

NBER WORKING PAPER SERIES

THE GENDER GAP IN SELF-PROMOTION

Christine L. Exley
Judd B. Kessler

Working Paper 26345
<http://www.nber.org/papers/w26345>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2019, Revised December 2019

The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Christine L. Exley and Judd B. Kessler. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Gender Gap in Self-Promotion
Christine L. Exley and Judd B. Kessler
NBER Working Paper No. 26345
October 2019, Revised December 2019
JEL No. C91,D90,J16

ABSTRACT

In applications, interviews, performance reviews, and many other environments, individuals are explicitly asked or implicitly invited to assess their own performance. In a series of experiments, we find that women rate their performance less favorably than equally performing men. This gender gap in self-promotion is notably persistent. It stays just as strong when we: eliminate gender differences in confidence about performance, eliminate incentives to self-promote, provide information about the average self-promotion of others, and make environments more ambiguous. Because of the prevalence of self-promotion opportunities, this self-promotion gap may contribute to persistent gender gaps in education and labor market outcomes.

Christine L. Exley
Harvard Business School
clexley@gmail.com

Judd B. Kessler
The Wharton School
University of Pennsylvania
3620 Locust Walk
Philadelphia, PA 19104
and NBER
judd.kessler@wharton.upenn.edu

1 Introduction

At various points in their educational and professional lives — in college and professional school applications, in job applications and interviews, in performance reviews — individuals are explicitly asked to report on their ability and performance. In myriad other contexts, individuals face implicit invitations or opportunities to talk about their ability and performance. In response to these explicit and implicit opportunities, individuals may subjectively convey their ability and performance, which we call their level of “self-promotion.”¹

Focusing on a sector that readers of this article may know well — academia — it is clear that the level of self-promotion may play a role at a variety of points in the career of an academic. Even before the academic job market, self-promotion can affect how one writes a graduate school application, how one talks about his or her research ideas and technical skills, whether one gets the attention of desired advisors, and how one is perceived in internal seminars. During the academic job market, self-promotion can have direct implications for how one’s work and seminar talks are received and how much of joint work is attributed to a candidate rather than coauthors. Self-promotion can continue to have impacts throughout an academic career, as the way academics talk about their research — and their contribution to the research — can affect citations, prominence, as well as tenure and promotion decisions.

Given the pervasiveness of self-promotion in educational and labor market environments, one might be worried about the potential for a gender gap in self-promotion. If women describe their performance and ability less favorably than equally capable men do, a gender gap in self-promotion might have implications for observed gender gaps in educational and labor market environments. However, there is scant research on how individuals describe their own performance and ability in strategic settings — in which self-assessments are communicated to others and can influence labor market outcomes — or about how such self-promotion varies with gender.²

In this paper, we examine self-promotion by having experimental participants complete an analytical task: answering 20 questions from the Armed Services Vocational Aptitude Battery (ASVAB). Participants report beliefs about the number of questions they answered correctly on the ASVAB test (our measure of “confidence”) and then respond to subjective, quantitative self-assessment questions about their performance (our measures of “self-promotion”). In our main study version, the *Public* version, participants are aware that one of their responses to a self-assessment question will be reported to a potential employer who will use that response — and only that re-

¹In standard parlance, self-promotion often has a negative connotation. We do not imply one here and instead follow literature from psychology that refers to more positive self-assessments as engaging in more self-promotion. For instance, [Rudman \(1998\)](#) defines self-promotion as follows: “Designed to augment one’s status and attractiveness, self-promotion includes pointing with pride to one’s accomplishments, speaking directly about one’s strengths and talents, and making internal rather than external attributions for achievements.” There is a literature that suggests that extensive self-promotion can lead to backlash, which we discuss below.

²As detailed later, our focus on strategic settings distinguishes us from much prior literature on beliefs in non-strategic environments (see the discussion in [Charness, Rustichini and Van de Ven \(2018\)](#)), and our focus on subjective rather than objective performance beliefs distinguishes us from much prior literature on gender differences.

sponse — to decide whether to hire them and how much to pay them. Data from employers confirm that self-promotion pays: more positive responses to these self-assessment questions increase the chance that participants are hired and the subsequent earnings they receive.

We document a large and persistent gender gap in self-promotion. Despite women performing slightly better than men on the analytical task, as documented in Section 3.1, women’s responses to the self-assessment questions suggest they performed less well than the men. The effects are large. For example, when asked to indicate their agreement on a scale from 0 to 100 with a statement that reads “I performed well on the test,” the average man in the *Public* version rates himself a 61 out of 100 and the average women rates herself a 46 out of 100 — 15 points (i.e., 25%) lower — despite the fact that the average performer in both groups answered 10 out of 20 ASVAB questions correctly. After documenting the gender gap in self-promotion in Section 3.2 — including the robustness of these results to several different types of self-assessment questions — the remainder of the paper examines underlying mechanisms of this gap and attempts to close it.

In Section 3.3, we show that the gender gap in self-promotion is *not* driven by a gender gap in confidence about performance. Consistent with related literature discussed towards the end of our Introduction, we observe a gender gap in confidence, as measured by participants’ beliefs about the number of ASVAB questions they answered correctly. We show, however, that the self-promotion gap persists when we provide participants with *perfect information* about their absolute and relative *past performance* on the ASVAB and then ask participants to self-assess that same *past performance*. We find that women still engage in less self-promotion than men, even when both are told that they answered the same number of questions correctly and both are told their exact place in the relative distribution (e.g., when both are told that they answered 15 out of 20 questions correctly and that their score was better than 80%, and worse than 12%, of prior participants). We emphasize that while we also ask self-promotion questions about potential future performance, one of the reasons we ask about — and generally focus on — *past performance* is that we can provide feedback to fully eliminate the potential role of confidence in past performance.³

In Section 3.4, we show that the gender gap in self-promotion is *not* driven by the strategic incentives in our *Public* version. In the *Private* version, participants answer the same self-assessment questions privately: there are no employers that participants have an incentive to impress, and how participants answer the self-assessment questions cannot influence their payments. While responses are less favorable in this version — evidence that participants are engaging in more self-promotion when incentivized to do so — we find a similarly sized gender gap. These results suggest no gender differences in willingness to inflate self-assessments for personal gain and highlight that the gender gap in self-promotion is *not* driven by gender differences in beliefs or preferences regarding potential employers, such as beliefs about how employers’ hiring decisions are influenced by self-promotion or preferences over employer earnings. As in the *Public* version, the gender gap persists in the *Private*

³In addition to being asked to make self-assessments about their past performance, workers know that any employer who hires them will be paid according to that same past performance.

version even after we inform participants of their absolute and relative performance.

In Sections 3.5 and 3.6, we further demonstrate the robustness of the gender gap in self-promotion. In Section 3.5, we present results from the *Private (Information about Others)* version. This version is nearly identical to our *Private* version except that participants are provided with the average level of self-promotion engaged in by others along with the information on their own performance. The gender gap in self-promotion is not attenuated by providing participants with information about others' self-promotion decisions, which highlights that the gender gap is *not* driven by a gender difference in beliefs about how much self-promotion is typical. In Section 3.6, we present results from the *Ambiguous* version, which introduces some chance of “being caught” if participants inflate their self-assessments too much. This version is identical to our *Public* version except that participants are told that information about their actual performance (i.e., how many questions they answered correctly on the ASVAB) *could be* communicated to employers along with their response to a self-assessment question about their performance. The possibility of true performance being communicated neither increases nor decreases the gender gap in self-promotion.

Taken together, our results document a large gender gap in self-promotion that arises in unambiguous and ambiguous labor market environments. In addition, we rule out potential explanations for this gap. We show that it is not driven by a gender difference in confidence, it is not driven by a gender difference in willingness to inflate self-assessments for personal gain, it is not driven by beliefs about how employers interpret self-promotion, and it is not driven by beliefs about the level of self-promotion chosen by others.⁴ This leaves open many avenues for future research on what drives a gender difference in self-promotion, several of which we provide direct evidence for or against in Section 4.

Our results make contributions to three related literatures. First, motivated by the well-documented empirical evidence on the gender gaps in pay and in occupational and industry representation (Goldin, 2014; Blau and Kahn, 2017), there is a large literature in experimental economics that documents gender gaps in important labor market decisions. Relative to men, women are less willing to select a challenging task (Gneezy and Rustichini, 2003), to negotiate (Small et al., 2007), to enter a competition (Niederle and Vesterlund, 2007; Croson and Gneezy, 2009; Niederle and Vesterlund, 2011; Niederle, 2016), to contribute their ideas or assume a “leadership” position in a group (Coffman, 2014; Born, Ranehill and Sandberg, 2018; Coffman, Flikkema and Shurchkov, 2019), to claim credit for their contributions (Isaksson, 2018), and to apply for a job or for a promotion (Bosquet, Combes and García-Peñalosa, 2019; Coffman, Collis and Kulkarni, 2019a).⁵ We add

⁴In Sections 4.1 through 4.7, we provide evidence against other potential causes of the gender gap in self-promotion. We provide evidence against men and women having: different standards for good work (see Section 4.1) or different beliefs about how much they deserve to earn from the study (see Section 4.6). We also provide evidence against our results reflecting differences in expected backlash (see Section 4.2), consistency (see Section 4.3), differences in lying aversion (see Section 4.5), or gender being correlated with other demographics (see Section 4.7).

⁵Two literatures that can help to explain some of these gender gaps include the literature on gender differences in confidence, which we discuss below, and the literature on the gender differences in risk, as reviewed in Eckel and Grossman (2008) and Croson and Gneezy (2009).

self-promotion to the list of behaviors with gender gaps that might contribute to disparities in labor market outcomes. In addition to its direct effects, self-promotion may also interact with other labor market decisions, including those listed above. That is, self-promotion — and in particular how individuals subjectively describe their performance — can affect how well one performs conditional on entering a negotiation, whether one is hired conditional on applying for a job, and whether one is selected to represent their group conditional on speaking up.⁶

Second, our results contribute to the literature on the gender gap in confidence in male-typed domains (Lundeberg, Fox and Punčochař, 1994; Niederle and Vesterlund, 2007; Coffman, 2014; Bordalo et al., 2018; Isaksson, 2018; Coffman, Collis and Kulkarni, 2019b) as well as the literature on potential reasons for its persistence, such as gender differences in how individuals update their beliefs (Ertac, 2011; Mobius et al., 2014; Buser, Gerhards and Van der Weele, 2018; Coutts, 2018).⁷ Given the importance of gender stereotypes in explaining these gender differences (Coffman, 2014; Bordalo et al., 2018; Coffman, Flikkema and Shurchkov, 2019), we follow much of this literature by focusing on a male-typed domain to align with gender gaps in labor market outcomes that motivate our study. Our paper replicates the gender gap in confidence in a male-typed domain and then deviates from the extant literature in two important ways.

We differ from almost all of the prior literature by focusing on a *strategic setting*, in which beliefs are communicated to others who can affect participants’ payoffs.⁸ As in many environments of interest outside the lab, participants in the *Public* version of our study may want to inflate their self-assessments about how well they performed, since potential employers will use these self-assessments to determine whether to hire them and how much to pay them. We are aware of only two other papers that elicit beliefs about performance in a strategic setting.⁹ Reuben, Sapienza and Zingales (2014) shows that, in a setting where gender is known, men more than women inflate estimates about how well they performed on a task when these estimates are provided to “employers” who may hire them for a future task. Charness, Rustichini and Van de Ven (2018) finds that men, but not women, inflate the likelihood that they scored in the top half of their group on a cognitive ability task when doing so might deter other participants from entering a competition with them.¹⁰

⁶The impact of self-promotion on these outcomes has been proposed, and to a limited extent investigated, in prior work. Our contribution is to focus on subjective descriptions themselves by eliciting them in a quantifiable manner as our main outcome of interest, avoiding potential confounds arising from a particular decision environment. For example, the communications considered in the literature on negotiations (e.g., Exley, Niederle and Vesterlund (Forthcoming)) or group decision-making contexts (e.g., Coffman, Flikkema and Shurchkov (2019)) are neither limited to a description of own’s performance nor easily quantifiable. In focusing on gender differences in the *use* of subjective descriptions when asked about performance, we also differ from work on gender differences in *responses* to subjective or ambiguous descriptions, such as those involved job advertisements when deciding whether to apply (Leibbrandt and List, 2015; Abraham and Stein, 2019; Coffman, Collis and Kulkarni, 2019a).

⁷For example, Coffman, Collis and Kulkarni (2019b) shows that increasing the informativeness of a signal about performance does not reduce the size of observed gender gaps in performance beliefs.

⁸Prior literature has typically elicited beliefs that are incentivized for accuracy but are not shared with others.

⁹Schwardmann and van der Weele (Forthcoming) and Soldà et al. (2019) examine a different form of strategic beliefs: individuals inflate *privately* held beliefs when incentivized to convince others of their high ability.

¹⁰There is also a related literature on gender differences in cheating (see, e.g., Dreber and Johannesson (2008), Erat and Gneezy (2012)), and Adams, Kuhn and Waddell (2019)).

We also differ from the prior literature — including both [Reuben, Sapienza and Zingales \(2014\)](#) and [Charness, Rustichini and Van de Ven \(2018\)](#) — by focusing on *subjective beliefs* rather than the objective beliefs that are typically elicited to measure confidence. Unlike objective beliefs about how many questions a participant answered correctly (as in [Reuben, Sapienza and Zingales \(2014\)](#)) or whether the participant is in the top half of performers (as in [Charness, Rustichini and Van de Ven \(2018\)](#)), subjective beliefs cannot be right or wrong (i.e., there is no correct response when a participant is asked to indicate agreement with the statement: “I performed well on the test”). When individuals communicate about their performance and ability in practice, they regularly report subjective beliefs.¹¹ A better understanding of subjective beliefs is important not only because they are often conveyed (in lieu of, or in conjunction with, objective beliefs) to potential employers, colleagues, supervisors, and other evaluators, but also because they operate differently from objective beliefs. We observe a gender gap in self-promotion (i.e., subjective beliefs) that persists even when we fully eliminate the gender gap in confidence (i.e., objective beliefs) by informing participants of both their absolute and relative performance.

Third, our results contribute to psychological literature related to self-promotion. This literature largely focuses on gender-specific backlash in response to self-promotion — for an example in the context of negotiations, see [Bowles, Babcock and Lai \(2007\)](#); for a review paper, see [Rudman and Phelan \(2008\)](#) — and there is noticeably little work on gender differences in self-promotion itself.¹² Not only does our paper provide evidence of a gender gap in self-promotion per se, we show that this gap arises in a strategic environment *absent gender-specific backlash concerns*. Employers in our *Public* version do not observe participant gender and only learn how a participant answers one self-assessment question. Similarly, the potential additional information that employers learn in our *Ambiguous* version is restricted to a participant’s actual performance. While gender-specific backlash is not relevant in our study, we speculate that the existence of gender-specific backlash outside of our study could cause our gender gap in self-promotion if it has led women to internalize the risks of engaging in self-promotion. One related question for future work is how a gender difference in self-promotion might respond to the potential for gender-specific backlash, for example if information about gender were communicated to employers along with self-assessments.¹³

The rest of the paper proceeds as follows. Section 2 describes our design. Section 3 documents the gender gap in self-promotion and our main results on mechanisms and robustness. Section 4

¹¹This distinction is related to work about verifiable versus unverifiable signals of support as in [Kessler \(2017\)](#).

¹²Indeed, [Moss-Racusin and Rudman \(2010\)](#), which documents gender differences in self-promotion in a simulated job interview, writes: “to our knowledge, researchers have yet to examine gender differences in self-promotion.”

¹³See [Alston \(2019\)](#) for evidence on how women, when given the opportunity to hide their gender on a job application for a male-typed job, are keen to do so. If future work makes the gender of workers known, self-promotion could also be examined in contexts where gender discrimination has been documented (for reviews, see [Riach and Rich \(2002\)](#) and [Blau and Kahn \(2017\)](#); for recent evidence, see [Reuben, Sapienza and Zingales \(2014\)](#); [Milkman, Akinola and Chugh \(2015\)](#); [Baert, De Pauw and Deschacht \(2016\)](#); [Bohnet and Bazerman \(2016\)](#); [Sarsons \(2017a,b\)](#); [Bohren, Imas and Rosenberg \(2019\)](#); [Bohren et al. \(2019\)](#); [Coffman, Exley and Niederle \(2019\)](#); [Kessler, Low and Sullivan \(2019\)](#)). For evidence about how others respond differently to men versus women in the language they use, see [Bohren, Imas and Rosenberg \(2018\)](#).

discusses additional potential mechanisms and directions for future research. Section 5 concludes.

2 Design and Data

In October 2018, we recruited 900 workers on Amazon Mechanical Turk (MTurk) to complete one of three main versions of our study: the *Public* version, the *Private* version, or the *Ambiguous* version. In November 2019, we recruited 600 workers on MTurk to complete one of two main versions of our study: the *Private (Replication)* version or the *Private (Information about Others)* version. The *Private (Replication)* version was — except for very slight formatting differences — identical to the *Private* version we ran in 2018. We ran this version to allow for a clean comparison to the new *Private (Information about Others)* version and to replicate our prior 2018 results.¹⁴

Each participant was randomized into one of our study versions and received a guaranteed \$2 completion fee for the 20-minute study. In addition, one out of the four parts was randomly selected to determine a possible bonus payment for each participant. After participants completed all four parts of the study, they took a short follow-up survey that collected demographic information.

Each of the five study versions are detailed in the following Sections 2.1 through 2.5. In Section 2.1, we describe the *Public* version of our study, upon which the other versions are based. In Section 2.2, we describe the *Private* version, highlighting how it differs from the *Public* version. In Section 2.3, we describe the *Private (Replication)* version of our study, which is exactly the same as the *Private* version except for a small formatting change in how self-promotion questions appear. In Section 2.4, we describe the *Private (Information about Others)* version of our study, highlighting how it differs from the *Private (Replication)* version. In Section 2.5, we describe the *Ambiguous* versions of our study, highlighting how it differs from the *Public* version. In Section 2.6, we describe — and present results from — an additional study version, involving 300 participants in the role of “employers,” who are relevant for the *Public* and *Ambiguous* versions of our study.¹⁵

2.1 The *Public* Version

The *Public* version of our study has four parts, described in sequence as follows (see Appendix B.1 for screenshots).

¹⁴To be eligible for any study version, workers must have previously completed at least 100 tasks on MTurk with a 95% or better approval rating from prior employers, and workers must be working from an United States IP address. In each of the three study versions run in 2018, the median age is 33–34 years old, the median educational attainment is a Bachelor’s Degree, and the percentage of male participants varies from 46%–52%. In each of the two study versions run in 2019, the median age is 33 years old, the median educational attainment is a Bachelor’s Degree, and the percentage of male participants varies from 59%–61%. While participants were required to correctly answer understanding questions at various points to proceed in the study, no participants were excluded from our data analysis.

¹⁵In addition, we use data from 100 participants from a prior study who completed the same ASVAB test described below, in order to provide information to study participants on their relative performance. We also analyze data from 399 MTurk workers who evaluated free-response comments generated by study participants (as described below). Including these participants, this paper involves a total of 2,299 study participants.

Part 1: ASVAB and Confidence

In Part 1 of the study, participants are asked to take a test comprising of 20 multiple choice questions from the Armed Services Vocational Aptitude Battery (ASVAB). They have up to 30 seconds to answer each question, and there are four questions each from the following five categories: General Science, Arithmetic Reasoning, Math Knowledge, Mechanical Comprehension, and Assembling Objects. Participants are informed that “In addition to being used by the military to determine which jobs armed service members are qualified for, performance on the ASVAB is often used as a measure of cognitive ability by academic researchers.” If Part 1 is randomly selected for payment, a participant’s bonus payment is equal to 5 cents times the number of ASVB questions answered correctly.

After participants answer the 20 ASVAB questions, and before they continue to part 2, they are asked: “Out of the 20 questions on the test you took in part 1, how many questions do you think you answered correctly?” This question is not incentivized, and participants can select any number from 0 to 20.¹⁶

Part 2: Uninformed Self-Promotion

In part 2, participants are asked five questions about their performance on the ASVAB. Participants are told that if part 2 is randomly selected for payment, one of the responses to one of the questions will be shared with another MTurk participant called their “employer.” The employer will see the response to the randomly selected question — and only that question (i.e., not any of the other responses or any information about actual performance) — and will determine whether to hire them and how much to pay them if hired.

If an employer chooses not to hire a participant, the participant will earn a bonus of 25 cents, and the employer will earn a bonus of 100 cents. If an employer chooses to hire a participant, the employer will choose a wage between 25 and 100 cents, which will be the bonus for the participant. The employer’s bonus payment will then equal: 100 cents minus the wage paid to the participant plus 5 cents times the number of questions the participant answered correctly on the ASVAB.¹⁷

To encourage participants to reflect on their performance, the first question in part 2 is a free-response question that states: “Please describe how well you think you performed on the test that you took in part 1 and why.” The remaining four questions are the subjective, quantitative self-assessment questions that we analyze for the remainder of the paper.¹⁸

¹⁶One could have imagined providing monetary incentives for answering this question correctly. However, in many cases in practice — particularly when individuals are asked to make self-assessments — they are likely to form such beliefs about performance in the absence of such monetary incentives. In addition, we avoid concerns about the accuracy of belief elicitation because our experiment (as discussed below) controls for confidence by design, rather than statistically, which avoids any potential concerns about noise in belief elicitation.

¹⁷Note that employer earnings are based on the number of correct answers that the participant completed previously. This means that participants do not have to complete additional tasks and the decision environment avoids any potential uncertainty that might arise about future performance.

¹⁸The free-response question can also theoretically be interpreted as providing an opportunity for self-promotion. Analyzing this free-response data is fraught, however, as the text is hard to evaluate and can convey information such as gender and competence that makes measuring self-promotion per se difficult. Nevertheless, we attempt to

The next two questions elicit self-assessments about how well the participants performed on the ASVAB test in more quantifiable ways. The first states: “Please indicate how well you think you performed on the test you took in part 1.” It allows participants to select from one of the following six answers: terrible, very poor, neutral, good, very good, and excellent. The second elicits a more continuous measure, asking participants to indicate the extent to which they agree, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: “I performed well on the test I took in part 1.” We note that these two questions elicit self-assessments about *past performance* and, in particular, the past performance that will affect the earnings of employers if a worker is hired.

The final two questions elicit self-assessments when there is also room for participants to hold preferences and beliefs about a related, hypothetical job. Using the same 0 to 100 scale described above, these questions ask participants to indicate the extent to which they agree with the following statements: “I would apply for a job that required me to perform well on the test I took in part 1” and “I would succeed in a job that required me to perform well on the test I took in part 1.”

Part 3: Informed Self-Promotion

In part 3, participants are asked precisely the same five questions about their performance on the ASVAB, and participants are told that if part 3 is randomly selected for payment, one of the answers to one of the questions will be shared with their employer.

Before answering these questions, however, participants learn precise information about their *absolute* and *relative* performance on the ASVAB test. In particular, participants are told exactly how many of the 20 questions they answered correctly (i.e., their absolute performance) and they are compared to 100 other participants who were asked the same ASVAB questions as part of a prior study and told how many of those participants answered more questions correctly and how many answered fewer questions correctly (i.e., their relative performance). As an attention check, participants must also correctly report how many of the 20 ASVAB questions they answered correctly before proceeding to answer the questions in part 3. Note that providing relative performance information as well as absolute performance information serves to eliminate potential gender differences in beliefs about where a score falls in the performance distribution.

Part 4: Deservingness Question

In Part 4, participants are asked one question that measures deservingness for earnings from our experiment: “Out of a maximum amount of 100 cents, what amount of bonus payment, in cents, do you think you deserve for your performance on the test you took in part 1?” If this part is randomly selected as the part that counts, their bonus payment equals whatever amount they indicate from 0 to 100 cents. This question allows us to control for a potential gender difference in deservingness (i.e., how much participants believe they deserve to earn from the study) or in the desire to earn money from the experiment. As highlighted in Appendix Section 4.6, we find no statistically significant

learn what we can from this data by having 399 MTurk participants evaluate the responses, and we summarize those findings in Appendix A.3.

(nor economically meaningful) gender differences on reports of deservingness, which means gender differences observed in response to the self-assessment questions cannot be attributed to differential deservingness or desire to earn money from the study.

2.2 The *Private* Version

The *Private* version of our study proceeds exactly as the *Public* version except that participants provide their part 2 and part 3 self-assessments in a non-strategic, non-incentivized setting. In particular, there is no mention of any “employer,” and participants are told that if part 2 or part 3 is randomly selected as the part-that-counts, their bonus payment will equal 25 cents regardless of how they answer the self-assessment questions.

