Appendix A. Experimental Materials

A.1. Experimental Instructions UpWork Field Experiment (under separate cover)

https://perma.cc/YTQ6-CYUC

A.2. Experimental Instructions Prolific Replication Study (under separate cover)

https://perma.cc/C5CF-P6XF

A.3. Experimental Instructions Real Job Ads Study (under separate cover)

https://perma.cc/2DZK-LHG8

A.4. Experimental Instructions MTurk Simulated Labor Market Study (under separate cover)

https://perma.cc/J33L-LWJM

Appendix B. Additional Tables and Figures

Table B1. Summary Statistics for UpWork Field Experiment

	All Freelancers in Dataset			Qualified Freelancers Only		
	Men	Women	p-value	Men	Women	p-value
Requested Hourly Rate	44.0	30.8	p<0.001	48.9	39.3	p<0.05
Hours Worked on UpWork	323	562	p=0.07	304	540	p=0.14
Jobs Worked on UpWork	13.7	15.6	p=0.49	11.8	22.0	p<0.05
Total Tests Displayed	6.18	7.32	p=0.002	6.62	7.62	p=0.29
Available Less than 30hrs/wk	0.18	0.22	p=0.12	0.14	0.20	p=0.26
Available More than 30hrs/wk	0.44	0.41	p=0.50	0.36	0.36	p=0.99
Available as Needed	0.37	0.34	p=0.36	0.49	0.40	p=0.18
College Degree	0.74	0.72	p=0.53	0.84	0.84	p=0.91
MBA Degree	0.14	0.08	p=0.001	0.21	0.16	p=0.25
Other Graduate Degree	0.20	0.21	p=0.47	0.28	0.28	p=0.93
Web/Mobile/Software Development	0.20	0.08	p<0.001	0.16	0.13	p=0.53
IT & Networking	0.08	0.005	p<0.001	0.07	0.010	p<0.05
Data Science & Analytics	0.18	0.11	p=0.001	0.21	0.19	p=0.67
Engineering & Architecture	0.04	0.01	p<0.001	0.07	0.02	p<0.10
Design & Creative	0.19	0.15	p=0.09	0.17	0.15	p=0.64
Writing	0.32	0.45	p<0.001	0.34	0.46	p<0.05
Translation	0.05	0.06	p=0.41	0.04	0.14	p<0.01
Legal	0.05	0.05	p=0.77	0.05	0.04	p=0.81
Administrative Support	0.25	0.48	p<0.001	0.22	0.47	p<0.001
Customer Service	0.04	0.11	p<0.001	0.02	0.06	p<0.10
Sales & Marketing	0.15	0.16	p=0.89	0.18	0.18	p=0.91
Accounting & Consulting	0.22	0.13	p<0.001	0.29	0.20	p<0.10
Analytical Skills Score	3.73	3.57	p<0.001	4.35	4.30	p=0.20
Time Taken on Analytical Test (minutes)	50.5	48.05	p=0.08	47.9	48.1	p=0.97
Management Skills Score	3.55	3.42	p<0.001	3.96	3.95	p=0.58
Time Taken on Management Test (minutes)	19.79	20.65	p=0.14	19.7	19.6	p=0.97
Proportion Qualified by Test Score	0.29	0.18	p<0.001			
Proportion in Analytical Skills Dataset	0.41	0.45	p=0.2	0.33	0.41	p=0.16
N	531	552		154	102	

Notes: p-values from binary variables are from two-tailed test of proportions. Continuous variables use two-tailed t-tests.

