Signaling Confidence

Elif E. Demiral[*] Johanna Mollerstrom[†]

December 2023

We study gender differences in confidence, and the impact of signaling confidence, on employment likelihood in an experimental hiring market. We document that moderate, as opposed to low, levels of confidence enhance the chances of being hired, while excessive confidence, characterized by high performance estimates and high certainty, diminishes employment prospects. Men display higher levels of confidence than women, and this gender gap widens in forward-looking scenarios, where performance estimates are provided ex ante rather than ex post. These findings serve as a cautionary reminder of the potential negative consequences of excessive confidence. Hence, encouraging both men and women to signal performance predictions that are aligned with reality instead of simply advising women to be more confident may well be a safer way to foster equity in labor market outcomes.

Keywords: gender, confidence, employment, experiment

JEL codes: C90, D90, J70, M51

We are grateful to seminar and conference participants at the Harvard Kennedy School, 2024 ASSA Meetings, and 2022 Economic Science Association North American Meetings for excellent comments and suggestions. We gratefully acknowledge funding for the experiments from AJ Taylor Distinguished Professorship Awards at Austin Peay State University.

[*]College of Business, Austin Peay State University, 601 College St., Clarksville, TN 37044, USA. E-mail: demiralsaglame@apsu.edu. Corresponding author.

[†]Interdisciplinary Center for Economic Science, George Mason University, 4400 University Dr., Fairfax, VA 22030, USA, and Research Institute of Industrial Economics (IFN), Stockholm, Sweden. E-mail: jmollers@gmu.edu.

1. Introduction

Gender disparities in economic outcomes persist despite ongoing efforts to promote equality with women frequently receiving lower wages, and encountering restricted career prospects. This gender imbalance can, among other things, be linked to gender difference in traits like willingness to ask (Hernandez-Arenaz and Iriberri, 2019; Exley et al., 2020) and compete (Niederle and Vesterlund, 2007). These traits, in turn, are believed to be related to a person's confidence (Niederle and Vesterlund, 2007; Charness et al., 2022). Extensive research does indeed highlight gender differences in confidence, with women generally displaying lower confidence than men. Women have been found to be more prone than men to underestimate their abilities (Deaux and Farris, 1977; Lenney, 1977; Coffman, 2014; Mobius et al., 2014; Bordalo et al., 2019), attribute success to external factors (Apicella et al., 2020; Shastry et al., 2020), engage less in self-promotion (Exley and Kessler, 2022) and fail to take credit for their accomplishments (Isaksson, 2018; Sarsons et al., 2021). These patterns often emerge in educational and professional contexts, where opportunities to establish and signal confidence are prevalent.

We are interested in how individuals communicate their own confidence as a way to signal skills in the job market. Thereby we add the perspective of employers' *demand* for confidence to an existing body of research that primarily focuses on the *supply* side. Previous work has shown that the gender confidence gap is associated with lower wage expectations (Reuben et al., 2017), a lower likelihood of engaging in competitive environments (Buser et al., 2014), and a reduced propensity to speak up (Coffman, 2014), which can all be significant contributors of systemic inequalities in the workforce (Blau and Khan, 2017; Bohnet et al., 2021). Encouraging women to lean in and display greater confidence is a common strategy to

mitigate this gap, but the labor market consequences of such signaling of confidence are understudied.

To investigate the dynamics of confidence and the implications of signaling confidence for hiring decisions, we use a simulated hiring environment. We run a controlled experiment involving two distinct groups of participants: workers and firms. The primary objective is to understand if and how male and female workers utilize confidence as a signaling tool, and to determine the influence that displaying confidence at different levels has on their likelihood of being hired.

Our treatments vary along two dimensions. The first treatment variation distinguishes between backward- and forward-looking confidence. In the backward-looking treatments, workers predict (and report their confidence in) their performance in a preceding task. In contrast, the forward-looking treatments require individuals to forecast their expected performance before undertaking the task. While existing research on confidence primarily focuses on the backward version (Bordalo et al., 2019), forward-looking confidence may well be even more relevant in certain contexts, such as during a job search, where applicants are expected to assess and project their future performance based on job requirements. Utilizing job postings with clear performance expectations is a common practice to attract applicants with suitable abilities (Leon, 2016), and in addition to reflecting on one's previous performance, job candidates are often expected to emphasize their expected future performance in job applications (Harvard Business Review, 2020). As the job search and recruitment processes involve signaling of both confidence measures, comparing how men and women establish confidence under these two conditions can shed light on phenomena we observe in the labor market, such as the gender disparities in the likelihood of applying for jobs (Coffman et al., 2023).

The second dimension along which we vary our experimental treatments regards whether confidence is assessed with or without the knowledge that the information will be available to recruiting firms. This allows us to understand how individuals display confidence with and without public observability.

Participants acting as workers in our experiment perform a simple counting task, and we measure their confidence in their performance in this task along different dimensions. This approach draws on interdisciplinary research (c.f. Fischhoff et al., 1977; Moore and Healy, 2008) and allows for a more comprehensive understanding of confidence. In economics, confidence is often assessed merely by having participants predict the number of correct answers they provide in a task. Participants who overestimate (underestimate) their performance relative to their actual performance are typically classified as overconfident (underconfident). While this is insightful regarding individuals' self-evaluation of performance, it may overlook other important aspects of confidence.

Our measure of confidence is inspired by Moore and Healy (2008) who provide a comprehensive framework that encompasses three dimensions of (over)confidence: performance estimation, performance comparison with others, and precision in beliefs. For instance, a person may appear excessively confident in their performance estimation but not have a strong sense of certainty regarding that predicted number. On the other hand, another person reporting the same estimated performance with greater precision could be considered more confident, according to this augmented measure (Fischhoff et al., 1977). Hence, while the predicted number of tasks solved, by oneself and by others, constitutes our first confidence measure, the degree of certainty in estimated performance (Hangya et al., 2016; Pouget et al., 2016) is our second measure. Additionally, we ask participants to provide an interval within which they believe their

performance will fall. This establishes our third measure of confidence, allowing us to capture the range of performance expectations.