2.3 The *Private (Replication)* Version

The *Private (Replication)* version of our study proceeds exactly as the *Private* version except that the formatting of self-promotion questions asked in part 2 and part 3 is slightly different to allow room for the additional information of the form that is introduced in the *Private (Information about Others)* version described next.

2.4 The *Private (Information about Others)* Version

The *Private (Information about Others)* version proceeds exactly as the *Private (Replication)* version of our study except that participants are provided with additional information when making their part 3 self-assessments. In particular, each of their four self-promotion questions now includes a message that reads: “Also note that, among participants in a prior study who scored the same as you on the test, the average answer to this question was: [insert relevant average answer].” The *Private (Information about Others)* version was designed to investigate whether the gender gap in self-promotion would persist when participants learned about the average self-promotion of others.

2.5 The *Ambiguous* Version

The *Ambiguous* version of our study proceeds exactly as the *Public* version except that participants are told that there is *some chance* that their employers will be informed of how many questions they answered correctly on the ASVAB test.¹⁹ We suspected that this chance would create a form of deterrence from overstating performance on the self-assessment questions. Since employers are free to choose any wage they want, participants may fear retribution through a low wage if they overstate how well they did in response to the self-assessment questions. It could have also created an incentive to appear modest by purposefully understating performance on the self-assessment questions. Regardless, the main goal for the *Ambiguous* version was designed to capture some of the ambiguity present in labor market settings, where applicants or employees are aware that signals about true performance may be available to employers.

¹⁹As implied by the name of the *Ambiguous* version, this chance is left ambiguous in the experimental instructions. In practice, there was a 1% chance we would run a version in which employers were informed of this additional information. This resulted in us not running such a version.

2.6 The *Employers* Version

In November 2018, we recruited 300 workers on MTurk to complete the *Employers* version of our study using the same criteria as in the main study versions (see footnote 14). Each employer received a guaranteed \$1.50 completion fee for the 15-minute study. In addition, two of their decisions, out of the 21 decisions in the study, was randomly selected to determine a possible bonus payment for them and for associated “workers,” participants in the *Public* and *Ambiguous* study versions. After employers completed all decisions of the study, they took a short follow-up survey that collected demographic information.

For each decision, employers are informed that they must decide whether to hire a worker, and, if so, how much to pay that worker. Payoffs from hiring a worker are as described above. If an employer chooses not to hire a worker, the employer earns a bonus of 100 cents and the worker earns a bonus of 25 cents. If an employer chooses to hire a worker, the employer must also choose a wage between 25 and 100 cents. The worker will receive that wage, and that employer’s bonus payment will equal 100 cents minus the wage paid to the worker plus 5 cents times the number of questions the worker answered correctly on the ASVAB test. The only information employers receive about a worker before hiring them is how the worker answered one of the four self-assessment questions.

Employers make hiring and wage decisions via the strategy method. That is, they respond to each of the six possible answers to the Likert-scale question and five randomly selected answers (i.e., numbers from 0 to 100) to the each of the three other questions.²⁰ Employers respond to each of these 21 answers to the self-assessment questions by deciding whether to hire a worker who gave this answer and how much to pay the worker if hired. Employers are told that two of the 21 answers will be selected for payment and that they will be paired with a worker who gave that answer. For these two answers, employers are paid based on their hiring and wage decisions and — if they chose to hire the worker who gave the answer — the worker’s performance on the ASVAB test.

As expected, self-promotion pays. Employers respond to more positive self-assessments by being more likely to hire workers and by paying them more. Table 1 shows how higher answers on each of the four self-assessment questions affects wages given to workers. In all specifications, the coefficient on *Assessment* is positive and significant. Columns (1), (3), and (4) show that the wage given to workers increases by an average of 0.21 or 0.22 cents more for every point participants add to their self-assessment on the 100-point scale. Column (2) shows that the wage given to workers increases by 4.26 cents for each increase on the six-point Likert scale. These results highlight that participants have an incentive to inflate their self-assessments to increase their expected study earnings.

²⁰As noted above, these questions ask participants to state their agreement with the following statements: “I performed well on the test I took in part 1,” “I would apply for a job that required me to perform well on the test I took in part 1,” and “I would succeed in a job that required me to perform well on the test I took in part 1.” Employers face all hiring decisions related to one self-assessment question before moving on to the next self-assessment question, but the order in which they face answers to each self-assessment question is randomized.

Table 1: *Employer Version*, Wage Regressions

	<i>Performance</i>	<i>Performance-Bucket</i>	<i>Willingness to Apply</i>	<i>Success</i>
<i>Assessment</i>	0.21*** (0.02)	4.26*** (0.27)	0.22*** (0.02)	0.21*** (0.02)
Constant	22.70*** (0.75)	18.95*** (0.70)	21.94*** (0.61)	22.76*** (0.78)
N	1490	1788	1490	1490

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are clustered by employer. Results are from OLS regressions of the wage received by the participant (25 cents if not hired and the chosen wage if hired). *Assessment* is the assessment provided by each participant in the assessment question noted in that column. *Performance* indicates the extent of each participant’s agreement (from 0 – 100) with the following statement: “I performed well on the test I took in part 1.” *Performance-Bucket* indicates which Likert-scale response (coded from 1 for the lowest to 6 for the highest) a participant selects when asked to “indicate how well you think you performed on the test in part 1.” *Willingness to Apply* indicates the extent of each participant’s agreement (from 0 – 100) with the following statement: “I would apply for a job that required me to perform well on the test I took in part 1.” *Success* indicates the extent of each participant’s agreement (from 0 – 100) with the following statement: “I would succeed in a job that required me to perform well on the test I took in part 1.” Data are from the hiring decisions in the *Employer* version.

3 Results

In this section, we present results from our five main study versions: the *Public*, *Private*, *Private (Replication)*, *Private (Information about Others)*, and *Ambiguous* versions. In Section 3.1, we show that women slightly outperform men on the ASVAB test and show that, nevertheless, we replicate a large gender gap in confidence (i.e., women report that they answered fewer questions correctly than equally performing men). In Section 3.2, we document a gender gap in the responses to the self-assessment questions elicited in part 2 of the *Public* version of our experiment — which we call the gender gap in self-promotion — and show that it is present for all four of our self-assessment questions. In Section 3.3, we show that the self-promotion gap persists in part 3 of the *Public* version of our study — after participants have been informed of their absolute and relative performance — demonstrating that the gender gap in self-promotion is not driven by the gender gap in confidence. In Section 3.4, we show that the gender gap in self-promotion persists in the *Private* version of our study, demonstrating that it is not driven by differences in willingness to inflate self assessments for personal gain or by any differences in beliefs or preferences surrounding employers. In Section 3.5, we show that the results from the *Private (Replication)* version indeed replicate results from the *Private* version and that the gender gap in self-promotion persists in the *Private (Information about Others)* version, demonstrating that it is not mitigated by participants knowing the average level of self-promotion of others with their same score. In Section 3.6, we show that the gender gap in self-promotion persists in the *Ambiguous* version of our study, suggesting it is robust to environments where performance information may become available to employers.

3.1 Performance and Confidence

In this section, we show that women, on average, outperform men on the ASVAB test but that we nevertheless replicate a gender gap in confidence.²¹ Since participants answer the 20 ASVAB questions and report their objective performance beliefs before they know which of the study versions to which they have been randomized, we pool the results from all study versions in presenting the results.²²

Panel A of Figure 1 shows CDFs of the number of ASVAB questions answered correctly by male participants and by female participants. On average, women answer 10 questions correctly and men answer 9.27 questions correctly. The mean difference is statistically significant ($p < 0.01$) and the distributions are statistically significantly different (a Kolmogorov–Smirnov test yields $p < 0.01$).

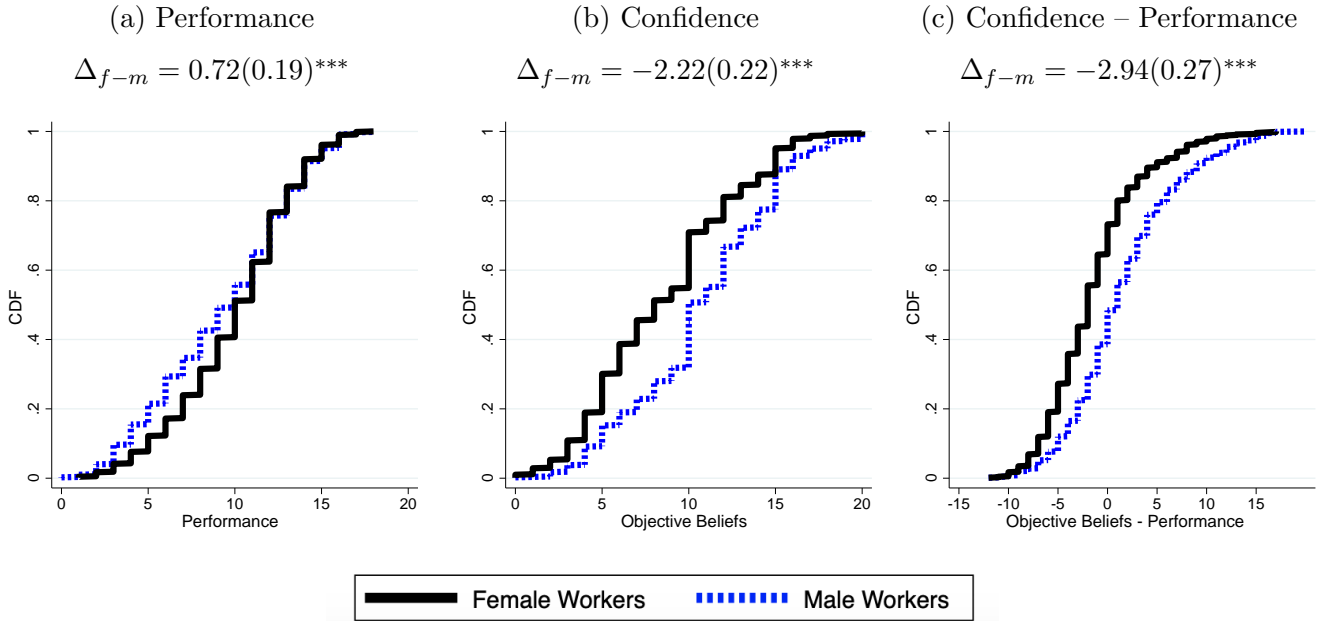
Despite women performing better than men, Panels B and C of Figure 1 replicate a common finding in the literature: women, relative to men, believe they performed worse on the test. Panel B shows raw performance beliefs (i.e., confidence). On average, men believe they answered 10.74 questions correctly while women believe they answered only 8.52 questions correctly. The mean difference is statistically significant ($p < 0.01$), and the distributions are statistically significantly different (a Kolmogorov–Smirnov test yields $p < 0.01$). Panel C shows the difference between confidence and performance and shows that the gender gap in confidence persists. Again, the mean difference is statistically significant ($p < 0.01$), and the distributions are statistically significantly different (a Kolmogorov–Smirnov test yields $p < 0.01$). Looking at where the CDFs cross 0, we see that the gender gap in confidence is driven both by more women than men underestimating their actual performance and more men than women overestimating their actual performance.

Table 2 presents the corresponding regression results. Column 1 shows that women outperform men on the ASVAB (the coefficient on *Female* is positive and statistically significant), and the remaining columns confirm the statistically significant gender gaps in confidence, including when considering the raw data only (Column 2), when controlling for performance with dummies for each possible score (Column 3), and when the outcome variable directly captures the difference between confidence and performance (Column 4). In the latter three columns, the coefficient on *Female* is negative, large, and statistically significant.

²¹That we find a gender gap in confidence reassures us that our environment is not particularly different from the settings that researchers have explored in related work.

²²As shown in Appendix A.1, across all study versions, the gender gap in confidence persists even though it is never the case that men significantly outperform women on average.

Figure 1: Performance and Confidence



Δ_{f-m} indicates the average gender difference of the associated outcome (i.e., the average outcome among women minus that among men). The associated robust standard error is shown in parentheses and the statistical significance is indicated as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table note applies to other figures that include Δ_{f-m} estimates.

Table 2: Performance and Confidence Regressions

DV:	<i>Performance</i>	<i>Confidence</i>		<i>Confidence - Performance</i>
	(1)	(2)	(3)	(4)
<i>Female</i>	0.72*** (0.19)	-2.22*** (0.22)	-2.13*** (0.21)	-2.94*** (0.26)
Constant	9.27*** (0.14)	10.74*** (0.15)		1.47*** (0.19)
N	1500	1500	1500	1500
Performance FEs	No	No	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The SEs are robust. Results are from OLS regressions of the noted dependent variable (DV). *Performance* equals the number of questions a participant correctly answered out of the 20 ASVAB questions. *Confidence* equals the number of questions a participant believes he or she correctly answered. *Female* is an indicator for the participant being female. Performance FEs are dummies for each possible performance on the ASVAB. Data are from all five study versions.

3.2 Is there a gender gap in self-promotion?

How do participants respond to the self-assessment questions? We turn first to the *Public* version of the experiment in which participants are told that one of their responses will be shared with a potential employer, and that this is all the employer will know when making a hiring and wage decision. Figure 2 shows responses to the four quantitative self-assessment questions in part 2 of the *Public* version (i.e., before participants learn their absolute and relative performance).

All four panels show large gender gaps in responses to the self-assessment questions. Women engage in less self-promotion than men. Panel A shows results from the question that asks participants to respond to the statement “I performed well on the test I took in part 1” on a scale from 0 (entirely disagree) to 100 (entirely agree). Women provide statistically significantly lower responses ($p < 0.01$ for the t-test and the Kolmogorov–Smirnov test). We obtain similar results in Panel B for the six-point Likert scale question: “Please indicate how well you think you performed on the test you took in part 1” ($p < 0.01$ for the t-test and the Kolmogorov–Smirnov test). Panels C and D show results from the self-assessment questions that allow participants to hold preferences and beliefs about a related, hypothetical job. Participants respond to the statements “I would apply for a job that required me to perform well on the test I took in part 1” (Panel C) and “I would succeed in a job that required me to perform well on the test I took in part 1” (Panel D) on a scale from 0 (entirely disagree) to 100 (entirely agree). We again see statistically significant differences in response to both questions ($p < 0.01$ for both t-tests and both Kolmogorov–Smirnov tests).

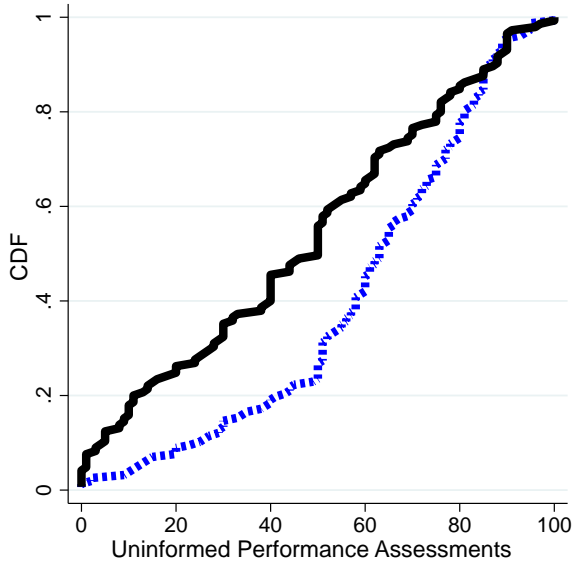
The first two columns of Table 3 confirm the statistical significance of these gender gaps, including when controlling for performance on the ASVAB.²³ Column (1) presents results when considering the raw data only, and column (2) controls for performance with dummies for each possible test score 0 to 20. As expected, the coefficient on *Female* remains negative, large, and statistically significant for all four self-assessment questions.

²³The *Performance-Bucket Assessments* results are also robust to Ordered Probit regression specifications.

Figure 2: *Public Version, Self-Promotion*

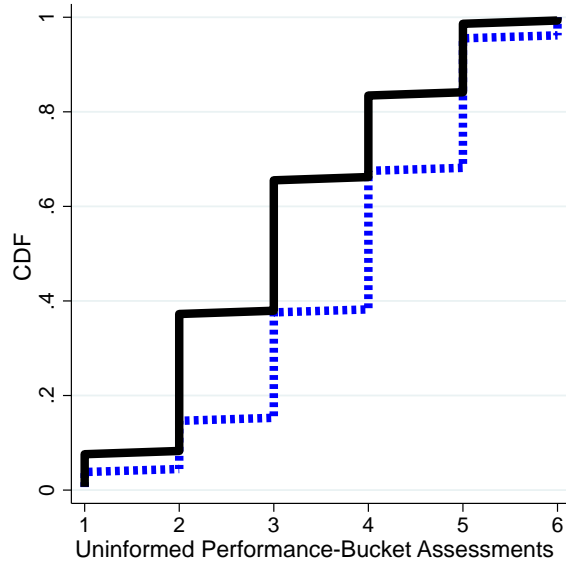
(a) Performance Assessments

$$\Delta_{f-m} = -14.77(3.07)^{***}$$



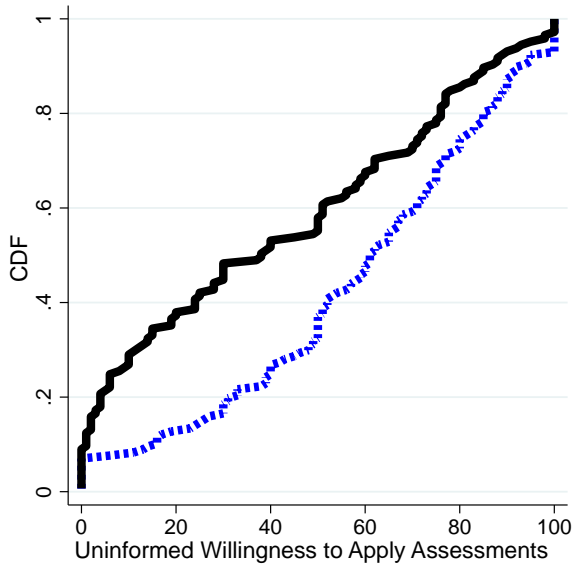
(b) Performance-Bucket Assessments

$$\Delta_{f-m} = -0.73(0.14)^{***}$$



(c) Willingness to Apply Assessments

$$\Delta_{f-m} = -17.70(3.55)^{***}$$



(d) Success Assessments

$$\Delta_{f-m} = -17.95(3.57)^{***}$$

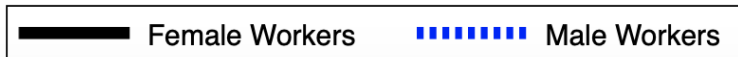
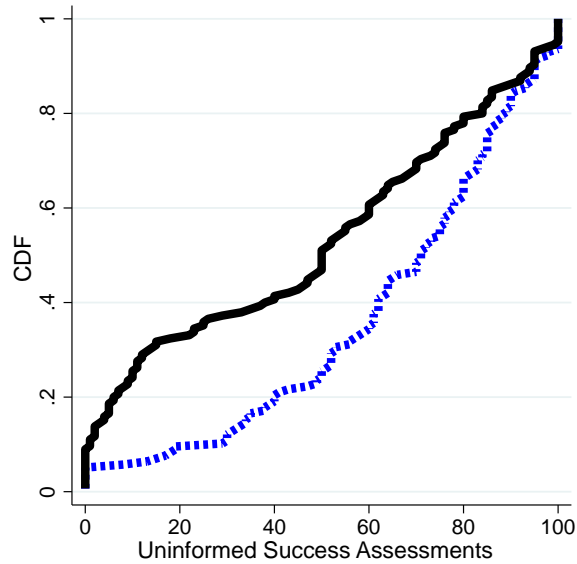


Table 3: *Public Version*, Self-Promotion Regressions

Data:	Uninformed Self-Promotion		Informed Self-Promotion	
	(1)	(2)	(3)	(4)
DV = Performance Assessments				
<i>Female</i>	-14.77*** (3.07)	-12.68*** (2.96)	-10.97*** (3.27)	-7.01** (2.90)
Constant	60.52*** (1.88)		59.43*** (2.11)	
DV = Performance-Bucket Assessments				
<i>Female</i>	-0.73*** (0.14)	-0.59*** (0.13)	-0.62*** (0.15)	-0.40*** (0.13)
Constant	3.81*** (0.10)		3.85*** (0.11)	
DV = Willingness to Apply Assessments				
<i>Female</i>	-17.70*** (3.55)	-15.31*** (3.46)	-15.55*** (3.72)	-10.73*** (3.40)
Constant	58.63*** (2.27)		60.30*** (2.44)	
DV = Success Assessments				
<i>Female</i>	-17.95*** (3.57)	-15.09*** (3.46)	-16.19*** (3.73)	-11.73*** (3.30)
Constant	65.04*** (2.15)		65.00*** (2.33)	
N	302	302	302	302
Performance FEs	No	Yes	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the noted dependent variable (DV). *Performance Assessments* indicate the extent of each participant’s agreement (from 0 – 100) with the following statement: “I performed well on the test I took in part 1.” *Performance-Bucket Assessments* indicates which Likert-scale response (coded from 1 for the lowest to 6 for the highest) a participant selects when asked to “indicate how well you think you performed on the test in part 1.” *Willingness to Apply Assessments* indicate the extent of each participant’s agreement (from 0 – 100) with the following statement: “I would apply for a job that required me to perform well on the test I took in part 1.” *Success Assessments* indicate the extent of each participant’s agreement (from 0 – 100) with the following statement: “I would succeed in a job that required me to perform well on the test I took in part 1.” *Female* is an indicator for female participants. Performance FEs are dummies for each possible performance out of the 20 questions on the ASVAB. Data in columns in (1) and (2) are from uninformed self-assessment questions elicited in part 2 of the *Public* version and data in columns (3) and (4) are from informed self-assessment questions elicited in part 3 of the *Public* version.

3.3 Is the gender gap in self-promotion driven by confidence?

In this section, we investigate whether the gender gap in self-promotion identified in Section 3.2 is driven by a gender gap in confidence about performance. We do this by analyzing the responses from part 3 of our *Public* version, in which participants are informed about their absolute and relative performance on the ASVAB test and then answer the same self-assessment questions about their performance.

Columns (3) and (4) of Table 3 report on data from part 3 of the *Public* version (see Appendix Figure A.6 for the corresponding distributions). Comparing the responses from part 2 to part 3, we see that the coefficients on *Female* are *directionally* smaller in part 3. That said, informing participants of their absolute and relative performance does not statistically significantly alter the self-promotion gap: the coefficients on *Female* in part 2 and part 3 are never statistically significantly different. Moreover, it is clear that the coefficient on *Female* remains negative, large, and statistically significant. The gender gap in self-promotion persists even when participants are fully informed of their absolute and relative performance.

A key feature of our design is that we control for the role of confidence *by design*, by providing participants with *perfect information* about absolute and relative *past performance*. This approach has a number of advantages. First, fixing beliefs eliminates the need to elicit and statistically control for beliefs, which has the potential for measurement error and for omitted variable bias (further discussed in Section 4.8). Second, fixing beliefs eliminates concerns about how participants update objective beliefs. Our approach contrasts with prior studies that are explicitly concerned with how participants update their objective beliefs and thus purposefully provide noisy performance information. Third, fixing beliefs about past performance (on the ASVAB) and then asking questions about the same past performance eliminates concerns about predicting future performance from past performance. Our approach contrasts with prior studies that provide information about past performance and ask about future performance, which cannot fully account for participants’ uncertainty around how they will perform in the future.²⁴

Our results highlight that the self-promotion gap is not driven by confidence. We provide participants with *perfect information* about their absolute and relative *past performance* before they proceed to part 3 and then ask participants to make self-assessments about their *past performance* in part 3, thus eliminating the role of confidence about performance in driving our results.

It is worth noting that only two of the self-promotion questions are completely about *past* performance: (i) *Performance Assessments* that result from participants indicating the extent of their agreement (from 0 – 100) with the following statement: “I performed well on the test I took in part 1,” and (ii) *Performance-Bucket Assessments* that result from participants choosing a Likert-scale response (coded from 1 for the lowest to 6 for the highest) when asked to “indicate how well you think you performed on the test in part 1.”