Table B2. Test of Balance for UpWork Field Experiment

	Control		Positive		Normative	
	Men	**Women**	**Men**	**Women**	**Men**	**Women**
Requested Hourly Rate	42 (2.61)	31 (2.24)	45 (3.41)	31 (1.67)	45 (3.28)	30 (1.97)
Hours Worked on UpWork	345 (70.31)	466 (107.46)	353 (94.59)	664 (306.25)	276 (83.69)	548 (117.93)
Jobs Worked on UpWork	11 (1.91)	19 (5.95)	19 (4.1)	17 (2.88)	12 (2.04)	12 (1.66)
Total Tests Displayed	6.1 (0.39)	7.3 (0.46)	6.5 (0.45)	7.4 (0.45)	6 (0.48)	7.2 (0.45)
Available Less than 30hrs/wk	0.13 (0.03)	0.24 (0.03)	0.22 (0.03)	0.22 (0.03)	0.2 (0.03)	0.2 (0.03)
Available More than 30hrs/wk	0.45 (0.04)	0.34 (0.04)	0.45 (0.04)	0.44 (0.04)	0.41 (0.04)	0.47 (0.04)
Available as Needed	0.4 (0.04)	0.38 (0.04)	0.33 (0.04)	0.32 (0.03)	0.37 (0.04)	0.32 (0.04)
College Degree	0.73 (0.03)	0.76 (0.03)	0.72 (0.04)	0.71 (0.03)	0.76 (0.03)	0.69 (0.04)
MBA Degree	0.17 (0.03)	0.083 (0.02)	0.12 (0.03)	0.082 (0.02)	0.12 (0.02)	0.062 (0.02)
Other Graduate Degree	0.2 (0.03)	0.24 (0.03)	0.2 (0.03)	0.16 (0.03)	0.19 (0.03)	0.24 (0.03)
Web/Mobile/Software Development	0.19 (0.03)	0.1 (0.02)	0.21 (0.03)	0.067 (0.02)	0.21 (0.03)	0.073 (0.02)
IT & Networking	0.082 (0.02)	0.0056 (0.01)	0.074 (0.02)	0.0051 (0.01)	0.091 (0.02)	0.0056 (0.01)
Data Science & Analytics	0.19 (0.03)	0.083 (0.02)	0.17 (0.03)	0.13 (0.02)	0.19 (0.03)	0.12 (0.02)
Engineering & Architecture	0.022 (0.01)	0.017 (0.01)	0.037 (0.01)	0.01 (0.01)	0.07 (0.02)	0.0056 (0.0057)
Design & Creative	0.18 (0.03)	0.15 (0.03)	0.2 (0.03)	0.12 (0.02)	0.18 (0.03)	0.18 (0.03)
Writing	0.31 (0.03)	0.43 (0.04)	0.35 (0.04)	0.48 (0.04)	0.3 (0.03)	0.45 (0.04)
Translation	0.038 (0.01)	0.061 (0.02)	0.061 (0.02)	0.041 (0.01)	0.043 (0.01)	0.085 (0.02)
Legal	0.038 (0.01)	0.044 (0.02)	0.043 (0.02)	0.051 (0.02)	0.054 (0.02)	0.051 (0.02)
Administrative Support	0.23 (0.03)	0.53 (0.04)	0.29 (0.04)	0.46 (0.04)	0.23 (0.03)	0.46 (0.04)
Customer Service	0.044 (0.02)	0.1 (0.02)	0.031 (0.01)	0.11 (0.02)	0.043 (0.02)	0.11 (0.02)
Sales & Marketing	0.18 (0.03)	0.13 (0.03)	0.13 (0.03)	0.16 (0.03)	0.15 (0.03)	0.18 (0.03)
Accounting & Consulting	0.21 (0.03)	0.13 (0.03)	0.25 (0.03)	0.11 (0.02)	0.2 (0.03)	0.14 (0.03)
Analytical Skills Score	3.72 (0.05)	3.61 (0.06)	3.67 (0.07)	3.54 (0.05)	3.69 (0.07)	3.54 (0.05)
Minutes on Analytical Test	53 (1.75)	46 (1.6)	50 (1.8)	47 (1.6)	49 (1.61)	51 (1.75)
Management Skills Score	3.6 (0.04))	3.4 (0.03)	3.5 (0.04)	3.5 (0.03)	3.6 (0.03)	3.4 (0.03)
Minutes on Mgmt. Test	21 (0.77)	21 (0.74)	20 (0.62)	21 (0.8)	19 (0.6)	20 (0.77)
Prop. Qualified by Test Score	0.30 (0.034)	0.20 (0.030)	0.25 (0.034)	0.20 (0.028)	0.32 (0.034)	0.16 (0.027)
Prop. Analytical Skills Dataset	0.45 (0.036)	0.42 (0.037)	0.36 (0.038)	0.48 (0.036)	0.42 (0.036)	0.45 (0.037)
Frequency	182	180	163	195	186	177

Notes: Standard errors displayed in parentheses.

Table B3. Replication of UpWork Table 1 Using Logistic Regression

	Logistic Regression Predicting Decision to Apply to Expert Job						
	All Participants			All Unqualified		All Qualified	
	I	II	III	IV	V	VI	VII
Positive	-0.25	-0.45	-0.37	-0.54*	-0.74*	0.37	0.041
Treatment	(0.24)	(0.28)	(0.33)	(0.29)	(0.44)	(0.48)	(0.57)
Normative	-0.28	-0.78**	-0.27	-0.85***	-0.57	0.87*	0.14
Treatment	(0.24)	(0.30)	(0.32)	(0.31)	(0.43)	(0.46)	(0.54)
Female	-0.31	-0.28	-0.37	-0.10	-0.070	-0.68	-2.37**
	(0.22)	(0.22)	(0.33)	(0.27)	(0.39)	(0.44)	(1.00)
Qualified		-0.66					
		(0.44)					
Positive x		0.63					
Qualified		(0.52)					
Normative x		1.42***					
Qualified		(0.51)					
Female x			0.24		0.37		1.62
Positive			(0.48)		(0.58)		(1.21)
Female x Normative			-0.031		-0.63		2.79**
			(0.48)		(0.64)		(1.16)
Controls	Y	Y	Y	Y	Y	Y	Y
Observations	1083	1083	1083	827	827	242	242

Notes: Qualified candidates 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 (> 30 hrs/wk, < 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.

Table B4. Replication of UpWork Table 1 excluding the 14 observations who applied to both jobs initially

	OLS Predicting Decision to Apply to Expert 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 candidates 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 (> than 30 hrs/wk, < 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.

