of applicants to the expert job, both by qualified applicants (Panel a) and unqualified applicants (Panel b). Adding qualifications directionally grows the pool of qualified candidates, disproportionately drawing in qualified women. At the same time, we receive fewer applications from unqualified applicants to the expert level job. Thus, from a hiring perspective, the impact of the qualifications on our potential pool seem quite positive: a larger, more diverse pool of qualified applicants, and fewer unqualified applicants.

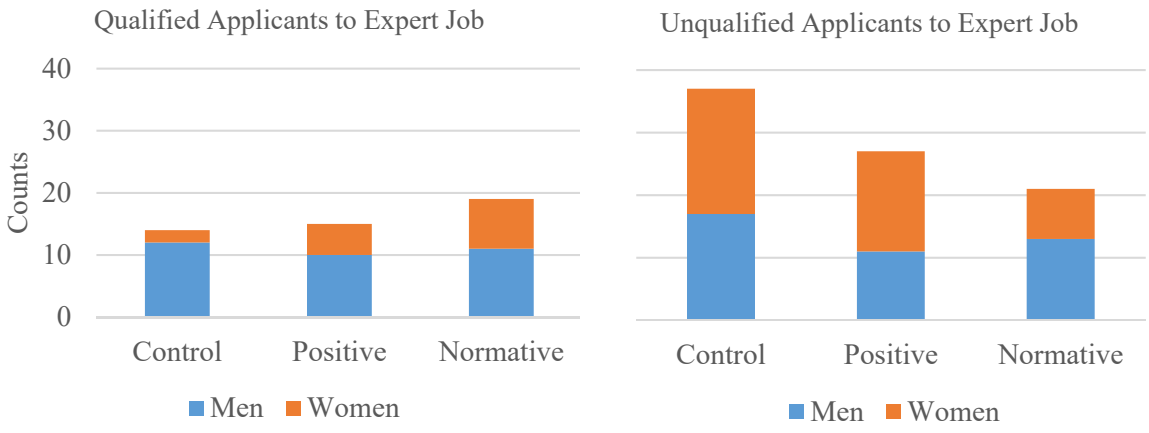

Figure 5. Total Applications to Expert Level Job

V. Discussion

A large literature explores the factors that contribute to gender gaps in labor market outcomes. Within this rich literature, however, supply-side decisions focused on when individuals choose to put themselves forward for different opportunities are understudied. This paper tackles this important question, asking whether there are gender differences in application decisions.

Across three complementary contexts, we explore the extent to which men and women view themselves as qualified for a given opportunity. We find convincing evidence for two distinct components of these judgements. First, as shown in previous work, women view themselves as less capable than equally talented men in male-typed domains. Second, holding fixed their belief about their own ability, women and men seem to have different beliefs about what "the bar" is. That is, women seem to believe it is more challenging *for any particular person* to be qualified for a given opportunity. This novel finding informs the policy intervention we test in the laboratory and in the field.

We show that exogenously providing clearer, more objective information about the required qualifications for a position can help to close the gender gap. This information intervention has heterogeneous effects. Among women, more clearly stated qualifications help to better sort applicants: truly qualified women become more likely to apply, while unqualified women are marginally less willing to apply. Among men, clearly stated qualifications have no significant impact on application behavior. In our field setting, the result of the information intervention is a larger and more diverse pool of qualified applicants.

Our results suggest that there may be soft touch employer interventions that can improve the diversity of their applicant pool, even if candidate beliefs about own ability are unchanged. This seems like a promising and low-cost path to explore.

VI. References

"2019 Salesforce Equal Pay Update." 2019. Salesforce Blog. April 2, 2019. https://www.salesforce.com/blog/2019/04/equal-pay-update-2019.html.

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Chapter 48 Race and Gender in the Labor Market." In *Handbook of Labor Economics*, 3:3143–3259. Elsevier. https://doi.org/10.1016/S1573-4463(99)30039-0.

Bohnet, Iris, Alexandra van Geen, and Max Bazerman. 2016. "When Performance Trumps Gender Bias: Joint vs. Separate Evaluation." *Management Science* 62 (5): 1225–34. https://doi.org/10.1287/mnsc.2015.2186.

Bordalo, Pedro, Katherine B. Coffman, Nicola Gennaioli, and Andrei Shleifer. 2018. "Beliefs about Gender." *American Economic Review* forthcoming.

Coffman, Katherine B. 2014. "Evidence on Self-Stereotyping and the Contribution of Ideas." *The Quarterly Journal of Economics* 129 (4): 1625–60. https://doi.org/10.1093/qje/qju023.

Coffman, Katherine B., Manuela R. Collis, and Leena Kulkarni. 2018. "Gender and Belief Updating." *Working Paper*.