We recruit our second set of participants to act as firms, which enables us to explore the demand for confidence in a hiring process. The firms are tasked with making one hiring decision per screen on five screens. On every screen two workers are presented, who differ in one aspect of their confidence and/or gender.

We report two main results. First, not only do women consistently exhibit lower levels of confidence in their performance than men, but the gender difference in confidence becomes even more pronounced in the forward looking treatments when workers predict their performance before engaging in the task. This, in combination with the fact that job listings generally involve forward looking performance expectations, may be part of the explanation for why men tend to apply for jobs more frequently than equally qualified women (cf. Coffman et al., 2023).

Second, while moving from a low to a moderate level of confidence increase the likelihood of employment, workers who display excessive levels of confidence in the sense of signaling high performance estimates with high certainty, are least likely to be hired. While this trait of excessive confidence is more commonly exhibited by men, it is occasionally exhibited also by women. This finding highlights the potential negative consequences of encouraging women to adopt stereotypically male-like traits, such as displaying higher levels of confidence. An alternative policy could be to encourage men and women to reflect and reevaluate, and moderate their confidence to better reflect their true capabilities.

Our research contributes to several strands of literature. To start, we corroborate the findings of Exley and Kessler (2022), who identify a robust gender gap in subjective self-

promotion. Additionally, we contribute to a commencing body of work examining the perceptions of confidence and its influence on hiring decisions. Here, Falk and Zimmermann (2017) study the role of consistency as a signal of skills. They document that subjects who are more consistent in their predictions of an outcome are indeed more skilled and that the consistency in itself is valued by the subjects acting as firms. Our study also complements a concurrent project undertaken by Exley and Nielsen (2022), focusing on the impact of the gender confidence gap on the demand for female workers. They find that subjects acting as evaluators in an online experiment fail to account for the gender confidence gap in their hiring choices, even when they anticipate the presence of it.

We also contribute to the research investigating the potential backlash associated with assertive traits in the labor market. A seminal study by Bowles et al. (2007) illuminates how women may face penalties when attempting to negotiate, a phenomenon contributing to their reluctance to ask for more. In a related vein, Exley et al. (2020) document how "lean in" policies to promote negotiations among women may not be beneficial, as women strategically choose when to ask and when to refrain from doing so. Finally, Demiral and Mollerstrom (2022) investigate another well-documented gender difference – women's lower inclination to compete – and examine the implications of competitiveness on employability. The findings demonstrate that candidates displaying a willingness to compete against others are perceived less favorably by evaluators and are less likely to be hired by firms, than candidates who primarily compete against themselves. This current paper also contributes to these findings highlighting the potential harm in advocating policies that encourage women to adopt behaviors traditionally associated with men.

In sum, our study takes a multifaceted approach to understanding confidence and assesses its role in generating gender gaps and its impact on employment likelihood. We utilize a range of confidence measures and by studying both the supply of and the demand for confidence through a hiring experiment, we provide a comprehensive perspective on the dynamics at play. From here, the paper proceeds as follows: Section 2 outlines the experimental design and procedures, whereas Sections 3 and 4 present the results for workers and firms, respectively. Section 5 contains robustness checks and Section 6 concludes.

2. Experimental Design and Procedures

We recruited 1,618 participants from the online labor platform Prolific in April 2022. Of these, 797 subjects acted as workers and 821 subjects acted as firms. The selection criteria required participants to be residents of the UK or the USA, with a minimum approval rate of 90 percent. We implemented the experiment using Qualtrics. Each participant received a fixed fee for their participation[1] plus a possible bonus payment determined by their decisions in the experiment. Participants reported their gender at the beginning of the study to prevent misreporting and we integrated the gender question with other demographic questions to avoid priming participants. After completing the experiment, participants answered a survey and provided self-reported risk and competition preferences. The instructions for the experiment can be found in Appendix A.

The experiment involved two groups of participants representing different sides of a labor market: "workers" and "firms." The workers completed a simple counting task and provided self-evaluations of their performance. The firms, on the other hand, were tasked with making a

[1] Participants acting as workers received $1 for their participation in a 6-minute-long study. Firms were paid $0.85 for a 5-minute-long study. After the bonus earnings, this corresponded to a minimum average hourly earning of $12 for both groups.

total of five hiring decisions among pairs of workers who systematically differed in gender and/or in their certainty of their performance estimates.

Workers revealed their confidence in their own and others' performance by answering multiple questions. This method involved participants predicting their own (and others') performance and also indicating their degree of certainty in those predictions. We utilized this augmented technique for two reasons. First, while using multiple measures of confidence is underutilized in economic experiments, it has been widely employed in fields such as psychology and neuroscience. Drawing on this established methodology (c.f. Fischhoff et al., 1977; Hangya et al., 2016; Pouget et al., 2016) allowed us to capture a broader categorization of confidence. With these multiple measures, for example, a decision-maker can be classified as *excessively confident* when they simultaneously overestimate performance and report high precision in that estimate (Fischhoff et al., 1977; Moore and Healy, 2008). Second, utilizing multiple measures of confidence was a deliberate design choice for our firm study, where we examined perceptions of confidence. In this context, firms' hiring decisions considered workers' predicted scores and the degree of certainty expressed in those predictions.

We implemented two treatment variations, resulting in four experimental treatments. The treatments were randomly assigned at the individual level. The first treatment variation focused on the type of confidence individuals established, specifically whether it pertained to their past or future expected performance. This distinction allowed us to differentiate between backward-looking and forward-looking confidence.