Table B5. Summary Statistics for Prolific Replication Study

	Full Dataset			Qualified Participants Only		
	Men	Women	p-value	Men	Women	p-value
Year of Birth	1984	1987	p<0.001	1985	1990	p<0.001
Attended high school in US	0.97	0.98	p=0.72	0.98	0.98	p=0.73
White	0.79	0.82	p=0.16	0.80	0.84	p=0.07
Black or African American	0.068	0.081	p=0.23	0.044	0.052	p=0.49
Asian	0.12	0.099	p=0.11	0.14	0.12	p=0.25
Latino or Latina	0.069	0.071	p=0.87	0.064	0.056	p=0.51
Indigenous	0.020	0.010	p=0.04	0.013	0.007	p=0.25
No Formal Education	0.010	0.005	p=0.23	0.009	0.005	p=0.43
Secondary Education	0.015	0.016	p=0.84	0.012	0.008	p=0.56
High School Education	0.24	0.28	p=0.03	0.22	0.30	p<0.01
Community or Technical College Education	0.12	0.13	p=0.26	0.091	0.088	p=0.85
College Education	0.39	0.39	p=0.83	0.43	0.38	p=0.09
Graduate Degree	0.19	0.15	p=0.02	0.20	0.18	p=0.52
Education Missing	0.042	0.031	p=0.20	0.047	0.041	p=0.58
Approval Rate	0.996	0.996	p=0.67	0.997	0.997	p=0.72
Total Jobs	718	533	p<0.001	709	476	p<0.001
ASVAB Score	5.96	5.64	p<0.001	7.21	6.99	p<0.001
Positive Treatment	0.30	0.32	p=0.32	0.29	0.34	p=0.11
Normative Treatment	0.36	0.33	p=0.09	0.36	0.31	p=0.06

Notes: p-values from binary variables are from two-tailed test of proportions. Continuous variables use two-tailed t-tests.

Table B6. Balance Check for Prolific Replication Study

	Control		Positive		Normative	
	Men	Women	Men	Women	Men	Women
Year of Birth	1984	1987	1984	1987	1984	1986
Attended high school in US	0.98	0.98	0.97	0.98	0.98	0.97
White	0.77	0.83	0.81	0.78	0.79	0.83
Black or African American	0.066	0.066	0.064	0.086	0.074	0.094
Asian	0.13	0.097	0.099	0.11	0.13	0.088
Latino or Latina	0.081	0.084	0.07	0.078	0.057	0.05
Indigenous	0.024	0.018	0.02	0.0055	0.017	0.0055
No Formal Education	0.018	0.0051	0.0058	0.0083	0.005	0.0028
Secondary Education	0.0079	0.018	0.026	0.014	0.012	0.017
High School Education	0.23	0.3	0.22	0.24	0.25	0.28
Community or Technical College Education	0.1	0.13	0.13	0.13	0.12	0.14
College Education	0.39	0.35	0.4	0.43	0.39	0.39
Graduate Degree	0.19	0.16	0.19	0.15	0.18	0.14
Education Missing	0.055	0.033	0.029	0.028	0.04	0.033
Approval Rate	1	1	1	1	1	1
Total Jobs	687	564	714	519	749	514
Avg ASVAB Score	5.9	5.6	6	5.7	6	5.5
Proportion Qualified	0.61	0.53	0.59	0.55	0.61	0.51
N	381	392	343	361	404	362

Table B7. Replication of Prolific Table 2 Using Logistic Regression

	Logistic Regression Predicting Decision to Apply to Expert Job						
	All Participants			All Unqualified		All Qualified	
	I	II	III	IV	V	VI	VII
Positive	0.53***	-0.12	0.42**	-0.085	-0.36	0.92***	0.94***
Treatment	(0.12)	(0.20)	(0.17)	(0.20)	(0.28)	(0.15)	(0.22)
Normative	0.42***	-0.53**	0.42***	-0.52**	-0.57**	0.93***	0.97***
Treatment	(0.12)	(0.21)	(0.16)	(0.21)	(0.28)	(0.15)	(0.21)
Female	-0.53***	-0.54***	-0.61***	-0.42**	-0.64**	-0.58***	-0.54***
	(0.099)	(0.10)	(0.17)	(0.17)	(0.27)	(0.13)	(0.20)
Qualified		0.024					
Indicator		(0.22)					
Positive x		1.02***					
Qualified		(0.25)					
Normative x		1.42***					
Qualified		(0.25)					
Female x			0.24		0.57		-0.048
Positive			(0.24)		(0.39)		(0.30)
Female x			0.00058		0.11		-0.078
Normative			(0.24)		(0.41)		(0.30)
Controls	Y	Y	Y	Y	Y	Y	Y
Observations	2243	2243	2243	967	967	1276	1276

Notes: Qualified candidates are those with a test score greater than or equal to the advertised threshold (5.5). Controls are ASVAB test score, Prolific approval rate, total studies completed on Prolific, indicator for college as highest education attainment, indicator for graduate degree as highest educational attainment, indicator for community or technical school education, indicator for Black or African American, indicator for Latino or Latina, indicator for Asian, indicator for attending high school in the United States, and year of birth. * indicates $p<0.10$, ** indicates $p<0.05$, ***indicates $p<0.01$, ****indicates $p<0.001$.

