WHEN TO APPLY?

Katherine B. Coffman*

Manuela R. Collis

Leena Kulkarni

January 2021

Abstract: Labor market outcomes depend, in part, upon an individual's willingness to put herself forward for different opportunities. We use laboratory and field experiments to explore gender differences in willingness to apply for higher return, more challenging work. We find that, in male-typed domains, women view themselves as less qualified for a given opportunity, both because of differences in beliefs about own ability and in beliefs about where the bar is. We provide evidence that reducing ambiguity surrounding required qualifications reduces the gender gap in willingness to apply among qualified applicants, increasing both the diversity and talent 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 2020). 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 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 and training to pursue, which jobs to apply to, and when to put themselves up for promotion. Gender differences at these crucial decision nodes could be a 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 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 opportunities, and how this impacts their decision about 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.

[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).

Outside of controlled experiments, many of these decisions about when to apply may be made in the face of (anticipated) bias, making it hard to isolate the role for willingness to apply 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 candidate perceptions and decisions, absent employer biases. We ask whether women are as likely as men to see themselves as qualified for challenging, higher-paying positions, and whether they apply at similar rates conditional on their degree of qualification.

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, both in objective terms (Niederle and Vesterlund 2007, Coffman 2014; Bordalo et al. 2019) and subjective terms (Exley and Kessler 2020). In the field, Murciano-Goroff (2020) finds that conditional on having the same level of skill, female software engineers are less likely to self-report that skill on their resume compared to men. This suggests that even if men and women both have the same skills and 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), as could differences in preferences for more challenging work (Niederle and Yestrumskas 2008).

A few clever field experiments have explored some factors that impact male and female job-seekers' probability of application. 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, in a field experiment with a high-skilled population, Samek (2019) finds that competitive compensation schemes deter women more so than men. 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 desirability of 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.

Closest to our work is a simultaneous project by Abraham and Stein (2020) exploring how the language used in job postings impacts the application behavior of men and women. In a large randomized control trial, they vary the language around how demanding and intense the required qualifications for a given position are. In particular, their treatment "softens qualifications," removing optional qualifications from the posting and using less demanding language for remaining qualifications. They find that when qualifications are softened, more individuals apply, and it reduces the skills gap between male and female applicants. That is, in comparison to the control treatment, it is no longer case that the female applicants that apply are significantly more skilled than the male applicants.

Our paper builds on this body of important work by attempting to understand better the decision of whether to apply through a series of three controlled experiments. Our experiments span different subject populations – Ivy League students, Amazon Mechanical Turk workers, and job candidates seeking work on an employment website – and different experimental approaches. Through our use of mixed populations and mixed methods, we aim to identify phenomena and policy prescriptions that may be generalizable to a range of real world contexts of interest.

Our first goal is to explore 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 undergraduate students in the laboratory 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 controlled experiment on Amazon Mechanical Turk, where we can directly measure the relevant talent of each individual in a simulated labor market environment, and ask them incentivized questions about their believed abilities.

Across these settings, we find significant gender differences in two different types of beliefs. First, women hold more pessimistic beliefs about their own abilities in male-typed domains. Second, we find evidence of gender differences in beliefs about what "the bar" is for a given opportunity. 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. While the first factor has been well-documented in previous work, the second factor is a novel and important finding.

We then ask how these differences in beliefs translate into application decisions. In both our controlled Amazon Mechanical Turk experiment and in a separate field experiment with an employment website, we create "promotion" opportunities to which participants can apply. We find that in our controlled online setting, qualified women are directionally less willing than qualified men to apply for these opportunities, though the effect is not significant. Within our field experiment on an employment website, qualified women are significantly less likely to apply than qualified men.

Finally, we turn to policy solutions. 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 reduce ambiguity around what is needed in order to receive the promotion, and we ask whether this helps to reduce the gender gap in beliefs and application rates among qualified applicants.[2]

We find that both the gender gap in perceived qualification and the gender gap in application rates among qualified applicants are reduced when the desired qualifications for the opportunity are less ambiguous.[3] In particular, clearer qualifications help to reduce gender gaps in perceived qualification level, in believed chances of receiving the position and, in our field setting, in application rates to the position among qualified applicants. In our field experiment, we find that very few qualified women apply for an advanced job in a baseline condition. Once we provide more information on expectations of what is needed to receive the advanced job, the gender gap is reduced, creating a larger, more gender diverse pool of qualified applicants.

II. Conceptual Model of the Application Decision

In this section, we build a simple conceptual model of the phenomenon we study. We consider an individual's decision of whether or not to apply for a position – such as a new job or a promotion. Denote the payoff to applying and receiving the position (i.e., salary or recognition) as A_1, the payoff to applying and not receiving the position as A_0, the cost of applying as c, and the perceived probability of receiving the position conditional on applying as P. Define the payoff to not applying, the outside option, as A.[4] Then, a risk neutral individual will apply for the promotion if and only if:

$$P(A_1) + (1 - P)(A_0) - c \geq A$$

This simple expression illustrates several factors that could drive a gender difference in willingness to apply. First, there could be gender differences in real or perceived payoffs or costs. For instance, it could be the case that men expect a new opportunity to be more lucrative, women face higher costs of applying, or women have better outside options. Second, because of the uncertainty around receiving the position, more risk averse individuals will be less likely to apply, holding all else fixed. If, consistent with past literature, women are more risk averse than men, this could lead to a gender difference in willingness to apply (Croson and Gneezy 2009, Niederle 2016). While these channels merit further study, they are not the main focus of our project. In our controlled experiment, we either hold fixed or measure payoffs, costs, and risk preferences, allowing us to account for any role they play in explaining behavior.

[2] This is in contrast to softening the qualifications, a la Abraham and Stein (2020); rather, we introduce increased clarity and specificity to the qualifications.

[3] This finding parallels the evidence from the negotiation literature showing that reducing situational ambiguity reduces gender differences in negotiation outcomes (Bowles, Babcock, and McGinn 2005).

[4] In many cases, we may have that $A_0 = A$, consistent with a rejected applicant simply returning to their outside option. But, we allow for the case where applying but not receiving the position changes the payoff relative to the outside option; for instance, a candidate may have to give up their current position in order to apply for a new one.

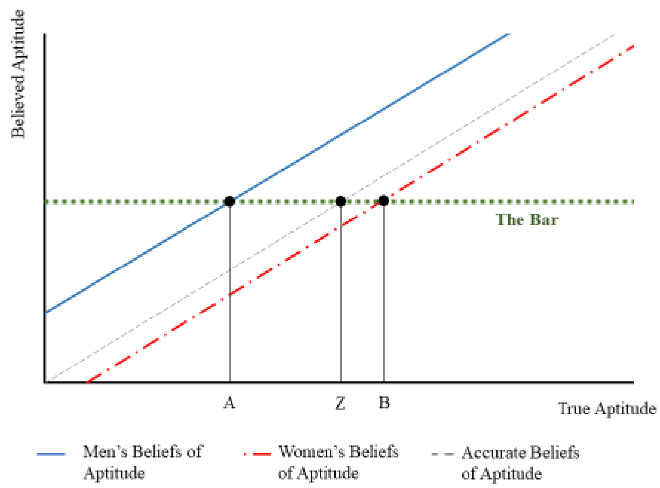
Instead, we will focus on *P,* the perceived probability of receiving the position, conditional on applying. We argue that this perceived probability is a function of two distinct beliefs. The first is the individual's belief of her own position-specific aptitude. The second is the individual's belief of what the bar is; that is, what is the aptitude level that is required to receive the position. Our assumption is that *P* is weakly increasing in the difference between the two, that is, the extent to which one's own perceived aptitude exceeds the perceived bar.

In this framework, gender differences in *P,* and corresponding application choices, can emerge from two different types of belief gaps. To make this concrete, we choose a simple example and illustrate each channel in Figure 1, Panels (a) and (b). For the figure, assume P=1 if perceived own aptitude is weakly greater than the bar, and P=0 otherwise, and that payoffs and costs are such that all individuals will choose to apply if and only if P>0.

The first channel is differences in beliefs of own aptitude, illustrated in Panel (a). If there is a gender difference in beliefs of own aptitude (conditional on true aptitude), then holding all else fixed, this can produce a gender gap in the believed probability of receiving the position among similarly well-qualified men and women. In our figure, the x-axis plots true aptitude for the position, while the y-axis plots self-perceived aptitude. We illustrate a gender gap in perceived abilities, such that conditional on true ability, men are overconfident and women are underconfident. To focus on the role of differences in beliefs of own aptitude, we assume that both men and women correctly forecast where the bar is for the position, illustrated as the horizontal dashed green line. With this pattern of beliefs, men with true aptitudes between points A and Z will apply despite not being qualified: while their self-perceived talents are above the bar, their true talents are not. Women with true aptitudes between Z and B will choose not to apply, despite being qualified. For all aptitudes between A and B, we observe a gender gap in application decisions: men apply while women do not.

The second type of belief gap that can generate a gender gap in application decisions stems from beliefs about where the bar is. We illustrate this in Panel (B) of Figure 1. To focus on the role of beliefs about the bar, here we assume that both men and women hold accurate beliefs about their own aptitudes. But, women perceive the bar to be higher than it truly is (illustrated as the double-dashed orange line), while men perceive it to be lower than it truly is. In this case, men with true aptitudes between X and Z will apply despite not being qualified, while women with true talents between Z and Y will not apply, despite being qualified. This again produces a gender gap: for all aptitudes between X and Y, men apply while women do not.

Panel (a): Gender Differences in Beliefs of Own Aptitude



Panel (b): Gender Differences in Beliefs of Where The Bar Is

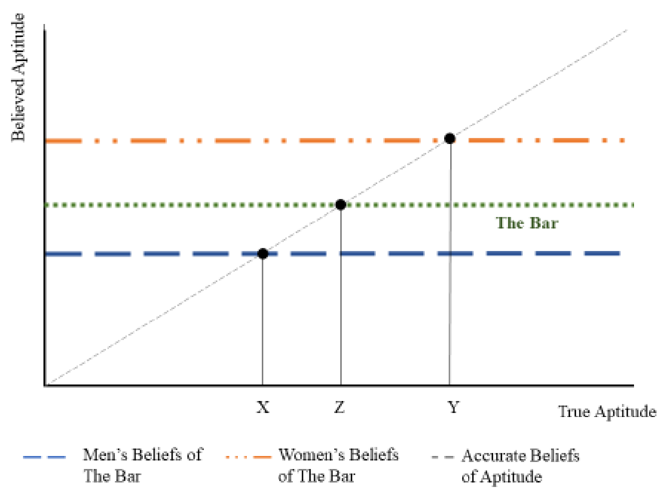


Figure 1. How Beliefs Can Drive Gender Gaps in Application Rates

While these are just two simplified cases, they help to illustrate the basic premise of our analysis. Our goal going forward is to document what these beliefs look like, how they translate into application decisions, and whether there are effective policy levers for reducing gender gaps driven by differences in beliefs. In particular, motivated by the case in Panel (B) of Figure 1, we will explore how interventions aimed at providing improved information about where the bar is can help to close the gender gap.

We assert that gender differences in beliefs of where the bar is are most likely to arise in cases where there is ambiguity. When it is not clear where the bar is, there is more room for interpretation, generating the possibility that different individuals will reach different conclusions. In our experiments, we will explore what happens when this ambiguity is reduced. Note that better revealing where the bar is will have different impacts on different candidates, depending upon their beliefs about themselves. Among candidates whose believed aptitude is below the revealed bar, we expect that clearer information on the bar should not increase their willingness to apply: the information simply reveals to them that their perceived aptitude is below the bar, if they did not already think that. But, among workers whose believed ability is above the revealed bar, the information should weakly increase willingness to apply.

Returning to Figure 1, Panel (B), suppose information about the bar succeeded in bringing both men's and women's beliefs of the bar closer to the true bar. We would see that those unqualified men between X and Z should now be less likely to apply, while the qualified women between Z and Y should now be more likely to apply. As a result, the gender gap in application rates should narrow.

III. Motivating Evidence from Impressions of Real Job Advertisements

Methods

Our goal in our first experiment 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. Our participants were provided by the Computer Lab for Experimental Research (CLER) at Harvard Business School, and were primarily Harvard undergraduates.

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 (13 cases). In addition, a coding error omitted 9 entry-level ads from the study (no participants were randomly assigned to view them). This left us with 28 job ads: 4 entry-level and 24 mid-level. Three examples of ads used in the experiment are illustrated in Figure B5, Figure B6, and Figure B7 in the Appendix. More samples are available in the Online Materials Appendix.

In the first part of the experiment, 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 28.[5]

We limit participants' time viewing the ads for two reasons. First, we wanted to ensure timely completion of the experiment, 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" by re-reading the ad when asked questions. We give 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 provide answers based upon their holistic (but maybe imperfect) 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. *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. The third question encourages participants to reflect on the qualifications for this particular ad. In Appendix B, Figures B1 – B4, we present some illustrative evidence on the types of responses we receive to this question, with a focus on gender differences.

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] 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 experiment, participants view one randomly-selected ad from a given bucket, ensuring that no participant sees the same ad twice in one part of the experiment.

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 ads in the experiment. 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 experiment, we collect additional data from *the same set of* participants about each of these 28 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 experiment is nearly identical to the first. Participants complete four rounds. Within each round, they are given 2 minutes to view one randomly-selected ad from the 28. Note that this randomization operates independently from the randomization in the first part; thus, participants could be randomly assigned the same ad in both experiments, 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 dependent on the gender congruence of the domain. If beliefs of own ability are a key driver in predicting beliefs of how well-qualified someone is, 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 sample feels qualified for that particular ad. This is important given how heterogeneous the various postings are. 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 opening (suggesting it was more something about themselves)?