The second treatment variation involved modifying the environment in which confidence was revealed. In the private treatment, workers were incentivized based on the accuracy of their predictions relative to their actual performance, without any explicit incentives to display

(over)confidence. In the public treatment, workers were not only incentivized to make accurate predictions but were also informed that their decisions would be observed by another group of participants acting as firms. The possibility of being hired by these firms could result in an additional bonus of 25 cents. This setup created a situation where confidence could be strategically used in a way that was not possible in the private condition. The experimental design is illustrated in Figure 1, and we explain each treatment in more detail below.

Figure 1: Experimental Conditions

	Workers		**Firms**	
	Private	**Public**	**Private Hiring**	**Public Hiring**
Forward-looking Confidence	Subjects privately reveal confidence about future performance	Subjects publicly reveal confidence about future performance	Hire from two workers who privately signaled confidence about future performance	Hire from two workers who publicly signaled confidence about future performance
Backward-looking Confidence	Subjects privately reveal confidence about past performance	Subjects publicly reveal confidence about past performance	Hire from two workers who privately signaled confidence about past performance	Hire from two workers who publicly signaled confidence about past performance

2.1. Worker Data

In the worker experiment, participants completed two stages: a real-effort task and a confidence elicitation part. The order in which participants performed the two parts of the experiment was determined by whether they were randomized into a backward or a forward-looking treatment. For our real effort task, participants counted zeros in tables filled with zeros and ones (Apicella et al., 2017). This task demonstrates little to no gender gap in performance. Moreover, it is suitable for an online setting where cheating can be a concern. Workers were paid 10 cents for each table they counted correctly within 90 seconds, with no penalty for answering

incorrectly. The number of available tables was 15, and participants on average answered five correctly.

During the confidence elicitation part of the experiment, workers responded to five questions related to their own performance on the counting zero task and their predictions about the performance of others. To ensure accurate responses, we incentivized two of these answers based on precision. The first confidence question (Q1) asked workers to predict their own performance in the counting task. Augmenting Q1, the second question (Q2) focused on the level of certainty participants felt about their estimated performance. Workers indicated their certainty in their estimates on a 5-point Likert scale. The third question (Q3) captured the participants' predicted range of performance by asking for the minimum and maximum values within which they believed their performance fell.[2]

The fourth and fifth questions were designed to capture the workers' beliefs about others' performance. In the fourth confidence question (Q4), subjects guessed how many tables a randomly picked other participant would have counted correctly. The fifth question (Q5) measured their certainty in their prediction of others' scores. We describe all confidence questions in Table 1. The answers to questions Q1 and Q4 were incentivized based on accuracy with each correct estimate paying 10 cents. The experiment concluded with a questionnaire.

In the backward confidence treatment, workers first completed the real-effort task and subsequently answered the confidence questions. In the forward confidence treatment, the order was reversed. Participants began by reviewing a sample table and answering the five confidence

[2] The second and third questions indicate a similar notion regarding confidence: a degree of certainty in predicted performance. To validate that participants comprehended the instructions correctly and genuinely expressed their beliefs, we assessed the correlation between the responses to these two questions. We find the correlation coefficient is -0.13 (p=0.00), suggesting lower certainty (Q2) indeed results in a wider interval for guessed performance (Q3). This correlation underscores the consistency in participants' responses and their understanding of the instructions.

questions regarding their estimated performance in the task involving similar tables. Afterward, they completed the actual counting zero task.

Table 1: Confidence Questions

Question No.	***Confidence Question***	***Answer Scale***
Q1	How many tables do you think you *solved/will solve* correctly in Part *1/2* within 90 seconds? (You will get an additional 10 cents if your estimate is correct.)	Integer with the condition of (0, 15)
Q2	You have estimated that you *solved/will solve* **XX correct tables.** How certain are you of your performance estimate?	On a 5-point scale between Not at all Certain – Extremely Certain
Q3	Can you please provide an interval of your performance that you are very certain of? (i.e., a minimum and a maximum estimate of your performance)	Slider (0, 20) for the Min (0, 20) for the Max with the condition of Max>Min
Q4	Now imagine a randomly chosen participant in Part *1/2*. How many tables do you think **this randomly chosen participant *solved/will solve* correctly in Part *1/2*** within 90 seconds? (You will get an additional 10 cents if your estimate is correct.)	Integer with the condition of (0, 15)
Q5	You estimated that the randomly chosen participant *solved/would solve* **XX correct tables.** How certain are you of your estimate regarding **the other person's performance?**	On a 5-point scale between Not at all Certain – Extremely Certain

Notes: Italics indicate backward/forward confidence variation.

The design descriptions above pertain to the private version. In the public treatment, the difference was the conditions under which workers revealed their confidence. In addition to making accurate predictions, workers in the public treatment knew that being hired by the firms entitled them to receive an additional 25 cents as a hiring bonus. Similar to the private treatment,

workers in the public treatment also varied in their confidence type, with some revealing backward-looking confidence and others revealing forward-looking confidence.

2. 2. Firm Data

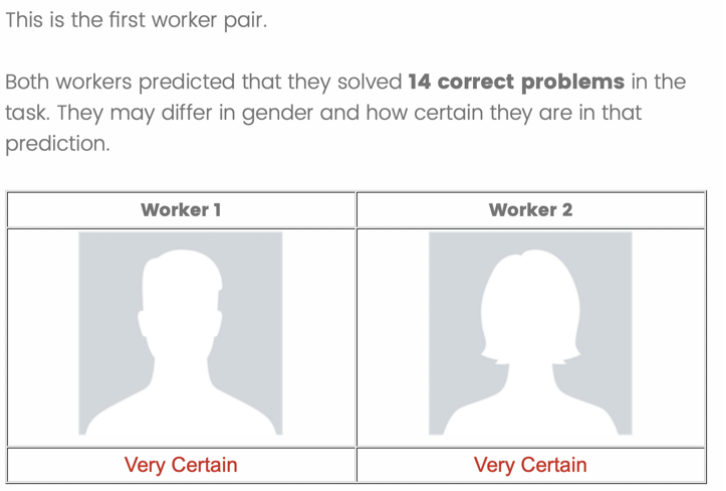
In a separate experiment, we recruited participants to act as firms. The firms were presented with five pairs of workers, and their main task was to hire one worker from each pair. Before making their hiring decisions, the firms were provided with information about the nature of the experiment and were shown a sample task that the workers had completed. The firms were also informed that, in the real-effort task, workers had solved an average of five tables correctly.