The other two of our self-assessments also appeal to potential future events: (i) *Willingness to Apply Assessments* that result from participants indicating the extent of their agreement (from 0 – 100) with the following statement: “I would apply for a job that required me to perform well on the test I took in part 1,” and (ii) *Success Assessments* that result from participants indicating the extent of their agreement (from 0 – 100) with the following statement: “I would succeed in a job that required me to perform well on the test I took in part 1.” It could be the case that for these

²⁴Prior studies that provide information on one task and ask about another, even closely related and already completed, task cannot fully account for participants’ uncertainty around their performance on the other task.

two self-promotion questions, participants are also thinking about potential future performance even though we are asking about their past performance. We are fine with this interpretation and thus point to our results from the two self-promotion questions that are completely about past performance as our strongest evidence against our results reflecting confidence about performance.

3.4 Is the gender gap in self-promotion driven by incentives to inflate self-assessments?

In this section, we investigate data from the *Private* version of our study to test whether the gender gap in self-promotion persists in the absence of an incentive to make favorable subjective assessments to a potential employer. If the self-promotion gap were due to men inflating their assessments more than women in response to such incentives, the gender gap would shrink — or disappear — in the *Private* version.

Table 4 reports results from the *Private* version and shows that the gender gap in self-promotion

Table 4: *Private Version*, Self-Promotion Regressions

Data:	Uninformed Self-Promotion		Informed Self-Promotion	
	(1)	(2)	(3)	(4)
DV = Performance Assessments				
<i>Female</i>	-15.80*** (2.88)	-13.46*** (2.93)	-10.96*** (3.14)	-8.01*** (2.88)
Constant	53.95*** (1.99)		50.82*** (2.26)	
DV = Performance-Bucket Assessments				
<i>Female</i>	-0.68*** (0.13)	-0.56*** (0.13)	-0.44*** (0.15)	-0.33** (0.14)
Constant	3.52*** (0.09)		3.48*** (0.11)	
DV = Willingness to Apply Assessments				
<i>Female</i>	-19.26*** (3.33)	-17.57*** (3.51)	-16.33*** (3.56)	-13.25*** (3.53)
Constant	53.74*** (2.41)		52.48*** (2.56)	
DV = Success Assessments				
<i>Female</i>	-18.60*** (3.46)	-16.46*** (3.61)	-16.13*** (3.61)	-13.15*** (3.53)
Constant	57.60*** (2.44)		54.98*** (2.55)	
N	304	304	304	304
Performance FEs	No	Yes	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from the same specifications noted in Table 3. Data in columns in (1) and (2) are from uninformed self-assessment questions elicited in part 2 of the *Private* version and data in columns (3) and (4) are from informed self-assessment questions elicited in part 3 of the *Private* version.

remains strong and significant in the absence of incentives to inflate subjective assessments (see Appendix Figures A.7 and A.8 for the corresponding distributions). Columns (1) and (2) report on data from part 2 of the *Private* version, and columns (3) and (4) report on data from part 3 of the *Private* version. Note that the constant terms in Table 4 are smaller than the corresponding constant terms in Table 3.²⁵ This pattern suggests that participants do inflate their subjective assessments in response to incentives (i.e., they engage in more self-promotion in the *Public* version than the *Private* version). However, the coefficients on *Female* in Table 4 are similar in size to the coefficients in Table 3, revealing that the gender gap in self-promotion is not driven by a differential response to incentives.

3.5 Is the gender gap in self-promotion mitigated by information on others’ self-promotion?

In the prior sections, we ruled out two potential drivers of the gender gap in self-promotion: confidence (see Section 3.3) and responses to strategic incentives (Section 3.4). Guided by prior research, we ex-ante viewed these as the most likely explanations.

As detailed in our Introduction, however, a defining characteristic of our work relative to much of the prior literature is our focus on a subjective outcome. Even when participants are fully informed about their absolute and relative performance on the test they took in part 1, it is not clear how they should subjectively describe their performance in response to self-assessment questions. As such, a remaining potential driver of our results could relate to men and women holding different beliefs about how much self-promotion is typical.²⁶

Our *Private (Information about Others)* version decreases the scope for potential differences in beliefs about the extent of self-promotion, since each question in part 3 also informs participants of the average response provided by prior participants with the same performance as them. This information proves ineffective at closing the gender gap. Table 5 confirms that the gender gap persists, and if anything strengthens, in the *Private (Information about Others)* version (see Appendix Figures A.9 and A.10 for the corresponding distributions) compared to the results from the *Private* version to the contemporaneous results from the *Private (Replication)* version (see Appendix Table A.6 and Appendix Figures A.11 and A.12).

²⁵In regressions when we include data from both the *Private* and *Public* versions (absent performance controls), the *Public* coefficient yields a p -value < 0.05 for all questions when participants are uninformed or informed (except for the uninformed apply question, which has $p = 0.14$).

²⁶Related work suggests the existence of gender differences in beliefs about what is required for certain jobs. Coffman, Collis and Kulkarni (2019a) shows that gender differences emerge when individuals are asked to assess the qualifications necessary from subjective job descriptions.

Table 5: *Private (Information about Others) Version*, Self-Promotion Regressions

Data:	Uninformed Self-Promotion		Informed Self-Promotion	
	(1)	(2)	(3)	(4)
DV = Performance Assessments				
<i>Female</i>	-19.95*** (3.21)	-15.14*** (3.28)	-13.78*** (3.14)	-11.93*** (3.15)
Constant	60.65*** (1.93)		59.49*** (1.87)	
DV = Performance-Bucket Assessments				
<i>Female</i>	-1.09*** (0.16)	-0.80*** (0.16)	-0.74*** (0.16)	-0.62*** (0.16)
Constant	3.96*** (0.10)		3.94*** (0.10)	
DV = Willingness to Apply Assessments				
<i>Female</i>	-21.75*** (3.61)	-16.93*** (3.71)	-19.67*** (3.44)	-16.39*** (3.42)
Constant	61.29*** (2.09)		61.18*** (2.05)	
DV = Success Assessments				
<i>Female</i>	-21.11*** (3.61)	-15.62*** (3.71)	-19.21*** (3.50)	-15.77*** (3.58)
Constant	63.06*** (2.07)		62.70*** (2.00)	
N	298	298	298	298
Performance FEs	No	Yes	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from the same specifications noted in Table 3. Data in columns in (1) and (2) are from uninformed self-assessment questions elicited in part 2 of the *Private (Information about Others)* version and data in columns (3) and (4) are from informed self-assessment questions elicited in part 3 of the *Private (Information about Others)* version.

3.6 Does the gender gap in self-promotion persist in a more ambiguous environment?

As a final test of the robustness of the gender gap in self-promotion (rather than a specific mechanism), we investigate data from the *Ambiguous* version of our study to explore whether the gender gap in self-promotion persists when participants are aware that information about their actual performance could be communicated to employers along with a response to a subjective question. Such ambiguity may be present in environments of interest outside the lab, where signals about performance and ability may be available when individuals engage in self-promotion.

Table 6 reports results from the *Ambiguous* version and shows that the gender gap in self-promotion remains strong and significant in the presence of such ambiguity (see Appendix Figures A.13 and A.14 for the corresponding distributions). Following the structure of the other tables, columns (1) and (2) report on data from part 2 of the *Ambiguous* version and columns (3) and (4)

report on data from part 3 of the *Ambiguous* version. The coefficients on *Female* in Table 6, while directionally smaller than the coefficients observed in Tables 3 and 4, are still negative, large, and statistically significant, suggesting that the gender gap in self-promotion is robust to the presence of ambiguity.

Table 6: *Ambiguous Version*, Self-Promotion Regressions

Data:	Uninformed Self-Promotion		Informed Self-Promotion	
	(1)	(2)	(3)	(4)
DV = Performance Assessments				
<i>Female</i>	-9.79*** (2.90)	-9.15*** (2.93)	-7.76** (3.09)	-7.24** (2.83)
Constant	59.35*** (1.96)		58.26*** (2.18)	
DV = Performance-Bucket Assessments				
<i>Female</i>	-0.52*** (0.13)	-0.47*** (0.13)	-0.42*** (0.14)	-0.36*** (0.14)
Constant	3.76*** (0.09)		3.77*** (0.11)	
DV = Willingness to Apply Assessments				
<i>Female</i>	-14.43*** (3.28)	-12.82*** (3.29)	-10.46*** (3.44)	-9.11*** (3.38)
Constant	60.29*** (2.15)		58.03*** (2.33)	
DV = Success Assessments				
<i>Female</i>	-10.70*** (3.30)	-9.24*** (3.32)	-9.60*** (3.40)	-8.07** (3.29)
Constant	63.45*** (2.18)		62.78*** (2.30)	
N	294	294	294	294
Performance FEs	No	Yes	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from the same specifications noted in Table 3. Data in columns in (1) and (2) are from uninformed self-assessment questions elicited in part 2 of the *Ambiguous* version and data in columns (3) and (4) are from informed self-assessment questions elicited in part 3 of the *Ambiguous* version.

4 Discussion

That the gender gap in self-promotion persists when there are strategic incentives to misreport (Sections 3.2 and 3.3), when strategic incentives to misreport are eliminated (Section 3.4), when information on the average self-promotion of others is provided (Section 3.5), and in more ambiguous environments (Section 3.6) highlights why the gender gap in self-promotion may be pervasive across contexts. That the findings in all of these contexts persist even after participants are provided with perfect information on their past performance and asked about their past performance in part 3 makes clear why the gender gap in self-promotion may be persistent across time. It also makes

clear that the gender gap in self-promotion is *not* simply another manifestation of a gender gap in confidence about performance.²⁷

Our results leave open opportunities for future work to identify the underlying causes of this self-promotion gap. To help guide that future work, this section discusses additional results from our data. Each of the following sections considers potential mechanisms that might be related to the gender gap in self-promotion and provides evidence from our data suggesting why they might be worth exploring in future work or why we believe they are unlikely to be relevant in driving our results. In addition, Section 4.8 demonstrates why it was important that we controlled for confidence through our experimental design (rather than statistically, as is often the approach in prior work).

4.1 Do women and men have different beliefs about what constitutes “good” work?

One potential reason a gender gap in self-promotion could arise is if men and women have different standards for good work. Such a difference might manifest as women believing that scoring a 15 out of 20 (and being in the 88th percentile) is only “good” and worth a 70/100 on the 100-point scale while men believe such a score is “very good” and worth an 85/100. On one hand, this could be viewed as synonymous with how we define the self-promotion gap: women describe their performance less favorably, even when they know their absolute and relative performance. On the other hand, we provide three sets of evidence against men and women having different standards for good work.

First, as shown in the *Private (Information about Others)* version, our results persist even when workers are provided with information about the average level of self-promotion of others with the same performance, which one would expect to narrow any potential gender gap in how men and women map performance to self-promotion. Second, our results are not driven by workers’ beliefs or concerns about what *employers* are likely to view as good work, since the gender gap in self-promotion persists in our *Private* versions in which there are no employers.²⁸ Third, male and female *employers* do not respond differently to self-promotion. There are no gender differences in how employers respond to self-assessments in determining wages paid to workers in either level or slope (as shown in Appendix Table A.7), suggesting that men and women likely have similar beliefs about how to map self-promotion to performance.²⁹

²⁷Even if it were, however, we would view it as an important manifestation worthy of investigation given the scant research on how individuals subjectively describe their own performance.

²⁸Coffman, Collis and Kulkarni (2019a) provide evidence of gender differences in applicants’ perceptions of what employers are looking for, suggesting a role for such beliefs when workers are responding to subjective job descriptions.

²⁹Of course, using employer data to make an assessment of workers’ mappings of performance to self-promotion is potentially confounded by differences in prosocial preferences by gender, as male and female employers might differ in how much they wish to reward other experimental participants (conditional on the same expected performance). That said, such gender differences in preferences would need to perfectly offset another gender difference in order for there to be no gender difference in the mapping of self-promotion to wages (in either level or slope).

4.2 Can differential concerns about backlash explain our results?

In line with the literature on backlash mentioned in our Introduction, a gender gap in self-promotion could reflect women, relative to men, expecting more backlash from employers if they engage in self-promotion. This could arise due to men and women having different beliefs about backlash in general. It could also arise because of fears about gender-specific backlash: women facing more backlash *because* they are women. In our study, concerns about gender-specific backlash are absent because employers neither learn the gender — nor any other identifiable characteristics — of workers. Data from our *Ambiguous* version helps to speak about concerns about backlash more generally. In that version, employers might learn about actual performance and could then engage in backlash by choosing to “punish” workers who seem to have inflated their self-assessments. That the gender gap in self-promotion is similarly sized in the *Ambiguous* version as in the *Public* version suggests that differential beliefs about backlash are not driving the results of our study.

While our study eliminated the possibility of gender-specific backlash by design, as we note in the Introduction, gender-specific backlash could contribute to our results if it has led women — more so than men — to internalize the costs of engaging in self-promotion. Such internalization could potentially even affect self-promotion in the *Private* versions of our study where no employers are present. We consequently view exploring how self-promotion responds to environments that make backlash concerns more relevant as an interesting and important avenue for future work.

4.3 Does consistency contribute to the persistent gender gap in self-promotion?

When considering gender gaps in self-promotion that may arise in the labor market, consistency could play a role. Individuals may assess their performance initially and then have repeated opportunities to engage in self-promotion about it. Initial self-promotion — which could take place before participants get information about their performance — could then affect subsequent self-promotion. More generally, self-promotion at one point in time may influence self-promotion at a later point in time as individuals progress through their schooling and careers.

To allow for the potentially important role of consistency, which is unavoidable in labor market environments, we designed an experiment in which participants in our study always answer the set of self-assessment questions first without information about their performance and then again after they were informed of their absolute and relative performance. That said, evidence for one manifestation of consistency — an unwillingness to change one’s answers to self-assessment questions between part 2 and part 3 — is not supported by our data. As shown in Appendix Table A.8, when participants learn that their performance was different than the belief they reported in Part 1, 79%–89% of participants (when considering male and female participants together or separately) change their answers to each of the three self-assessment questions on the 0 to 100 scale, and 47–49% of participants change their Likert-scale selection.³⁰ In addition, as shown in Appendix Table A.9,

³⁰There is also substantial switching between part 2 and part 3 by the approximately 10% of participants who

the correlation between part 2 and part 3 responses is large, but substantially less than 1, for each of the four self-assessment questions; to the extent it differs by gender, women’s responses are less correlated across part 2 and part 3 than men’s responses are.

4.4 Does confidence about factors other than performance contribute to a self-promotion gap?

As detailed in Section 3.3, the gender gap in self-promotion is not driven by the gender gap in confidence about performance. To reiterate this point, consider responses from part 3 of our study versions in which participants are asked to indicate the extent of their agreement (from 0 – 100) with the following statement: “I performed well on the test I took in part 1.” Confidence about past performance on the test in part 1 is not relevant because participants have received perfect information on both their absolute and relative performance. Confidence about future performance is not relevant because participants are clearly asked about their past performance.

Could other forms of “confidence” — not about performance per se — play a role? Since prior literature has focused on confidence about performance, the potential relevance of other forms of confidence does not diminish the novelty of our results. That said, we provide evidence against the self-promotion gap being driven by several other forms of confidence. Consider the results from part 3 of the *Private (Information about Others)* version. First, confidence about how employers will respond to self-promotion (and resulting payoffs) is not relevant because no employers are involved and workers know they receive a fixed payment regardless of how they answer. Second, confidence about what the task entails or what success in the task requires is not relevant because participants have already completed the task. Third, confidence about how others self-promote is muted because participants have been told the average self-promotion of others with the same score as them.

Consequently, we conclude that the self-promotion gap is not driven by a gender gap in confidence about performance (i.e., what researchers typically refer to when considering a gender gap in confidence) or by potential gender gaps in several other forms of “confidence.” We note, however, that those willing to describe confidence broadly may conceive of other potential gender gaps that we have not ruled out.³¹ Such potential gaps may be hard to measure — which may be why they have not previously been studied — but they may be interesting to consider, nevertheless. We leave open the exploration of these potential “confidence” gaps — and whether they contribute to the gender gap in self-promotion — as avenues for future work.

4.5 Can differences in aversion to lying explain our results?

Building off of the seminal paper on deception (Gneezy, 2005), Dreber and Johannesson (2008) show that men are more likely than women to lie in order to achieve a small financial gain.³² A

correctly guessed their true performance in part 1.

³¹For example, even conditional on how they scored, women might be less confident than men about how much they were “responsible” for their score on the ASVAB (e.g., how much was attributable to luck in which questions were asked or luck in whether they answered particular questions correctly).

³²This gender difference has been replicated in other studies (see, e.g., Erat and Gneezy (2012) and Houser, Vetter and Winter (2012)) and has been shown to depend on: whether the lie is harmful or prosocial towards others (Erat

natural question, then, is whether the gender gap in self-promotion relates to gender differences in lying. On one hand, we note that participants cannot explicitly lie in our study. There is no correct or true answer to the self-assessment questions because the questions are inherently subjective. On the other hand, it could be that purposefully inflating self-assessments in order to increase the expected financial gain from employers is similar to lying. To the extent that lying plays a role in self-assessments, we find evidence against this driving our results. As detailed in Section 3.4, both men and women inflate their self-assessments in the *Public* version relative to self-assessments in the *Private* version. However, men do *not* inflate their self-assessments more than women, and the gender gap in self-promotion is just as large in the *Private* version as in the *Public* version. Consequently, we think it is unlikely that the self-promotion gap is due to differences in aversion to lying.

4.6 Do women believe they deserve (or want to earn) less money from the study?

One reason that women might engage in less self-promotion is that they do not believe they deserve to earn as much money from the study as the men do. First, we note that this channel could not explain the persistence of the gender gap in the *Private* versions of our study, in which self-promotion cannot affect study earnings. Further evidence against this channel comes from part 4 of the study, which directly asks participants how much money they believe they deserve for participating in the study in an incentivized way (i.e., if part 4 is randomly selected for payment, participants earn the amount they state). Appendix Figure A.15 and Appendix Table A.10 show that, if anything, women believe they deserve to earn more from the study than men, and there is no statistically significant difference in how men and women respond to this question once we control for performance. Consequently, we think it is unlikely that the self-promotion gap is due to differences in deservingness or a desire for study earnings.

4.7 Is there a relationship between self-promotion and other demographic characteristics?

Motivated by prior literature, we designed our experiments to investigate the relationship between gender and self-promotion. If other subject demographics are correlated with gender in our data, however, part of the gender difference could arise due to differences in self-promotion across these demographic groups. While we had no priors on these relationships, Appendix Table A.11 presents results that include the other demographic characteristics we observe from our follow-up survey. In doing so, we observe that extent of self-promotion significantly decreases with age, significantly increases with education, and significantly increases with the extent to which individuals feel favorably about the Republican party. Nevertheless, the gender gap in self-promotion remains as strong when we additionally control for these demographic characteristics.

and Gneezy, 2012), the size of the stakes (Childs, 2012), and culture (Gylfason, Arnardottir and Kristinsson, 2013).

4.8 Is there a relationship between self-promotion and confidence?

As detailed in Section 3.3, we show that confidence about performance does not drive the self-promotion gap. We eliminate the role of confidence through our experimental design and find that the self-promotion gap persists. That said, there could still be a *correlation* between confidence and self-promotion.

Appendix Table A.12 reports on the regressions that present our main results on self-promotion when including performance fixed effects in columns (1) and (4) and then expands these specifications to consider the relationship with confidence (i.e., a subject’s report of the number of questions they answered correctly, provided before part 2 in all our study versions). We add a linear control for confidence in columns (2) and (5), and we add dummies for each possible confidence in columns (3) and (6). Both when participants are uninformed about their performance in part 2 and when participants are perfectly informed about their performance in part 3, our measure of confidence is positively correlated with self-promotion and decreases the coefficient on *Female*. Several implications are therefore of note.

First, controlling for confidence statistically can lead to misleading estimates. If we had only controlled for confidence statistically when participants were not fully informed about their performance (i.e., in part 2), we could have inaccurately concluded that confidence is a substantial driver of the self-promotion gap. Meanwhile, by controlling for confidence in our design, we observe that confidence does not explain the self-promotion gap. Such conflicting results should serve as a guide (and a warning) for future researchers.

Second, that there is a significant correlation between confidence and self-promotion, even when participants are fully informed about their performance (i.e., in part 3), makes clear that our measure of confidence is picking up something other than a causal relationship with self-promotion. For instance, this correlation could reflect improper measurement of confidence, due to the two variables being correlated with some omitted variable, or could reflect the confidence gap is — at least in part — driven by the self-promotion gap (i.e., the causal relationship may go in the opposite direction). This relationship is worth exploring in future work.

5 Conclusion

We have documented a gender gap in self-promotion. When communicating to potential employers, women systematically provide less favorable assessments of their own past performance and potential future ability than equally performing men. The gap is not a function of the gender gap in confidence; we find that it persists when participants are perfectly informed of their absolute and relative performance on the relevant task. The gap is not driven by differential responses to an incentive to inflate self-assessments; we find that it is present, and that it is just as large, in a study version absent employers in which participants do not have an incentive to make favorable reports. In addition, we find that the gap is robust to the provision of information on self-assessments made by others and to an ambiguous environment where information about true performance — along

with the self reports — might become known to employers.

We focus our work on self-promotion because we view it as an understudied behavior that could have important implications for labor market outcomes. Among other contexts, individuals are often explicitly invited to engage in self-promotion: in applications to educational institutions, in job applications, in job interviews, and in performance reviews. Many additional environments provide implicit opportunities to engage in self-promotion (e.g., when casually discussing work or work related issues with colleagues or superiors, when discussing private contribution to group work, and when advocating for oneself in the workplace).

The frequency of opportunities to engage in self-promotion means that it has the potential to interact with other gender gaps that have been observed in the literature. For example, prior work has demonstrated a gender gap in a willingness to negotiate. Self-promotion can have direct consequences for whether negotiation is successful, so a gender gap in self-promotion might contribute to differential outcomes for men and women who choose to negotiate. Similar arguments can be made for job applications (a gender gap in self-promotion might affect the likelihood a candidate will be hired), and whether one is selected to represent a group (a gender gap in self-promotion might affect how successful one is when they choose to speak up). Thus, we view future research on how self-promotion interacts with other labor market decisions as particularly promising.

In considering future work, we also note that two other avenues would be interesting to pursue. First, given the prevalence of self-promotion opportunities, how can the gender gap in self-promotion be mitigated? While our *Private (Information about Others)* results cast doubt on the effectiveness of simply providing more information on others' self-promotion, we wonder whether making unambiguous the “appropriate” level of self-promotion in a given environment might prove effective.³³

Second, do employers respond differently to the self-promotion of men and women in settings, unlike those examined in our studies, in which gender is known? On one hand, if employers make their own mental corrections because they accurately expect that men — more than women — inflate their self-assessments, this may mitigate any gender gap in labor market outcomes that arises from a gender gap in self-promotion. However, [Reuben, Sapienza and Zingales \(2014\)](#) provides evidence against the empirical relevance of this possibility. That paper finds that men, more than women, tend to inflate their performance estimates, but it finds that employers do not (fully) account for this. On the other hand, the literature on backlash (discussed in the Introduction) makes clear the potential for greater backlash for women relative to men, which could exacerbate gender gaps in self-promotion and their associated impact on labor market outcomes.

³³Indeed, there is prior literature that shows how clarifying the norms of the decision environment can mitigate gender gaps. For instance, [He, Kang and Lacetera \(2019\)](#) shows that the gender gap in willingness to enter competition is eliminated when individuals must “opt-out” of a competition rather than “opt-in” to a competition; [Coffman, Collis and Kulkarni \(2019a\)](#) shows that the gender gap in willingness to apply to an advanced job is eliminated when individuals are provided with clear guidance as to the conditions under which they should apply (e.g., if they have scored above some threshold on a skills-assessment test); and a reduction in ambiguity has mitigated gender gaps in the negotiations literature ([Bowles and McGinn, 2008](#); [Mazei et al., 2015](#); [Leibbrandt and List, 2015](#)).