Finally, the fourth question speaks to the main mechanism we will test in our other controlled experiments: does the amount of ambiguity surrounding the desired qualifications matter? We think of increased objectivity, increased clarity, and increased specificity as reducing the amount of ambiguity around where the bar is. The hypothesis is that as qualifications become less ambiguous, holding all else fixed, the gender gap in how well-qualified individuals believe they are will be reduced, in line with our conceptual model.

Results

In total 200 participants completed the two experiments as part of bundle sessions at the Computer Lab for Experimental Research at Harvard Business School, of which 197 provided information on their gender.[6] We provide summary statistics on our participants in Appendix Table B1. Most of our participants are college students. In Appendix Table B2, we provide summary statistics on our job ads, including major industry sector.[7]

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). This gender gap is approximately 9% of the mean of how well-qualified individuals feel. 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 experiments within the session, and demographic controls. To increase interpretability, we create z-scores for the variables that were elicited with a scale. Controlling for

[6] 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 (non-mandatory) demographic questions that are asked of all participants in “bundle sessions” in the laboratory. This laboratory format we used bundles our project with short projects from other researchers. These bundle sessions are administered by the laboratory and target 200 participants in a single week. The placement of our experiments with respect to these other projects was varied across session, though note that the two parts of our project are always placed in the same order (Experiment 1 and then Experiment 2, as described above) and appear consecutively, with no other projects in between.

[7] We assign each ad to a Bureau of Labor Statistics major industry sector, including: educational services, financial activities, health care and social assistance, information, leisure and hospitality, manufacturing, professional and business services, state and local government, and transportation and warehousing.

these ad fixed effects and demographics, we estimate that women rate themselves approximately 0.18 standard deviations less qualified than men (p<0.10, Column I). In Column II, we include more information about the ads. In particular, we use the assessments provided by our participants in the second part of the experiment, as to stereotype, prestige, objectivity of qualifications, and average qualification level of a typical Harvard undergraduate. For each ad, we take the average of the ratings provided by all raters who saw that ad for each characteristic.[8] Then, we take the z-score to capture where this particular ad falls relative to the full set of ads on that characteristic. We also include a dummy for whether the ad was for an entry-level position, and dummies for the major industry sector of the ad. Controlling for these ad characteristics does not have a large impact on our estimate of the gender gap (Column II).

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

	OLS Predicting How Well-Qualified an Individual Feels for Job Opening (z-score)			OLS Predicting What Percentage of Qualifications an Individual Believes She Possesses (0 – 100 scale)		
	I	II	III	IV	V	VI
Female	-0.18*	-0.21**	-0.22**	-5.52**	-6.59**	-6.70**
	(0.093)	(0.095)	(0.095)	(2.61)	(2.65)	(2.64)
Average rating of Male Stereotype for this Ad (z-score)		-0.25***	-0.23**		-8.78***	-8.58***
		(0.081)	(0.090)		(2.24)	(2.46)
Average Rating of Prestige for this Ad (z-score)		0.090	0.15**		2.73*	4.21**
		(0.055)	(0.071)		(1.49)	(1.84)
Average Rating of Objectivity of Stated Qualifications for this Ad (z-score)		-0.17***	-0.27***		-4.36***	-6.38***
		(0.051)	(0.069)		(1.51)	(2.05)
Average Belief of How Well-Qualified Average Undergrad Would be for this Ad (z-score)		0.048	0.028		1.10	-0.16
		(0.064)	(0.077)		(1.76)	(2.14)

[8] We note that there are no significant differences in how men and women rate these ads on average in terms of stereotype, prestige, or objectivity of qualifications. And, across ads, the average male and average female ratings are highly correlated along each of these dimensions.

Female x Avg. Male Stereotype (z-score)			-0.034			-0.45
			(0.068)			(1.94)
Female x Avg. Prestige (z-score)			-0.11			-2.53
			(0.072)			(1.99)
Female x Avg. Objectivity (z-score)			0.17**			3.54
			(0.077)			(2.26)
Female x Avg. of Avg. Qualified (z-score)			0.030			2.15
			(0.079)			(2.15)
Entry Level Opening Dummy		0.34***	0.34***		6.86**	6.96**
		(0.12)	(0.12)		(3.24)	(3.24)
Ad Fixed Effects	Yes	No	No	Yes	No	No
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Order of Experiment within Session	Yes	Yes	Yes	Yes	Yes	Yes
Observations (Clusters)	784 (196)	784 (196)	784 (196)	784 (196)	784 (196)	784 (196)
Adjusted R-squared	0.155	0.136	0.139	0.139	0.112	0.113

Notes: Controls are fixed effects for each ad in Columns I and IV, 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 experiments fell within the session. In columns without ad fixed effects, we include dummies for major industry sector. * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.

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 other 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: we estimate that a one standard deviation increase in how objectively stated the qualifications are reduces men's perceived qualification by 0.27 standard deviations. This effect is significantly less pronounced for women, serving to decrease the average gender gap in perceived qualification level.[9]

[9] While not a focus of our project, one could ask how our two decoy questions – what the expected salary would be and how appealing the opening is – predict how well-qualified an individual feels. In particular, one could ask whether differential expectations of wages lead to differences in feelings of how well-qualified one is, or application behavior (Neal and Johnson (1996)). We trim the salary data to exclude the bottom 5% and top 5% of estimates. We

These results are very similar when we use the fraction of qualifications that a participant believes they possess (Columns IV – VI). On average, conditional on ad and individual characteristics, women believe they possess roughly 5pp fewer of the qualifications than men do (male average: 50%, female average: 45%, $p<0.05$).Again, as desired qualifications become more objective, men assess themselves to be less well-qualified on average; this effect is directionally weaker for women ($p=0.12$ on the interaction, Column VI). Conditional on believing they have the same fraction of qualifications, men and women rate themselves as equally well-qualified (Appendix Table B3).

Interestingly, we see significant gender differences in how individuals rate the average qualification level of a typical Harvard undergraduate for a given ad in Experiment 2. Men rate the typical Harvard undergraduate as being significantly more well-qualified than women do. Men believe the average student at Harvard would have a qualification rating of 6.55 (SD 2.45) on average, while women report the average student would have a rating of just 5.68 (SD 2.57), $p<0.01$ in a regression that clusters at individual level. When an individual provides an assessment of how qualified she, herself, is for a given opening, this assessment is a function of her beliefs about her own abilities and also her beliefs about where the bar is. When we instead ask individuals to assess how well-qualified the average Harvard undergraduate would be for a given opening, individual beliefs about own aptitudes should be much less relevant. The fact that we see a gender gap in assessing the average qualification level for others suggests that men and women may have differing beliefs about where the bar is. We will dig into this more in our next two experiments.

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 the actual qualification levels of 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. We also do not observe incentivized application decisions. To move forward on these important issues, we move to the controlled environment of an incentivized online experiment.

IV. Evidence from a Controlled Experiment

Overview of Design

find that (i) there are no gender differences in wage expectations in our data, and (ii) salary estimates are not a significant predictor of how well-qualified an individual feels. But, individuals do indicate feeling significantly more qualified for more appealing openings: a one standard deviation increase in the appeal of the opening corresponds to a 0.6 standard deviation increase in how well-qualified an individual feels. Neither of these effects vary by gender.

We conduct our experiment on Amazon Mechanical Turk (MTurk).[10] The general structure of the experiment is as follows. Before we run the main experiment, which we call the "worker experiment," we recruit 10 MTurk employers to make contingent hiring decisions. We use these contingent decisions to later incentivize worker decisions, and to inform the crafting of qualifications in the main worker experiment. Next, we move to the "worker experiment." In Round 1 of the worker experiment, we collect data on participant aptitude in a diagnostic test and elicit participant beliefs about their aptitudes. Then, we confront workers with a decision; they are asked to decide whether to apply for a promotion for Round 2 of the experiment. We also directly elicit their perceived probability of promotion conditional on applying. Then, participants complete Round 2. Finally, on the back end, after the completion of the "worker experiment," the researchers use random matching to allocate workers from the main experiment to employers from the preliminary employer experiment to determine outcomes and payoffs. We provide a visual overview of the "worker experiment" in Figure 2. This is followed by a detailed description of each stage.

Round 1

In Round 1 of the worker experiment, participants take an assessment test that covers general science, arithmetic reasoning, math knowledge, mechanical comprehension, and assembling objects. We draw the questions from the Armed Services Vocational Aptitude Battery (ASVAB). These questions have a simple multiple-choice format, several of the categories we use are rather difficult to quickly "Google" answers to, and they cover stereotypically male-typed domains. We choose to study male-typed domains to better proxy the male-typed environments that have been historically characterized by female under-representation and lack of advancement outside of the laboratory (i.e. business, STEM). 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.

---

[10] The study was restricted to workers with a United States based IP address who had completed at least 100 tasks (called Human Intelligence Tasks, or HITs) and had an approval rating by previous MTurk requesters of at least 95%. The study contains understanding questions and a participant must answer those understanding questions correctly in order to complete the study.

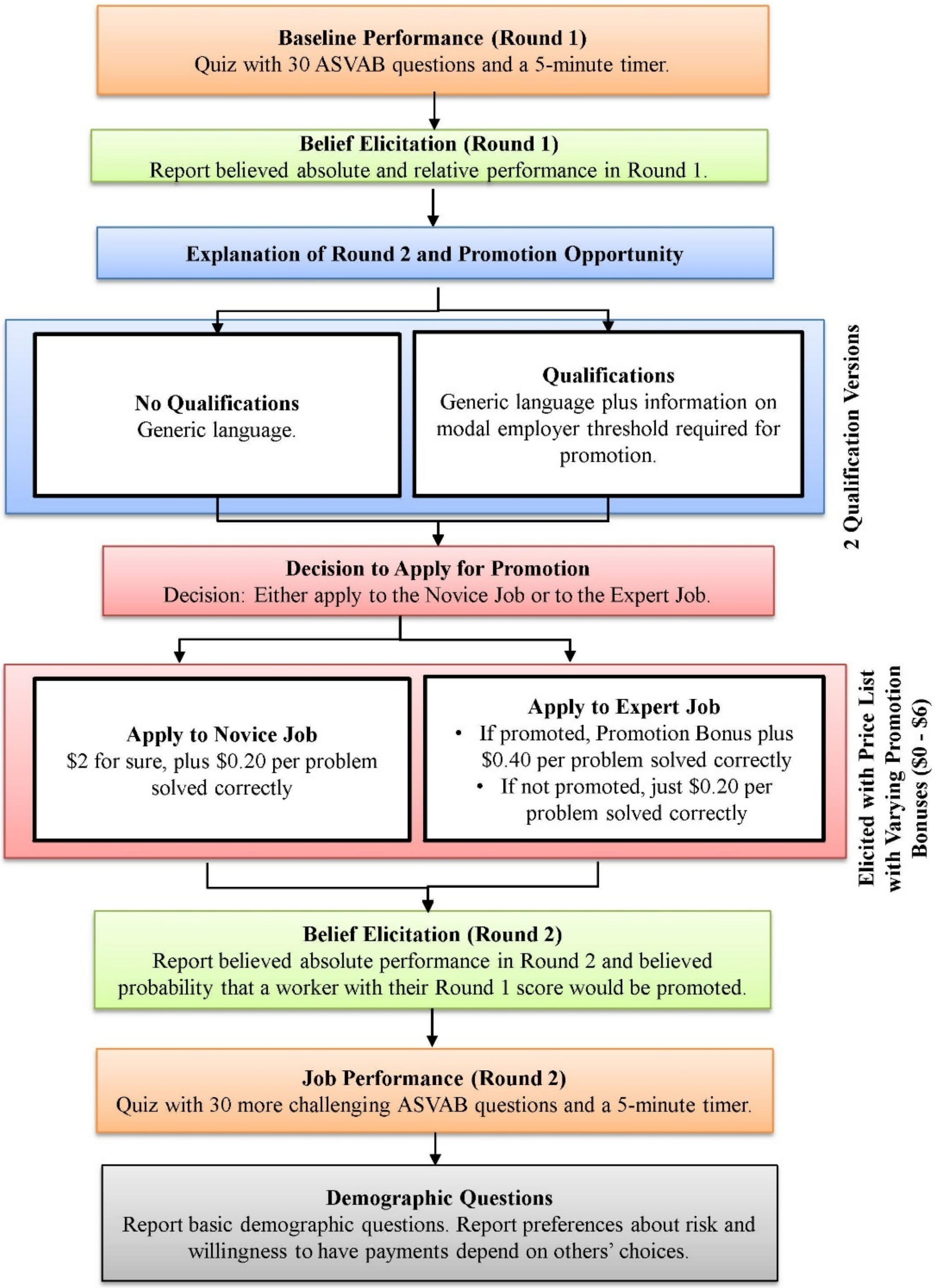


Figure 2. Design of Worker Experiment

Beliefs of Round 1 Performance

After completing Round 1, each participant is asked to guess their score –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%).[11]

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 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, the default option is that they will receive \$0.20 per problem solved correctly.