The firms then learned about the gender of each worker in the pairs through avatar images (for similar uses of avatar images please see Gangadharan et al., 2016 and Mengel, 2020). They were also provided with information about the workers' predictions of their performance in the counting task, which was the same for both workers in each pair. Additionally, the firms were informed about the level of certainty expressed by the workers regarding their performance predictions, and this certainty level varied.

We implemented the strategy method for the firms, and the gender, confidence, and certainty of each worker pair were randomly determined. As a result, the representation of workers to the firms was not influenced by the observed weights in the actual worker data. For each firm, one of the five hiring decisions was randomly selected for payment. The payment for the firm was determined by the number of correctly answered tables provided by the hired worker in the counting task. The firms received a bonus of 10 cents for each correct answer provided by the hired worker.

Like the workers, the firms also participated in one of the four treatments of the study. Figure 1 illustrates an example hiring screen, and detailed instructions can be found in the Appendix.

Figure 1: Example Hiring Screen



3. Worker Results

We use the worker data to investigate the gender gaps in confidence and how different treatment variations may influence it. Our findings reveal an overall gender gap in confidence, with men reporting higher levels of performance estimates, greater certainty in their predictions, and more optimism about the range of their performance estimates. We do not find a gender difference in how men and women evaluate others' performance.

The public treatment results in a moderate increase in confidence but does not affect men and women differently. However, the gender gap in confidence becomes more pronounced in the context of forward-looking confidence, as men report even higher performance estimates than women (and relative to the backward-looking condition).

Of the 15 available tables, workers, on average, answer 4.54 (SE=0.06) correct tables in the task. Women perform slightly better and answer 0.23 more tables than men (M=4.42, SE=0.10

for men and M=4.65, SE=0.08 for women, p=0.07, for the gender difference in performance)[3]. Although the task results in a slight female advantage, the gender difference in confidence has the opposite sign, as summarized in the findings below.

Result 1 (Gender gap in performance estimates): *Male workers report higher performance estimates than women.*

When asked to predict their task performance, men report that they solve 6.19 (SE=0.15) tables correctly, while women estimate 5.57 (SE=0.13) tables, on average (p=0.00 for the gender difference).

Both genders overestimate their performance relative to their actual scores. However, men display a higher degree of overprediction and predict solving 1.77 (SE=0.15) more correct tables than they actually do, whereas women estimate solving 0.92 (SE=0.13) more correct tables than their actual scores (p=0.00 for the gender gap in overprediction). Hence women provide more accurate estimates that align closer with their actual performance.

Result 2 (Gender certainty gap): *Male workers report higher levels of certainty in their performance estimates than women.*

In addition to reporting higher performance estimates, male workers exhibit more certainty in their performance predictions than women. The second question measures certainty in score estimates on a scale from 1 to 5. Overall, male workers report higher levels of certainty in their performance predictions. On average, men's certainty is 2.74 (SE=0.04), while women's certainty

[3] All tests are two-sided t-tests, unless noted otherwise.

is slightly lower at 2.57 (SE=0.05, gender difference is 0.18 p=0.01, t-test; p=0.04 Chi-square test).

An OLS regression confirms the above finding. The gender certainty gap persists when estimated performance is controlled for, as Specification 2 in Table 2 illustrates. Given the same performance prediction in Q1 as men, women workers report a 0.17 points lower certainty (p=0.00).

Table 2: Gender Difference in Certainty in Estimates

	(1) Certainty	(2) Certainty
Female	-0.176***	-0.171***
	(0.064)	(0.065)
Constant	2.743***	2.701***
	(0.044)	(0.093)
Est. Perf. Control	No	Yes
Observations	797	797
R-squared	0.009	0.010

*Notes: OLS regression results. Dependent variable is Certainty level coded 1 to 5, with 5 being extremely certain. Female is a dummy indicating the gender of the worker. Est. Perf. is the performance prediction in Q1, ranging between 1 and 15. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.*

Result 3 (Gender gap in performance interval): *Women report more pessimistic estimates than men for their performance's lower and upper bounds.*

When predicting the minimum (lower bound) and maximum (upper bound) levels of their performance, men estimate they would solve at least 5.14 (SE=0.18) correct tables. In contrast, women estimate a minimum of 4.53 (SE=0.18) correct tables (p=0.02, for the gender difference). Similarly, a gap appears for the upper bounds. Here, men predict an upper bound of 10.45 (SE=0.23) correct tables, and women indicate they would give a maximum of 9.69 (SE=0.25) correct answers (p=0.02 for the gender difference).

This finding is intriguing considering women's slight advantage in actual performance in our setting. On average, women's actual performance (M=4.65) falls within their estimated range (p=0.53 for the difference between the actual score and the lower bound), indicating that their provided intervals align with their performance levels. However, men's actual performance (M=4.42) falls significantly below their average predicted lower bound (p=0.00), further highlighting the substantial overestimation in their predictions.

There is, however, no significant gender gap in the interval's width. The difference between women's lower and upper bounds is 5.16 (SE=0.18), and this number is 5.32 (SE=0.17) for men (p=0.52 for the gender difference). In sum, men report higher performance estimates, are more certain of those estimates, and set a performance interval at higher levels than their actual performance than women. These findings are summarized in Table 3.