References

- Abraham, Lisa, and Alison Stein.** 2019. “Words Matter: Experimental Evidence from Job Applications.” *Working Paper*.
- Adams, Nathan R., Michael A. Kuhn, and Glen R. Waddell.** 2019. “Confidence and Contrition: Is Cheating Internalized in Performance Assessments?” *Working Paper*.
- Alston, Mackenzie Alston.** 2019. “The (Perceived) Cost of Being Female: An Experimental Investigation of Strategic Responses to Discrimination.” *Working Paper*.
- Baert, Stijn, Ann-Sophie De Pauw, and Nick Deschacht.** 2016. “Do employer preferences contribute to sticky floors?” *ILR Review*, 69(3): 714–736.
- Blau, Francine D., and Lawrence M. Kahn.** 2017. “The Gender Wage Gap: Extent, Trends, and Explanations.” *Journal of Economic Literature*, 55(3).
- Bohnet, Iris, Alexandra van Geen, and Max Bazerman.** 2016. “When Performance Trumps Gender Bias: Joint Versus Separate Evaluation.” *Management Science*, 62(5): 1225–1234.
- Bohren, Aislinn, Alex Imas, and Michael Rosenberg.** 2018. “The Language of Discrimination: Using Experimental versus Observational Data.” *AEA Papers and Proceedings*, 108(169–74).
- Bohren, J Aislinn, Alex Imas, and Michael Rosenberg.** 2019. “The Dynamics of Discrimination: Theory and Evidence.” *American Economic Review*.
- Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G. Pope.** 2019. “Inaccurate Statistical Discrimination.” *NBER Working Paper*.
- Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer.** 2018. “Belief about Gender.” *American Economic Review*.
- Born, Andreas, Eva Ranehill, and Anna Sandberg.** 2018. “A man’s world? – The impact of a male dominated environment on female leadership.” *Working Paper*.
- Bosquet, Clément, Pierre-Philippe Combes, and Cecilia García-Peñalosa.** 2019. “Gender and promotions: evidence from academic economists in France.” *The Scandinavian Journal of Economics*, 121(3): 1020–1053.
- Bowles, Hannah Riley, and Kathleen L. McGinn.** 2008. “Gender in job negotiations: a two-level game.” *Negotiation Journal*, 24(4): 393–410.
- Bowles, Hannah Riley, Linda Babcock, and Lei Lai.** 2007. “Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask.” *Organizational Behavior and Human Decision Processes*, 103(1): 84–103.

- Buser, Thomas, Leonie Gerhards, and Joël J Van der Weele.** 2018. “Measuring responsiveness to feedback as a personal trait.” *Journal of Risk and Uncertainty*, 56(2): 165–192.
- Charness, Gary, Aldo Rustichini, and Jeroen Van de Ven.** 2018. “Self-confidence and strategic behavior.” *Experimental Economics*, 21(1): 72–98.
- Childs, Jason.** 2012. “Gender differences in lying.” *Economics Letters*, 114(2): 147–149.
- Coffman, Katherine Baldiga.** 2014. “Evidence on Self-Stereotyping and the Contribution of Ideas.” *The Quarterly Journal of Economics*, 129(4): 1625–1660.
- Coffman, Katherine B., Christine L. Exley, and Muriel Niederle.** 2019. “When gender discrimination isn’t about gender.” *Working Paper*.
- Coffman, Katherine B., Manuela R. Collis, and Leena Kulkarni.** 2019a. “When to Apply?” *Working Paper*.
- Coffman, Katherine, Clio Bryant Flikkema, and Olga Shurchkov.** 2019. “Gender Stereotypes in Deliberation and team decisions.” *Working Paper*.
- Coffman, Katherine, Manuela Collis, and Leena Kulkarni.** 2019b. “Stereotypes and Belief Updating.” *Working Paper*.
- Coutts, Alexander.** 2018. “Good news and bad news are still news: Experimental evidence on belief updating.” *Experimental Economics*, 1–27.
- Croson, Rachel, and Uri Gneezy.** 2009. “Gender Differences in Preferences.” *Journal of Economic Literature*, 47(2): 448–474.
- Dreber, Anna, and Magnus Johannesson.** 2008. “Gender differences in deception.” *Economics Letter*, 99(1).
- Eckel, Catherine C., and Philip J. Grossman.** 2008. “Men, Women and Risk Aversion: Experimental Evidence.” In *Handbook of Experimental Economics Results*. 1061–1073.
- Erat, Sanjiv, and Uri Gneezy.** 2012. “White Lies.” *Management Science*, 50(4): 723–733.
- Ertac, Seda.** 2011. “Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback.” *Journal of Economic Behavior & Organization*, 80(3): 532–545.
- Exley, Christine L., Muriel Niederle, and Lise Vesterlund.** Forthcoming. “Knowing When to Ask: The Cost of Leaning-in.” *Journal of Political Economy*.
- Gneezy, Uri.** 2005. “Deception: The Role of Consequences.” *The American Economic Review*, 95(1): pp. 384–394.

- Gneezy, Uri, Muriel Niederle, and Aldo Rustichini.** 2003. "Performance in competitive environments: Gender differences." *The Quarterly Journal of Economics*, 118(3): 1049–1074.
- Goldin, Claudia.** 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review*, 104(4): 1091–1119.
- Gylfason, Haukur Freyr, Audur Arna Arnardottir, and Kari Kristinsson.** 2013. "More on gender differences in lying." *Economic Letters*, 119(1): 94–95.
- He, Joyce, Sonia Kang, and Nicola Lacetera.** 2019. "Leaning in or not leaning out? Opt-out choice framing attenuates gender differences in the decision to compete." *Working Paper*.
- Houser, Daniel, Stefan Vetter, and Joachim Winter.** 2012. "Fairness and cheating." *European Economic Review*, 56: 1645–1655.
- Isaksson, Siri.** 2018. "It Takes Two: Gender Differences in Group Work." *Working Paper*.
- Kessler, Judd B.** 2017. "Announcements of support and public good provision." *American Economic Review*, 107(12): 3760–87.
- Kessler, Judd B, Corinne Low, and Colin D Sullivan.** 2019. "Incentivized Resume Rating: Eliciting Employer Preferences without Deception." *American Economic Review*.
- Leibbrandt, Andreas, and John A. List.** 2015. "Do women avoid salary negotiations? Evidence from a large-scale natural field experiment." *Management Science*, 61(9): 2016–2024.
- Lundeberg, Mary A, Paul W Fox, and Judith Punčochař.** 1994. "Highly confident but wrong: Gender differences and similarities in confidence judgments." *Journal of educational psychology*, 86(1).
- Mazei, Jens, Joachim Hüffmeier, Philipp Alexander Freund, Alice F. Stuhlmacher, Lena Bilke, and Guido Hertel.** 2015. "A meta-analysis on gender differences in negotiation outcomes and their moderators." *Psychological Bulletin*, 141(1): 85–104.
- Milkman, Katherine L, Modupe Akinola, and Dolly Chugh.** 2015. "What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations." *Journal of Applied Psychology*, 100(6).
- Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat.** 2014. "Managing Self-Confidence: Theory and Experimental Evidence." *Working Paper*.
- Moss-Racusin, Corinne A, and Laurie A Rudman.** 2010. "Disruptions in women's self-promotion: The backlash avoidance model." *Psychology of women quarterly*, 34(2): 186–202.
- Niederle, Muriel.** 2016. "Gender." In *Handbook of Experimental Economics*. Vol. 2, , ed. John Kagel and Alvin E. Roth, 481–553. Princeton University Press.

- Niederle, Muriel, and Lise Vesterlund.** 2007. “Do Women shy away from competition? Do men compete too much?” *Quarterly Journal of Economics*, 122(3): 1067–1101.
- Niederle, Muriel, and Lise Vesterlund.** 2011. “Gender and Competition.” *Annual Review of Economics*, 3: 601–630.
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales.** 2014. “How stereotypes impair women’s careers in science.” *Proceedings of the National Academy of Sciences*, 111(12): 4403–4408.
- Riach, P. A., and J. Rich.** 2002. “Field Experiments of Discrimination in the Market Place.” *The Economic Journal*, 112(483).
- Rudman, Laurie A.** 1998. “Self-promotion as a risk factor for women: the costs and benefits of counter-stereotypical impression management.” *Journal of personality and social psychology*, 74(3).
- Rudman, Laurie A, and Julie E Phelan.** 2008. “Backlash effects for disconfirming gender stereotypes in organizations.” *Research in organizational behavior*, 28(6-79).
- Sarsons, Heather.** 2017a. “Interpreting Signals in the Labor Market: Evidence from Medical Referrals.” *Working Paper*.
- Sarsons, Heather.** 2017b. “Recognition for Group Work: Gender Differences in Academia†.” *American Economic Review: Papers & Proceedings*, 107(5): 141–145.
- Schwardmann, Peter, and Joël van der Weele.** Forthcoming. “Deception and Self-Deception.” *Nature Human Behavior*.
- Small, Deborah A., Michele Gelfand, Linda Babcock, and Hilary Gettman.** 2007. “Who goes to the bargaining table? The influence of gender and framing on the initiation of negotiation.” *Journal of personality and social psychology*, 93(4): 583–599.
- Soldà, Alice, Changxia Ke, Lionel Page, and William von Hippel.** 2019. “Strategically delusional.” *Working Paper*.

A Appendix

A.1 Performance and Confidence by Study Version

This section reports on performance and confidence data by study version. In all versions, there is no gender gap in performance but there is a gender gap in confidence.

Figure A.1: *Public Version*, Performance and Confidence

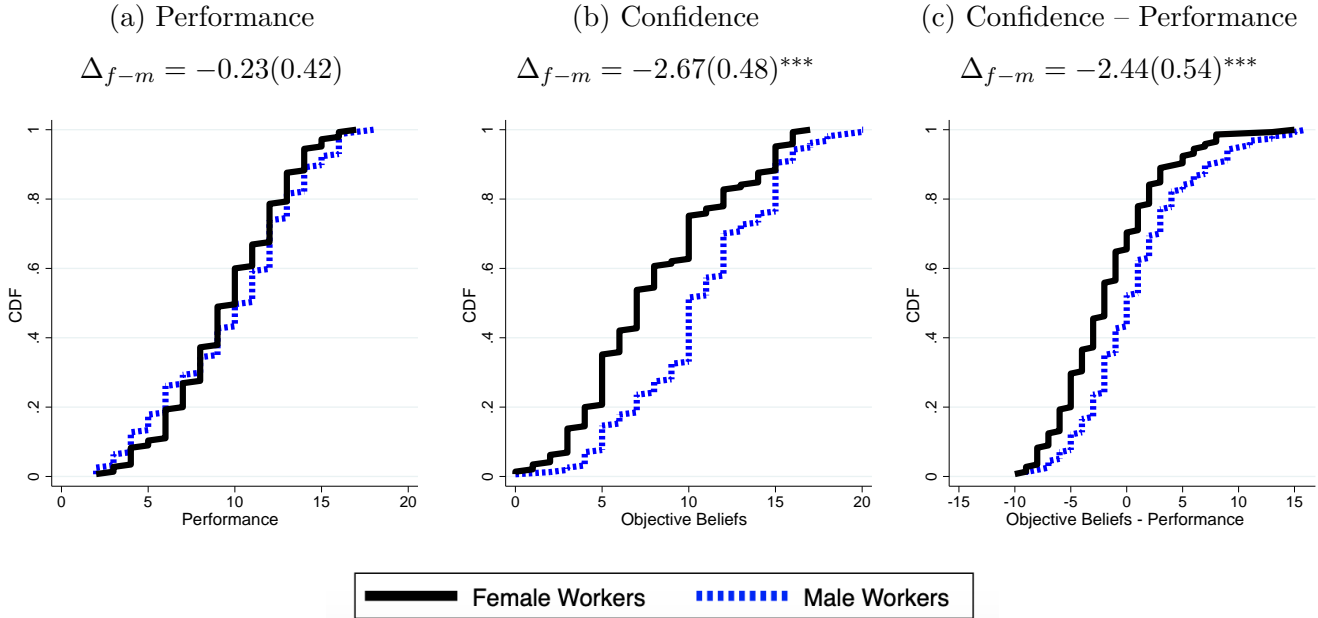


Table A.1: In the *Public Version*, Performance and Confidence Regressions

DV:	<i>Performance</i>	<i>Confidence</i>		<i>Confidence – Performance</i>
	(1)	(2)	(3)	(4)
<i>Female</i>	-0.23 (0.42)	-2.67*** (0.48)	-2.20*** (0.45)	-2.44*** (0.54)
Constant	9.84*** (0.31)	10.67*** (0.32)		0.83** (0.39)
N	302	302	302	302
Performance FEs	No	No	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of detailed in Table 2. Data are from the *Public* version.

Figure A.2: *Private Version*, Performance and Confidence

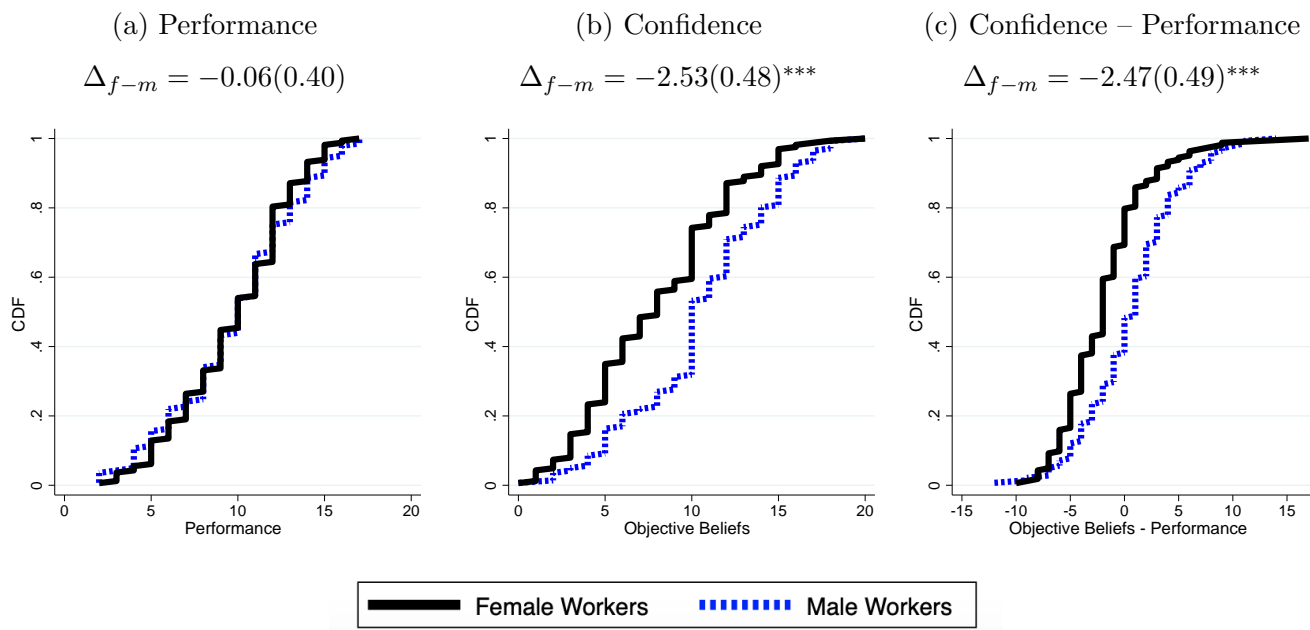


Table A.2: In the *Private Version*, Performance and Confidence Regressions

DV:	<i>Performance</i>	<i>Confidence</i>		<i>Confidence – Performance</i>
	(1)	(2)	(3)	(4)
<i>Female</i>	-0.06 (0.40)	-2.53*** (0.48)	-2.36*** (0.46)	-2.47*** (0.49)
Constant	9.84*** (0.31)	10.50*** (0.35)		0.65* (0.37)
N	304	304	304	304
Performance FEs	No	No	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of detailed in Table 2. Data are from the *Private* version.

Figure A.3: *Ambiguous Version*, Performance and Confidence

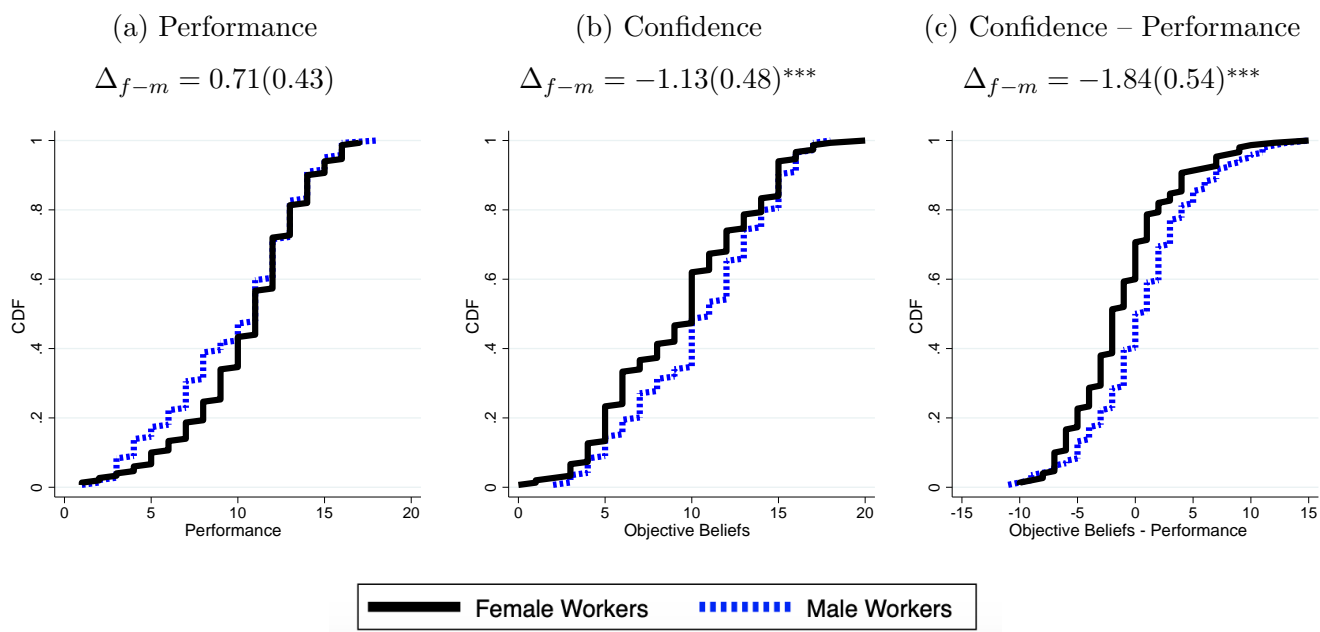


Table A.3: In the *Ambiguous Version*, Performance and Confidence Regressions

DV:	<i>Performance</i>	<i>Confidence</i>		<i>Confidence - Performance</i>
	(1)	(2)	(3)	(4)
<i>Female</i>	0.71 (0.43)	-1.13** (0.48)	-1.06** (0.46)	-1.84*** (0.54)
Constant	9.78*** (0.33)	10.53*** (0.33)		0.75* (0.40)
N	294	294	294	294
Performance FEs	No	No	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of detailed in Table 2. Data are from the *Ambiguous* version.

Figure A.4: *Private (Replication)*, Performance and Confidence

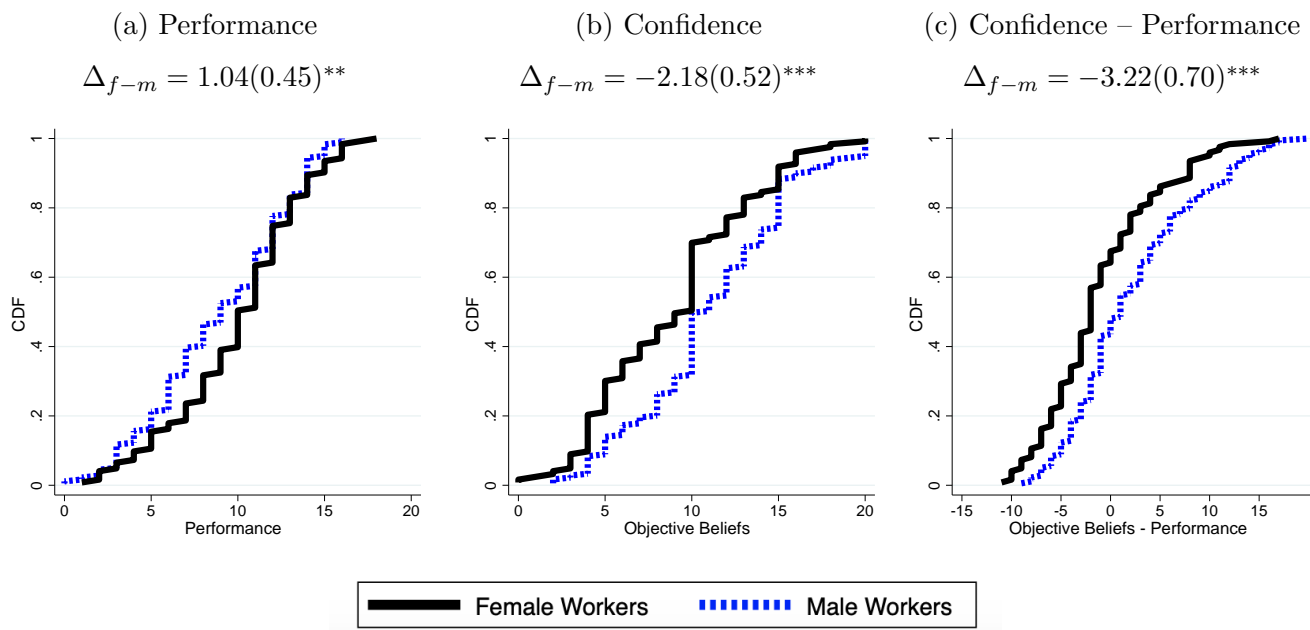


Table A.4: In the *Private (Replication)* Version, Performance and Confidence Regressions

DV:	<i>Performance</i>	<i>Confidence</i>		<i>Confidence – Performance</i>
	(1)	(2)	(3)	(4)
<i>Female</i>	1.04** (0.45)	-2.18*** (0.52)	-1.94*** (0.53)	-3.22*** (0.70)
Constant	8.95*** (0.30)	11.12*** (0.33)		2.17*** (0.47)
N	302	302	302	302
Performance FEs	No	No	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of detailed in Table 2. Data are from the *Private (Replication)* version.

Figure A.5: *Private (Information about Others) Version*, Performance and Confidence

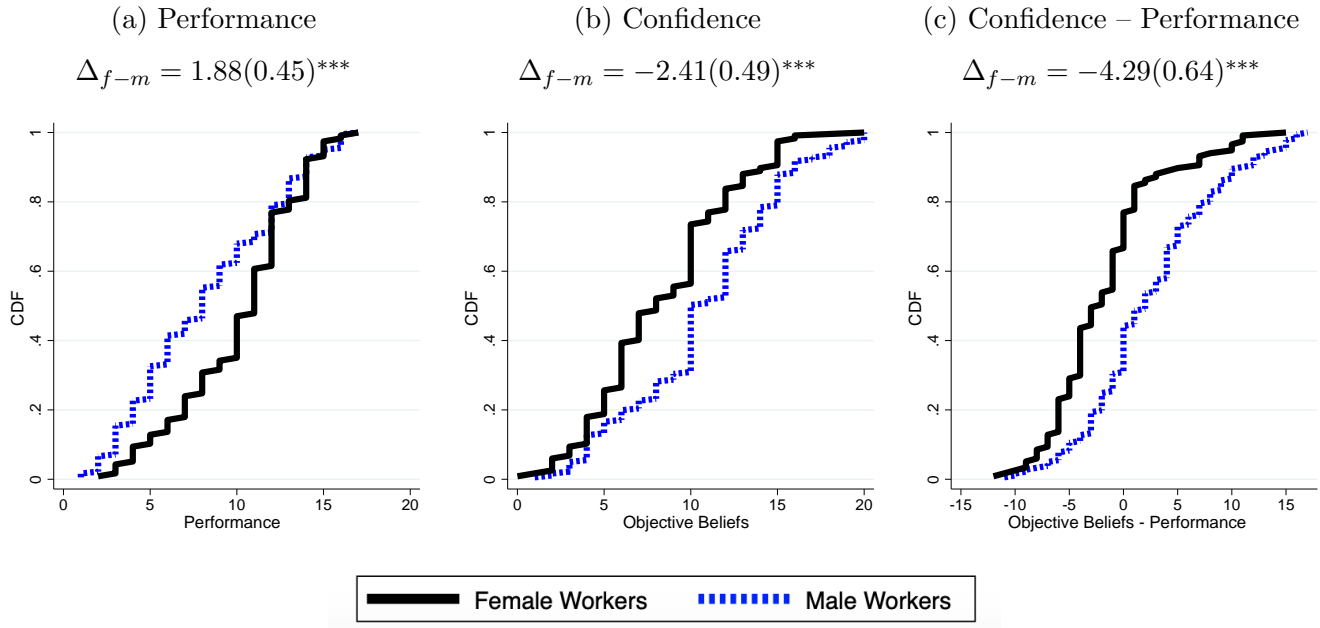


Table A.5: In the *Private Version (Information about Others)*, Performance and Confidence Regressions

DV:	<i>Performance</i>	<i>Confidence</i>		<i>Confidence – Performance</i>
	(1)	(2)	(3)	(4)
<i>Female</i>	1.88*** (0.45)	-2.41*** (0.49)	-2.29*** (0.52)	-4.29*** (0.64)
Constant	8.25*** (0.31)	10.79*** (0.33)		2.54*** (0.44)
N	298	298	298	298
Performance FEs	No	No	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of detailed in Table 2. Data are from the *Private (Information about Others)* version.