Prior to completing Round 2, they have to make a decision about whether they want to apply for an "expert-level promotion". This "expert-level promotion" is an increase in compensation. Participants are presented with two options about how to be paid for Round 2 performance. Importantly, within each option, the problems faced in Round 2 are identical. Participants are explicitly told that the set of questions will be the same regardless of the option they chose. They simply choose how to be compensated:

"**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, choosing between Option 1 (accepting the novice job) and Option 2 (applying for the expert-level job) over a range of possible promotion bonuses. Within the price list, we

[11] 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 account for posterior beliefs of Round 1 score and for feedback treatment assignment in the analysis that follows.

vary the size of the promotion bonus from $0 to $6, in increments of $0.50. Participants are aware that one row of the price list will be randomly-selected as the decision-that-counts, and that we will use their decision in that row to determine their application decision and associated payoffs. In Figure 3, we reproduce the price list used to elicit these decisions (full materials are available in Appendix A).

	Your Decision	
	Accept novice job (Receive $0.20 per correct answer plus $2)	Apply for expert-level job (Receive $0.40 per correct answer plus Promotion Bonus IF PROMOTED; Receive $0.20 per correct answer IF NOT PROMOTED)
Promotion Bonus of $0	○	○
Promotion Bonus of $0.50	○	○
Promotion Bonus of $1.00	○	○
Promotion Bonus of $1.50	○	○
Promotion Bonus of $2.00	○	○
Promotion Bonus of $2.50	○	○
Promotion Bonus of $3.00	○	○
Promotion Bonus of $3.50	○	○
Promotion Bonus of $4.00	○	○
Promotion Bonus of $4.50	○	○
Promotion Bonus of $5.00	○	○
Promotion Bonus of $5.50	○	○
Promotion Bonus of $6.00	○	○

Figure 3. Price List Used to Elicit Application Decisions

While choosing to apply for promotion outside of our experiment typically entails applying both for higher compensation *and* different, more challenging work, we hold fixed the nature of the work. While this sacrifices some external validity, it comes with a number of advantages. First, by ensuring that all participants complete the same Round 2 problems, we can measure the returns to being promoted for each participant, absent any selection. Second, we can rule out explanations for not applying for promotion related to a distaste or disinterest in doing the work (i.e. if women apply for promotion less than men, it cannot be because they simply want to avoid doing harder problems). This way, we can better focus on our main channel of interest: beliefs.

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 to workers who choose Option 1, the novice job. 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 apply for the expert-level job only if the worker believes she has sufficient probability of receiving it. In addition, because receiving the promotion entails a higher per-problem solved correctly wage ($0.40 versus $0.20), the returns to applying for and receiving the promotion are larger for more talented participants. We believe both of these features reflect many promotion opportunities outside of the laboratory.

We also structure the promotion opportunity in a way that mimics some of the uncertainty about the probability of receiving the promotion that would be present in the field. 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, prior experience, and potential), but also dependent on the discretion of a potential "employer" with only imperfect knowledge of the candidate's true capacity for success.

To achieve this, we recruited 10 other MTurk workers in a separate experiment to serve as employers. We recruit these employers in advance of running the worker experiment, and we ask them to make contingent decisions about what types of workers they would choose to "promote." We explained the worker experiment to the employers.

We told the employers that they would be randomly paired with *one worker who applied for the expert-level promotion*. 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. Of course, at the time of the employer experiment, no workers had yet been recruited to complete the experiment. So, we ask employers to make contingent decisions, and we use these contingent decisions to execute their hiring decisions once the full worker experiment had been run. We show employers a series of possible Round 1 performances (i.e. 3 problems solved correctly, 4 problems solved correctly, etc.), and we ask whether they would want to hire a worker with that Round 1 performance.[12] They make a series of binary decisions, covering all possible Round 1 scores.

---

[12] It is worth noting that employers make their decisions solely on the basis of the employee's Round 1 score. The employers are not provided with any other information such as age or gender.

We use these employer decisions to execute promotion decisions for all workers who apply to the expert-level promotion. Workers who apply are divided evenly and randomly among the 10 employers. Then, each employer's contingent decisions are used to determine whether each worker is promoted or not.[13,14]

Treatment Intervention

Our key treatment variation is varying how ambiguous the desired requirements for promotion are for workers, between-subject. In our control treatment, "No Qualifications", we provide workers with a short information section entitled, "Should I apply?". We remind them of the details of Round 2 and we tell them about the incentives employers faced when making their hiring decisions. Employees are told that the only information provided to the employer is their Round 1 score.

Participants in our "Qualifications" treatment receive the same language, but with one additional sentence that aims to reduce ambiguity about the bar: "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."[15] We argue that the key question workers must wrestle with is, "what test score do I need in order to get promoted?" Relative to the "No Qualifications" treatment, we argue that workers in the "Qualifications" treatment have a clearer, more specific, and more objective answer to this key question, reducing ambiguity as to where the bar is.

Beliefs about Promotion

After completing their application decisions, we ask 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 (A_1 in our model). They receive \$0.10 if they guess their Round 2 score exactly correctly.[16] The second, unincentivized question 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, the key measure, P, in our model.

---

[13] And, one of the matched workers for each employer is randomly selected to determine the employer's payoffs. Employers are aware of this payment rule.

[14] 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 one of the 10 employers 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 promoted. Both workers and employers have complete information on this process. See Appendix A for full instructions.

[15] Indeed, this threshold is informed by employer decisions. A Round 1 score of 10 is the lowest score at which at least 5 of the 10 employers in our employer experiment were willing to hire a worker.

[16] In principle, a participant could "game" this by guessing "0" and then intentionally not answering any questions correctly in Round 2, earning the bonus for guessing correctly. However, given the incentives to correctly solving problems, at least \$0.25 per problem solved, this does not seem like a strategy that participants are likely to employ.

Round 2

Participants then complete the Round 2 problems. 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 experiment might be distorted by an MTurk-specific skepticism about having payments be less transparent.

Hypotheses

Before turning to results, we take a moment to connect this experiment to our model. Recall that we argue that a risk neutral individual will apply if:

$$P(A_1) + (1 - P)(A_0) - c > A$$

In our analysis below, we look at application decisions, and ask whether there is a gender difference in willingness to apply conditional on measured ability. If there is, we can ask which of the factors in this model seem to be driving it. Our main focus is on the role of *P*, the believed probability of promotion. In our conceptual model, we argued that *P* is increasing in the extent to which own perceived aptitude exceeds the perceived bar. We explore directly whether there are gender differences in *P*, and whether these differences are well-explained by differences in beliefs of own aptitude (beliefs of Round 1 score). If they are not, this suggests a role for differences in beliefs of where the bar is in producing the gender gap. In particular, it may be the case that women believe the bar is higher.

How should *P* influence application rates? Following the model and plugging in the corresponding values from our setting, we have that a risk neutral individual will apply if:

$$P(0.4\,R2 + Bonus) + (1 - P)(0.2R2) > \ 0.2R2 + 2$$

where *R2* is the participant's expected Round 2 score and "Bonus" is the value of the promotion bonus for that row of the price list. Thus, a risk neutral participant will apply if her perceived probability of promotion is greater than $\frac{2}{0.2R2+Bonus}$. Individuals who expect to perform better in Round 2 will demand a lower perceived probability of promotion, all else equal; more risk averse individuals will demand a higher one. In line with this, we can ask whether Beliefs of Round 2 score help to explain any gender gap in willingness to apply. We can also ask whether elicited risk preferences contribute to a gender gap in application rates. We can also control for each of these factors in attempting to isolate the role of P.

Finally, we analyze the impact of our Qualifications treatment. In our model section, we hypothesized that by making the bar less ambiguous, we might reduce the gender gap in beliefs of where the bar was. In doing so, we expect to decrease any gender gap in perceived probability of promotion, and in willingness to apply. The reduction in the gender gap should be driven by two distinct shifts that occur when the bar is less ambiguous: unqualified workers should be less likely to apply, while qualified workers should be more likely to apply.

Results

The experiment was conducted in May 2018 with 1,502 workers. Table 2 provides summary statistics on the workers.[17] 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 54th percentile, while women rank in the 46th percentile on average ($p<0.001$).

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
Prop. Assigned to Qualifications Treatment	0.49	0.50	0.78
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.

Beliefs and Decisions

We start by exploring our control treatment, where participants receive less guidance on where the bar is for promotion. We ask every participant to estimate their chances of being promoted, conditional on

[17] Summary statistics restricted to the set of qualified workers are provided in Appendix Table B4.

applying. As hypothesized, we find that women's beliefs of their probability of being promoted 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. In Table 3, we use regression analysis to probe this. When we condition on true aptitude, as measured by Round 1 performance, women believe they are significantly less likely to be promoted conditional on applying (Table 3, Column I, 7pp, $p<0.001$).

Our model suggests that this gap could be driven by two factors. The first factor is beliefs of own ability. Indeed, we find gender differences in self-assessments. 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, conditional on having the same true score, ($p<0.001$), and believes she places 7.2pp worse in the distribution of performers ($p<0.001$) (see Appendix Table B5). In Table 3, Column II, we control for absolute beliefs of own ability and ask whether they explain the gender gap in believed probability of promotion. While beliefs of own aptitude are highly predictive of beliefs about promotion probability, they do not explain much of the gender gap, which remains 6pp ($p<0.001$).

Table 3. Gender Differences in Believed Probability of Promotion

OLS Predicting Believed Probability of Promotion (0 – 100pp) *No Qualifications Treatment*			
	I	II	III
Female	-7.17***	-5.69***	-3.14**
	(1.64)	(1.60)	(1.41)
Round 1 Score	1.78***	0.31	0.55***
	(0.18)	(0.27)	(0.17)
Belief of Rd. 1 Score		2.02***	
		(0.28)	
Belief of Rd. 1 Rank			50.7***
			(2.96)
Controls	Y	Y	Y
Observations	759	759	759
Adjusted R-squared	0.175	0.226	0.408

Notes: Controls are fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US, as well as 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.

In Column III, we ask whether beliefs of relative ability explain the gender gap. When we control for beliefs of relative ability, we see more explanatory power, suggesting they capture something important about this decision (Column III). This is worth noting, given that this is *not* a competitive environment (one applicant's decision to apply or ability to receive the promotion has no impact on another's). But, it seems reasonable that a person who believes their performance compares quite favorably to others will also view themselves as having a good chance of being promoted. In this way, beliefs of relative ability may reflect a mix of both beliefs about self and beliefs about what the bar is, hinting at whether the individual feels "good enough."

And yet, even accounting for differences in self-assessments, whether they be absolute or relative, there remains a gender gap in beliefs about the probability of promotion, suggesting differences in assessments of what qualifications are required for promotion. While we do not directly elicit beliefs of where the bar is from participants, we can give a sense of these beliefs by plotting believed probability of promotion against beliefs of Round 1 score. Our model proposes that believed probability of promotion increases in the distance between self-perceived aptitude and the perceived bar. This suggests that the greater the believed probability of promotion is, conditional on believed Round 1 score, the lower one perceives the bar to be. In Figure 4, we present this data for our no qualifications treatment, splitting by gender. The data suggests uncertainty about where the bar is: average beliefs hover close to 50%, and even within a given believed aptitude (Round 1 score), there is much variation. Participants with believed scores less than 7 perceive their chances of promotion as roughly a third, while participants with believed scores greater than 10 perceive their chances of promotion as better than 50-50. For most scores, women are directionally less confident in their chances of promotion than men, echoing Table 3. In sum, there is uncertainty about where the bar is, and in the face of that uncertainty, women seem to expect the bar to be somewhat higher than men do. This sets the stage for our qualifications intervention.

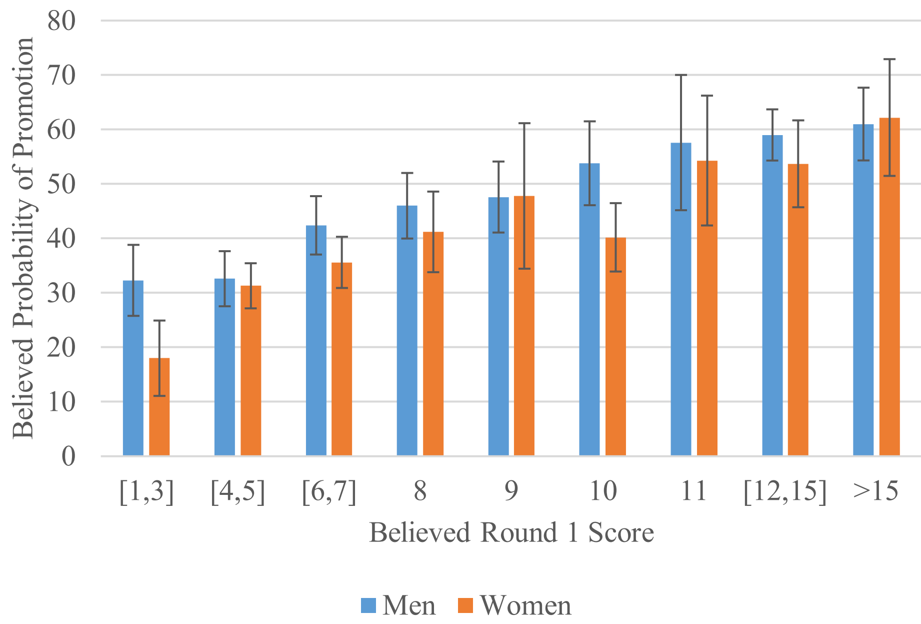


Figure 4. Believed Probability of Promotion, Conditional on Believed Score
No Qualifications Treatment

Next we consider our key behavioral outcome: willingness to apply for promotion at different wages. Conditional on applying, the average minimum promotion bonus 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 promotion bonus willingness to accept of 650 cents, 50 cents more than the maximum promotion bonus we offered. With this coding, the average min. promotion bonus required to induce a man to apply is 325 cents, while for women it is 344 cents (p=0.21).