Table B8. Replication of Prolific Table 2 Restricted to Only Candidates "Close" to Cutoff

	OLS Predicting Decision to Apply to Expert Job						
	Participants with Scores of 5 or 6			Participants with Scores of 5		Participants with Scores of 6	
	I	II	III	IV	V	VI	VII
Positive	0.078*	-0.020	0.030	-0.011	-0.091	0.19***	0.18**
Treatment	(0.040)	(0.054)	(0.059)	(0.051)	(0.077)	(0.059)	(0.085)
Normative	0.085**	-0.12**	0.082	-0.11**	-0.19**	0.26***	0.26***
Treatment	(0.039)	(0.055)	(0.058)	(0.051)	(0.080)	(0.056)	(0.077)
Female	-0.068**	-0.044	-0.099*	-0.026	-0.12	-0.066	-0.072
	(0.033)	(0.032)	(0.057)	(0.042)	(0.073)	(0.048)	(0.081)
Qualified		0.046					
Indicator		(0.054)					
Positive x		0.21***					
Qualified		(0.077)					
Normative x		0.37***					
Qualified		(0.075)					
Female x			0.089		0.14		0.030
Positive			(0.081)		(0.10)		(0.12)
Female x			0.0061		0.13		-0.0085
Normative			(0.079)		(0.10)		(0.11)
Controls	Y	Y	Y	Y	Y	Y	Y
Observations	877	877	877	430	430	447	447
	0.016	0.103	0.016	0.006	0.006	0.052	0.048

Notes: Qualified candidates are those with a test score greater than or equal to the advertised threshold (5.5). Controls are Prolific approval rate, total studies completed on Prolific, indicator for college as highest education attainment, indicator for graduate degree as highest educational attainment, indicator for community or technical school education, indicator for Black or African American, indicator for Latino or Latina, indicator for Asian, indicator for attending high school in the United States, and year of birth. * indicates p<0.10, ** indicates p<0.05, ***indicates p<0.01, ****indicates p<0.001.

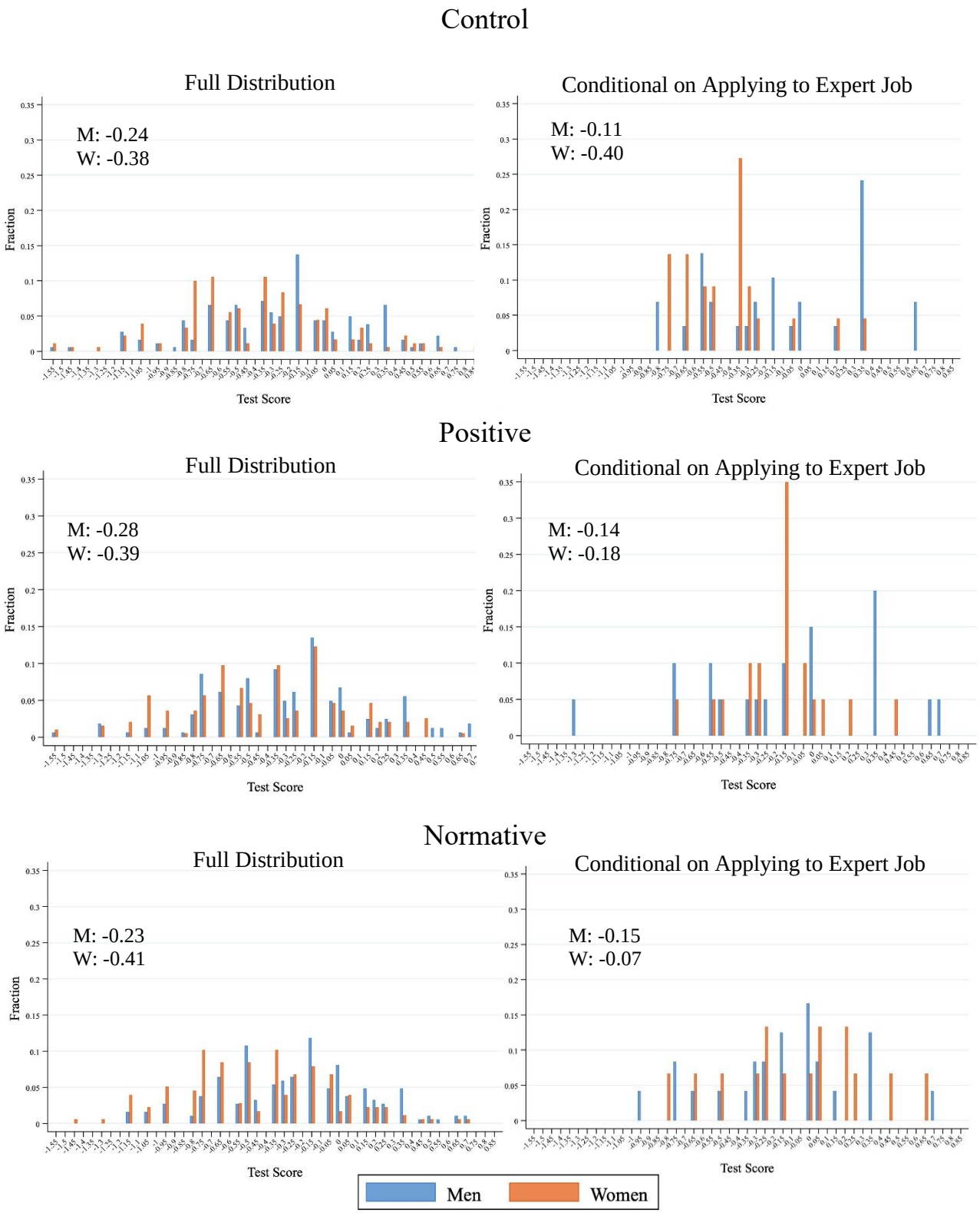

Figure B1. Distributions of Test Scores by Treatment for UpWork

Notes: Test scores have been normalized by subtracting out the threshold for being Qualified. Means are presented for men (M) and women (W) within each panel.