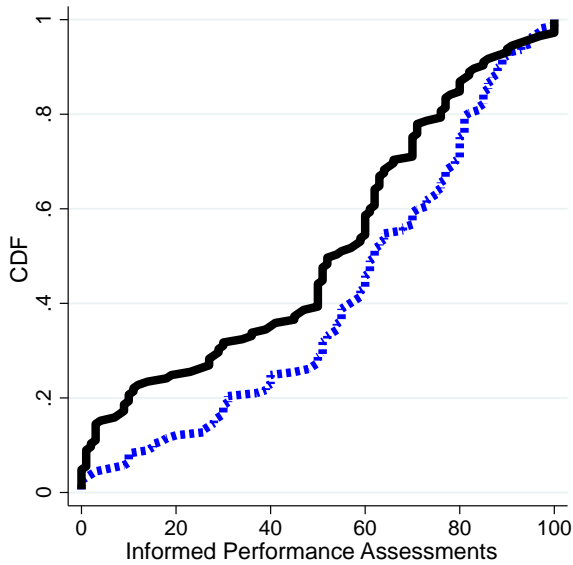
A.2 Additional Figures and Tables on Self-Promotion

A.2.1 Public Version

Figure A.6: *Public Version*, Informed Self-Promotion

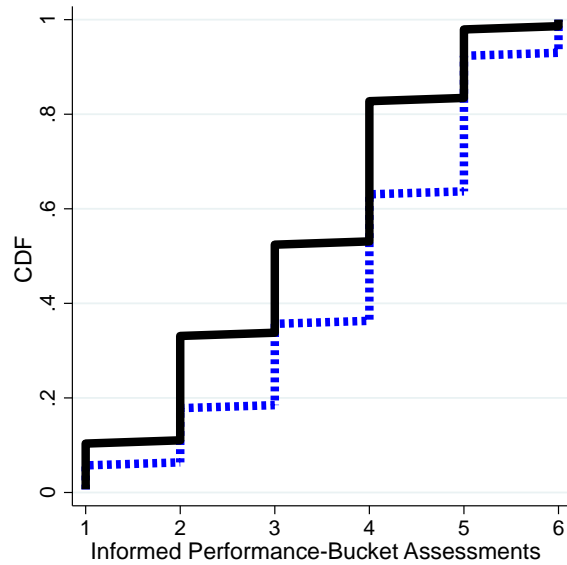
(a) Performance Assessments

$$\Delta_{f-m} = -10.97(3.27)^{***}$$



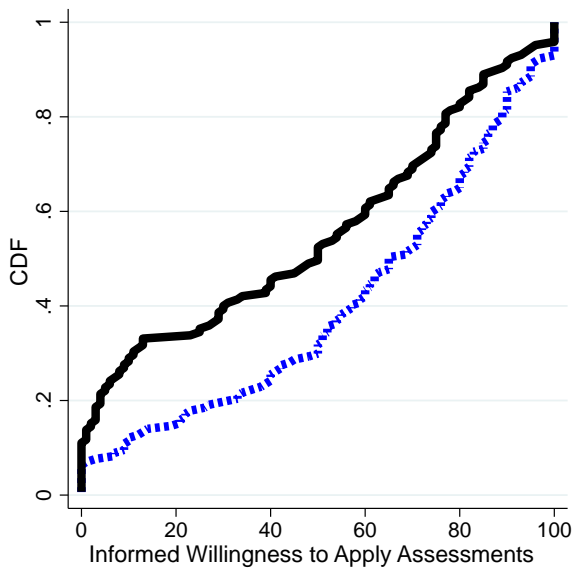
(b) Performance-Bucket Assessments

$$\Delta_{f-m} = -0.62(0.15)^{***}$$



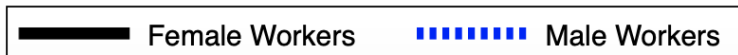
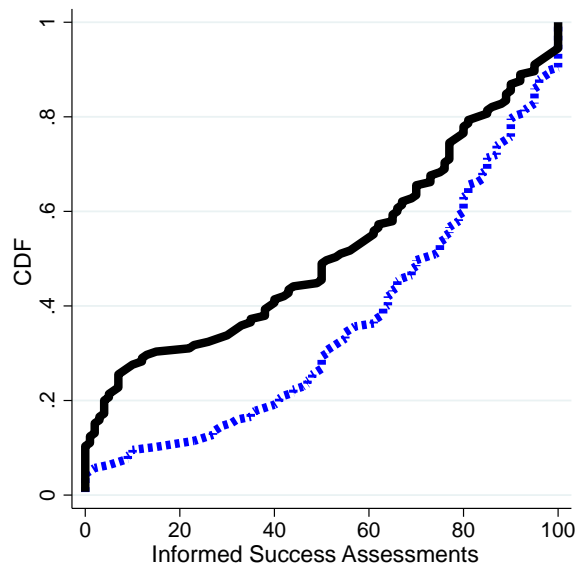
(c) Willingness to Apply Assessments

$$\Delta_{f-m} = -15.55(3.72)^{***}$$



(d) Success Assessments

$$\Delta_{f-m} = -16.19(3.73)^{***}$$

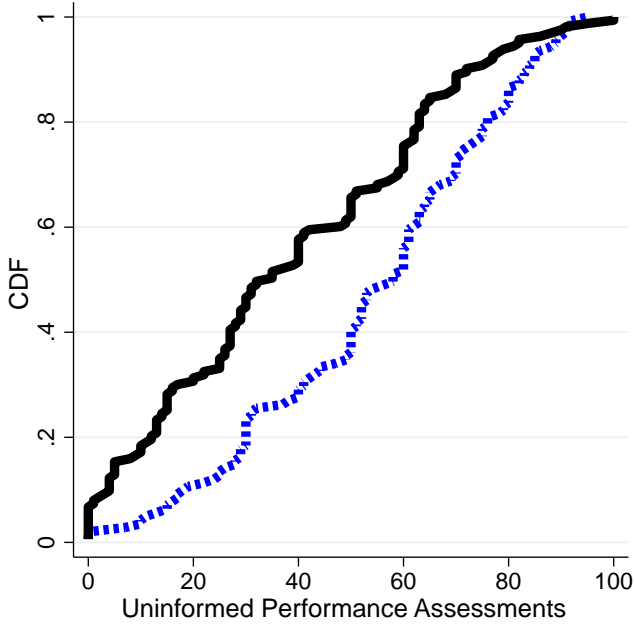


A.2.2 Private Version

Figure A.7: *Private Version*, Uninformed Self-Promotion

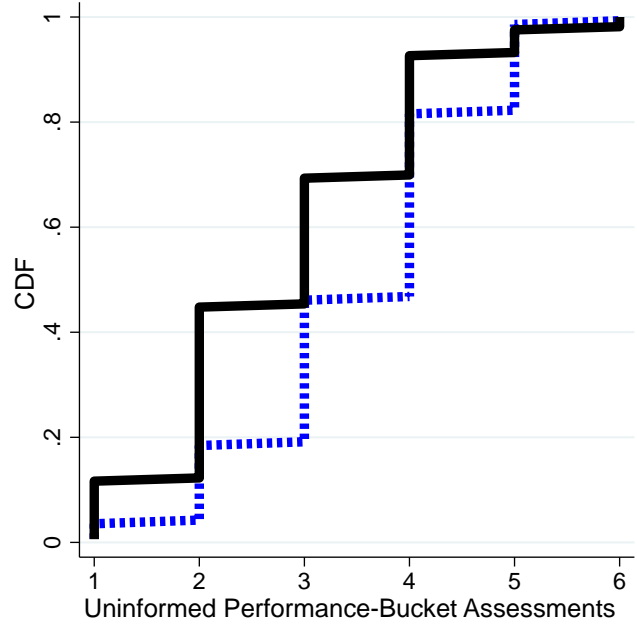
(a) Performance Assessments

$$\Delta_{f-m} = -15.80(2.88)^{***}$$



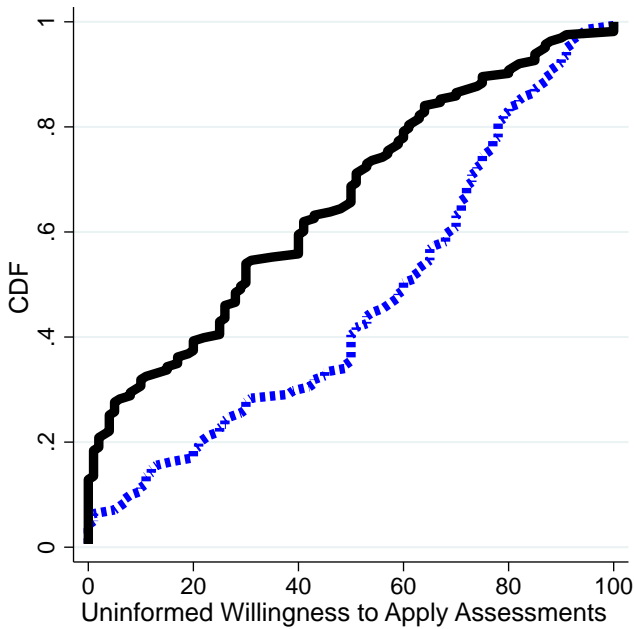
(b) Performance-Bucket Assessments

$$\Delta_{f-m} = -0.68(0.13)^{***}$$



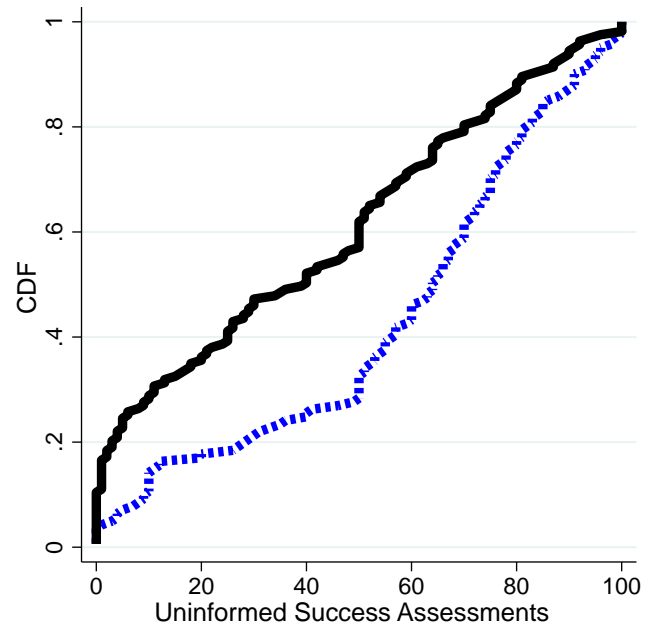
(c) Willingness to Apply Assessments

$$\Delta_{f-m} = -19.26(3.33)^{***}$$



(d) Success Assessments

$$\Delta_{f-m} = -18.60(3.46)^{***}$$

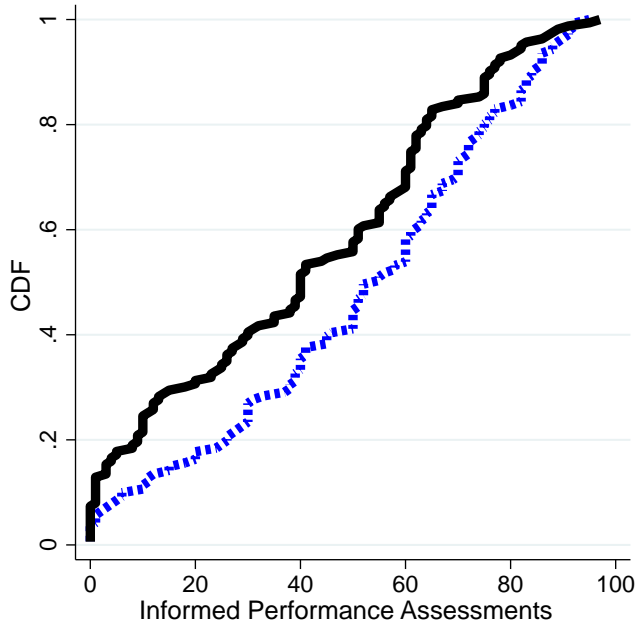


— Female Workers - - - - - Male Workers

Figure A.8: *Private Version*, Informed Self-Promotion

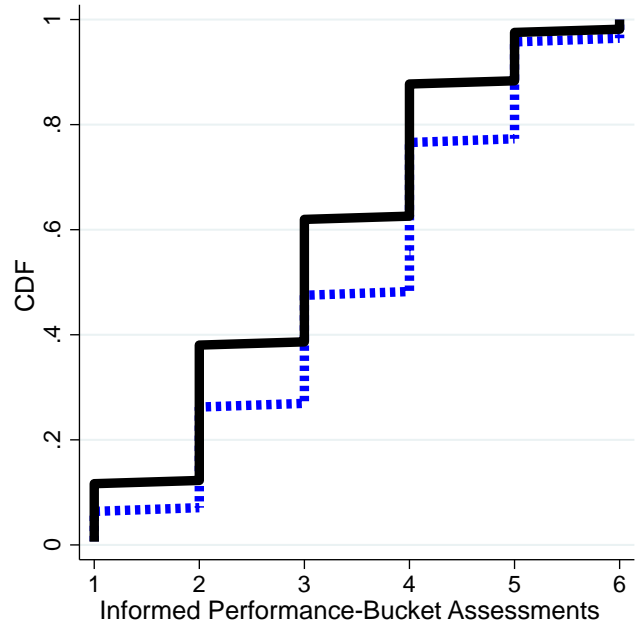
(a) Performance Assessments

$$\Delta_{f-m} = -10.96(3.14)^{***}$$



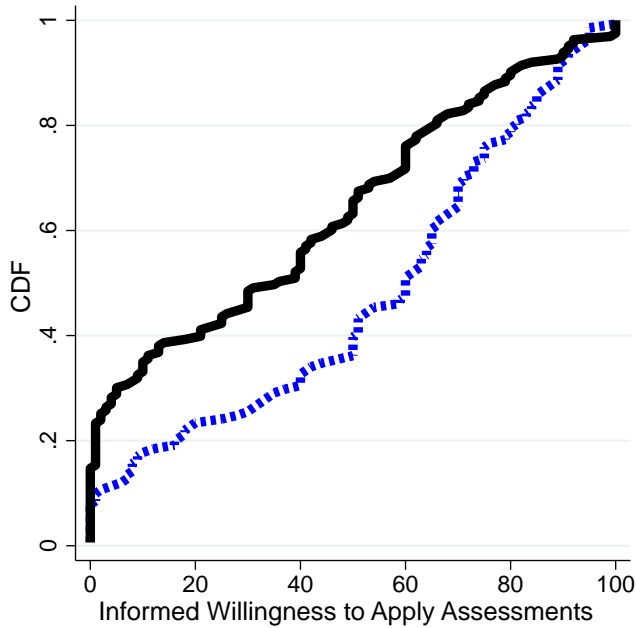
(b) Performance-Bucket Assessments

$$\Delta_{f-m} = -0.44(0.15)^{***}$$



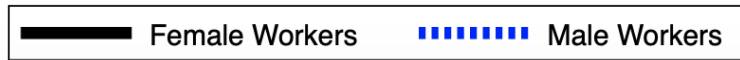
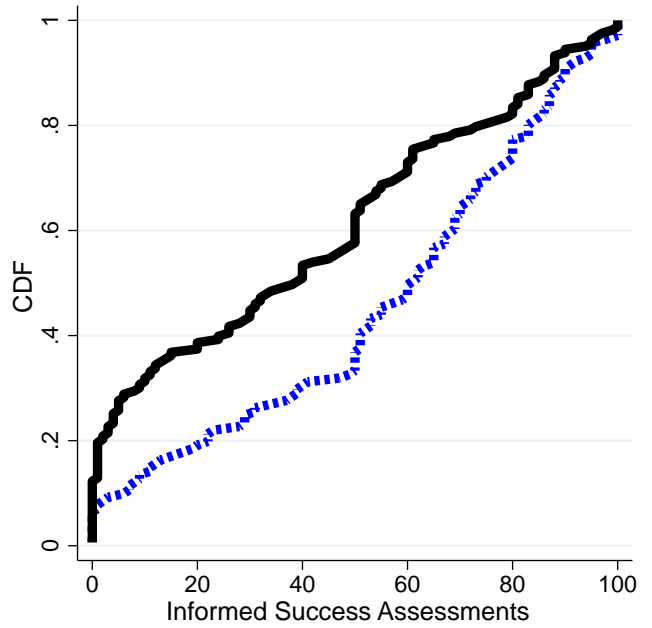
(c) Willingness to Apply Assessments

$$\Delta_{f-m} = -16.33(3.56)^{***}$$



(d) Success Assessments

$$\Delta_{f-m} = -16.13(3.61)^{***}$$

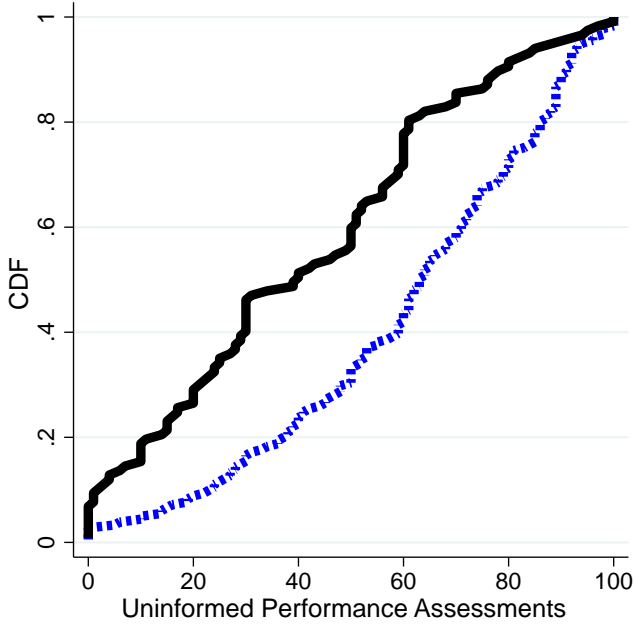


A.2.3 Private (Information about Others) Version

Figure A.9: *Private (Information about Others) Version, Uninformed Self-Promotion*

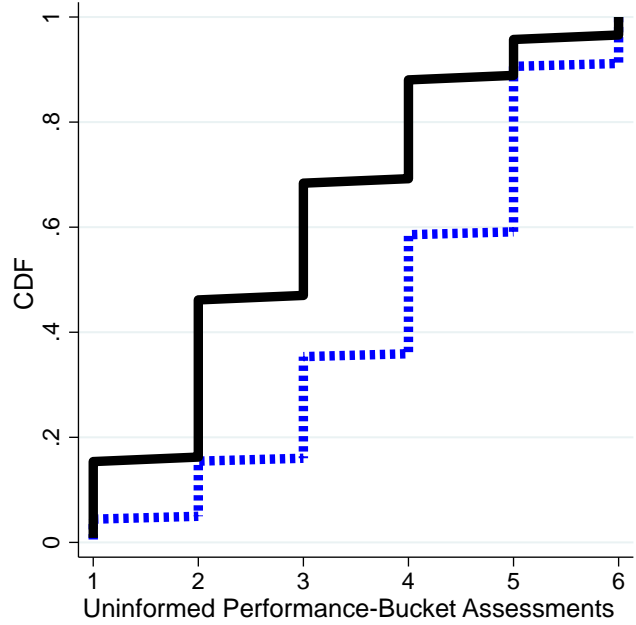
(a) Performance Assessments

$$\Delta_{f-m} = -19.95(3.21)^{***}$$



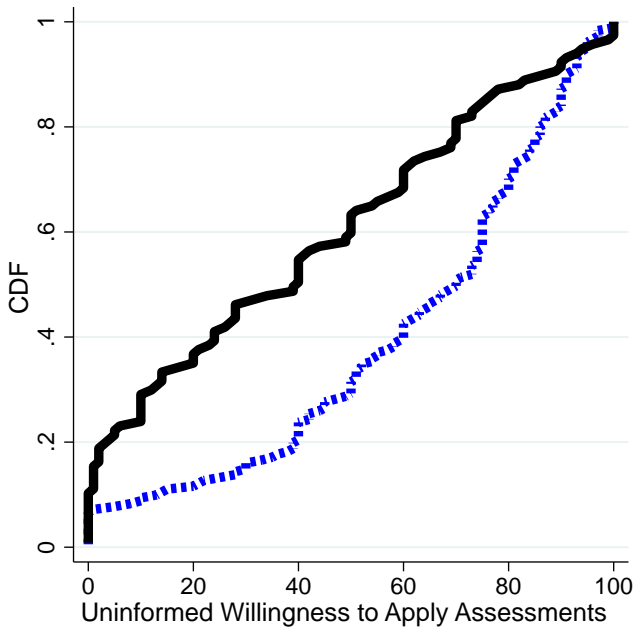
(b) Performance-Bucket Assessments

$$\Delta_{f-m} = -1.09(0.16)^{***}$$



(c) Willingness to Apply Assessments

$$\Delta_{f-m} = -21.75(3.61)^{***}$$



(d) Success Assessments

$$\Delta_{f-m} = -21.11(3.61)^{***}$$

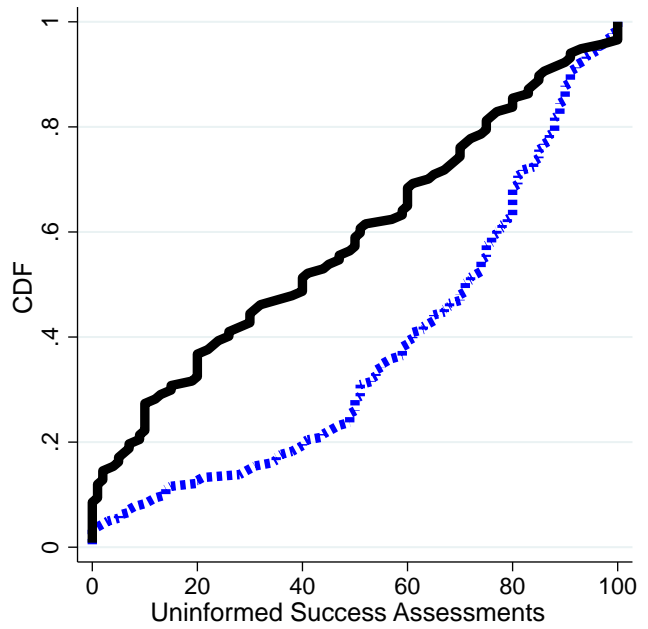
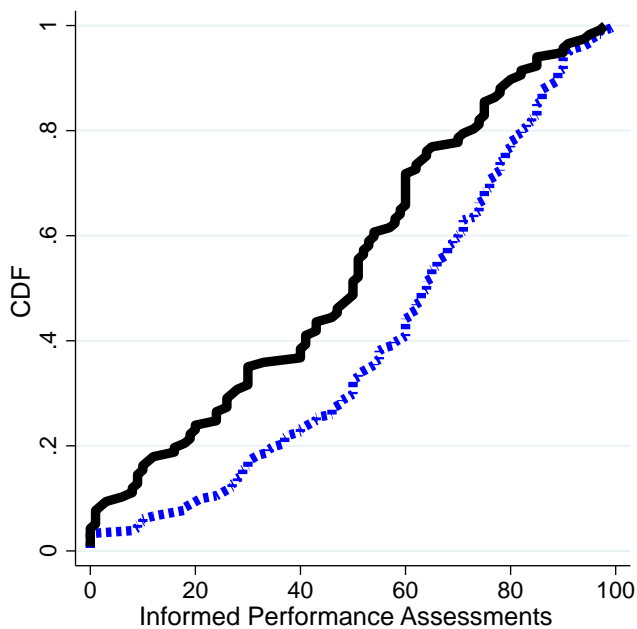


Figure A.10: *Private (Information about Others) Version, Informed Self-Promotion*