Coffman, Katherine B, Christine L. Exley, and Muriel Niederle. 2018. "When Gender Discrimination Is Not about Gender." *Working Paper*.

Del Carpio, Lucia and Maria Guadalupe. 2018. "More Women in Tech? Evidence from a Field Experiment addressing Social Identity." *Working Paper.*

Dubey, Alpana, Kumar Abhinav, Mary Hamilton, and Alex Kass. 2017. "Analyzing Gender Pay Gap in Freelancing Marketplace." In *Proceedings of the 2017 ACM SIGMIS Conference on Computers and People Research*, 13–19. SIGMIS-CPR '17. New York, NY, USA: ACM. https://doi.org/10.1145/3084381.3084402.

Exley, Christine and Judd Kessler. 2019. "The Gender Gap in Self-Promotion." Working paper.

Fernandez, Roberto M., and Marie Louise Mors. 2008. "Competing for Jobs: Labor Queues and Gender Sorting in the Hiring Process." *Social Science Research* 37 (4): 1061–80. https://doi.org/10.1016/j.ssresearch.2007.10.003.

Flory, Jeffrey A., Andreas Leibbrandt, and John A. List. 2015. "Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions." *The Review of Economic Studies* 82 (1): 122–55. https://doi.org/10.1093/restud/rdu030.

Foong, Eureka, Nicholas Vincent, Brent Hecht, and Elizabeth M. Gerber. 2018. "Women (Still) Ask For Less: Gender Differences in Hourly Rate in an Online Labor Marketplace." *Proc. ACM Hum.-Comput. Interact.* 2 (CSCW): 53:1–53:21. https://doi.org/10.1145/3274322.

Gee, Laura K. 2018. "The More You Know: Information Effects on Job Application Rates in a Large Field Experiment." *Management Science* 65 (5): 2077–94. https://doi.org/10.1287/mnsc.2017.2994.

Ginther, Donna K., and Shulamit Kahn. 2009. “Does Science Promote Women? Evidence from Academia 1973 - 2001.” In *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, edited by Richard B. Freeman and Daniel L. Goroff. A National Bureau of Economic Research Conference Report. Chicago: University of Chicago Press.

Heilman, Madeline E. 2001. “Description and Prescription: How Gender Stereotypes Prevent Women’s Ascent Up the Organizational Ladder.” *Journal of Social Issues* 57 (4): 657–74. https://doi.org/10.1111/0022-4537.00234.

Ibarra, Herminia, Nancy M. Carter, and Christine Silva. 2010. “Why Men Still Get More Promotions Than Women.” *Harvard Business Review*, September 1, 2010. https://hbr.org/2010/09/why-men-still-get-more-promotions-than-women.

Murciano-Goroff, Raviv. 2018. “Missing Women in Tech: The Labor Market for Highly Skilled Software Engineers.” *Working Paper*, October, 65.

Niederle, Muriel, and Lise Vesterlund. 2007. “Do Women Shy Away From Competition? Do Men Compete Too Much?” *The Quarterly Journal of Economics* 122 (3): 1067–1101. https://doi.org/10.1162/qjec.122.3.1067.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. “How Stereotypes Impair Women’s Careers in Science.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (12): 4403–8.

Riach, P. A., and J. Rich. 2002. “Field Experiments of Discrimination in the Market Place.” *The Economic Journal* 112 (483): F480–518.

Samek, Anya Savikhin. 2015. “A University-Wide Field Experiment on Gender Differences in Job Entry Decisions.” *Working Paper*. https://doi.org/10.2139/ssrn.2579257.

“Skills Tests.” n.d. Upwork Help Center. Accessed February 4, 2019. http://support.upwork.com/hc/en-us/articles/211063198-Skills-Tests.

Upwork. n.d. “Upwork.” Accessed December 5, 2018. https://www.upwork.com.

Zahidi, Saadia, and Herminia Ibarra. 2010. “The Corporate Gender Gap Report 2010.” World Econonomic Forum.

A. Appendix

Experiment Materials (under separate cover):

https://drive.google.com/file/d/17XGTRn-Ukle-gG8CRtsRbUfPmCaO-y_P/view?usp=sharing

B. Appendix

Table B1. Summary Statistics for Study 1 and Study 2

	Men	Women	P-value from test of proportions
White	0.28	0.32	0.51
Black	0.16	0.11	0.34
Asian	0.35	0.34	0.93
Latino	0.08	0.07	0.90
Multiracial	0.12	0.09	0.48
Middle East	0.01	0.01	0.89
Is a Student	0.80	0.87	0.22
Average Age	23.93	23.81	0.78
Highest obtained Education			
High School	0.12	0.05	0.05
Some College	0.30	0.29	0.80
Bachelor's Degree	0.36	0.49	0.06
Advanced Degree	0.21	0.18	0.51
Is fluent in English	0.99	0.99	0.89
Order of Study within Session	0.48	0.50	0.81