Table 3: Gender Differences in Confidence Measures

	Men	**Women**	**p-value for the Gender Diff.**
(Over)prediction	1.77	0.92	0.00
	(0.15)	(0.13)	
Certainty	2.74	2.57	0.01
	(0.04)	(0.05)	
Lower Bound	5.14	4.53	0.02
	(0.18)	(0.18)	
Upper Bound	10.45	9.69	0.02
	(0.23)	(0.25)	
Perf. Interval Width	5.32	5.16	0.52
	(0.17)	(0.18)	

Notes: Standard errors in parentheses. P-values from t-test.

Result 4 (No gender gap in beliefs about others): *Female and male workers report similar predictions regarding other people's performance and are equally certain about these predictions.*

Turning our focus to the final two questions, which measure predictions of other people's performance, we find no significant gender difference. Both men and women believe that a randomly chosen participant would answer 6.57 (SE=0.09) correct tables (p=0.72 for the gender difference in estimates). Moreover, men and women report similar levels of certainty regarding their estimates, with an average rating of 1.20 (SE=0.03) on a scale from 1 (not at all certain) to 5 (extremely certain) (M=1.25 for men and M=1.15 for women, p=0.11 for the gender difference in certainty).

We now examine the impact of treatment variation on confidence through regression analyses. Initially, we examine how forward vs. backward guessing influences workers' overprediction and certainty levels and its potential effects on gender differences. Subsequently, we analyze the role of public signaling in shaping confidence.

Result 5 (Larger gender confidence gap with forward guessing): *With forward guessing, the gender gap in performance estimates becomes more pronounced.*

As reported in Specifications 1 and 2 of Table 4, forward reporting of estimated (future) performance overall increases overprediction but also further exacerbates gender differences. The gender gap in overprediction is 0.29 (SE=0.17) correct tables with the backward version (p=0.08 for the gender difference) but rises to 1.36 (SE=0.35) correct tables with the forward guessing (p=0.00 for the gender difference). Men react to forward-guessing by excessively overpredicting their performance (diff=1.68 between backward and forward treatments for men,

p=0.00) whereas the boost in overprediction for women with forward guessing is more moderate (diff=0.61 across backward and forward treatments, p=0.02). Specification 2 in the OLS regression analysis of Table 4 shows this finding. Here, the difference in gender difference with forward-guessing is significant (Coeff = -1.25 for the *Female*Forward* interaction term, p=0.01 from the OLS regression).

Specifications 3 and 4 of Table 4 show that forward guessing lowers workers' certainty in their performance predictions (but not in different ways between the two genders). This also reassures us that participants are attentive to the experiment, as predictions about future performance should naturally lead to increased uncertainty about performance.

Result 6 (No impact of public reporting on gender confidence gap): *Both male and female workers express higher confidence in the public treatments, with no disparate effects on gender.*

Workers express higher levels of certainty (but not higher levels of performance) when they know that this information will be made available to hiring firms: While overprediction remains at the same level in the public and private treatments (treatment diff=0.09, SE=0.20, p=0.66), certainty in score estimates, on the other hand, increases significantly by 0.13 (SE=0.06) points (on a scale from 1-5, p=0.05, for the treatment effect) in the public treatments. Public signaling does not impact men and women disparately, however.

Table 4: Treatment Effects on Gender and Self-Confidence Levels

	(1) Overprediction	(2) Overprediction	(3) Certainty	(4) Certainty
Female	-0.825***	0.700	-0.184***	-0.473
	(0.192)	(0.853)	(0.063)	(0.312)
Forward Confidence	1.142***	2.753***	-0.379***	-0.514***
	(0.193)	(0.631)	(0.063)	(0.196)
Public Treatment	0.074	-0.310	0.134**	-0.024
	(0.193)	(0.633)	(0.063)	(0.196)
Female*Forward		-1.254**		0.086
		(0.500)		(0.175)
Female*Public		0.053		0.102
		(0.496)		(0.171)
Female*Forward*Public		0.124		0.003
		(0.239)		(0.081)
Constant	0.763	-1.086	3.298***	3.736***
	(0.472)	(1.132)	(0.158)	(0.399)
Observations	797	797	797	797
R-squared	0.064	0.074	0.058	0.059

*Notes: OLS regression results. Dependent variable is Overprediction level (difference between predicted performance and actual score) for Specifications 1 and 2, and Certainty (on a scale from 1-5) in performance estimates for Specifications 3 and 4. Female is a dummy indicating worker gender. Forward Confidence is a dummy indicating the treatment involved forward guessing. Public Treatment is a dummy indicating public confidence signaling. Female*Forward is an interaction of Female worker dummy and Forward Confidence, Female*Public is the interaction of Female worker dummy and Public Treatment. Female*Forward*Public is a three-way interaction of all these dummies. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

We now move on to the results of the firm experiment and how confidence and certainty are impacting hiring decisions.

4. Firm Results

In the firm study, we investigate how firms consider confidence when making hiring decisions, thus shedding light on the demand for confidence. We employ the strategy method to present firms with workers who vary in gender and certainty levels.

Each firm sees five worker pairs and makes one hiring decision in each pair. This gives us a total of 8,210 observations (4,105 hiring decisions). In each hiring screen, firms see two workers reporting the same performance estimates (where the estimated performance can be between 1 and 15). This means within each pair, the firms' hiring decisions are based on two remaining variables: certainty in performance estimates (on a Likert scale from 1 to 5) and/or gender.