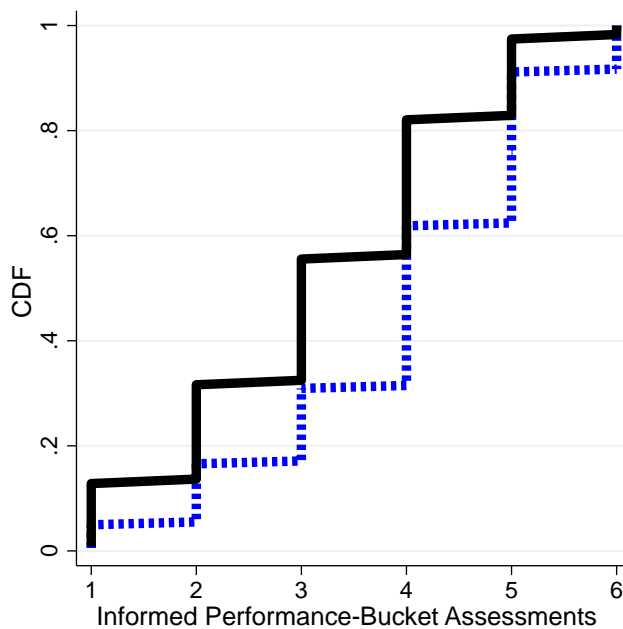
(a) Performance Assessments

$$\Delta_{f-m} = -13.78(3.14)^{***}$$



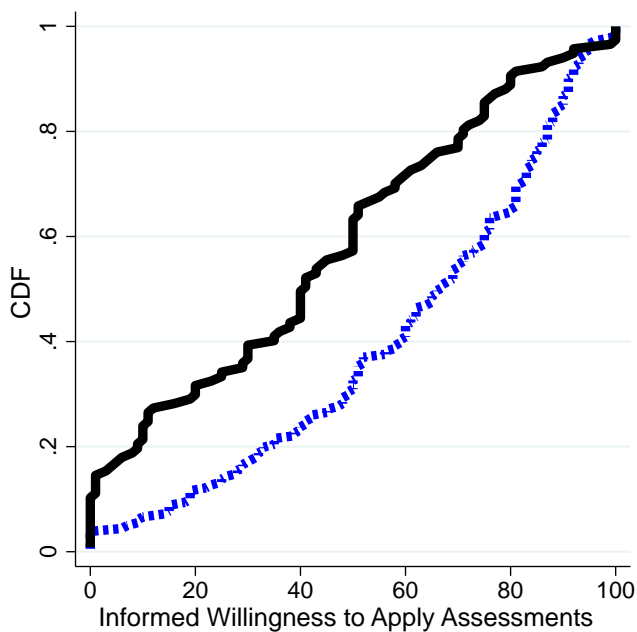
(b) Performance-Bucket Assessments

$$\Delta_{f-m} = -0.74(0.16)^{***}$$



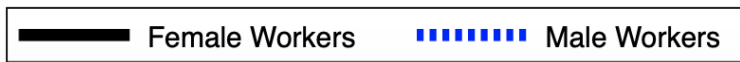
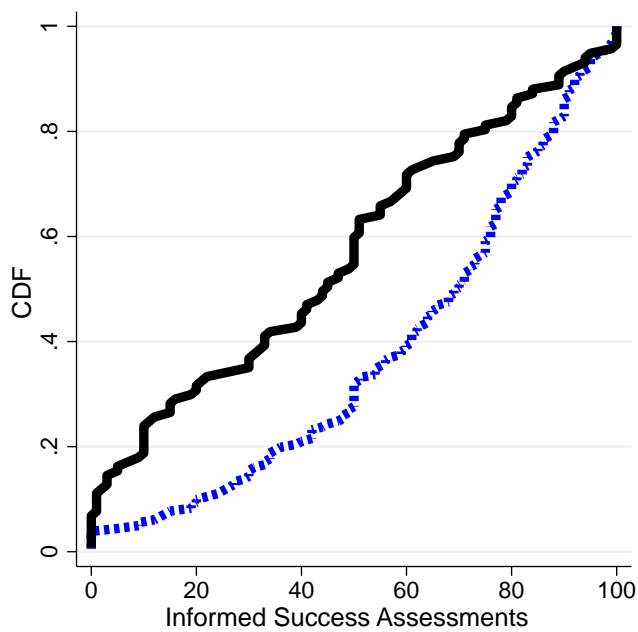
(c) Willingness to Apply Assessments

$$\Delta_{f-m} = -19.67(3.44)^{***}$$



(d) Success Assessments

$$\Delta_{f-m} = -19.21(3.50)^{***}$$



A.2.4 Private (Replication) Version

Table A.6: *Private (Replication) Version*, Self-Promotion Regressions

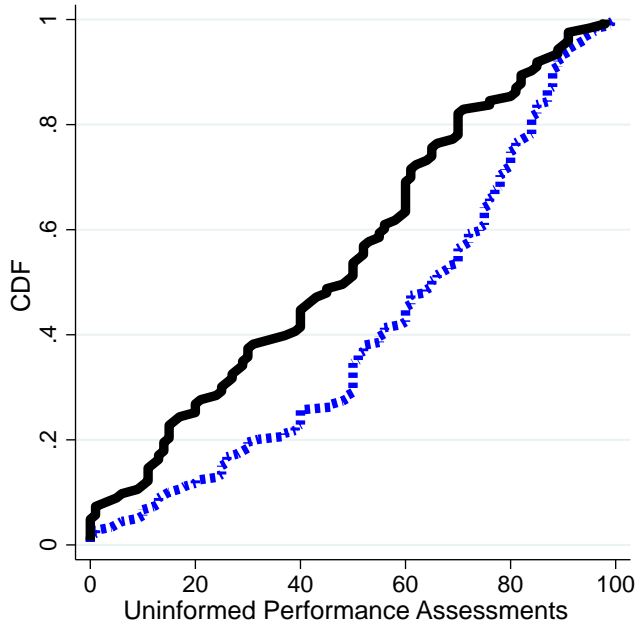
Data:	Uninformed Self-Promotion		Informed Self-Promotion	
	(1)	(2)	(3)	(4)
DV = Performance Assessments				
<i>Female</i>	-14.45*** (3.21)	-12.21*** (3.18)	-9.51*** (3.21)	-7.58** (3.18)
Constant	59.49*** (1.99)		57.73*** (1.94)	
DV = Performance-Bucket Assessments				
<i>Female</i>	-0.70*** (0.16)	-0.55*** (0.15)	-0.54*** (0.15)	-0.42*** (0.15)
Constant	3.87*** (0.10)		3.85*** (0.10)	
DV = Willingness to Apply Assessments				
<i>Female</i>	-19.28*** (3.48)	-17.25*** (3.54)	-16.18*** (3.51)	-14.15*** (3.53)
Constant	58.10*** (2.09)		59.99*** (2.13)	
DV = Success Assessments				
<i>Female</i>	-16.49*** (3.57)	-14.39*** (3.53)	-15.75*** (3.51)	-14.37*** (3.46)
Constant	62.16*** (2.00)		62.40*** (1.96)	
N	302	302	302	302
Performance FEs	No	Yes	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from the same specifications noted in Table 3. Data in columns in (1) and (2) are from uninformed self-assessment questions elicited in part 2 of the *Private (Replication)* version and data in columns (3) and (4) are from informed self-assessment questions elicited in part 3 of the *Private (Replication)* version.

Figure A.11: *Private (Replication) Version*, Uninformed Self-Promotion

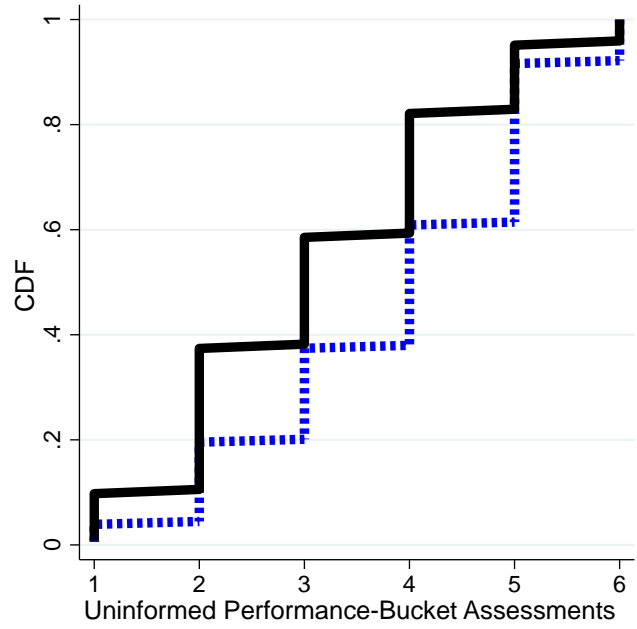
(a) Performance Assessments

$$\Delta_{f-m} = -14.05(3.21)^{***}$$



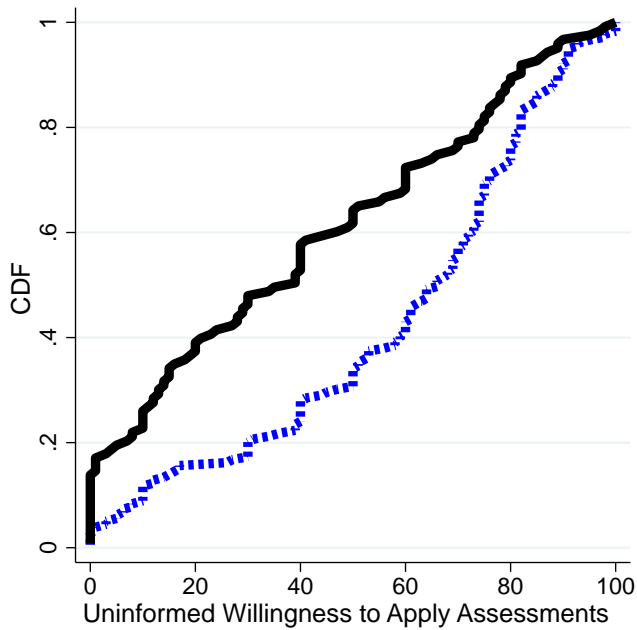
(b) Performance-Bucket Assessments

$$\Delta_{f-m} = -0.70(0.16)^{***}$$



(c) Willingness to Apply Assessments

$$\Delta_{f-m} = -19.28(3.48)^{***}$$



(d) Success Assessments

$$\Delta_{f-m} = -16.49(3.57)^{***}$$

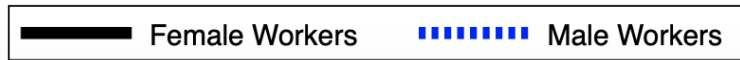
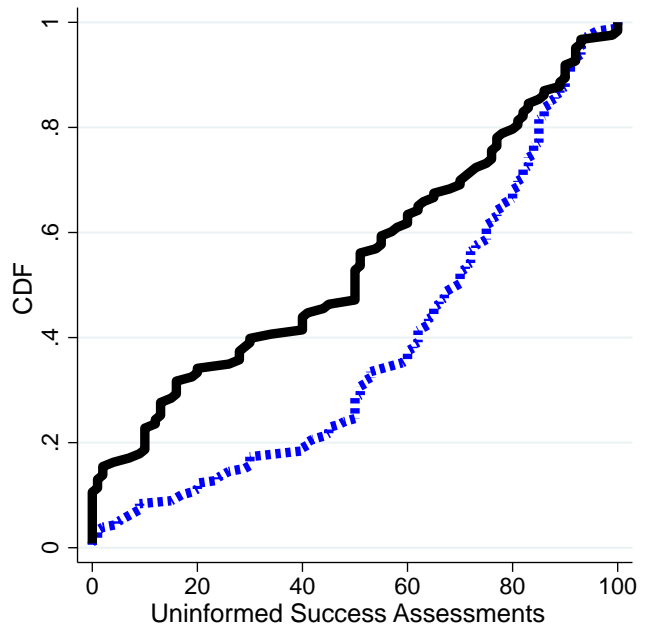
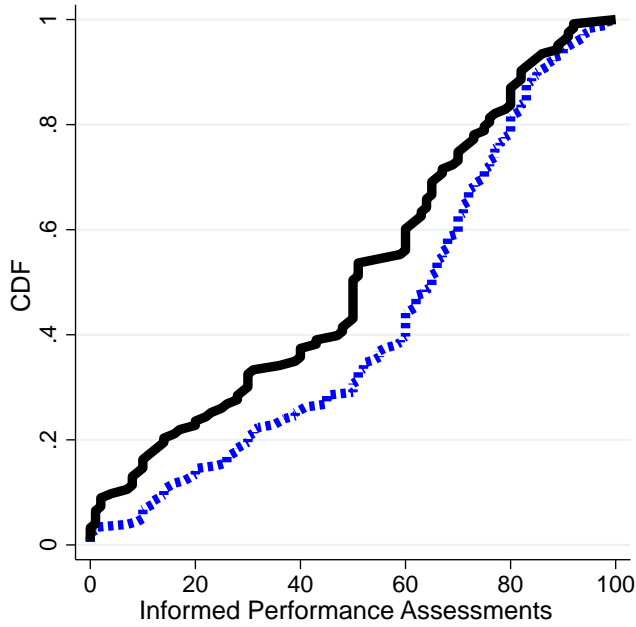


Figure A.12: *Private (Replication) Version*, Informed Self-Promotion

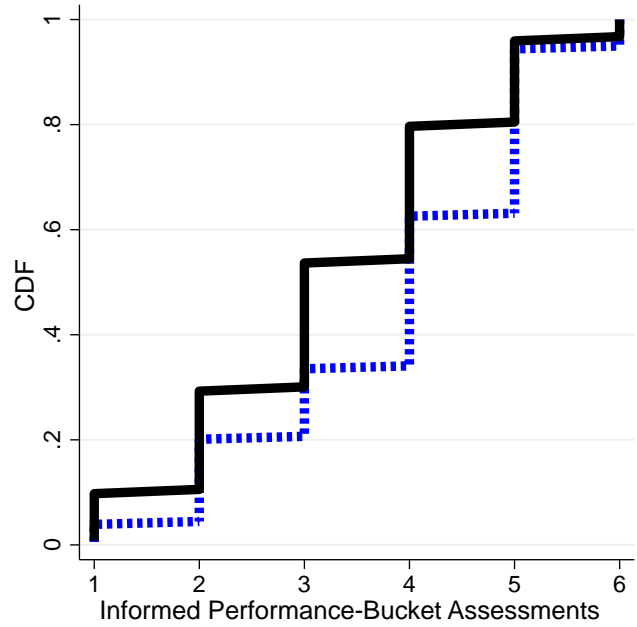
(a) Performance Assessments

$$\Delta_{f-m} = -9.51(3.21)^{***}$$



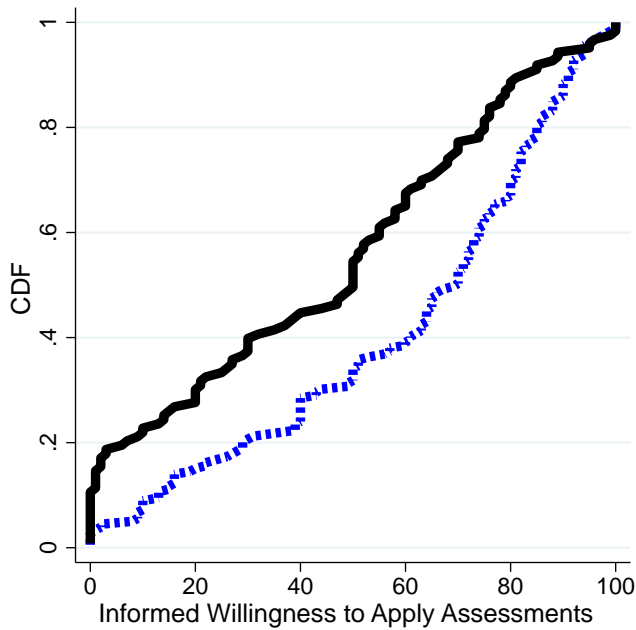
(b) Performance-Bucket Assessments

$$\Delta_{f-m} = -0.54(0.15)^{***}$$



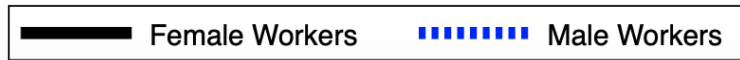
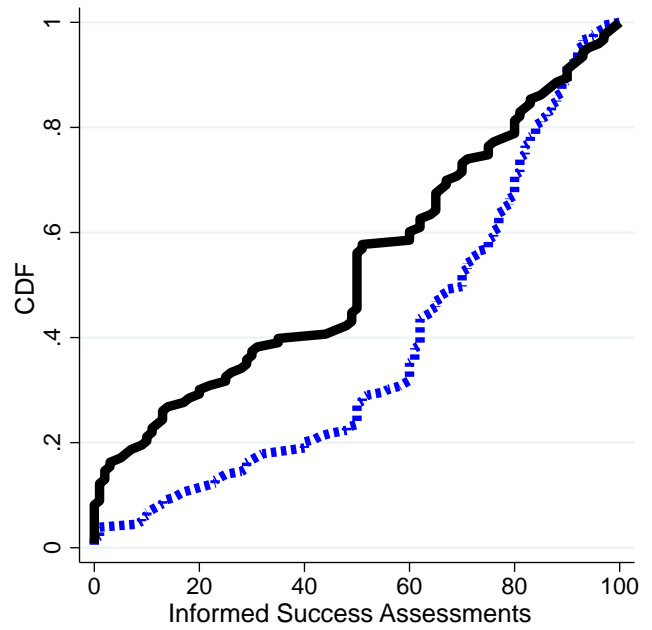
(c) Willingness to Apply Assessments