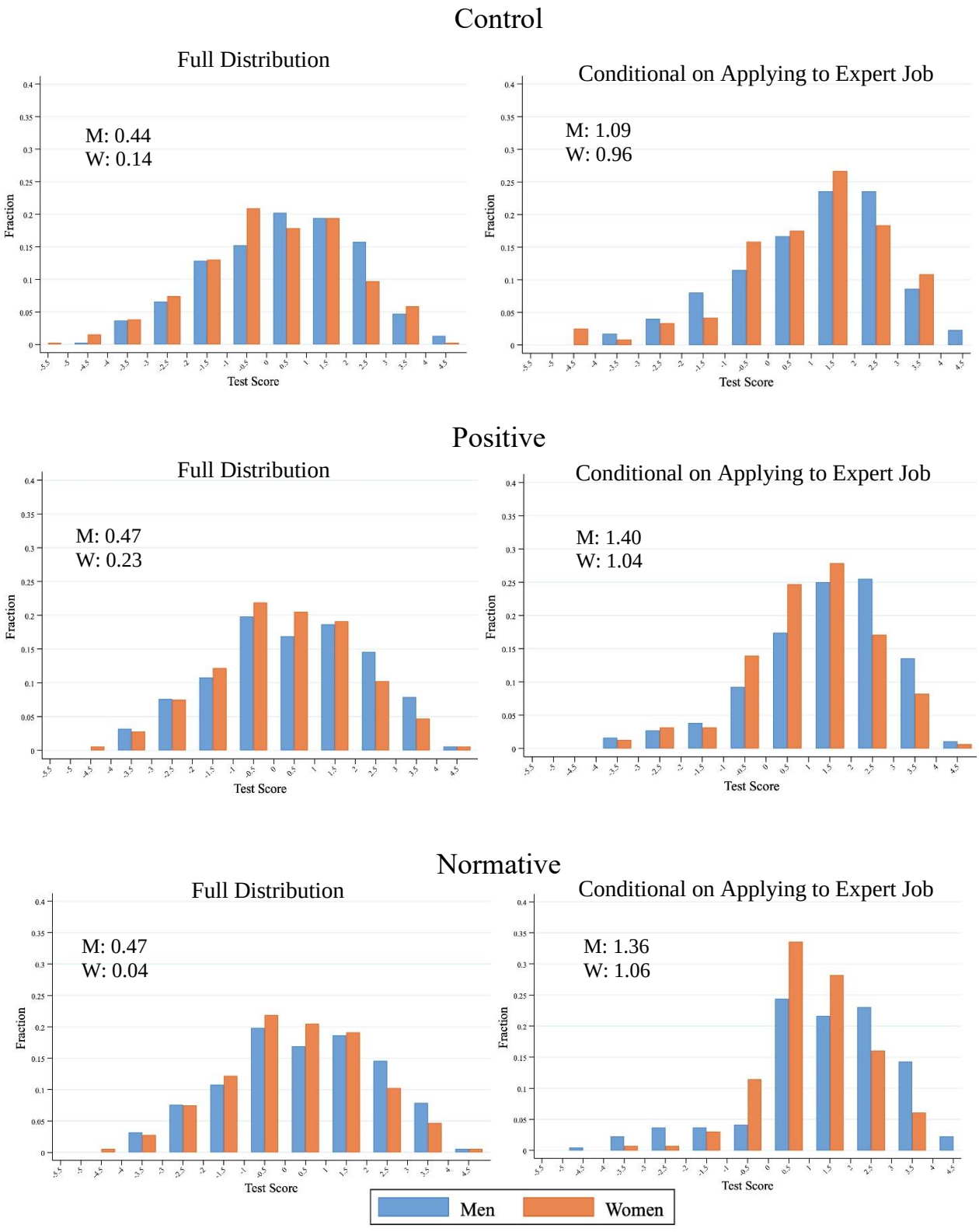

Figure B2. Distributions of Test Scores by Treatment for Prolific

Notes: Test scores have been normalized by subtracting out 5.5, the threshold for being Qualified. Means for men (M) and women (W) are presented within each panel.

Figure B3. Beliefs of "The Bar" on Prolific

Appendix C. Additional Studies

Appendix C provides detailed report of the two additional experiments we ran. In the interest of space and cohesion, we omitted these studies from the main text. When we made the decision to run the pre-registered replication of the UpWork study in Fall 2021, we pre-committed to reporting that follow-up in the main text and moving these studies to the Appendix. Please see our pre-registration for more details (AEARCTR-0008223).

C1. Impressions of Real Job Advertisements

In this simple laboratory study, we seek to document the beliefs potential job-seekers' hold about how well-qualified they are for different openings. By varying the job advertisements, we ask how beliefs about qualification level vary with different features, including how objective, specific, and clear the desired qualifications are.