Table 4 predicts the minimum promotion bonus at which someone was willing to apply for promotion for our No Qualifications treatment.[18] Conditional on Round 1 performance, women demand directionally larger promotion bonuses in order to apply, but the difference is not significant (Column I). Our conceptual model laid out three factors that should predict the decision to apply: perceived payoffs and costs, risk preferences, and believed probability of promotion conditional on applying. In Columns II - IV, we assess the role of each of these in predicting choices, controlling for beliefs of Round 2 score, risk preferences,

[18] In Appendix Table B6, we replicate this analysis using only the 696 participants in the control treatment who made monotonic choices. The results are quite similar.

and the believed probability of promotion.[19] As expected, each of these factors predict willingness to apply. Finally, in Column V, we include each of these factors and its interaction with the female indicator, asking whether any of these factors are more predictive for women than men. While we estimate that beliefs of Round 2 score and risk preferences are similarly important for men and women, we find that the effect of believed probability of promotion on the decision to apply is nearly twice as large for women as for men.

One question is why we observe a gender gap in believed probability of promotion but no corresponding gender gap in application rates. In our model, this could be the case if women were less risk averse than men, or perceived lower returns to promotion conditional on applying. In our sample, women are no less risk averse than men. While they do perceive lower returns to being promoted (women have lower forecasts of Round 2 score), this does not explain the puzzle. Controlling for beliefs of Round 2 score does not produce a gender gap in willingness to apply (see Column II).

We collected data on a few other arguably less interesting factors. 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 experiment. When asked on a 1-7 scale how reluctant they would be to have their payment depend upon chance or the decisions of others, with 7 being extremely reluctant, the average response is similar for men and women (4.4 and 4.5, respectively). Both of these measures are 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 own ability, and in particular relative ability. We estimate that, conditional on ability, women are directionally less willing to apply, but the gender gap is not significant.

[19] 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 B5, Column IV).

Table 4. Willingness to Apply for Promotion

	OLS Predicting Minimum Promotion Bonus at Which Applied *No Qualifications Treatment*				
	I	II	III	IV	V
Female	9.57	4.22	8.00	-7.90	73.3*
	(15.4)	(15.4)	(15.2)	(15.1)	(38.8)
Round 1 Score	-9.51***	-5.47***	-9.53***	-5.17***	-3.04
	(1.71)	(2.08)	(1.69)	(1.76)	(2.01)
Beliefs of Rd. 2 Score		-8.70***			-4.26
		(2.58)			(2.94)
Took Common			-69.0***		-67.0***
Risk Gamble			(14.8)		(19.4)
Believed Prob.				-2.43***	-1.68***
Of Promotion				(0.33)	(0.45)
Female x					-1.20
Belief of Rd. 2					(4.50)
Female x					-24.2
Risk Gamble					(28.7)
Female x Believed					-1.53**
Prob. Of Prom.					(0.67)
Controls	Y	Y	Y	Y	Y
Observations	759	759	759	759	759
Adjusted R-squared	0.044	0.058	0.071	0.107	0.147

Notes: Controls are fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US, as well as 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.

Does More Information on The Bar Help?

We hypothesize that reducing ambiguity about required qualifications could reduce the gender gap in believed probability of promotion, and application rates. We implement this in our controlled experiment and test it, randomly assigning half of participants to receive additional information on where the bar is.

In Table 5, we ask how this treatment impacts believed probability of promotion. We predict the worker's perceived chances of being promoted from her treatment assignment, controlling for her demographics, her true Round 1 score, her beliefs about her performance, and her risk preferences. We find that, on average, our treatment significantly reduces 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 (see Table 5, Column I). Note that this is true controlling 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), not through changing beliefs of own ability. Thus, overall, stated qualifications significantly reduce the gender gap in believed probability of promotion.

Of course, we expect heterogeneous effects depending upon whether a participant actually possesses the desired qualification (or believes she possesses the desired qualification). So, in Columns II – V, we split the sample across unqualified and qualified by participants.[20] We do this both by actual Round 1 score (i.e. truly qualified or not, Columns II and IV), and believed Round 1 score (i.e. believes themselves to be qualified or not, Columns III and V).

As we expect, the qualifications treatment reduces the believed probability of being promoted significantly among the unqualified group. But, this reduction is directionally larger for men than women (see Columns II and III). Among qualified participants, the reduction in the gender gap achieved by the stated qualifications is significant. Men's beliefs about their likelihood of receiving the promotion are directionally lower when there are clearly stated qualifications, while qualified women's beliefs increase significantly, eliminating the gender gap completely.

[20] If we use 11 rather than 10 as the cutoff score (reflecting a different reading of the stated qualifications required), results are qualitatively similar. See Appendix Table B7.

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
Qualification	-3.55***	-4.47**	-5.54***	-2.09	-0.21
Treatment	(1.32)	(2.04)	(1.76)	(1.71)	(1.97)
Female	-2.31*	-2.25	-2.15	-2.52	-3.68
	(1.38)	(1.97)	(1.71)	(1.92)	(2.34)
Female x Qual.	3.88**	2.19	2.93	5.44**	6.12*
Treatment	(1.92)	(2.80)	(2.42)	(2.62)	(3.20)
Controls	Y	Y	Y	Y	Y
Observations	1502	706	920	796	582
Adjusted R-squared	0.429	0.285	0.306	0.408	0.340

Notes: Controls are Round 1 score, beliefs of Round 1 score – absolute and relative, beliefs of Round 2 score, risk preferences, fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US as well as 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.

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.[21] We estimate no significant impact of the qualifications on unqualified participants, nor any gender differences among these participants. Among qualified participants, the effects are also quite noisily estimated; if anything, it seems that the stated qualifications directionally increase the gender gap in willingness to apply.

[21] In Appendix Table B8, we repeat this analysis using only those participants who made monotonic choices.

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

OLS Predicting Minimum Promotion Bonus at Which Applied					
	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
Qualification	-8.70	-10.1	7.20	-10.4	-32.5*
Treatment	(14.2)	(24.4)	(20.8)	(16.7)	(18.2)
Female	-4.33	1.74	7.57	-11.1	-23.4
	(14.9)	(23.7)	(20.2)	(18.8)	(21.5)
Female x Qual.	12.7	20.7	-8.56	8.17	52.1*
Treatment	(20.7)	(33.6)	(28.6)	(25.6)	(29.4)
Controls	Y	Y	Y	Y	Y
Observations	1502	706	920	796	582
Adjusted R-squared	0.099	0.043	0.061	0.090	0.069

Notes: Controls are Round 1 score, beliefs of Round 1 score – absolute and relative, beliefs of Round 2 score, risk preferences, fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US as well as 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.

What can we make of these results? 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 information on the qualifications required for promotion helps to reduce this gap. In particular, more 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.

V. Using Qualifications in the Field

In this section, we report the results of a field experiment testing the idea that less ambiguous qualifications can help to attract more talented female applicants.[22] From August to November 2017, 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.). Upwork facilitates match-making between freelance workers and potential employers. To implement our field experiment, 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 and have the option to retake a test after 180 days. For each test, Upwork provides information on the number of freelancers who have already taken this test, and their corresponding scores. 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

[22] This experiment was registered in the AEA RCT Registry (post-completion) and the unique identifying number is: AEARCTR-0006522.

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 whether they are 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, and the total number of tests they have chosen to display on their profile.[23] On top of that, we enter into the dataset an indicator of freelancer gender.[24]

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.[25]

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 "expert level job". Both types of jobs 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

[23] 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 self-reported capabilities at the job category level.

[24] Gender determinations were done as follows. First, we made three predictions of the gender using three different methods: (i) one member of the research team assigned gender based on name and photo prior to treatment assignment. Next, we assigned gender with use of the (ii) 1990 Census (IPUMs) and (iii) 1940-1970 Social Security Administration (SSA) name files. For (ii) and (iii), a name is assigned if 90 percent of individuals with the freelancer's stated name are classified as either male or female by the data source. In most cases, the three sources were in agreement. When all three were not or when we couldn't make a prediction based on IPUMs and/or SSA, we had another researcher code the gender (blind to treatment and results, n=231). In those cases, we go with the gender given by the majority of the predictors (of the two researchers, the Census, and the SSA), with a minimum of two predictors having to be in agreement. Otherwise, we drop the observation (n=9).

[25] Freelancers are able to re-take these tests, provided 180 days have passed since the last time they took the test. We do not observe the number of times a freelancer took the test.

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. 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]”.

Our treatments are informed by language that is common to application and admissions contexts. Our positive treatment reflects language used in some college and graduate school admissions. While top schools rarely issue strict cutoffs or qualifications, they often provide information on what typical scores look like for successful applicants. For instance, MIT undergraduate admissions provides the Middle 50% score range of admitted students for SAT and ACT scores (see: https://perma.cc/6LA2-HPMH). [26] Our normative treatment borrows language common to many job advertisements, where employers often “invite” applicants with particular qualifications to apply (see, for instance, this posting from Deloitte: https://perma.cc/Z4ZA-Q9L3). [27]

In our view, both treatments increase the objectivity, specificity, and clarity of desired qualifications relative to the control, honing in on our key hypothesis. While the positive treatment simply describes the qualification, the normative treatment takes things a step farther: it explicitly encourages applicants with the qualification to apply. An applicant worried about whether applying is the socially appropriate or right thing to do may be reassured by the normative treatment. In this way, the normative treatment may be a more aggressive intervention relative to the positive treatment. By studying both, we hope to increase the applicability of our results to a wider set of contexts and commonly used wordings in application and admissions contexts.

[26] MIT Admissions. 2021. “Admissions Statistics.” MIT Admissions. 2021. https://mitadmissions.org/apply/process/stats/ ; Permalink: https://perma.cc/6LA2-HPMH.

[27] Indeed.com. 2020. “Data Science Team Leader.” https://www.indeed.com/viewjob?jk=ccdabcc52f59cd3e&tk=1eptrgvj2qfuo800&from=serp&vjs=3; Permalink: https://perma.cc/Z4ZA-Q9L3.

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.[28]

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 unique and 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 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 experiment. 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.

By construction, all workers in our sample have completed and displayed either the Management Skills or Analytical Skills test; what does this mean for selection into our sample? Upwork actively encourages their freelancers to complete skills test (Upwork 2020). Freelancers can earn points for every addition they make to their profile. Such additions can be in the form of a profile photo, employment history, or skills tests. Freelancers who have earned enough points receive a badge ("Rising Star" or "Top Rated"). From conversations among Upwork freelancers, there seems to be some consensus that skills tests are mostly

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

valuable to freelancers who are newer to the platform (Upwork Community 2019); freelancers take the tests to help establish a positive reputation before they have completed jobs or earned ratings on the site. We cannot find evidence that these concerns seem to be different by gender. [29] To the extent that we are selecting on some characteristic, this selection is the same across randomly-assigned treatment condition.

Suppose that, on the extensive margin, having taken and chosen to display a skills test is capturing something important that we are selecting on. It seems fair to assume that this might also be true on the intensive margin -- that is, the total number of skills tests taken and displayed indicates an increasing extent of that "something important." From freelancer profiles, we record the total number of tests taken and displayed by each worker in our sample, and we control for it in our analysis. In this way, we are at least partially accounting for variation in this characteristic.

Results

Table 7 provides descriptive statistics of the freelancers in our sample.[30] Men and women vary in many dimensions in our sample. Womenhave more experience on Upwork and are more likely to advertise Writing skills, Administrative Support skills, and Customer Service skills. Men, on the other hand, 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. This could reflect true differences in skills, though we should caution that Murciano-Goroff (2020)'s finds that women are less likely to advertise skills on resumes in the tech domain, even given the same level of experience and skill.

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).

[29] Unfortunately, in 2019 Upwork retired skills tests; thus, at the time of drafting the paper we were unable to conduct a systematic comparison of workers with and without skills tests displayed.
[30] Appendix Table B9 provides summary statistics for only qualified freelancers (those whose test scores are at or above our cutoff score).

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	323	562	$p=0.07$
Jobs Worked on Upwork	13.7	15.6	$p=0.49$
Total Tests Displayed	6.18	7.32	$p=0.002$
Available Less than 30hrs/wk	0.18	0.22	$p=0.12$
Available More than 30hrs/wk	0.44	0.41	$p=0.50$
Available as Needed	0.37	0.34	$p=0.36$
College Degree	0.74	0.72	$p=0.53$
MBA Degree	0.14	0.08	$p=0.001$
Other Graduate Degree	0.20	0.21	$p=0.47$
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	$p=0.77$
Administrative Support	0.25	0.48	$p<0.001$
Customer Service	0.04	0.11	$p<0.001$
Sales & Marketing	0.15	0.16	$p=0.89$
Accounting & Consulting	0.22	0.13	$p<0.001$
Analytical Skills Score	3.73	3.57	$p<0.001$
Time Taken on Analytical Test (minutes)	50.5	48.05	$p=0.08$
Management Skills Score	3.55	3.42	$p<0.001$
Time Taken on Management Test (minutes)	19.79	20.65	$p=0.14$
Proportion Qualified by Test Score	0.29	0.18	$p<0.001$
Proportion in Analytical Skills Dataset	0.41	0.45	$p=0.2$
N	531	552	

Overall, 20% of men and 18% of women in our sample apply to one of our job postings.[31] This aggregate rate is relatively constant across the three treatments, with 20% of men and 19% women applying in the control, 21% of men and 18% of women applying in the Positive treatment, and 20% of women and 17% of women applying in the Normative treatment. Of the 209 participants who apply to our job postings, most (130) apply to the expert-level job.[32]

However, the key question is how application rates to the expert job vary by treatment and by qualification level. Our prediction is that reactions to the treatments should vary by qualification level: reduced ambiguity in desired qualifications should increase the likelihood of qualified applicants applying, while decreasing the likelihood of unqualified applicants applying. To explore this question, we look at the unconditional probability of applying to the expert level job, depending upon treatment, true qualification level, and gender.