Result 7 (Firms hire workers with higher certainty estimates): *On average, workers who signal higher certainty estimates are more likely to be hired.*

On average, firms hire workers with a higher certainty estimate (on a scale from 1-5, diff=0.79 between hired and non-hired workers, SE=0.03, p=0.00). The regression analysis in Table 5 further confirms this. Each one-point increase in the certainty level results in a 9.9 percent increase in hiring likelihood (p=0.00, from the OLS regression). As Specification 2 of Table 5 summarizes, this result holds in all treatments.

Table 5: Treatment Effects on Hiring Interest

	(1) Hired	(2) Hired
Certainty Level	0.099***	0.099***
	(0.004)	(0.004)
Public Treatment		0.001
		(0.013)
Forward Confidence		-0.001
		(0.013)
Certainty*Public*Forward		0.001
		(0.006)
Constant	0.203***	0.203***
	(0.012)	(0.016)
Observations	8,210	8,210
R-squared	0.078	0.078

*Notes: OLS regression results. Dependent variable is the dummy indicating that the firm hired the worker. Certainty level is certainty in performance prediction, coded 1 to 5, with 5 being extremely certain. Forward Confidence is a dummy indicating the treatment involved forward guessing. Public Treatment is a dummy indicating public confidence signaling. Certainty*Public*Forward is a three-way interaction of Certainty Level, Public Treatment dummy and Forward Confidence dummy. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Result 8 (Women are more likely to be hired): *Firms hire women workers more often.*

There is a stronger interest in hiring female workers. Women are hired 55.9 percent (SE=0.7) of the time, whereas men are hired 44.1 percent (SE=0.07) of the time (gender diff=11.7 percent, SE=1.1, p=0.00). According to Specification 2 of Table 6, this seems to be primarily driven by the fact that female firms are more likely to hire women. Here, the coefficient of the interaction variable *Female_Worker*Female_Firm* suggests that female firms account for the 6.2 percent (p=0.00, from the OLS regression) of this greater hiring interest toward women workers. Specification 3 in Table 6 confirms that this greater hiring interest towards female workers holds in all treatments.

Table 6: Worker Gender and Hiring Interest

	(1)	(2)	(3)

	Hired	Hired	Hired
Female Worker	0.117***	0.024	0.117***
	(0.011)	(0.035)	(0.011)
Public			0.000
			(0.014)
Forward			-0.001
			(0.015)
Female*Public*Forward			0.002
			(0.013)
Female Firm		-0.094***	
		(0.035)	
Female_Worker*Female_Firm		0.062***	
		(0.022)	
Constant	0.324***	0.464***	0.324***
	(0.017)	(0.055)	(0.021)
Observations	8,210	8,210	8,210
R-squared	0.014	0.015	0.014

*Notes: OLS regression results. Dependent variable is the dummy indicating that the firm hired the worker. Female Worker is a dummy indicating the worker was female. Public is a dummy indicating the treatment featured public confidence signaling. Forward is a dummy indicating the treatment involved forward confidence. Female*Public*Forward is a three-way interaction of all dummies. Female Firm is a dummy indicating firm's gender. Female_Worker*Female_Firm is an interaction of worker and firm gender. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Result 9 (Excessive confidence results in retribution): *Firms are less likely to hire a worker who signals an excessively high performance estimate with a high certainty level.*

Each worker signaled a performance prediction between 1 to 15 (which was the same for both workers by experimental design), and a certainty estimate for that prediction (on a five-point scale: Not at all Certain (1) – Extremely Certain (5)). First, we categorize performance predictions into bins and test how the interaction of these two variables impacts the hiring decisions of the firms. We group performance prediction into three categories: lower-bin (workers give performance prediction between 0-5), middle-bin (workers give performance prediction between 6-10), and upper-bin (workers give performance prediction between 11-15). Next, using an OLS regression, we treat performance prediction and certainty levels as

categorical variables and specify a full factorial interaction of all confidence variables – the main effects of each performance prediction, certainty levels, and their interaction.

As outlined in Table 7, high levels of performance estimates coupled with high certainty result in significantly lower employment likelihood. Firms' hiring interest significantly declines for workers whose performance estimate is in the higher-bin and if they reported a high-level certainty. This results in workers reporting a high-performance guess (upper-bin) with extreme certainty being least likely to be hired (coeff= - 24.7 percent, p=0.00, from the OLS regression).[4]

Table 7: Negative Role of Excessive Confidence on Hiring Interest

	(1)

[4] This finding is also robust with a full factorial analysis without creating bins for performance. Table B.1 in the Appendix outlines that high performance estimates with high certainty levels consistently result in declining hiring likelihood.

	Hired
Public	0.006
	(0.002)
Forward	0.001
	(0.001)
Certainty	0.128***
	(0.000)
Female Worker	0.122***
	(0.011)
Female Firm	-0.001
	(0.004)
2.Performance Pred. (Middle Bin)	0.026***
	(0.001)
3.Performance Pred. (Higher Bin)	0.105***
	(0.000)
2.Certainty (Slightly Certain)	0.086***
	(0.001)
3.Certainty (Moderately Certain)	0.137***
	(0.000)
4.Certainty (Very Certain)	0.065***
	(0.000)
5.Certainty (Extremely Certain)	omitted
2.Performance Pred. (Middle Bin) #2.Certainty (Slightly Certain)	0.021***
	(0.001)
2.Performance Pred. (Middle Bin) #3.Certainty (Moderately Certain)	-0.028***
	(0.001)
2.Performance Pred. (Middle Bin) #4.Certainty (Very Certain)	-0.042***
	(0.001)
2.Performance Pred. (Middle Bin) #5.Certainty (Extremely Certain)	-0.068***
	(0.000)
3.Performance Pred. (Higher Bin) #2.Certainty (Slightly Certain)	-0.023***
	(0.001)
3.Performance Pred. (Higher Bin) #3.Certainty (Moderately Certain)	-0.097***
	(0.000)
3.Performance Pred. (Higher Bin) #4.Certainty (Very Certain)	-0.154***
	(0.000)
3.Performance Pred. (Higher Bin) #5.Certainty (Extremely Certain)	-0.247***
	(0.000)
Constant	-0.129***
	(0.023)
Observations	8,210
R-squared	0.112