Table B2. The Gender Gap in Perceived Qualification for Real Job Openings

OLS Predicting the Fraction of Qualifications that a Participant Believes They Possess			
	I	II	III
Female	-5.52** (2.61)	-5.52** (2.61)	-54.30* (32.46)
Average rating of Male Stereotype for this Ad		32.68 (38.15)	33.19 (38.46)
Average Rating of Prestige for this Ad		-9.83 (7.97)	-7.45 (8.28)
Average Rating of Objectivity of Stated Qualifications for this Ad		2.15 (0.67)	-1.56 (5.61)
Average Belief of How Well-Qualified Undergrad Would be for this Ad		22.33 (14.73)	21.06 (14.78)
Fem x Stereotype			1.11 (9.31)
Fem x Prestige			-4.04 (3.79)
Fem x Objectivity			6.25* (3.47)
Fem x Avg. Qualified			3.50 (2.74)
Ad Fixed Effects	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes
Order of Study within Session	Yes	Yes	Yes
R-squared	0.19	0.19	0.19
Clusters (Obs.)	196 (784)	196 (784)	196 (784)

Notes: Controls are fixed effects for each ad, fixed effects for each race category, fixed effects for each education category, age, a dummy for fluent in English, and a dummy indicating where our pair of studies fell within the session. * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.

Table B3. The Gender Gap in Perceived Qualification for Real Job Openings

OLS Predicting How Well-Qualified an Individual Feels for Job Opening	
	I
Female	-0.01 (0.11)
Fraction of Qualifications that a Participant Believes They Possess	0.08**** (0.00)
Ad Fixed Effects	Yes
Demographic Controls	Yes
Order of Study within Session	Yes
R-squared	0.80
Clusters (Obs.)	196 (784)

Notes: Controls are fixed effects for each ad, fixed effects for each race category, fixed effects for each education category, age, a dummy for fluent in English, and a dummy indicating where our pair of studies fell within the session. * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.

Table B4. Believed Performance by Gender

OLS Predicting Believed Performance				
	I Pre-Signal Belief of Absolute Score	II Pre-Signal Belief of Relative Performance	III Post-Signal Belief of Absolute Score	IV Post-Signal Belief of Round 2 Score
Female	-0.70**** (0.16)	-0.07**** (0.01)	-0.72**** (0.15)	-0.58**** (0.15)
True Round 1 Score	0.59**** (0.02)	0.02**** (0.00)	0.76**** (0.02)	0.46**** (0.02)
Demographic Controls	Yes	Yes	Yes	Yes
Signal Treatment	No	No	Yes	Yes
Qualification Treatment	No	No	No	Yes
R-squared	0.45	0.23	0.64	0.38
Observations	1,502	1,502	1,502	1,502

Notes: Controls are their true Round 1 Score, fixed effects for each race category, fixed effects for each education category, a dummy for attendance of High School in the US, fixed effects for the Signal treatment versions (60 versus 90 percent to see true Round 1 Score as signal), fixed effects for Qualification treatment versions (no, vague, or clearly stated qualifications). * indicates $p<0.10$, ** indicates $p<0.05$, ***indicates $p<0.01$, ****indicates $p<0.001$.

Table B5. Application rates excluding the 15 observations who applied to both jobs

OLS Predicting Decision to Apply to Expert Job						
	I	II	III	IV	V	VI
	Qualified Men	Qualified Women	All Qualified	Unqualified Men	Unqualified Women	All Unqualified
Positive Treatment	0.007	0.10	0.043	-0.064*	-0.033	-0.063*
	(0.087)	(0.081)	(0.080)	(0.037)	(0.033)	(0.036)
Normative Treatment	0.013	0.25***	0.028	-0.022	-0.075**	-0.026
	(0.080)	(0.091)	(0.073)	(0.037)	(0.033)	(0.035)
Female			-0.17*			0.0004
			(0.084)			(0.036)
Female x Positive			0.068			0.035
			(0.12)			(0.049)
Female x Normative			0.25**			-0.044
			(0.12)			(0.049)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.17	0.34	0.14	0.12	0.085	0.077
N	154	104	258	383	447	830

Notes: Excludes individuals who applied to both postings. Qualified applicants are those with a test score greater than or equal to the advertised threshold. Controls are posted hourly rate, hours worked, jobs worked, total tests posted, normalized test score, time taken to complete the test, college degree dummy, MBA dummy, other graduate degree dummy, dummies for each category of availability (less than 30 hrs/wk, more than 30 hrs/wk, as needed), dummies for each self-reported skill, and a dummy for being in the second wave of experiment.