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 5 demonstrates the results by gender and treatment. We see that the fraction of all qualified men who apply to the expert level job is quite steady across treatment, ranging from 19 – 23%. The application decisions of qualified women, however, vary more widely across treatment, with the lowest rate of application in the control – 6% -- and the highest rate of application in the normative treatment – 29%. In terms of these raw means, we see suggestive evidence that qualified women respond to more clearly stated qualifications in a way that qualified men do not.

[31] Of the 209 participants who apply to our position, 14 did not apply to strictly one job. For all 14 participants who either (i) failed to specify which of the two jobs they wished to apply to, or (ii) explicitly applied to both jobs, our research team contacted them via the Upwork platform after their initial application and asked them to clarify which job they were choosing to apply to. Nine of those 14 individuals specified one application decision (intermediate or expert level). Four participants remained unspecified in their choice and 1 participant remained an applicant for both jobs. We code these five workers as having applied to the intermediate job *and* as having applied to the expert job. Table B10 in the appendices consists of a robustness check of the results presented in Table 8 where we drop all the 14 observations. The results remain directionally unchanged.

[32] In the control treatment, 16% of all men and 12% of all women in our sample apply to the expert-level job. Applications to the expert-level job are directionally lower in our two qualification treatments. In the Positive treatment, 12% of men and 10% of women apply to the expert-level job, and in the normative treatment, 13% of men and 9% of women apply to the expert-level job.

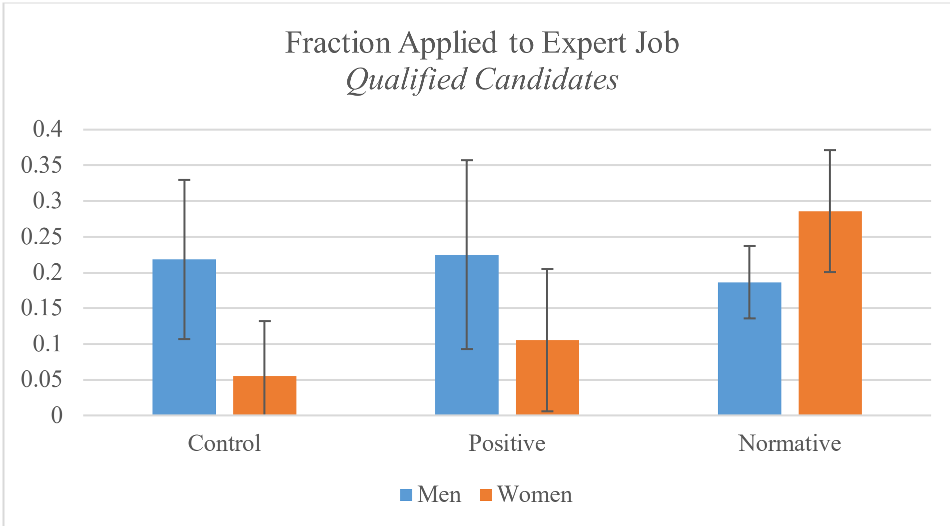


Figure 5. Fraction of Qualified Applicants that Apply to Expert Job

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. See Figure 6. 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, though differences are not very large.

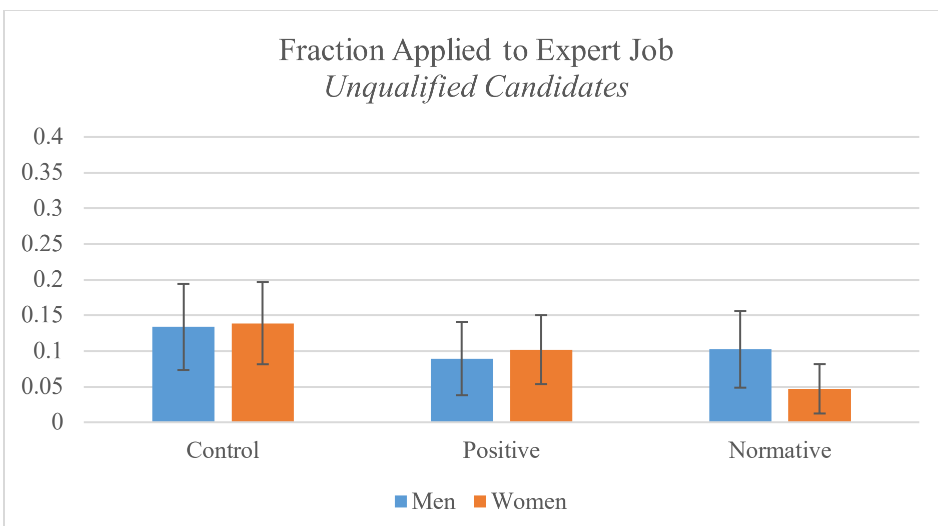


Figure 6. Fraction of Unqualified Applicants that Apply to Expert Job

In Table 8, we use regression analysis to explore these questions formally. 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.[33] We start by analyzing the full sample in Columns I - III. Consistent with the raw data, when we do not condition on qualification level, we see that overall our treatments have no significant impact on qualification rates for men or women (Column I, Column III). Of course, this may mask any competing patterns across unqualified and qualified candidates. In fact, in Column II, we show that relative to the control treatment, both the Positive and Normative treatment decrease application rates among unqualified applicants, while increasing them among qualified applicants. These effects are directional in the Positive treatment, and significant in the Normative treatment (we cannot reject no differences between the two qualification treatments).

In Columns IV - V, we analyze the decisions of unqualified applicants (those with test scores less than the advertised threshold). Overall, we find that both qualification treatments decrease application rates to the expert-level job (Column IV). We estimate that men's decisions are not significantly impacted by our treatments. For women, we estimate that, relative to the control treatment, the normative treatment deters applications from unqualified women by 9pp ($p<0.01$). However, in an interacted model, we cannot reject that the deterrence effect is of a similar size for men and women (Column V).

In Columns VI - VII, we focus on qualified applicants. Overall, we estimate that our two qualifications treatments directionally increase the rate at which qualified applicants apply to our expert-level job. Consistent with Figure 5, 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. Note that in our control, qualified women are 20pp less likely to apply than qualified men. Both treatments directionally reduce this gap. Qualified women are 10pp more likely to apply in our positive treatment relative to the control ($p=0.24$), and 28pp more likely to apply in our normative treatment relative to the control ($p<0.01$).

Table 8. Application Rates to Expert Level Job

	OLS Predicting Decision to Apply to Expert-Level Job						
	All Participants			All Unqualified		All Qualified	
	I	II	III	IV	V	VI	VII
Positive	-0.026	-0.043	-0.039	-0.046*	-0.057	0.044	0.0067
Treatment	(0.024)	(0.027)	(0.035)	(0.026)	(0.038)	(0.061)	(0.081)
Normative	-0.030	-0.067**	-0.033	-0.070***	-0.044	0.098	-0.00076
Treatment	(0.024)	(0.028)	(0.034)	(0.026)	(0.038)	(0.060)	(0.074)

[33] Results are unchanged if we exclude their self-reported skills as dummies.

Female	-0.029	-0.026	-0.039	-0.0062	0.0023	-0.075	-0.20**
	(0.021)	(0.021)	(0.035)	(0.023)	(0.038)	(0.055)	(0.086)
Qualified		-0.057					
		(0.046)					
Positive x		0.066					
Qualified		(0.057)					
Normative x		0.15***					
Qualified		(0.056)					
Female x			0.024		0.020		0.10
Positive			(0.049)		(0.052)		(0.12)
Female x Normative			0.0050		-0.047		0.28**
			(0.048)		(0.052)		(0.12)
Controls	Y	Y	Y	Y	Y	Y	Y
Observations	1083	1083	1083	827	827	256	256
Adj. R-squared	0.035	0.039	0.034	0.037	0.037	0.012	0.026
p-value Pos. = Norm.	0.87	0.86	0.39	0.35	0.72	0.39	0.93
p-value: Pos. x Above = Norm. x Above			0.15				
p-value: Fem Pos. = Fem Norm.		0.70			0.20		0.16

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. * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.

Overall, our treatments directionally increase the rate at which qualified candidates apply to our expert job, and directionally decrease the rate at which unqualified candidates apply. In addition, the share of women in the pool of qualified candidates also increases. Thus, from a firm's perspective, the impact of the qualifications on the potential pool seem quite positive: a larger, more diverse pool of qualified applicants, and fewer unqualified applicants.

In our setting, the normative treatment has a directionally larger impact than the positive treatment, though we should note that the differences are not statistically significant. Why might the normative treatment have a larger impact on the behavior of qualified female candidates than the positive treatment? It may be that the explicit ask – inviting someone with that score to apply – more successfully overcomes hesitations about what the social norms, employer expectations, or right course of action is. This would be related to findings from the negotiation literature; Babcock et al (2005) document how "strong" situations "where everyone has the same understanding of how they are supposed to respond" produce smaller gender gaps in negotiation outcomes than "weak," more ambiguous situations. Our normative treatment may more successfully produce the type of "strong" situation that minimizes gaps in the willingness to apply context.

VI. 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 *in general* to be qualified for a given opportunity. This novel finding relates to the policy intervention we test in an online experiment and in the field.

We show that exogenously reducing ambiguity about the required qualifications for a position can help to close the gender gap in beliefs and, in our field setting, application rates among talented candidates. Our results suggest that there may be soft touch employer interventions that can improve the diversity of their applicant pool in male-typed domains, helping to draw in qualified female candidates. This seems like a promising and low-cost path to explore.

In future work, it would be useful to consider behavior also in more female-typed domains, to understand whether the patterns we observe generalize. Returning to our conceptual model, we expect that in more female-typed domains, the gender gap in beliefs of own ability should be reduced, if not reversed. It is much less clear how beliefs of the bar might differ in female-typed domains: does the gender-type of the domain also influence beliefs of where the bar is? This is an interesting question for further study. It could be that it is not women, in general, who are less likely to apply, but rather that individuals are less likely to apply in more gender incongruent areas.

Of course, many hiring decisions are substantially more complicated than those studied in our experiments, and may involve evaluating candidates across a range of dimensions, some qualitative and some quantitative. Our policy suggestion is most obvious to translate for quantitative dimensions: better or more clearly specifying desired years of experience, minimum GRE score, number of projects successfully completed in the past, etc. Assuming that indeed the employer has a bar in mind for these dimensions (that is, they only want to hire people above that bar), it seems that our type of intervention could be helpful. Candidates below that bar should be less likely to apply, and the employer may draw in qualified people who, for example, didn't realize that "extensive experience" meant X years. Assuming that performance on these quantitative dimensions is not systematically negatively correlated with more qualitative dimensions, better sorting on at least one dimension should weakly improve the pool of applicants.

While in Experiments 2 and 3 we analyzed quantitative cases where it was straightforward to specify a bar, our argument is that more general forms of ambiguity reduction around desired qualifications could produce similar effects, in line with Experiment 1. Consider a less quantitative example from hiring and promotion criteria within academia. Many universities demand scholarly excellence. In some cases, little more is said. In other cases, more specificity, objectivity, and/or clarity is provided. For instance, consider this language around tenure criteria at the University of Houston (https://perma.cc/JZ9P-GKPQ [34]):

> The candidate is expected to have a monographic book of sufficiently high quality and originality to serve as proof of the candidate's capacity for engaging in intellectually challenging interaction with peers, colleagues and students. At a minimum, the candidate should have a book accepted for publication, completed except possibly for minor revisions, by a reputable press. The candidate is expected to have had published or accepted for publication a minimum of four (4) articles/book chapters, excluding book reviews that demonstrate on-going research indicative of the candidate's successful transition from the postgraduate level of scholarship to that of the level expected of tenured faculty. Preference in this category will be given to articles published in or accepted by first-tier, refereed professional journals (in printed, on-line or other electronic medium) and book chapters published by reputable presses.

In our view, this illustrates that even in a complex setting where setting a single, quantitative bar would be infeasible, it may still be possible to meaningfully reduce ambiguity around desired qualifications, helping to better sort candidates.

[34] University of Houston. 2021. "Criteria and Guidelines for Promotion and Tenure." 2021. https://www.uh.edu/class/spanish/faculty/p-and-t-guidelines/index; Permalink: https://perma.cc/JZ9P-GKPQ.

While extrapolating from one context to others always presents challenges, we think our results are likely to offer useful lessons for many real world settings of interest, particularly because of our multi-method, multi-population approach. We observe common themes across controlled experimental settings with Ivy League undergraduates and Amazon Mechanical Turk workers and in a field experiment on an employment website. This suggests that our findings are not a function of a particular feature of any single experiment or sample. An important next step could be studying these questions in a more traditional, salaried employment setting. While candidates in these contexts are likely to have more experience with job application processes, one could also imagine this being a setting where learning is particularly difficult. If qualified candidates choose not to apply, they miss out not only on the job, but also on the opportunity for feedback about the bar.

VII. References

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

Abraham, Lisa and Alison Stein. 2020. Words Matter: Experimental Evidence from Job Applications. Working paper.

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. 2019. "Beliefs about Gender." *American Economic Review 109*(3), 739-73..