Notes: OLS regressions results. Dependent variable is the dummy indicating that the firm hired the worker. Public is a dummy indicating the treatment featured public confidence signaling. Forward is a dummy indicating the treatment involved forward confidence Certainty is level of certainty in estimated performance, coded 1 to 5, with 5 being extremely certain. Female Worker is a dummy indicating worker's gender. Female Firm is a dummy indicating firm's gender. Performance Pred. indicates the bin (lower, middle or upper) that the worker's

*performance predictions fall into. Robust standard errors clustered for each bin in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Taken together, we document that while moderate levels of confidence boost employability compared to low confidence, excessive levels result in a declining likelihood of being hired. We also document women being hired more often than men, an effect that is primarily driven by female firms.

5. Further Analysis and Robustness Checks

In this section, we explore whether specific design choices confound the findings and report the results of a range of other robustness checks.

5.1. Do the workers in the forward-confidence treatment calibrate their performances to guess correctly?

An interesting finding surfaces in our forward-confidence treatment: the gender gap in overprediction widens as men start displaying even higher levels of overprediction compared to women (and compared to the backward treatment). It is possible to argue that, due to the incentive to guess accurately, workers in the forward confidence treatment might have adjusted their actual performance to align with their preceding predicted scores.

To investigate whether this could be the case, we turn to our worker data and compare the actual performances of male and female workers between the forward and backward treatments. We find a significant performance boost with the forward guessing for women workers only. Specifically, forward guessing increases the believed performance for men (Mean diff=1.84, SE=0.28) more than it does for women (Mean diff=1.05, SE=0.25), but a significant boost in the performance only happens for women (Mean diff=0.44, SE=0.15,

p=0.00 for the treatment effect for women, Mean diff=0.16, SE=0.19, p=0.41 for men). The gender difference does not explain the increased gender gap in confidence with forward performance estimation.

This, of course, does not rule out other mechanisms that may take place with forward guessing. For example, forward guessing may be perceived as a goal-setting opportunity for our participants, which could explain why we observe a boost in the actual performances of women. Although these mechanisms may explain why women react to forward-guessing by improving their performance, the findings remain valid; the overprediction gender gap is enlarged with forward guessing.

5.2. Are the results robust after excluding the participants who did not get all attention questions correct?

To ensure data quality, our experiment utilizes several understanding questions about the instructions. Rather than enforcing the correct answer to proceed, we incentivize each correct answer with 5 cents. The fact that subjects can proceed without answering all questions correctly allows us to limit our dataset to those who answered all questions correctly for further robustness checks.

Workers answer a total of three understanding questions. As detailed in Appendix C.1., after excluding 156 workers who answered at least one of these questions incorrectly, the results in the worker study remain almost virtually unchanged both in magnitude and significance. Similarly, firms answer four understanding questions, and after dropping 120 firms who failed to answer all four questions correctly, the firm findings also persist.

Therefore, our results remain robust even after excluding participants who couldn't answer all understanding questions correctly.

5.3. Do the findings change after dropping workers whose score is zero in the task?

As mentioned earlier, we observe an imbalance in task performance across genders, with women displaying better performance. However, upon closer examination, we notice that this difference could be driven by the fact that more men fail to provide at least one correct answer to the counting problems. Specifically, ten of the 11 workers who do not provide a single correct answer in the task are male.

We test if the worker results remain consistent after excluding these observations. After excluding the workers who scored zero in the task, we find that the gender performance difference is no longer present. The female advantage declines to 0.12 more correct tables (from the earlier 0.23) and loses statistical significance ($p=0.30$ for the gender difference in performance). Furthermore, we conduct all the previous analyses and arrive at the same conclusions. These findings are reported under Appendix C.2. All worker results remain consistent and statistically significant even after restricting our analysis to subjects who provided at least one correct answer in the task.

6. Discussion

We conduct a hiring experiment to investigate various aspects of the gender gap in confidence and examine how signaling confidence influences employment likelihood. We recruit two groups of subjects representing different sides of the market: workers to study the supply of, and firms to understand the demand for, confidence.

In the worker experiment we confirm the existence of a significant gender gap in confidence with men consistently reporting higher performance estimates, displaying greater certainty in those estimates, and setting their performance intervals at higher levels than women. Notably, when confidence is assessed ex ante, i.e. before the task is conducted, the gender gap in performance estimates becomes even more pronounced.

The results from the firm experiment show how higher confidence can increase the likelihood of being hired, but only up to a point: firms are less likely to hire workers who signal excessively high-performance estimates together with high certainty levels.

In our setting, women not only predict performance levels closer to their actual abilities but refrain from displaying excessive levels of confidence. Whether women are more cautious in their confidence expression purely because this reflects their beliefs or if they also take strategic and social considerations into account is an area for future study. Likewise, it would be interesting to investigate if encouraging both men and women to adopt confidence levels better aligned with their actual capabilities could lead to more equitable hiring outcomes.

The work presented here opens other avenues for future research as well. First, while we observe a gender confidence gap that widens with forward guessing, we do not directly assess whether, and to what extent, employers are aware of and account for this gap in their hiring decisions.

Second, delving deeper into the underlying reasons behind the role of confidence in hiring decisions would be valuable. Our study shows that excessive confidence is met with retribution from employers. However, we do not differentiate whether this is due to firms perceiving the displayed confidence as unrealistic and expecting much lower performance (performance-driven) or simply finding the workers socially dislikable (social-driven). Future research can

employ evaluator studies, where independent evaluators rate workers on their performance and social skills. This approach would enable an investigation of the underlying mechanisms and offer more detailed insights into how perceptions of confidence influence hiring choices.