$$\Delta_{f-m} = -16.18(3.51)^{***}$$



(d) Success Assessments

$$\Delta_{f-m} = -15.75(3.51)^{***}$$

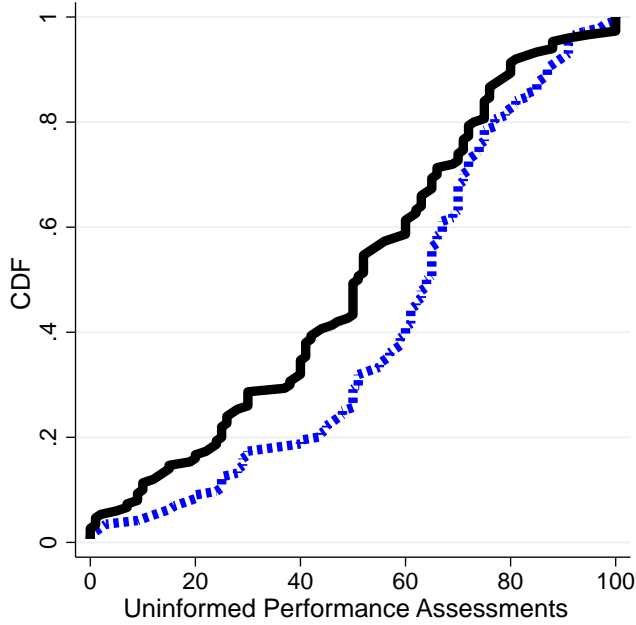


A.2.5 Ambiguous Version

Figure A.13: *Ambiguous Version*, Uninformed Self-Promotion

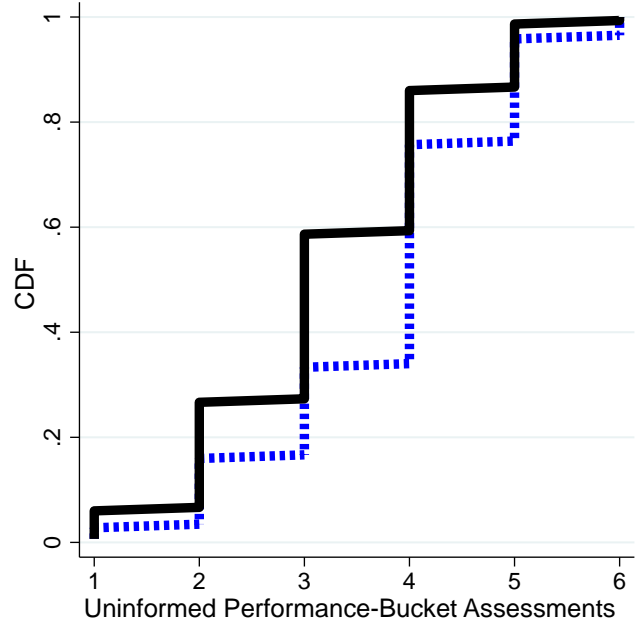
(a) Performance Assessments

$$\Delta_{f-m} = -9.79(2.90)^{***}$$



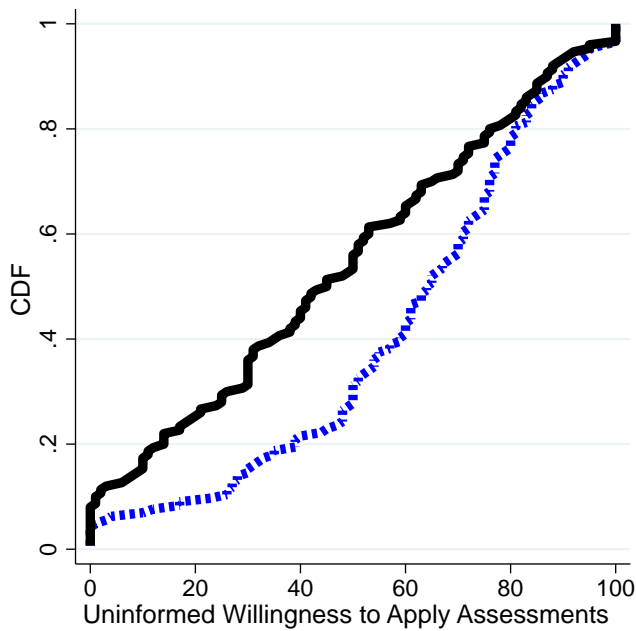
(b) Performance-Bucket Assessments

$$\Delta_{f-m} = -0.52(0.13)^{***}$$



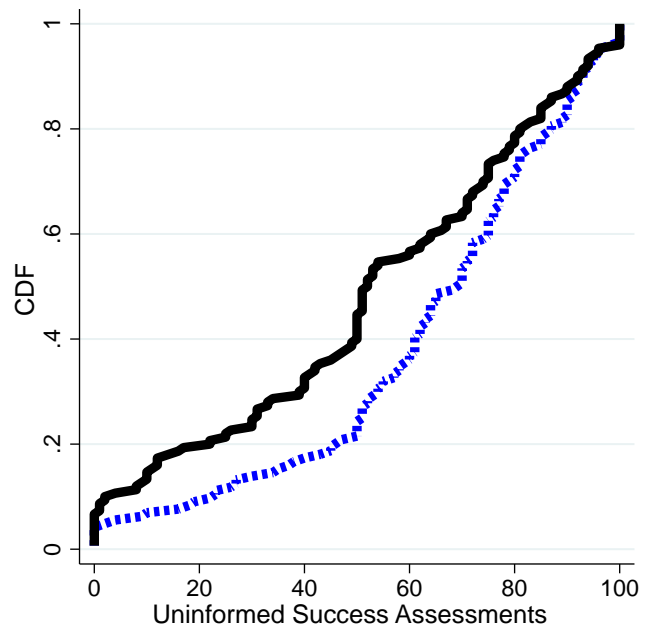
(c) Willingness to Apply Assessments

$$\Delta_{f-m} = -14.43(3.28)^{***}$$



(d) Success Assessments

$$\Delta_{f-m} = -10.70(3.30)^{***}$$

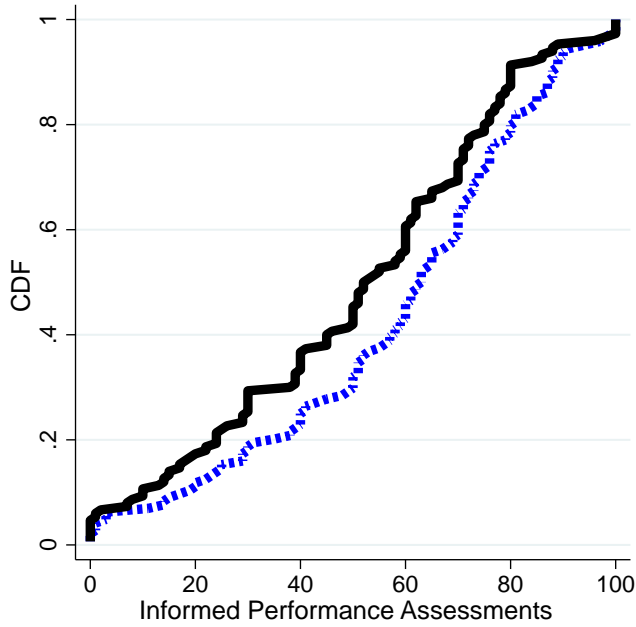


— Female Workers ····· Male Workers

Figure A.14: *Ambiguous Version*, Informed Self-Promotion

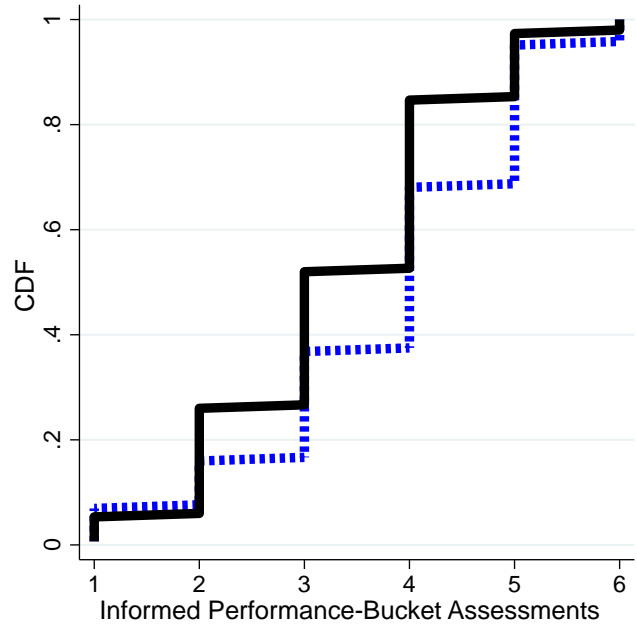
(a) Performance Assessments

$$\Delta_{f-m} = -7.76(3.09)**$$



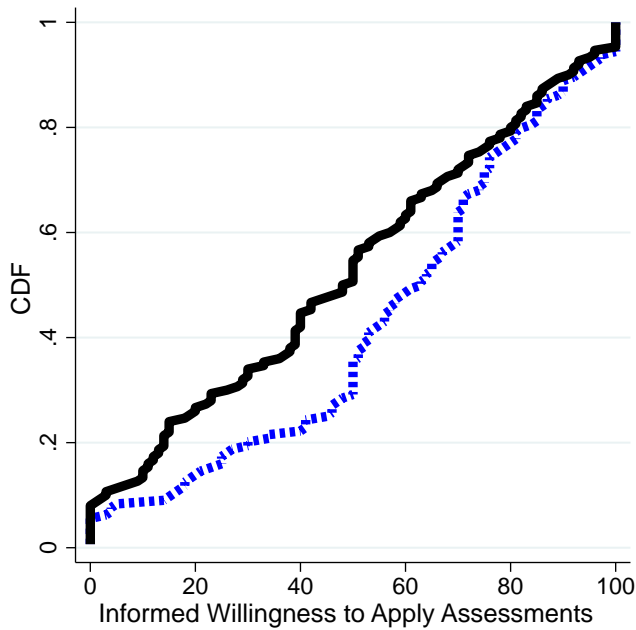
(b) Performance-Bucket Assessments

$$\Delta_{f-m} = -0.42(0.14)***$$



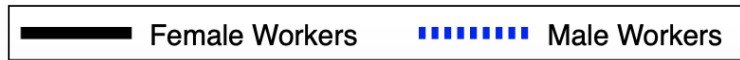
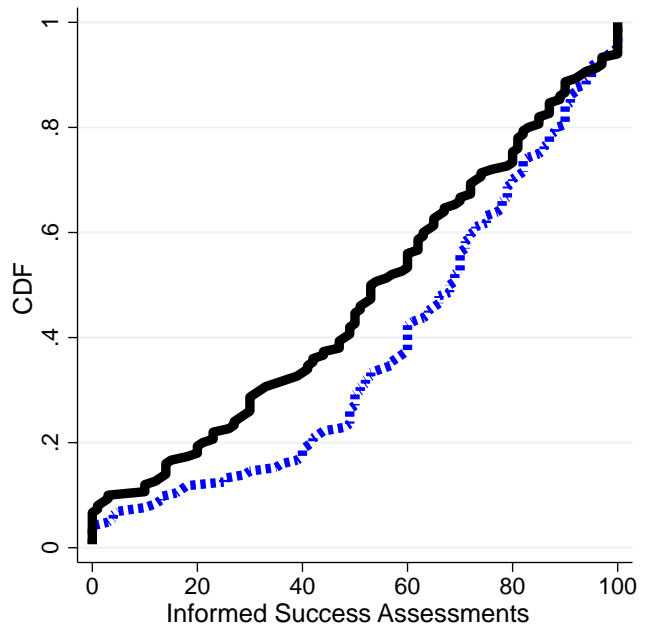
(c) Willingness to Apply Assessments

$$\Delta_{f-m} = -10.46(3.44)***$$



(d) Success Assessments

$$\Delta_{f-m} = -9.60(3.40)***$$



A.2.6 Additional Analyses Mentioned in Discussion Section 4

Table A.7: *Employer Version*, Wage Regressions

	<i>Performance</i>	<i>Performance-Bucket</i>	<i>Willingness to Apply</i>	<i>Success</i>
	(1)	(2)	(3)	(4)
<i>Assessment</i>	0.21*** (0.03)	4.23*** (0.41)	0.23*** (0.02)	0.22*** (0.02)
<i>Female*Assessment</i>	-0.01 (0.03)	0.06 (0.55)	-0.01 (0.03)	-0.02 (0.03)
<i>Female</i>	-1.30 (1.51)	-2.23 (1.40)	-1.37 (1.24)	-1.21 (1.56)
Constant	23.37*** (1.28)	20.11*** (1.08)	22.66*** (1.03)	23.39*** (1.22)
N	1490	1788	1490	1490

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are clustered by employer. Results are from OLS regressions of the wage received by the participant (25 cents if not hired and the chosen wage if hired). *Assessment* is the assessment provided by the participant in response to the self-assessment question noted in that column. See Table 3 for descriptions of each self-assessment question. *Female* is an indicator for a female employer. Data are from the hiring decisions in the *Employer* version.

Table A.8: *All Versions*: Consistency of Self-Promotion

Sample: Confidence – Performance:	Fraction of participants who change self-assessments					
	All Participants		Female Participants		Male Participants	
	$\neq 0$ (1)	$= 0$ (2)	$\neq 0$ (3)	$= 0$ (4)	$\neq 0$ (5)	$= 0$ (6)
<i>Performance Assessments</i>	0.88	0.76	0.87	0.72	0.89	0.80
<i>Performance-Bucket Assessments</i>	0.48	0.36	0.49	0.43	0.47	0.32
<i>Performance-Bucket Assessments</i>	0.82	0.70	0.79	0.67	0.84	0.72
<i>Success Assessments</i>	0.83	0.74	0.78	0.70	0.87	0.76
N	1360	140	637	61	723	79

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. These results show the fraction of participants who changed their answers to each self-assessment question after learning their performance information. See Table 3 for descriptions of each self-assessment question. Column 1 restricts to the set of participants who learn that their objective beliefs were initially wrong after learning their performance information. Column 2 restricts to the set of participants who learn that their objective beliefs were initially correct (i.e., their performance equals their believed performance). Data are from all versions.

Table A.9: *All Versions*, Consistency Regressions

	(1)	(2)
DV = Informed Performance Assessments		
<i>Female</i>	2.80 (2.18)	5.48*** (1.95)
<i>Uninformed Performance Assessment</i>	0.79*** (0.02)	0.77*** (0.02)
<i>Female*Uninformed Performance Assessment</i>	-0.04 (0.04)	-0.10*** (0.03)
Constant	10.79*** (1.63)	2.01 (2.25)
DV = Informed Performance-Bucket Assessments		
<i>Female</i>	0.19 (0.13)	0.38*** (0.13)
<i>Uninformed Performance Bucket Assessments</i>	0.78*** (0.02)	0.78*** (0.02)
<i>Female*Uninformed Performance Bucket Assessments</i>	-0.06 (0.04)	-0.12*** (0.04)
Constant	0.82*** (0.10)	0.71** (0.31)
DV = Informed Willingness to Apply Assessments		
<i>Female</i>	-0.97 (2.00)	0.11 (1.88)
<i>Uninformed Willingness to Apply Assessments</i>	0.82*** (0.02)	0.80*** (0.02)
<i>Female*Uninformed Willingness to Apply Assessments</i>	0.01 (0.03)	-0.02 (0.03)
Constant	10.59*** (1.65)	21.98*** (3.39)
DV = Informed Success Assessments		
<i>Female</i>	-3.04 (2.02)	-1.48 (1.85)
<i>Uninformed Success Assessments</i>	0.81*** (0.02)	0.80*** (0.02)
<i>Female*Uninformed Success Assessments</i>	0.03 (0.03)	-0.01 (0.03)
Constant	11.08*** (1.67)	-3.58 (2.49)
N	1500	1500
Performance FEs	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are clustered at the participant level. Results include the same dependent variable (DV) noted in Table 3. Also, *Female* is an indicator that the participant is female. Performance FEs are dummies for each possible performance on the ASVAB. Data are from all the versions.

Figure A.15: *All Versions*, Deservingness Measure

$$\Delta_{f-m} = 3.13(1.79)^*$$

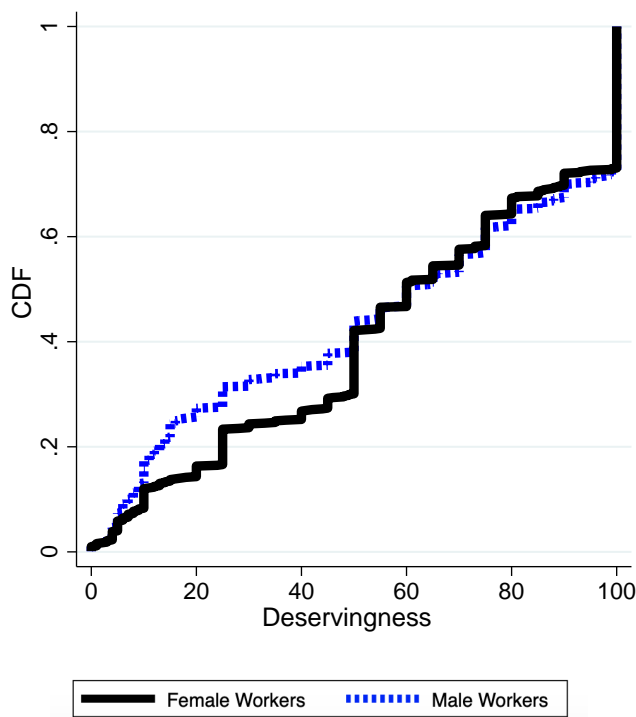


Table A.10: *All Versions*, Deservingness Regressions

	(1)	(2)
<i>Female</i>	3.13* (1.79)	-1.08 (1.53)
Constant	57.88*** (1.29)	
Performance FEs	No	Yes
N	1500	1500

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the deservingness measure in Part 4 and thus in response to the following question: “Out of a maximum amount of 100 cents, what amount of bonus payment, in cents, do you think you deserve for your performance on the test you took in part 1.” *Female* is an indicator for the a female participant. Performance FEs are dummies for each possible performance out of the 20 questions on the ASVAB. Data are from all versions.

Table A.11: *All Versions*, Self-Promotion Regressions with Other Demographic Controls

Sample:	Uninformed Self-Promotion		Informed Self-Promotion	
	(1)	(2)	(3)	(4)
DV = Performance Assessments				
<i>Female</i>	-13.40*** (1.36)	-11.92*** (1.30)	-9.42*** (1.31)	-8.14*** (1.26)
<i>Age</i>		-0.35*** (0.06)		-0.38*** (0.06)
<i>Education (1-9)</i>		4.50*** (0.47)		3.62*** (0.47)
<i>Republican Leaning (0-100)</i>		0.13*** (0.02)		0.13*** (0.02)
DV = Performance-Bucket Assessments				
<i>Female</i>	-0.64*** (0.06)	-0.57*** (0.06)	-0.47*** (0.06)	-0.42*** (0.06)
<i>Age</i>		-0.02*** (0.00)		-0.02*** (0.00)
<i>Education (1-9)</i>		0.23*** (0.02)		0.17*** (0.02)
<i>Republican Leaning (0-100)</i>		0.01*** (0.00)		0.01*** (0.00)
DV = Willingness to Apply Assessments				
<i>Female</i>	-16.94*** (1.53)	-15.25*** (1.49)	-14.10*** (1.51)	-12.65*** (1.47)
<i>Age</i>		-0.39*** (0.07)		-0.36*** (0.07)
<i>Education (1-9)</i>		4.80*** (0.53)		4.32*** (0.54)
<i>Republican Leaning (0-100)</i>		0.12*** (0.03)		0.13*** (0.03)
DV = Success Assessments				
<i>Female</i>	-15.29*** (1.56)	-13.47*** (1.51)	-13.97*** (1.51)	-12.29*** (1.47)
<i>Age</i>		-0.33*** (0.07)		-0.34*** (0.07)
<i>Education (1-9)</i>		5.39*** (0.55)		4.78*** (0.54)
<i>Republican Leaning (0-100)</i>		0.11*** (0.03)		0.11*** (0.03)
N	1500	1500	1500	1500
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results include the same dependent variables (DV) noted in Table 3. *Female* is an indicator for female participants. *Age* equals each participant's age. *Education (1-9)* is a number from 1 to 9 that corresponds with lower to higher levels of education. *Republican Leaning (0-100)* is a number from 0 to 100 that indicates the extent to which a participant indicated feeling favorably about the Republican party. Performance FEs are dummies for each possible performance on the ASVAB. Data in columns in (1) and (2) are from uninformed self-assessment questions elicited in part 2 of all the versions and data in columns (3) and (4) are from uninformed self-assessment questions elicited in part 3 of all the versions.

Table A.12: *All Versions*, Self-Promotion Regressions with Confidence Controls

Sample:	Uninformed Self-Promotion			Informed Self-Promotion		
	(1)	(2)	(3)	(4)	(5)	(6)
DV = Performance Assessments						
<i>Female</i>	-13.40*** (1.36)	-4.57*** (1.04)	-4.10*** (1.05)	-9.42*** (1.31)	-3.98*** (1.24)	-3.74*** (1.24)
<i>Confidence</i>		4.15*** (0.13)			2.56*** (0.15)	
DV = Performance-Bucket Assessments						
<i>Female</i>	-0.64*** (0.06)	-0.24*** (0.05)	-0.21*** (0.05)	-0.47*** (0.06)	-0.22*** (0.06)	-0.22*** (0.06)
<i>Confidence</i>		0.19*** (0.01)			0.12*** (0.01)	
DV = Willingness to Apply Assessments						
<i>Female</i>	-16.94*** (1.53)	-8.71*** (1.35)	-8.32*** (1.37)	-14.10*** (1.51)	-7.93*** (1.43)	-7.79*** (1.44)
<i>Confidence</i>		3.88*** (0.16)			2.91*** (0.16)	
DV = Success Assessments						
<i>Female</i>	-15.29*** (1.56)	-6.70*** (1.33)	-6.13*** (1.34)	-13.97*** (1.51)	-8.01*** (1.43)	-7.62*** (1.44)
<i>Confidence</i>		4.04*** (0.16)			2.80*** (0.17)	
N	1500	1500	1500	1500	1500	1500
Performance FEs	Yes	Yes	Yes	Yes	Yes	Yes
Confidence FEs	No	No	Yes	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results include the same dependent variable (DV) noted in Table 3. Also, *Female* is an indicator that the participant is female *Confidence* equals the number of questions a participant believes he or she correctly answered. Performance FEs are dummies for each possible performance on the ASVAB. Confidence FEs are dummies for each possible response to the question about how many ASVAB questions the participant correctly answered. Data in columns in (1) through (3) are from uninformed self-assessment questions elicited in part 2 of all the versions and data in columns (4) through (6) are from uninformed self-assessment questions elicited in part 3 of all the versions.

A.3 The Free-Response Evaluators Versions

In February 2019, we recruited 400 workers on MTurk to complete the *Free-Response Evaluators* versions of our study using the same criteria as in the main study versions (see footnote 14). We collected data from 399 workers.³⁴ Each participant received a guaranteed \$1.50 completion fee for the 15-minute study. In addition, one of their decisions, out of the 21 decisions in the study, was selected to determine a possible bonus payment for them, and if relevant, for an associated “worker.”³⁵ After evaluators completed all decisions of the study, they took a short follow-up survey that collected demographic information.

The evaluators were randomly assigned either to make 21 hiring decisions (n=198) or to make 21 sets of predictions (n=201). Before making each decision or set of predictions, the evaluator was provided with the text entered by a participant to the free-response question: “Please describe how well you think you performed on the test that you took in part 1 and why.” The free response either came from part 2 or part 3. Evaluators were randomly assigned these 21 free responses from the set of eligible free responses written by the participants from the three main versions of the study.³⁶

Evaluators assigned to make hiring decisions were asked whether they would like to hire the participant who provided that free response and, if so, how much to pay them. The payoffs for the evaluator and associated participant are the same as described in the *Employers* version.³⁷ While similar to the *Employers* version, there are many more possible free responses than answers to the quantitative self-promotion questions, which means our analysis on hiring decisions is underpowered relative to the *Employers* version, since we only have at most a few evaluators reacting to each free response.

Evaluators assigned to make predictions were instead asked to predict whether the participant who wrote the free response was male or female and how many questions, out of 20, that participant answered correctly on the ASVAB. The payoffs for evaluators are determined as follows. One of the two predictions from one of the 21 sets was randomly selected. If the prediction was correct, the evaluator received a bonus payment of 50 cents.³⁸

Relative to the *Employers* version, there are three important differences when considering the results

³⁴One worker was excluded from participation for having previously participated in the study but was counted as being recruited.

³⁵Each participant who completed the *Public* or *Ambiguous* versions of our study was matched with an employer from the *Employer* version of our study and received corresponding payoffs from their employers’ hiring decisions. By contrast, in the *Free-Response Evaluators* versions, only select workers from the *Public* and *Ambiguous* versions were matched with an evaluator and received corresponding payoffs, rather than everyone. Since we also wanted evaluators to provide data on the free responses from the *Private* version, evaluators were (accurately) told that one of their decisions would be selected to count but *not* that one of their decisions would be randomly selected to count (as this would have required putting 0% weight on free responses from the *Private* version in the randomization).

³⁶Each of our 1500 participants provided a response in part 2 and a response in part 3. Not all of the 3000 possible free responses were evaluated, however. First, the *Free-Response Evaluators* versions were run before the *Private (Replication)* and *Private (Information about Others)* versions were run, so free responses from those study versions did not yet exist. We consequently consider the 1800 free responses from the *Public*, *Private*, and *Ambiguous* versions. Second, a research assistant — blinded to sex and study version — deemed 130 of the 1800 potentially eligible free responses “ineligible” due to the answer not relating to the question asked or due to severe grammar and/or spelling issues that made an answer incomprehensible. Consequently, the evaluators were each randomly shown 21 free-responses from the set of 1670 eligible free responses.

³⁷As explained in footnote 35, however, free responses from the *Private* version were never selected for payment.

³⁸Unlike hiring decisions, the randomly selected prediction can come from a participant from any of our three main study versions.

in the *Free-Response Evaluators* versions. First, since there is no objective way to rank free-response answers, we cannot examine how hiring decisions or predictions vary as the responses improve (as we did when examining the impact of a one unit increase on the 0–100 scales in the *Employers* version). Second, while evaluators are not informed of the gender of the associated worker, they may be able to infer gender — to some degree — given how the free responses are written. Below, we test this hypothesis using data from the predictions. Third, as noted above, given the large number of possible free responses, we are underpowered to consider the effect of specific free responses.

For these reasons, we favor the analysis of our quantitative self-assessment questions presented in the main text to examine self-promotion. Here, however, we investigate the hiring decisions and predictions from the *Free-Response Evaluators* versions, to present several interesting (but inherently secondary) results. Given our power issues, we combine free responses from the three main study versions (i.e., the *Public*, *Private*, and *Ambiguous* versions).³⁹ In cases where multiple evaluators faced a decision about the same free response, we use the average decision provided by the evaluators (e.g., if a free response is predicted to be written by a female participant by one evaluator but a male participant by another evaluator, that participant is recorded as being predicted to be female with a 0.50 probability).

Appendix Table A.13 presents results from regressions testing whether the gender of the free response author affects the hiring decisions and predictions of evaluators. The structure of the table mirrors the structure of tables in the main text (e.g., columns (1) and (3) have no controls, and (2) and (4) have dummies for each level of performance). Panel A shows that evaluators predict that free responses provided by female participants come from lower-performing workers. This evidence is relatively consistent with our findings from the quantitative self-assessment questions since women appear to provide less favorable subjective descriptions of their performance. Panel B shows that, although these evaluators are not informed of the gender of the participant associated with the free response, evaluators can infer gender — to some degree — when viewing the responses. Evaluators are significantly more likely to predict that a response was written by a female participant when it was indeed written by a female participant. Panel C shows that the relationship between the gender of the worker and evaluators’ hiring decisions is inconclusive. Based on the free response answers, evaluators pay directionally, but not significantly, less to female worker. We note that there are several possible explanations for this last finding. For instance, a preference to hire workers believed to be higher performing (who are more likely to be male, per our first finding) may counteract a preference to hire workers believed to be female (who are more likely to be female, per our second finding). In other words, hiring decisions based off of the free responses may conflate performance beliefs and other preferences. As mentioned in footnote 18 in the main text of the paper, this difficulty with the free-response data contributes to our decision to focus our self-promotion analysis on the self-assessment questions we explore in the main text of the paper.

³⁹Note that an additional limitation is that approximately 10% of free responses were not evaluated either because they were deemed ineligible, as explained in footnote 36, or because they were never randomly selected to be shown to an evaluator. The results are qualitatively similar when restricting to the data from each of these three version, with one possible exception: the gender difference in the wage data is largely statistically insignificant but is sometimes directionally negative and sometimes directionally positive, depending on the study version.

Table A.13: Free Response Regressions

Sample:	Uninformed Free Responses		Informed Free Responses	
	(1)	(2)	(3)	(4)
Panel A: DV = Predicted Performance				
<i>Female</i>	-0.82*** (0.23)	-0.67*** (0.22)	-0.51** (0.24)	-0.35 (0.23)
Constant	12.16*** (0.17)		12.36*** (0.18)	
N	749	749	773	773
Panel B: DV = Predicted Probability Female				
<i>Female</i>	0.08*** (0.03)	0.08*** (0.03)	0.09*** (0.03)	0.09*** (0.03)
Constant	0.37*** (0.02)		0.34*** (0.02)	
N	749	749	773	773
Panel C: DV = Wage				
<i>Female</i>	-1.28 (0.82)	-1.44* (0.81)	-0.96 (0.99)	-0.66 (1.04)
Constant	33.58*** (0.60)		35.45*** (0.76)	
N	743	743	755	755
Performance FEs	No	Yes	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the noted dependent variable (DV). *Predicted Performance* equals the number of questions that an evaluator predicts a participant correctly answered out of the 20 ASVAB questions. *Predicted Probability Female* equals the probability with which an evaluator predicted a participant to have been female. *Wage* equals the wage given to the participant by an evaluator. *Female* is an indicator that the participant is a female. Performance FEs are dummies for each possible performance on the ASVAB. Data in columns in (1) and (2) are from uninformed free responses elicited in part 2 and data in columns (3) through (4) are from informed free responses elicited in part 3 of all three study versions run in our first wave of data collection: the *Public* version, the *Private* version, and the *Ambiguous* version.

B Experimental instructions

B.1 Instructions for *Public* version

Prior to participating in the study, participants must correctly answer a captcha and consent to participate in the study. At the end of the study, participants must complete a short follow-up survey to gather demographic information.

The study begins by informing each participant of the \$2 study completion fee and of the opportunity to earn additional payment for themselves. Figure B.1 shows how this payment information is explained along with the understanding question that the participant must answer correctly to proceed.

Figure B.1: Payment Information

Overview: This study will consist of 4 parts and a short follow-up survey. Part 1 is the longest, so you should expect to spend more time completing part 1 and less time completing each of the subsequent parts 2 - 4. Following certain instructions, you will be asked understanding questions. You must answer these understanding questions correctly in order to proceed to complete the study.

Your Payment: For completing this study, you are guaranteed to receive \$2 within 24 hours. In addition, one part out of the 4 parts will be randomly selected as the part-that-counts. Any amount you earn in the part-that-counts will be distributed to you as a bonus payment.

Understanding Question: Which of the following statements is true?

For completing this study, I will receive \$2 within 24 hours, but I do NOT have a chance of receiving any additional bonus payment.