Bowles, H. R., Babcock, L., & McGinn, K. L. 2005. "Constraints and triggers: Situational mechanics of gender in negotiation." *Journal of personality and social psychology*, *89*(6), 951.

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. 2019. "Stereotypes and Belief Updating." *Working Paper*.

Coffman, Katherine B, Christine L. Exley, and Muriel Niederle. 2020. "The Role of Beliefs in Driving Gender Discrimination." Forthcoming in *Management Science*.

Croson, Rachel and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature,* 47(2), 448 - 474.

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. 2020. "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. 2020. "Missing Women in Tech: The Labor Market for Highly Skilled Software Engineers." *Working Paper*.

Niederle, M. 2016. "Gender," in *The Handbook of Experimental Economics 2*, Kagel John, Roth Alvin E., eds. (Princeton, NJ: Princeton University Press, 2016).

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.

Niederle, Muriel and Alexandra H. Yestrumskas. 2008. "Gender Differences in Seeking Challenges: The Role of Institutions." *Working paper*.

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. 2019. "A University-Wide Field Experiment on Gender Differences in Job Entry Decisions." *Management Science* 65 (7): 3272-3281.

"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.

Upwork Community Forum. 2019. "Removing Skill Tests on July 16th." June 21, 2019. https://community.upwork.com/t5/Announcements/Removing-Skill-Tests-on-July-16th/m-p/609790#M31055; Permalink: https://perma.cc/K4S2-YJL6.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

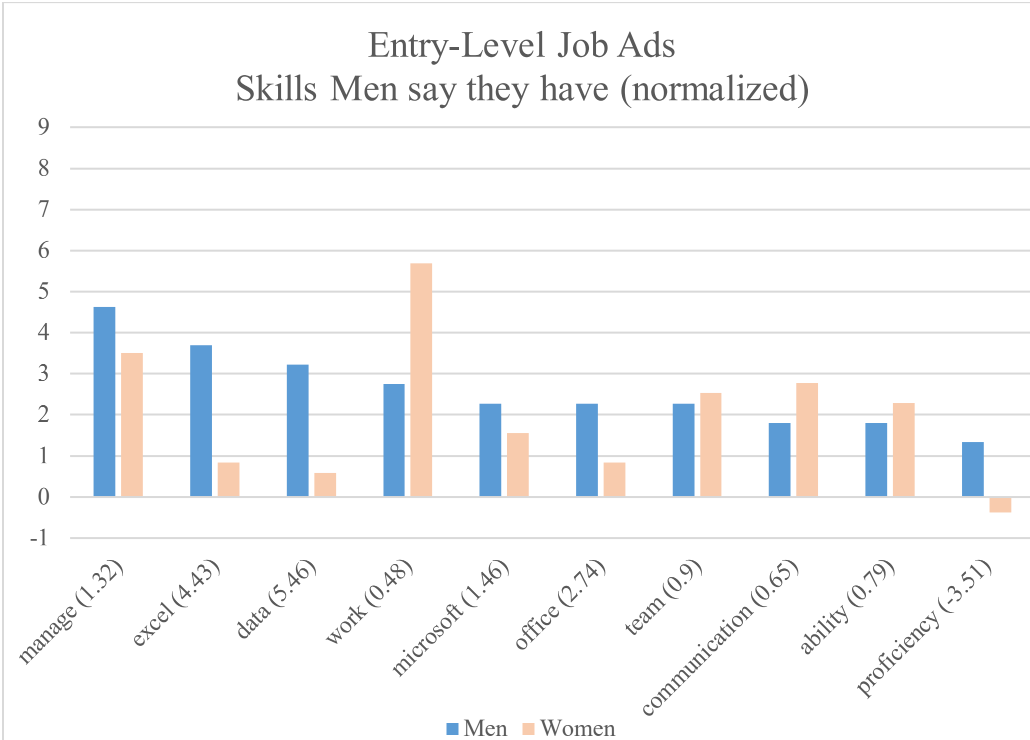
B. Appendix

Figures B1 to B4 show the normalized frequency of the top 10 mentioned words for men and the top 10 mentioned words for women. The number in brackets next to each skill indicates the gender ratio. To arrive at the first graph, we do the following. We collected all the skills men mentioned to have when they were given entry-level job ads. We cleaned this qualitative data to derive only the stem of each skill. This means we derive the root of each word, where for example "managing," "manage," and "managed" all become "manag." This approach will remove some of the meaning but will help us better understand which skills participants mentioned. We do the cleaning and stemming with the help of the R-packages "tm" (text mining) and "SnowballC" (text stemming). Given the strong left-skewed nature of the distribution, where there are a large number of skills which have been mentioned only once, we decided to limit the analysis to the 100 most often mentioned skills for men and the 100 most often mentioned skills for women. To compare men and women, we then normalized all the 100 most frequently mentioned skills. The normalized measure is a simple z-score.

The graphs show the 10 most frequently mentioned skills by gender. To make that graph more informative, we contrast these normalized frequencies with the normalized frequencies of the other gender. For the first graph, this means we go through the same cleaning and normalizing procedure for the 100 most often mentioned skills for women as we went through for men's responses. We then add the normalized frequency for women to each of the men's top 10 mentioned skills. This creates a contrasting bar for the bars of main interest. If there is no contrasting bar, it means that the skill was not on the list of the 100 most frequently mentioned skills for the other gender.

The number in brackets next to each skill indicates the gender ratio. It takes the normalized frequency of that skill (calculated as a simple Z-score) and compares it to the normalized frequency of the other gender (calculated as a simple Z-score). There are cases where the skill was not among the 100 most frequently mentioned skills for the contrasting gender. Those cases are labeled as "NA."

Figure B1. Frequently mentioned words for *possessed* skills for entry-level openings



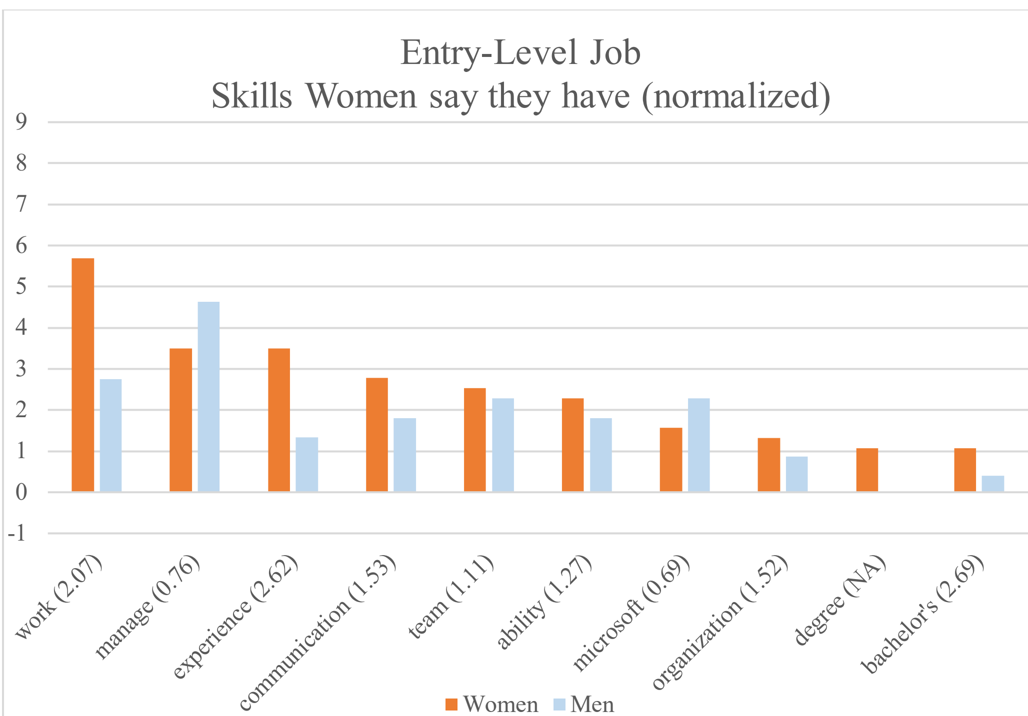
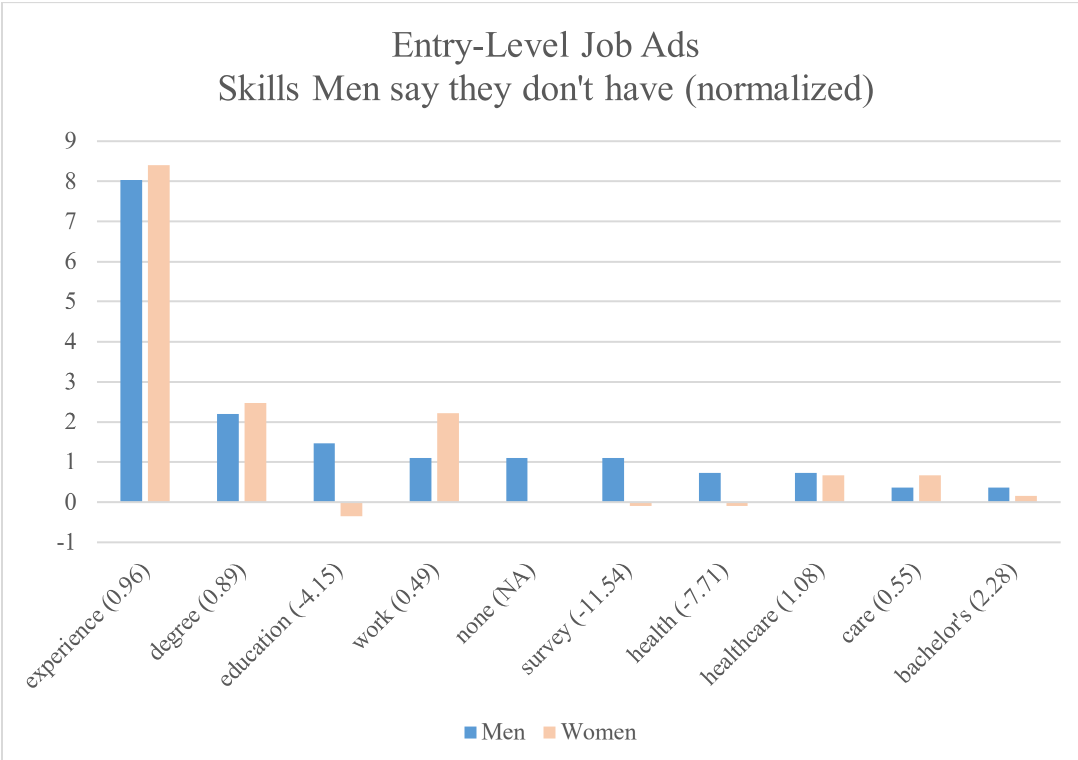


Figure B2. Frequently mentioned words for *non-possessed* skills for entry-level openings



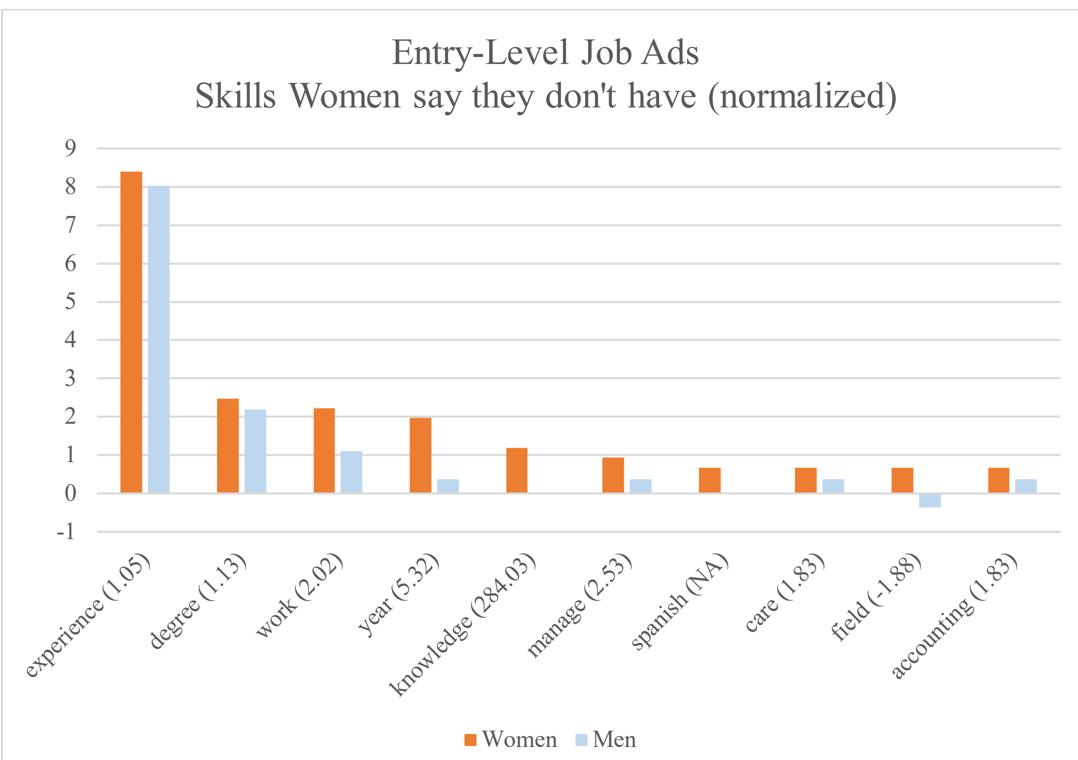
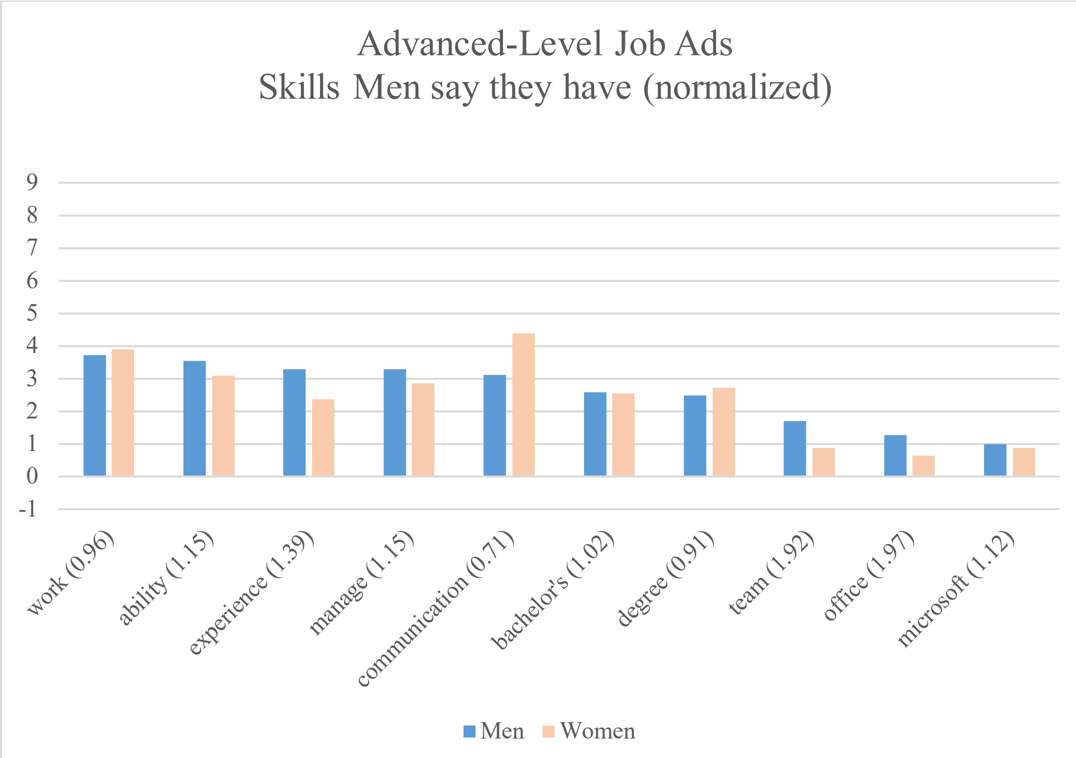