In April 2018, we constructed a random sample of real job postings from Indeed.com, representative of available online postings in the geographic area of our participants. 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.

We use these ads to collect beliefs from participants at the Computer Lab for Experimental Research (CLER) at Harvard Business School. In the first part of the experiment, we collect participant impressions of how well-qualified they feel 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.[1] We limit participants' time viewing the ads to ensure timely completion of the experiment, as completion in under 20 minutes was required for the laboratory format we took advantage of.

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?*

[1] 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.

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 provides a more quantitative assessment of those beliefs. The third question encourages participants to reflect on the qualifications for this particular ad, and provides insight into the types of skills participants list.

We also ask two "decoy" questions of our participants:

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

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

We include these questions so that participants do not become solely focused on qualifications as they read additional 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 volunteer ideas is dependent on the gender congruence of the domain. If beliefs of own aptitude 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 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 perceived prestige of the position.

The third question 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. Finally, the fourth question speaks to the hypothesis tested in the UpWork experiment: does the amount of ambiguity surrounding the desired qualifications matter for beliefs of how well-qualified individuals feel? In particular, does ambiguity contribute to a gender gap in these beliefs?

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.[2] We provide summary statistics on our participants and our job ads in Tables C1 and C2, respectively. More than 80% of our sample identifies themselves as a current student.

Table C1. Summary Statistics for Laboratory Participants for Real Job Ads Study

	Men	**Women**	**P-value from test of proportions**
White	0.28	0.32	0.51
Black or African American	0.16	0.11	0.34
Asian	0.35	0.34	0.93
Latino or Latina	0.08	0.07	0.90
Multiracial	0.12	0.09	0.48

[2] 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.

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

Notes: p-values from binary variables are from two-tailed test of proportions. Continuous variables use two-tailed t-tests.

Table C2. Summary Statistics on Job Ads

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 of Job Ads					
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

We find that, on average, men view themselves as marginally more well-qualified than women in our sample. 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 C3, we present a regression that explores the determinants of these ratings. To increase interpretability, we create z-scores for the variables that were elicited with a scale. Controlling for 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, believed qualifications of others, and objectivity of qualifications. For each ad, we take the average of the ratings provided by all raters who saw that ad for each characteristic.[3] 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 C3. 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)
Male Stereotype (z-score)		-0.25***	-0.23**		-8.78***	-8.58***
		(0.081)	(0.090)		(2.24)	(2.46)
Prestige (z-score)		0.090	0.15**		2.73*	4.21**
		(0.055)	(0.071)		(1.49)	(1.84)
Objectivity of Stated Qualifications (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)
Female x Male Stereotype			-0.034			-0.45

[3] 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.

			(0.068)			(1.94)
Female x Prestige			-0.11			-2.53
			(0.072)			(1.99)
Female x Objectivity			0.17**			3.54
			(0.077)			(2.26)
Female x Avg. of Avg. Qualified			0.030			2.15
			(0.079)			(2.15)
Entry Level 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 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 indicator, to ask whether any of the characteristics of the job opening impact the gender gap in believed qualification level. We find a role for ambiguity in shaping the gender gap. We find that more objectively stated job qualifications, as rated by our participants, reduce the gender gap in perceived qualification level.

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 in terms of the fraction of qualifications possessed; 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.

C2. MTurk Simulated Labor Market Study

In this online experiment, we examine both beliefs and application decisions in a simulated labor market. We exogenously vary the level of ambiguity in stated qualifications surrounding a promotion opportunity, and explore its impact on gender gaps in beliefs and behavior.

Overview of Design

We conduct our experiment on Amazon Mechanical Turk (MTurk).[4] 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 C1. 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, matching our field setting. Participants have 5 minutes to answer up to 30 questions. All 30 questions appear on the same page and can be answered in any order. If Round 1 is chosen for payment, participants receive $0.20 per question answered correctly.

Beliefs of Round 1 Performance

After completing Round 1, each participant is asked to guess their score –how many problems she solved correctly in Round 1 – and how they rank relative to other MTurkers who are completing the HIT. They

[4] 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.

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%).

We note that all workers 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 and its impact on beliefs is the focus of a different paper, Coffman, Collis, and Kulkarni (2021). For our purposes, the only important thing to note that is that our measure of beliefs of Round 1 score in our analysis below will be the *posterior* beliefs of these participants: the beliefs they hold after receiving the information. In principle, this could work against us finding large gender gaps in beliefs or application decisions. But, Coffman, Collis, and Kulkarni (2021) show that gender gaps in prior and posterior beliefs in this setting are actually quite similar, suggesting the inclusion of this feedback treatment is not significantly impacting our conclusions in a meaningful way.

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 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 C2, we reproduce the price list used to elicit these decisions (full materials are available in Appendix A).