Taken together, we are hopeful that our work here will be valuable in further examining the implications of diversity and inclusion policies. It is essential that such policies are designed in a way that ensures that they foster an equitable labor market without inadvertently perpetuating biases or imbalances, and our findings can help achieve this goal.

References:

Apicella, Coren L., Elif E. Demiral, and Johanna Mollerstrom. "No gender difference in willingness to compete when competing against self." *American Economic Review* 107, no. 5 (2017): 136-140.

Apicella, Coren L., Elif E. Demiral, and Johanna Mollerstrom. "Compete with others? No, thanks. With myself? Yes, please!" *Economics Letters* 187 (2020): 108878.

Blau, Francine D., and Lawrence M. Kahn. "The gender wage gap: Extent, trends, and explanations." *Journal of Economic Literature* 55, no. 3 (2017): 789-865.

Bohnet, Iris, Oliver Hauser, and Ariella Kristal. "Supply-and Demand-Side Effects in Performance Appraisals: The Role of Gender and Race." *HBR Working Paper RWP21-016* (2021).

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. "Beliefs about gender." *American Economic Review* 109, no. 3 (2019): 739-73.

Bowles, Hannah Riley, Linda Babcock, and Lei Lai. "Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask." *Organizational Behavior and Human Decision Processes* 103, no. 1 (2007): 84-103.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. "Gender, competitiveness, and career choices." *The Quarterly Journal of Economics* 129, no. 3 (2014): 1409-1447.

Charness, Gary, Lien Dao, and Olga Shurchkov. "Competing now and then: The effects of delay on competitiveness across gender." *Journal of Economic Behavior & Organization* 198 (2022): 612-630.

Coffman, Katherine Baldiga. "Evidence on self-stereotyping and the contribution of ideas." *The Quarterly Journal of Economics* 129, no. 4 (2014): 1625-1660.

Coffman, K.B., Collis, M. and Kulkarni, L. *Whether to Apply?*. Forthcoming in *Management Science* (2023).

Deaux, Kay, and Elizabeth Farris. "Attributing causes for one's own performance: The effects of sex, norms, and outcome." *Journal of Research in Personality* 11, no. 1 (1977): 59-72.

Demiral, Elif, and Johanna Mollerstrom. "Competitiveness and Employability." *Available at SSRN 4019667* (2022).

Exley, Christine L., Muriel Niederle, and Lise Vesterlund. "Knowing when to ask: The cost of leaning in." *Journal of Political Economy* 128, no. 3 (2020): 816-854.

Exley, Christine, and Kirby Nielsen. "The gender gap in confidence: Expected but not accounted for." *Available at SSRN 4352381* (2022).

Exley, Christine L., and Judd B. Kessler. "The gender gap in self-promotion." *The Quarterly Journal of Economics* 137, no. 3 (2022): 1345-1381.

Falk, Armin, and Florian Zimmermann. "Consistency as a signal of skills." *Management Science* 63, no. 7 (2017): 2197-2210.

Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein. "Knowing with certainty: The appropriateness of extreme confidence." *Journal of Experimental Psychology: Human Perception and Performance* 3, no. 4 (1977): 552.

Gangadharan, Lata, Tarun Jain, Pushkar Maitra, and Joseph Vecci. "Social identity and governance: The behavioral response to female leaders." *European Economic Review* 90 (2016): 302-325.

Hangya, Balázs, Joshua I. Sanders, and Adam Kepecs. "A mathematical framework for statistical decision confidence." *Neural Computation* 28, no. 9 (2016): 1840-1858.

Harvard Business Review, "*How to Get your Resume Noticed (And Out of the Trash Bin)*". https://hbr.org/2020/09/how-to-get-your-resume-noticed-and-out-of-the-trash-bin. Accessed 18 July 2023.

Isaksson, Siri. "It takes two: Gender differences in group work." *Unpublished manuscript* (2018).

Hernandez-Arenaz, Iñigo, and Nagore Iriberri. "A review of gender differences in negotiation." *Oxford Research Encyclopedia of Economics and Finance* (2019).

Lenney, Ellen. "Women's self-confidence in achievement settings." *Psychological Bulletin* 84, no. 1 (1977): 1.

Leon Rohr, Samuel. "Harnessing the power of the job description." *Human Resource Management International Digest* 24, no. 6 (2016): 8-11.

Mengel, Friederike. "Gender differences in networking." *The Economic Journal* 130, no. 630 (2020): 1842-1873.

Moore, Don A., and Paul J. Healy. "The trouble with overconfidence." *Psychological Review* 115, no. 2 (2008): 502.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. "Managing self-confidence." *NBER Working Paper* 17014 (2014).

Murciano-Goroff, Raviv. "Missing women in tech: The labor market for highly skilled software engineers." *Management Science* 68, no. 5 (2022): 3262-3281.

Niederle, Muriel, and Lise Vesterlund. "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics* 122, no. 3 (2007): 1067-1101.

Niederle, Muriel, and Lise Vesterlund. "Gender and Competition" *Annual Review of Economics* 3 (2011): 601-630.

Pouget, Alexandre, Jan Drugowitsch, and Adam Kepecs. "Confidence and certainty: distinct probabilistic quantities for different goals." *Nature Neuroscience* 19, no. 3 (2016): 366-374.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. "Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender." *The Economic Journal* 127, no. 604 (2017): 2153-2186.

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram. "Gender differences in recognition for group work." *Journal of Political Economy* 129, no. 1 (2021): 101-147.

Shastry, Gauri Kartini, Olga Shurchkov, and Lingjun Lotus Xia. "Luck or skill: How women and men react to noisy feedback." *Journal of Behavioral and Experimental Economics* 88 (2020): 101592.