Figure B3. Frequently mentioned words for *possessed* skills for advanced openings



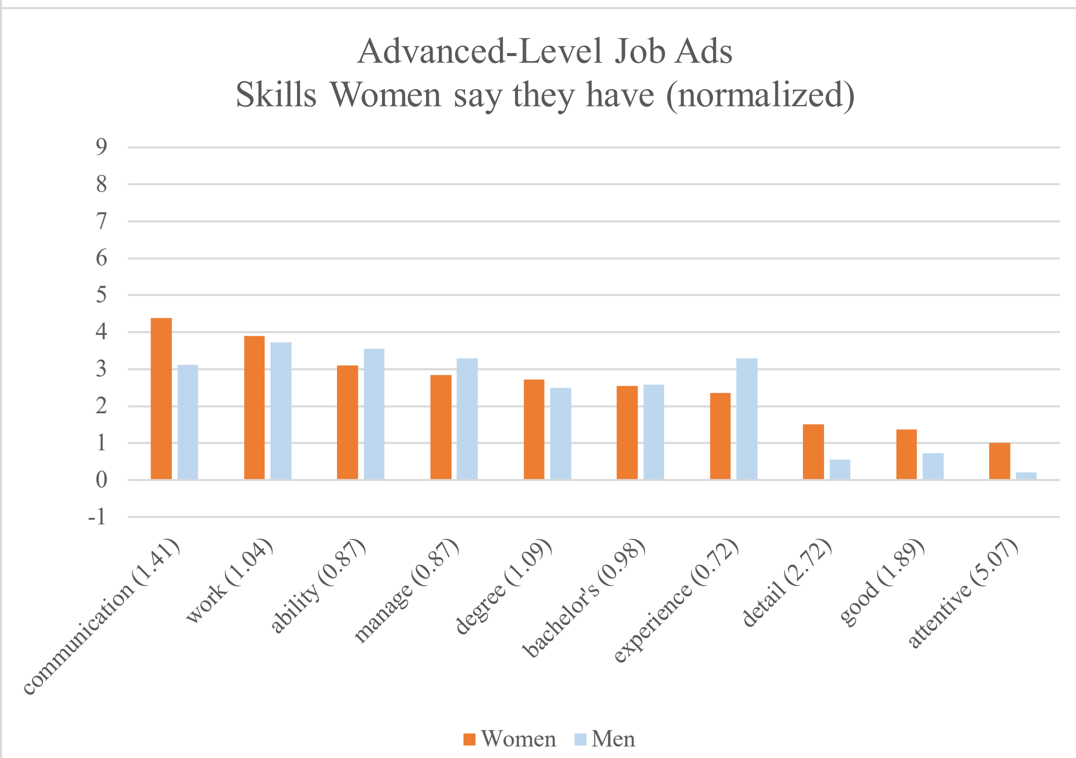
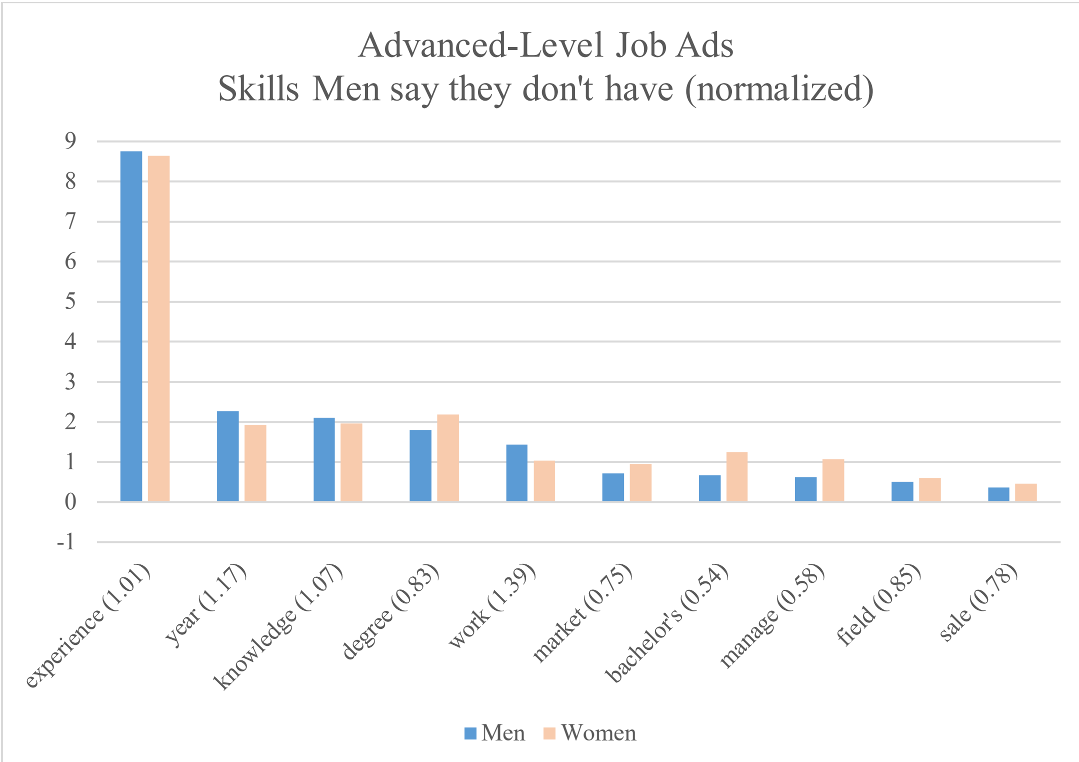


Figure B4. Frequently mentioned words *non-possessed* skills for advanced openings



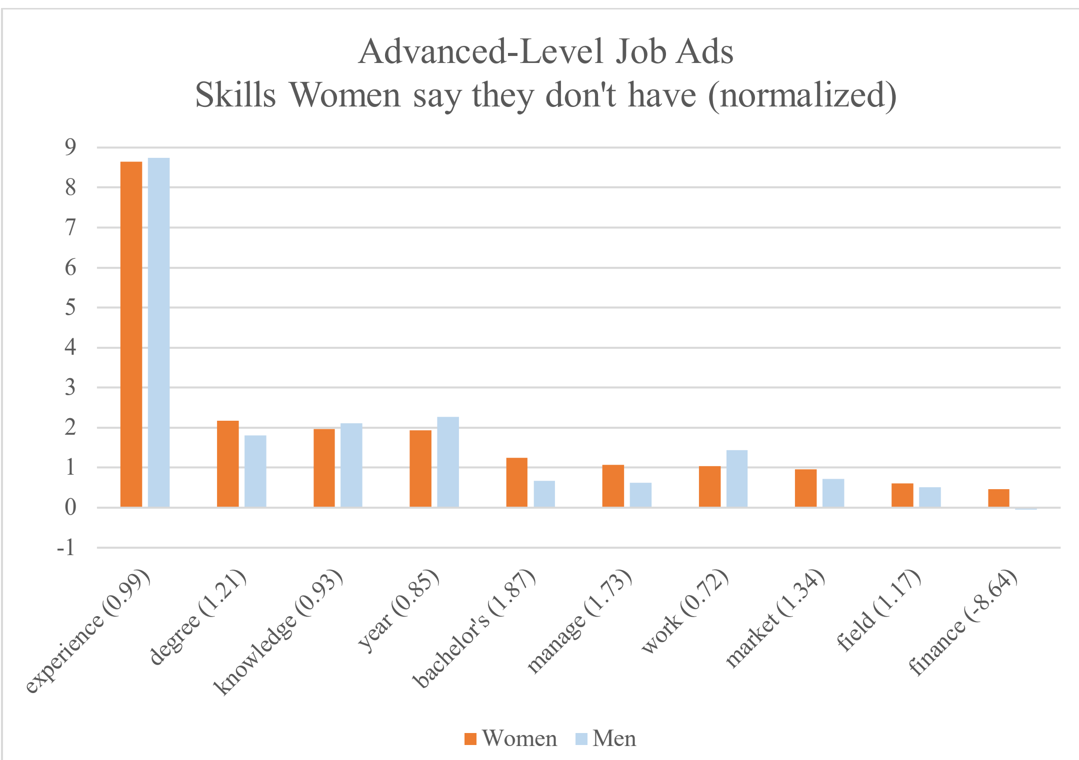


Table B1. Summary Statistics for Laboratory Participants in Experiment 1

	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
Humanities Major	0.10	0.18	0.14
Social Science Major	0.25	0.29	0.53
STEM Major	0.50	0.44	0.41
Is fluent in English	0.99	0.99	0.89
Order of Experiment within Session	0.48	0.50	0.81

Table B2. Summary Statistics on Job Ads in Experiment 1

Panel A: Data on Bureau of Labor Statistics Sector for Ads	
BLS Sector	**Percent of Ads**
Educational Services	7
Financial Activities	7
Health Care and Social Assistance	21
Information	14
Leisure and Hospitality	11
Manufacturing	7
Professional and Business Services	25
State and Local Government	4
Transportation and Warehousing	4

Panel B: Summary of Participant Assessments					
Variable	**Min.Value**	**25th pctl**	**75th pctl**	**Max. Value**	**Mean**
Individual Level Well-Qualified	1	2	6	10	4.41
Ad Level Male Stereotype	-0.26	-0.12	0.2	0.5	0.070
Ad Level Prestige	2.86	3.44	4.14	4.82	3.856
Ad Level Objectivity	5.14	6.37	7.38	7.7	6.857
Ad Level Avg. Qualified	4.57	5.44	6.64	7.23	6.077

Figure B5. Example 1 of Job Ad in Experiment 1

Marketing Representative

SP Plus Corporation 533 reviews - Boston, MA 02110

Job Function – Prepare financial reports, developed budgets, and performed variance analysis in accordance with business plan expectation. Compile periodic financial reporting packages for senior management. Oversee general accounting functions, including AR/AP, account reconciliation, and cash management. Administer all financial management systems, evaluating and integrating new applications.

Responsibilities

MAIN RESPONSIBILITIES

- Develop an acute understanding of SP+ core services as well as the clients, challenges, opportunities, and key business processes of the parking industry.
- Work with the team to develop innovative, actionable marketing strategies to acquire new business as well as retain existing business more effectively.
- Responsible for tasks and deliverables that are integral to the company's business development and expansion.
- Conduct market research that captures data on target market demographics, competitor initiatives, and industry trends.
- Create and send client/tenant letters (rate increases, sales letters, update letters to tenants, letters requested by the client).
- Coordinate, implement and manage various marketing initiatives (marketing programs, direct mailings, telemarketing campaigns, e-mail campaigns, website and mobile promotions, paid search campaigns, social media and partnership marketing) to achieve business objectives.
- Develop and maintain marketing collateral (brochures, flyers, handouts, signage, coupons and sales letters).
- Analyze marketing results/trends and make recommendations for improvement/implementation.
- Generate monthly marketing activity reports, charts and statistical tables.
- Implement annual marketing plans and develop marketing budgets accordingly.
- Interact with internal and external vendors.
- Research marketing media opportunities - print, radio, billboard, cross promotions, web site advertising and links.
- Report progress of marketing goals and strategies to Senior Manager and Client.
- Handle customer inquiries from yelp.
- Coordinate amenity programs and customer appreciation days.
- Networks with property managers/garage clients and attend trade association/neighborhood meetings as assigned.