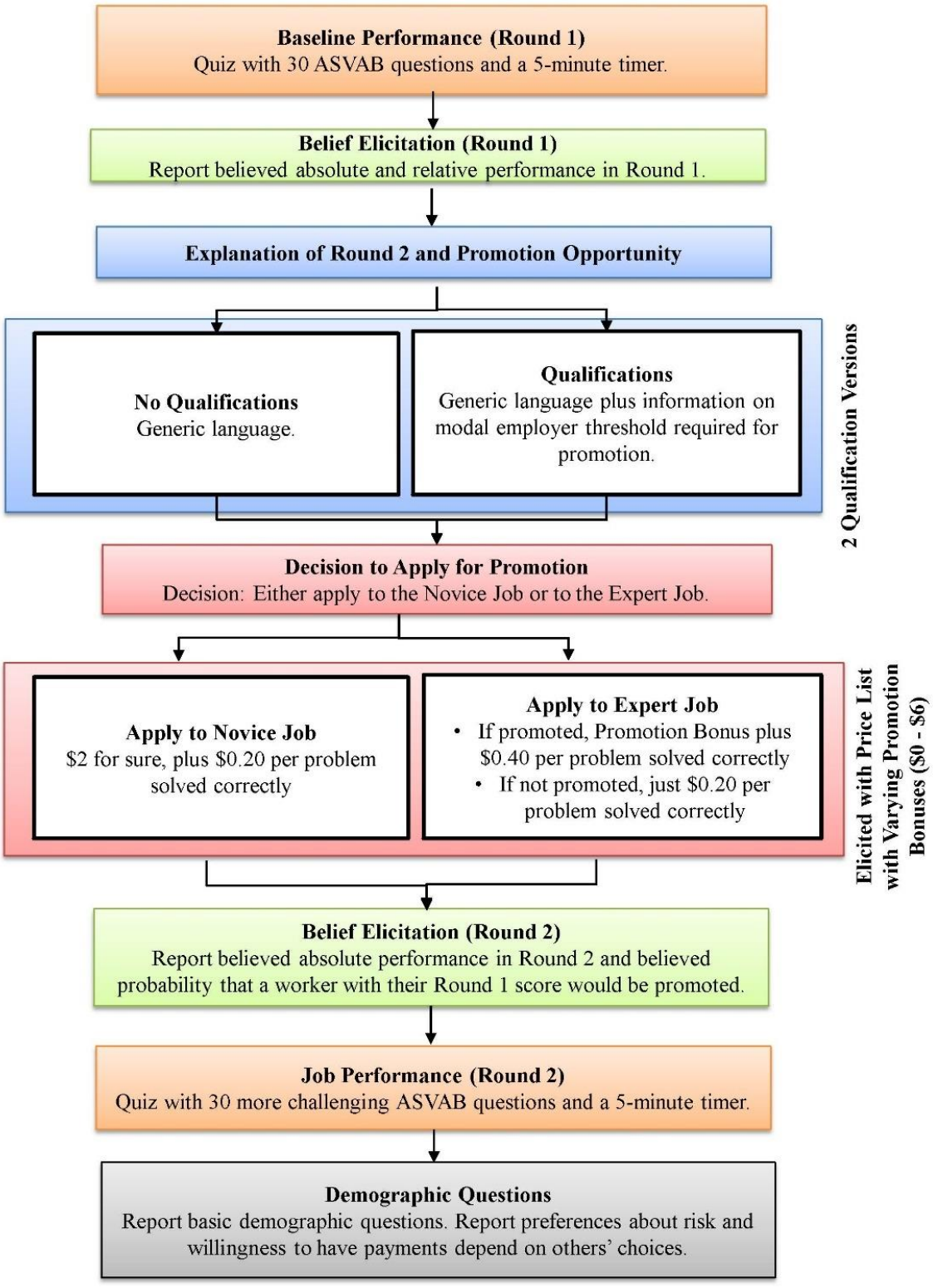

Figure C1. Design of Worker Experiment

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.

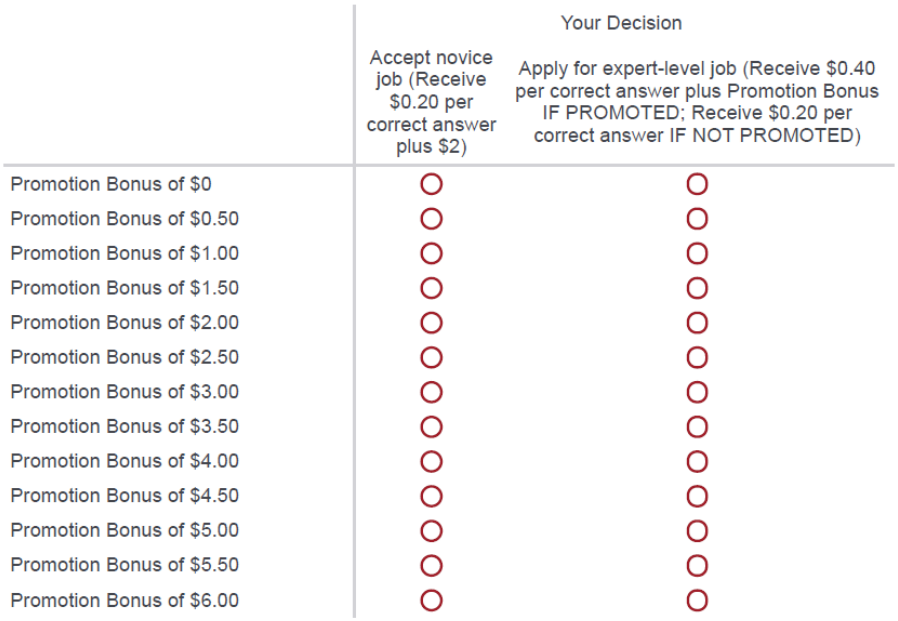

Figure C2. Price List Used to Elicit Application Decisions

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 explain the worker experiment to them.