Qualifications

MINIMUM QUALIFICATIONS

- 5+ years of marketing and/or account management experience is preferred.
- Bachelor Degree, preferably in Business or Marketing.
- Solid marketing background.
- Solid customer service skills are a must.
- Computer literate to include Microsoft Excel, Word, and PowerPoint.
- Knowledge in Adobe Illustrator, Photoshop and Acrobat is a plus
- Motivated self-starter who works well with minimal direction.

SP+ is an equal opportunity employer committed in policy and practice to recruit, hire, train, and promote, in all job classifications, without regard to race, color, religion, sex, age, national origin, citizenship status, marital status, sexual orientation, veteran status, disability or other classes protected by federal or state law. SP+ does not tolerate harassment of or retaliation against any employee or applicant on the basis of these characteristics, or because the individual exercised his or her EEO rights.

Company Info

Follow

Get job updates from SP Plus Corporation

SP Plus Corporation

533 reviews

SP+ is a diverse provider of professional parking, ground transportation, facility maintenance, security and event logistics services to...

Let Employers Find You

Thousands of employers search for candidates on Indeed

Upload Your Resume

Source: Indeed.com, **only page 1 displayed here**

Figure B6. Example 2 of Job Ad in Experiment 1

Client Service Associate

Fanning Personnel - Boston, MA 02116
$65,000 a year

Registered Investment Advisor west of Boston seeking a Client Service Associate! Will work closely with Senior Executive, prospective clients and existing clients in every aspect of the process and relationship management. The individual in this role will also act as the point of contact for client account maintenance and compliance requests. This person will serve as an integral component in client service and client management. To $65K + Profit Sharing

The Client Service Associate will:

- Communicate clearly via phone, email and written correspondence to establish and maintain quality client relationships
- Prepare detailed portfolio analysis of current client and potential client holdings
- Provide clients with the necessary documentation to create, transfer and maintain accounts
- Facilitate the timely, accurate transfer of client assets
- Maintain confidential client personal and account information in CRM Database
- Work closely with Operations to produce Quarter End Client Reporting
- Handle compliance related tasks
- Complete ad hoc special projects as assigned

The Client Service Associate possesses the following qualifications:

- Bachelor's degree
- Minimum 3 years' experience in financial services industry preferred
- Excellent customer service, communication, writing and interpersonal skills
- Advanced computer skills in Microsoft Excel, Word, PowerPoint
- CRM and portfolio accounting database experience a plus
- Ability to multitask with great attention to detail

Job Type: Full-time

Salary: $65,000.00 /year

Experience:

- Client Service: 3 years (Required)

Education:

- Bachelor's (Required)

20 hours ago - Save Job

Apply Now

Please review all application instructions before applying to Fanning Personnel.

Company Info

Follow

Get job updates from Fanning Personnel

Fanning Personnel

Fanning Personnel is a leading provider of staffing services to Boston's financial industry. For over 40 years we have had a track record of...

Let Employers Find You

Thousands of employers search for candidates on Indeed

Upload Your Resume

Source: Indeed.com, **only page 1 displayed here**

Figure B7. Example 2 of Job Ad in Experiment 1

Compensation Survey Associate

Pearl Meyer & Partners, LLC. - Wellesley, MA

Looking for an opportunity to gain exposure to the compensation area of human resources, play an important role as part of Pearl Meyer's Compensation Survey Team, and work with a very impressive portfolio of clients?

This unique and challenging role offers a chance to wear many hats. The Survey Team at Pearl Meyer publishes approximately 30 surveys annually on employee compensation, benefits and human resources practices. We are seeking candidates who thrive on working in a team environment with an aptitude for multitasking. From client management and engagement to data analysis, this role offers full exposure to the survey publication cycle.

As a Compensation Survey Associate, you will:

- Assist Survey Project Managers throughout the survey cycle including data analysis, report peer review, and client meeting coordination.
- Manage an assigned group of compensation survey clients. This includes assisting clients throughout the survey submission process, reviewing individual client data to ascertain data integrity, and advising clients regarding questions pertaining to survey report outputs.
- Maintain client databases and generate reports as needed.

The Survey Team at Pearl Meyer offers motivated employees the opportunity for career growth through extensive educational, training, and development opportunities. We offer a career ladder with opportunity for upward career movement.

This position may offer the occasional opportunity for travel in spring and fall for those who desire.

Additional Information:

All candidates must be authorized to work in the U.S.

Pearl Meyer is an EEO employer.

Requirements
Applicants should have a strong orientation towards client management and data analysis. This role is deadline driven and requires strong attention to detail. The ability to work in a collaborative, team environment is needed. Ideal candidates will have an interest in, or prior exposure to, the field of compensation and a desire to further their knowledge through Pearl Meyer's many training and educational opportunities.

Additional requirements and expectations include:

- A Bachelors degree
- Proficiency in Microsoft Excel
- Solid organizational and analytical skills
- The ability to work as an effective and reliable team member
- The ability to handle and safeguard confidential and sensitive information
- Working knowledge of Microsoft Windows environment and intermediate ability with MS Word and PowerPoint
- Experience using compensation surveys is desired but not required

Company Info

Follow

Get job updates from Pearl Meyer & Partners, LLC.

Let Employers Find You

Thousands of employers search for candidates on Indeed

Upload Your Resume

Source: Indeed.com, **only page 1 displayed here**

Table B3. The Gender Gap in Perceived Qualification for Real Job Openings in Experiment 1

OLS Predicting How Well-Qualified an Individual Feels for Job Opening	
	I
Female	-0.02 (0.11)
Fraction of Qualifications that a Participant Believes They Possess	0.08**** (0.002)
Ad Fixed Effects	Yes
Demographic Controls	Yes
Order of Experiment within Session	Yes
R-squared	0.80
Clusters (Obs.)	196 (784)

Notes: Controls are fixed effects for each race category, fixed effects for each education category, age, a dummy for majoring in humanities, a dummy for majoring in social science, a dummy for majoring in STEM, and a dummy for fluent in English. * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.

Table B4. Summary Statistics on Qualified Workers in Experiment 2

	Qualified Men	Qualified Women	P-value
White	0.82	0.86	0.15
Black	0.04	0.05	0.44
Asian	0.11	0.07	0.03
Attended HS in US	0.98	0.97	0.16
HS Only	0.09	0.06	0.17
Some College/Assoc.	0.32	0.29	0.26
Bachelors	0.42	0.47	0.12
Advanced Degree	0.17	0.18	0.76
Rd. 1 Score	14.0	13.2	<0.01
Rd. 2 Score	9.73	8.55	<0.001
Prop. Assigned to Qualifications Treatment	0.50	0.53	0.44
N	460	336	

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.

Table B5. Believed Performance by Gender in Experiment 2

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 B6. Experiment 2 Replication of Table 6 with Monotonic Participants Only

	OLS Predicting Minimum Promotion Bonus at Which Applied *No Qualifications Treatment* MONOTONIC PARTICIPANTS ONLY				
	I	II	III	IV	V
Female	9.97	5.06	7.66	-9.30	59.3
	(14.8)	(14.7)	(14.5)	(14.4)	(36.7)
Round 1 Score	-12.5***	-8.28***	-12.4***	-7.77***	-5.87***
	(1.63)	(2.03)	(1.60)	(1.67)	(1.93)
Beliefs of Rd. 2 Score		-9.06***			-2.90
		(2.62)			(3.07)
Took Common			-70.8***		-69.7***
Risk Gamble			(14.1)		(18.4)
Believed Prob.				-2.50***	-2.02***
Of Promotion				(0.32)	(0.44)
Female x					-2.98
Belief of Rd. 2					(4.40)
Female x					-22.0
Risk Gamble					(27.1)
Female x Believed					-0.98
Prob. Of Prom.					(0.64)
Controls	Y	Y	Y	Y	Y
Observations	696	696	696	696	696
Adjusted R-squared	0.094	0.108	0.125	0.168	0.211

Notes: Controls are fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US, as well as 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.

Table B7. Replication of Table 5 with Round 1 Score Threshold of 11 in Experiment 2

OLS Predicting Believed Probability of Promotion					
	All Participants	Unqualified Participants (Round 1 score < 11)	Unqualified Participants (Believed Round 1 score < 11)	Qualified Participants (Round 1 score ≥11)	Qualified Participants (Believed Round 1 score ≥11)
	I	II	III	IV	V
Qualification	-3.55***	-4.90***	-5.21***	-1.41	-0.37
Treatment	(1.32)	(1.84)	(1.62)	(1.89)	(2.26)
Female	-2.31*	-2.74	-3.08*	-1.73	-2.38
	(1.38)	(1.81)	(1.60)	(2.16)	(2.77)
Female x Qual.	3.88**	4.28*	3.94*	3.49	6.71*
Treatment	(1.92)	(2.55)	(2.24)	(2.94)	(3.84)
Controls	Y	Y	Y	Y	Y
Observations	1502	706	920	796	582
Adjusted R-squared	0.429	0.285	0.306	0.408	0.340

Notes: Controls are Round 1 score, beliefs of Round 1 score – absolute and relative, beliefs of Round 2 score, risk preferences, fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US as well as 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.

Table B8. Experiment 2 Replication of Table 8 with Monotonic Participants Only

OLS Predicting Minimum Promotion Bonus at Which Applied MONOTONIC PARTICIPANTS ONLY					
	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
Qualification	-4.71	0.93	13.3	-11.2	-31.7*
Treatment	(13.5)	(22.8)	(19.6)	(16.3)	(17.7)
Female	-7.66	0.63	-7.09	-14.0	-7.67
	(14.2)	(22.1)	(19.0)	(18.3)	(21.1)
Female x Qual.	9.77	10.8	-3.37	9.00	30.4
Treatment	(19.6)	(31.3)	(26.9)	(24.9)	(28.6)
Controls	Y	Y	Y	Y	Y
Observations	1379	635	838	744	541
Adjusted R-squared	0.160	0.077	0.095	0.130	0.115

Notes: Controls are Round 1 score, beliefs of Round 1 score – absolute and relative, beliefs of Round 2 score, risk preferences, fixed effects for each race category, fixed effects for each education category, and a dummy for attended high school in the US as well as 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.

Appendix Table B9. Summary statistics for Qualified Upwork Freelancers in Experiment 3

	Men	**Women**	
Requested Hourly Rate	48.9	39.3	p<0.05
Hours Worked on Upwork	304	540	p=0.14
Jobs Worked on Upwork	11.8	22.0	p<0.05
Total Tests Displayed	6.62	7.62	p=0.29
Available Less than 30hrs/wk	0.14	0.20	p=0.26
Available More than 30hrs/wk	0.36	0.36	p=0.99
Available as Needed	0.49	0.40	p=0.18
College Degree	0.84	0.84	p=0.91
MBA Degree	0.21	0.16	p=0.25
Other Graduate Degree	0.28	0.28	p=0.93
Web/Mobile/Software Development	0.16	0.13	p=0.53
IT & Networking	0.07	0.010	p<0.05
Data Science & Analytics	0.21	0.19	p=0.67
Engineering & Architecture	0.07	0.02	p<0.10
Design & Creative	0.17	0.15	p=0.64
Writing	0.34	0.46	p<0.05
Translation	0.04	0.14	p<0.01
Legal	0.05	0.04	p=0.81
Administrative Support	0.22	0.47	p<0.001
Customer Service	0.02	0.06	p<0.10
Sales & Marketing	0.18	0.18	p=0.91
Accounting & Consulting	0.29	0.20	p<0.10
Analytical Skills Score	4.35	4.30	p=0.20
Time Taken on Analytical Test (minutes)	47.9	48.1	p=0.97
Management Skills Score	3.96	3.95	p=0.58
Time Taken on Management Test (minutes)	19.7	19.6	p=0.97
Proportion in Analytical Skills Dataset	0.33	0.41	p=0.16
N	154	102	

Table B10. Replication of main results excluding the 14 observations who applied to both jobs initially in Experiment 3

	OLS Predicting Decision to Apply to Expert-Level Job						
	All Participants			All Unqualified		All Qualified	
	I	II	III	III	V	VII	VIII
Positive	-0.025	-0.044*	-0.038	-0.047*	-0.069*	0.055	0.046
Treatment	(0.023)	(0.026)	(0.034)	(0.024)	(0.037)	(0.060)	(0.080)
Normative	-0.016	-0.053**	-0.015	-0.056**	-0.030	0.12**	0.027
Treatment	(0.023)	(0.026)	(0.032)	(0.024)	(0.036)	(0.059)	(0.072)
Female	-0.028	-0.025	-0.036	-0.0077	-0.0056	-0.070	-0.16*
	(0.021)	(0.021)	(0.034)	(0.022)	(0.036)	(0.053)	(0.085)
Qualified		-0.051					
		(0.045)					
Positive x		0.075					
Qualified		(0.055)					
Normative x		0.15***					
Qualified		(0.054)					
Female x			0.025		0.041		0.035
Positive			(0.047)		(0.050)		(0.12)
Female x Normative			-0.0025		-0.048		0.26**
			(0.046)		(0.049)		(0.12)
Controls	Y	Y	Y	Y	Y	Y	Y
Observations	1069	1069	1069	816	816	253	253
Adj. R-squared	0.037	0.042	0.035	0.039	0.041	0.020	0.034

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. * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.