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. The employers are not provided with any other information such as gender. They make a series of binary decisions, covering all possible Round 1 scores.

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.[5]

Treatment Intervention

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.

[5] 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. 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. Both workers and employers have complete information on this process. See Appendix A for full instructions.

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."[6] 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. They receive $0.10 if they guess their Round 2 score exactly correctly. 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 speaks directly to their beliefs of how well-qualified they are.

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.

Results

The experiment was conducted in May 2018 with 1,502 workers. Table C4 provides summary statistics on the workers. We control for the demographic variables collected in the analysis that follows. Men outperform women on average in Round 1: 10.96 versus 9.65 problem solved correctly ($p<0.001$). Men on average rank in the 54^{th} percentile, while women rank in the 46^{th} percentile on average ($p<0.001$). Note that

[6] 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.

given the employer decisions and candidate performance, the average chance of receiving the promotion, conditional on applying, is 44%.

Table C4. Summary Statistics on Workers in the Online Simulated Labor Market

	All Participants			Qualified Participants Only		
	Men	Women	P-value	Men	Women	P-value
White	0.80	0.81	0.65	0.82	0.86	0.15
Black	0.06	0.09	0.08	0.04	0.05	0.44
Asian	0.10	0.06	0.01	0.11	0.07	0.03
Attended HS in US	0.98	0.96	0.05	0.98	0.97	0.16
HS Only	0.11	0.085	0.06	0.09	0.06	0.17
Some College/Assoc.	0.36	0.37	0.86	0.32	0.29	0.26
Bachelors	0.39	0.40	0.76	0.42	0.47	0.12
Advanced Degree	0.14	0.15	0.36	0.17	0.18	0.76
Rd. 1 Score	10.96	9.65	<0.001	14.0	13.2	<0.01
Rd. 2 Score	8.44	7.14	<0.001	9.73	8.55	<0.001
Prop. Assigned to Qualifications Treatment	0.49	0.50	0.78	0.50	0.53	0.44
N	798	704		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.

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 applying. 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 C5, 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, $p<0.01$). Of course, conditional on Round 1 performance, *true* likelihood of being promoted is the same.

Table C5. 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.

Why do women feel they are less likely to be promoted conditional on applying? Part of it seems related to beliefs about own aptitude. 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, ($p<0.01$), and believes she places 7.2pp worse in the distribution of performers ($p<0.01$) (see Appendix Table B7). In Table C5, 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 predictive of beliefs about promotion probability, they do not explain much of the gender gap, which remains 6pp ($p<0.01$).

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 candidate'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."

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 C6 predicts the minimum promotion bonus at which someone was willing to apply for promotion for our No Qualifications treatment. Conditional on Round 1 performance, there are no significant gender differences in willingness to apply (Column I). We have focused on beliefs of how well-qualified one is, but conceptually, there are several other factors that could also influence willingness to apply in this setting. In particular, risk preferences and expected returns to promotion (i.e. beliefs of Round 2 score) may matter. In Column II, we add these factors to the regression to explore the role they play in shaping application decisions. As expected, each of these factors predict willingness to apply. Finally, in Column III, 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. It could be that there are other meaningful factors that influence application decisions that we are not capturing, or that we are measuring these decisions with too much noise. We also collected data on a few other explanations. 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.

Table C6. Willingness to Apply for Promotion

OLS Predicting Minimum Acceptable Promotion Bonus *No Qualifications Treatment*			
	I	II	III
Female	9.57	-12.3	73.3*
	(15.4)	(14.8)	(38.8)
Round 1 Score	-9.51***	-3.06	-3.04
	(1.71)	(2.01)	(2.01)
Believed Probability of Promotion		-2.40***	-1.68***
		(0.34)	(0.45)
Took Common		-76.5***	-67.0***
Risk Gamble		(14.2)	(19.4)
Beliefs of		-4.73*	-4.26
Round 2 Score		(2.53)	(2.94)
Female x			-1.53**
Believed Prob. of Prom.			(0.67)
Female x			-24.2
Risk Gamble			(28.7)
Female x Belief			-1.20
of Round 2 Score			(4.50)
Controls	Y	Y	Y
Observations	759	759	759
Adjusted R-squared	0.044	0.142	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 Reducing Ambiguity Narrow the Gender Gap?

In Table C7, we ask how more information on where the bar is 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, a clearer bar directionally increases 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 better communicating where the bar is (see Column II), not through changing beliefs of own ability. Thus, overall, more information significantly reduces 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 participants. We do this both by actual Round 1 score (i.e. true score above the threshold, Columns II and IV), and believed Round 1 score (i.e. believed score above the threshold, Columns III and V).

As we expect, the qualifications treatment reduces the believed probability of being promoted significantly among the unqualified group. Qualified men's beliefs about their likelihood of receiving the promotion are directionally lower in the treatment than in the control, while qualified women's beliefs increase significantly, eliminating the gender gap.

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 C8 replicates Table C7, but using the behavioral dependent variable. 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 treatment directionally increases the gender gap in willingness to apply.

Table C7. 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.

Table C8. 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. 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.