For completing this study, I will receive \$2 within 24 hours, and I will also receive the amount I earn in the part-that-counts as additional bonus payment.

For completing this study, I will receive \$2 within 24 hours, and I will also receive the total amount I earn across all parts as additional bonus payment.

The instructions for Part 1 are displayed in Figures B.2 and an example of a corresponding ASVAB question is displayed in Figure B.3 (note that the timer in that screenshot indicates the participant has 23 seconds left to answer the question although the timer starts that count down at 30 seconds).

Figure B.2: Instructions for Part 1

Overview: This study will consist of 4 parts and a short follow-up survey. Part 1 is the longest, so you should expect to spend more time completing part 1 and less time completing each of the subsequent parts 2 - 4. Following certain instructions, you will be asked understanding questions. You must answer these understanding questions correctly in order to proceed to complete the study.

Your Payment: For completing this study, you are guaranteed to receive \$2 within 24 hours. In addition, one part out of the 4 parts will be randomly selected as the part-that-counts. Any amount you earn in the part-that-counts will be distributed to you as a bonus payment.

Understanding Question: Which of the following statements is true?

For completing this study, I will receive \$2 within 24 hours, but I do NOT have a chance of receiving any additional bonus payment.

For completing this study, I will receive \$2 within 24 hours, and I will also receive the amount I earn in the part-that-counts as additional bonus payment.

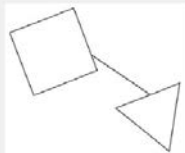
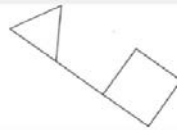
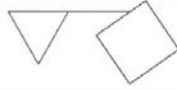
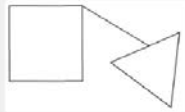
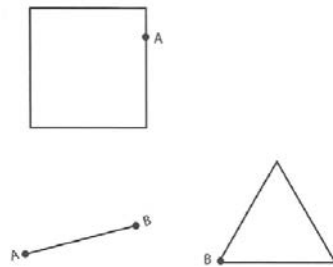
For completing this study, I will receive \$2 within 24 hours, and I will also receive the total amount I earn across all parts as additional bonus payment.

Figure B.3: Part 1: Example of ASVAB question

23

Question 1 out of 20:

ASSEMBLING OBJECTS: Given the following set of objects, please determine which answer choice shows how the objects will look once the parts are put together.



After completing the ASVAB questions in Part 1 but before proceeding to part 2, participants are asked to about their confidence as shown in Figure B.4.

Figure B.4: Confidence Question

Congrats! You have now completed part 1 out of 4.

Before pushing the arrow to proceed onto the next part in this study, please answer the following question.

Out of the 20 questions on the test you took in part 1, how many questions do you think you answered correctly?

Participants then receive instructions for part 2 (see Figure B.5), must correctly answer understanding questions about those instructions (see Figure B.6), and then are asked the self-assessment questions about their performance (see Figure B.7).

Figure B.5: Part 2 Instructions

Instructions for Part 2 out of 4:

In part 2, you will be asked several questions -- on the next page -- related to your performance on the test you completed in part 1.

One of your answers to these questions will be shown to "your part 2 employer," who will be another MTurk worker who completes a different version of this study. Your part 2 employer can decide whether to hire you and, if so, how much to pay you.

Prior to deciding whether to hire you and, if so, how much to pay you, your part 2 employer will NOT be informed of how many questions you answered correctly on the test in part 1.

If this part is randomly selected as the part-that-counts, the additional payment given to your part 2 employer and to you will be determined as follows:

- If your part 2 employer chooses NOT to hire you, your additional payment will equal 25 cents and your part 2 employer's additional payment will equal 100 cents.

- If your part 2 employer chooses to hire you, your additional payment will equal how much they choose to pay you, and your part 2 employer's additional payment will equal (i) 100 cents minus how much they choose to pay you, plus (ii) 5 cents times the number of questions you answered correctly on the test in part 1. Your part 2 employer can choose to pay you any amount between 25 cents and 100 cents.

Figure B.6: Part 2 Understanding Questions

Understanding Question: If this part is randomly selected as the part-that-counts, your additional payment...

will equal 25 cents for sure.

will equal 5 cents times the number of questions you answered correctly on the test in part 1.

will equal 25 cents if you are not hired and the amount your part 2 employer chooses to pay you if you are hired.

Understanding Question: When deciding how much to pay you, your part 2 employer will only know...

how many questions you answered correctly on the test you took in part 1.

how you answer one of the questions -- on the next page -- about your performance on the test you took in part 1.

how you answer all of the questions -- on the next page -- about your performance on the test you took in part 1.

Figure B.7: Part 2 Self-Assessment Questions about Performance

Now, please answer the five questions below to complete part 2. Note that, although the final three questions appear in the same block, they are three separate questions.

Please describe how well you think you performed on the test that you took in part 1 and why.

Please indicate how well you think you performed on the test you took in part 1.

Terrible	Very Poor	Neutral	Good	Very Good	Exceptional
----------	-----------	---------	------	-----------	-------------

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with each of the following statements:

Entirely	Strongly		Somewhat	Neither	Somewhat		Strongly	Entirely		
Disagree	Disagree	Disagree	Disagree	Disagree	Agree	Agree	Agree	Agree		
0	10	20	30	40	50	60	70	80	90	100

I performed well on the test I took in part 1.

I would apply for a job that required me to perform well on the test I took in part 1.

I would succeed in a job that required me to perform well on the test I took in part 1.

After completing part 2, participants are provided with perfect information on their absolute and relative performance as well as required to correctly report back their absolute performance as shown in Figure B.8.

Figure B.8: Absolute and Relative Performance Information

Congrats! You have now completed part 2 out of 4.

Before pushing the arrow to proceed onto the next part in this study, please read the information below on how well you performed on the test in part 1 and answer the corresponding understanding question.

You answered **0 questions correctly out of the 20 questions**. As a result, compared to 100 other participants who were asked the exact same questions as you were, you answered more questions correctly than 0 of them and fewer questions correctly than 100 of them.

Understanding Question: Out of the 20 questions on the test you took in part 1, how many questions did you answer correctly?

In part 3, participants are then provided with the same instructions (see Figure B.9), understanding questions (see Figure B.10), and self-assessment questions about their performance (see Figure B.11) as they were in part 2.

Figure B.9: Part 3 Instructions

Instructions for Part 3 out of 4:

In part 3, you will be asked several questions -- on the next page -- related to your performance on the test you completed in part 1.

One of your answers to these questions will be shown to "your part 3 employer," who will be another MTurk worker who completes a different version of this study. Your part 3 employer can decide whether to hire you and, if so, how much to pay you.

Prior to deciding whether to hire you and, if so, how much to pay you, your part 3 employer will NOT be informed of how many questions you answered correctly on the test in part 1 (even though you were informed of this information on the previous page).

If this part is randomly selected as the part-that-counts, the additional payment given to your part 3 employer and to you will be determined as follows:

- If your part 3 employer chooses NOT to hire you, your additional payment will equal 25 cents and your part 3 employer's additional payment will equal 100 cents.

- If your part 3 employer chooses to hire you, your additional payment will equal how much they choose to pay you, and your part 3 employer's additional payment will equal (i) 100 cents minus how much they choose to pay you, plus (ii) 5 cents times the number of questions you answered correctly on the test in part 1. Your part 3 employer can choose to pay you any amount between 25 cents and 100 cents.

Figure B.10: Part 3 Understanding Questions

Understanding Question: If this part is randomly selected as the part-that-counts, your additional payment...

will equal 25 cents for sure.

will equal 5 cents times the number of questions you answered correctly on the test in part 1.

will equal 25 cents if you are not hired and the amount your part 3 employer chooses to pay you if you are hired.

Understanding Question: When deciding how much to pay you, your part 3 employer will only know...

how many questions you answered correctly on the test you took in part 1.

how you answer one of the questions -- on the next page -- about your performance on the test you took in part 1.

how you answer all of the questions -- on the next page -- about your performance on the test you took in part 1.

Figure B.11: Part 3 Self-Assessment Questions about Performance

Now, please answer the five questions below to complete part 3. Note that, although the final three questions appear in the same block, they are three separate questions.

Please describe how well you think you performed on the test that you took in part 1 and why.

Please indicate how well you think you performed on the test you took in part 1.

Terrible	Very Poor	Neutral	Good	Very Good	Exceptional
----------	-----------	---------	------	-----------	-------------

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with the following statement:

Entirely	Strongly		Somewhat	Neither	Somewhat		Strongly	Entirely		
Disagree	Disagree	Disagree	Disagree	Disagree	Agree	Agree	Agree	Agree		
0	10	20	30	40	50	60	70	80	90	100

I performed well on the test I took in part 1.

I would apply for a job that required me to perform well on the test I took in part 1.

I would succeed in a job that required me to perform well on the test I took in part 1.

Finally, participants receive instructions about and are asked to answer the deservingness question in Part 4 (see Figure B.12)

Figure B.12: Part 4 Instructions and Deservingness Question

Instructions for Part 4 out of 4:

To complete part 4, please answer the one question below. If this part is randomly selected as the part-that-counts, your additional payment will equal whatever amount you answer in this question.

Out of a maximum amount of 100 cents, what amount of bonus payment, in cents, do you think you deserve for your performance on the test you took in part 1?

B.2 Instructions for the *Private* version

The *Private* version of the study proceeds in the same manner as the *Public* version of the study except for the instructions about part 2 and part 3, as participants are simply informed that they will receive 25 cents regardless of how they answer the self-assessment questions about their performance. See Figure B.13 for these instructions and the corresponding understanding question.

Figure B.13: The *Private* version: Part 2 Instructions and Understanding Question

Instructions for Part 2 out of 4:

In part 2, you will be asked several questions -- on the next page -- related to your performance on the test you completed in part 1.

If this part is randomly selected as the part-that-counts, your additional payment will equal 25 cents regardless of how you answer these questions. Thus, we ask that you please answer these questions carefully and honestly.

Understanding Question: If this part is randomly selected as the part-that-counts, your additional payment...

will equal 25 cents for sure.

will equal 5 cents times the number of questions you answered correctly on the test in part 1.

will depend on how you answer the questions -- on the next page -- about your performance on the test you took in part 1.

B.3 Instructions for the *Private (Replication)* version

The *Private (Replication)* version of the study proceeds in the same manner as the *Private* version of the study except for a slight formatting change in how three of the self-assessment questions appear to allow for room to introduce the additional information in the *Private (Information about Others)* version. See Figure [B.14](#) for the corresponding screenshot of the part 3 self-assessment questions (and note that this is identical to how they appear in part 2).

Figure B.14: The *Private (Replication)* version: Part 3 Self-Assessment Questions about Performance

Please describe how well you think you performed on the test that you took in part 1 and why.

Please indicate how well you think you performed on the test you took in part 1.

Terrible

Very Poor

Neutral

Good

Very Good

Exceptional

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with the following statement: "I performed well on the test I took in part 1."

Entirely	Strongly			Somewhat	Neither	Somewhat			Strongly	Entirely
Disagree	Disagree	Disagree		Disagree	Disagree	Agree	Agree		Agree	Agree
0	10	20	30	40	50	60	70	80	90	100

I performed well on the test I took in part 1.

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with the following statement: "I would apply for a job that required me to perform well on the test I took in part 1."

Entirely	Strongly			Somewhat	Neither	Somewhat			Strongly	Entirely
Disagree	Disagree	Disagree		Disagree	Disagree	Agree	Agree		Agree	Agree
0	10	20	30	40	50	60	70	80	90	100

I would apply for a job that required me to perform well on the test I took in part 1.

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with the following statement: "I would succeed in a job that required me to perform well on the test I took in part 1."

Entirely	Strongly			Somewhat	Neither	Somewhat			Strongly	Entirely
Disagree	Disagree	Disagree		Disagree	Disagree	Agree	Agree		Agree	Agree
0	10	20	30	40	50	60	70	80	90	100

I would succeed in a job that required me to perform well on the test I took in part 1.

B.4 Instructions for the *Private (Information about Others)* version

The *Private (Information about Others)* version of the study proceeds in the same manner as the *Private* version of the study except that, in part 3, additional information is provided on the average answer to each of the part 3 self-assessment questions from prior participants with the same score as the participant currently completing the study. See Figure [B.15](#) for the corresponding screenshot of the part 3 self-assessment questions.

Figure B.15: The *Private (Information about Others)* version: Part 3 Self-Assessment Questions about Performance for a Participant who Correctly Answered 10 out of 20 ASVAB Questions

Please describe how well you think you performed on the test that you took in part 1 and why.

Please indicate how well you think you performed on the test you took in part 1. Also note that, among participants in a prior study who scored the same as you on the test, the average answer to this question was: Neutral.

Terrible	Very Poor	Neutral	Good	Very Good	Exceptional
----------	-----------	---------	------	-----------	-------------

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with the following statement: "I performed well on the test I took in part 1."

Also note that, among participants in a prior study who scored the same as you on the test, the average answer to this question was: 38.

Entirely	Strongly		Somewhat	Neither	Somewhat		Strongly	Entirely		
Disagree	Disagree	Disagree	Disagree	Disagree	Agree	Agree	Agree	Agree		
0	10	20	30	40	50	60	70	80	90	100

I performed well on the test I took in part 1.

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with the following statement: "I would apply for a job that required me to perform well on the test I took in part 1."

Also note that, among participants in a prior study who scored the same as you on the test, the average answer to this question was: 37.

Entirely	Strongly		Somewhat	Neither	Somewhat		Strongly	Entirely		
Disagree	Disagree	Disagree	Disagree	Disagree	Agree	Agree	Agree	Agree		
0	10	20	30	40	50	60	70	80	90	100

I would apply for a job that required me to perform well on the test I took in part 1.

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with the following statement: "I would succeed in a job that required me to perform well on the test I took in part 1."

Also note that, among participants in a prior study who scored the same as you on the test, the average answer to this question was: 40.

Entirely	Strongly		Somewhat	Neither	Somewhat		Strongly	Entirely		
Disagree	Disagree	Disagree	Disagree	Disagree	Agree	Agree	Agree	Agree		
0	10	20	30	40	50	60	70	80	90	100

I would succeed in a job that required me to perform well on the test I took in part 1.

B.5 Instructions for the *Ambiguous* version

The *Ambiguous* version of the study proceeds in the same manner as the *Public* version of the study except for the instructions about part 2 and part 3, as participants are informed that there is some chance that their employer will learn their actual performance. See Figures B.16 and B.17 for these instructions and the corresponding understanding questions, respectively.

Figure B.16: The *Ambiguous* version: Part 2 Instructions

Instructions for Part 2 out of 4:

In part 2, you will be asked several questions -- on the next page -- related to your performance on the test you completed in part 1.

There is some chance that one of your answers to these questions will be shown to "your part 2 employer," who will be another MTurk worker who completes a different version of this study. Your part 2 employer can decide whether to hire you and, if so, how much to pay you.

Prior to deciding whether to hire you and, if so, how much to pay you, there is also some chance that your part 2 employer will be informed of how many questions you answered correctly on the test in part 1.

However, while your part 2 employer may learn one of your answers to the questions -- on the next page -- related to your performance on the test in part 1 and/or how many questions you answered correctly on the test in part 1, it is also possible that your part 2 employer will not learn any information related to your performance prior to deciding whether to hire you and, if so, how much to pay you.

If this part is randomly selected as the part-that-counts, the additional payment given to your part 2 employer and to you will be determined as follows:

- If your part 2 employer chooses NOT to hire you, your additional payment will equal 25 cents and your part 2 employer's additional payment will equal 100 cents.

- If your part 2 employer chooses to hire you, your additional payment will equal how much they choose to pay you, and your part 2 employer's additional payment will equal (i) 100 cents minus how much they choose to pay you, plus (ii) 5 cents times the number of questions you answered correctly on the test in part 1. Your part 2 employer can choose to pay you any amount between 25 cents and 100 cents.

Figure B.17: The *Ambiguous* version: Part 2 Self-Assessment Questions about Performance

Understanding Question: If this part is randomly selected as the part-that-counts, your additional payment...

will equal 25 cents for sure.

will equal 5 cents times the number of questions you answered correctly on the test in part 1.

will equal 25 cents if you are not hired and the amount your part 2 employer chooses to pay you if you are hired.

Understanding Question: When deciding how much to pay you, your part 2 employer will...

definitely know how many questions you answered correctly on the test you took in part 1.

definitely know how you answer all of the questions -- on the next page -- about your performance on the test you took in part 1.

will know nothing about your performance on the test in part 1, or instead will know one of your answers to the questions -- on the next page -- related to your performance on the test in part 1 and/or how many questions you answered correctly on the test in part 1.

B.6 Instructions for *Employer* version

Prior to participating in the study, participants must correctly answer a captcha and consent to participate in the study. At the end of the study, participants must complete a short follow-up survey to gather demographic information.

The study begins by informing each participant of the \$1.50 study completion fee and of the opportunity to earn additional payment for themselves. Figure B.18 shows how this payment information is explained. Figure B.19 shows the understanding questions that the participant must answer correctly to proceed.

Figure B.18: Payment Information

Overview:

This study will consist of 21 decisions and a short follow-up survey. For completing this study, you are guaranteed to receive \$1.50 within 24 hours. In addition, any additional payment you earn will be distributed to you as a bonus payment.

The Workers:

In a prior study, MTurk workers completed a test. On the test, they were asked to answer up to 20 questions from the Armed Services Vocational Aptitude Battery (ASVAB). Each question tested their aptitude in one of the following five categories: General Science, Arithmetic Reasoning, Math Knowledge, Mechanical Comprehension, and Assembling Objects. In addition to being used by the military to determine which jobs armed service members are qualified for, performance on the ASVAB is often used as a measure of cognitive ability by academic researchers.

Your Decisions:

For each of the 21 decisions, you will be matched with one worker from the prior study. You then must decide whether to hire that worker, and if so, how much to pay that worker.

After you make all of your 21 decisions, two decisions will be selected as a decision-that-counts.

In each decision-that-counts, the additional payment given to the worker in that decision and to you will be determined as follows:

- If you choose NOT to hire the worker, that worker's additional payment will equal 25 cents and your additional payment will equal 100 cents.

- If you choose to hire the worker, that worker's additional payment will equal how much you choose to pay them, and your additional payment will equal (i) 100 cents minus how much you choose to pay them, plus (ii) 5 cents times the number of questions that worker answered correctly on the test. You can choose to pay that worker any amount between 25 cents and 100 cents.

Figure B.19: Understanding Questions of Payment Information

Understanding Question: Which of the following statements is true?

For completing this study, I will receive \$1.50 within 24 hours, but I do NOT have a chance of receiving any additional bonus payment.

For completing this study, I will receive \$1.50 within 24 hours, and I will also receive the amount I earn in two decisions-that-count as additional bonus payment.

For completing this study, I will receive \$1.50 within 24 hours, and I will also receive the total amount I earn across all decisions as additional bonus payment.

Understanding Question: In each decision-that-counts, a worker's additional payment...

will equal 25 cents for sure.

will equal 25 cents if you do not hire that worker and 100 cents if you do hire that worker.

will equal 25 cents if you do not hire that worker and how much you choose to pay that worker if you do hire that worker.

Understanding Question: If you do NOT hire a worker in a decision-that-counts, your additional payment from that decision...

will equal 100 cents for sure.

will equal 100 cents **plus** 5 cents for each question that worker answered correctly on the test.

will equal 100 cents **plus** 5 cents for each question that worker answered correctly on the test **minus** the amount you choose to pay that worker.

Understanding Question: If you hire a worker in a decision-that-counts, your additional payment from that decision...

will equal 100 cents for sure.

will equal 100 cents **plus** 5 cents for each question that worker answered correctly on the test.

will equal 100 cents **plus** 5 cents for each question that worker answered correctly on the test **minus** the amount you choose to pay that worker.

Understanding Question: If you hire a worker in a decision-that-counts, your additional payment from that decision...

will not depend on how many questions that worker answered correctly on the test.

will be lower if that worker answered more questions correctly on the test.

will be higher if that worker answered more questions correctly on the test.

The 21 decisions that employers face involve four blocks. Three blocks relate to the three self-assessment questions that involve the 0 to 100 scale (i.e., the *Performance Assessment*, the *Willingness to Apply Assessment* and the *Succeed Assessment*), and each of these blocks involves five decisions that correspond to five randomly selected answers (i.e., numbers from 0 to 100). One block relates to the self-assessment question involving a six point Likert-scale (i.e., the *Performance-Bucket Assessment*), and this block involves six decisions that correspond to each of the six possible answers to that question. The order of these four blocks is randomized on the participant-level.

The instructions for, and examples of, decisions relating to the *Performance Assessments* are displayed in Figures B.20 and B.21, respectively.

Figure B.20: Instructions for *Performance Assessments* Decisions

Instructions for Decisions 1 - 5

In each decision below, you will learn how the worker in that decision answered a question in which they indicated the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: "I performed well on the test I took."

Recall that, in each decision-that-counts, the additional payment given to the worker in that decision and to you will be determined as follows:

- If you choose NOT to hire the worker, that worker's additional payment will equal 25 cents and your additional payment will equal 100 cents.

- If you choose to hire the worker, that worker's additional payment will equal how much you choose to pay them, and your additional payment will equal (i) 100 cents minus how much you choose to pay them, plus (ii) 5 cents times the number of questions that worker answered correctly on the test. You can choose to pay that worker any amount between 25 cents and 100 cents.

Figure B.21: *Performance Assessments* Decisions

Decision 1 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **6**, indicating **strong disagreement** with the following statement: "I performed well on the test I took." What would you like to do?

Decision 2 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **25**, indicating **disagreement** with the following statement: "I performed well on the test I took." What would you like to do?

Decision 3 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **56**, indicating **neither much disagreement nor agreement** with the following statement: "I performed well on the test I took." What would you like to do?

Decision 4 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **61**, indicating **agreement** with the following statement: "I performed well on the test I took." What would you like to do?

Decision 5 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **93**, indicating **strong agreement** with the following statement: "I performed well on the test I took." What would you like to do?

The instructions for, and examples of, decisions relating to the *Performance-Bucket Assessments* are displayed in Figures B.22 and B.23, respectively.

Figure B.22: Instructions for *Performance-Bucket Assessments* Decisions

Instructions for Decisions 6 - 11

In each decision below, you will learn how the worker in that decision answered a question in which they indicated whether they thought their performance on the test was terrible, very poor, neutral, good, very good, or exceptional.

Recall that, in each decision-that-counts, the additional payment given to the worker in that decision and to you will be determined as follows:

- If you choose NOT to hire the worker, that worker's additional payment will equal 25 cents and your additional payment will equal 100 cents.

- If you choose to hire the worker, that worker's additional payment will equal how much you choose to pay them, and your additional payment will equal (i) 100 cents minus how much you choose to pay them, plus (ii) 5 cents times the number of questions that worker answered correctly on the test. You can choose to pay that worker any amount between 25 cents and 100 cents.

Figure B.23: *Performance-Bucket Assessments* Decisions

Decision 6 out of 21: The worker in this decision indicated that their performance on the test was **terrible**. What would you like to do?

Decision 7 out of 21: The worker in this decision indicated that their performance on the test was **very poor**. What would you like to do?

Decision 8 out of 21: The worker in this decision indicated that their performance on the test was **neutral**. What would you like to do?

Decision 9 out of 21: The worker in this decision indicated that their performance on the test was **good**. What would you like to do?

Decision 10 out of 21: The worker in this decision indicated that their performance on the test was **very good**. What would you like to do?

Decision 11 out of 21: The worker in this decision indicated that their performance on the test was **exceptional**. What would you like to do?

The instructions for, and examples of, decisions relating to the *Willingness to Apply Assessments* are displayed in Figures B.24 and B.25, respectively.

Figure B.24: Instructions for *Willingness to Apply Assessments* Decisions

Instructions for Decisions 12 - 16

In each decision below, you will learn how the worker in that decision answered a question in which they indicated the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: "I would apply for a job that required me to perform well on the test I took."

Recall that, in each decision-that-counts, the additional payment given to the worker in that decision and to you will be determined as follows:

- If you choose NOT to hire the worker, that worker's additional payment will equal 25 cents and your additional payment will equal 100 cents.

- If you choose to hire the worker, that worker's additional payment will equal how much you choose to pay them, and your additional payment will equal (i) 100 cents minus how much you choose to pay them, plus (ii) 5 cents times the number of questions that worker answered correctly on the test. You can choose to pay that worker any amount between 25 cents and 100 cents.

Figure B.25: *Willingness to Apply Assessments* Decisions

Decision 12 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **18, indicating strong disagreement** with the following statement: "I would apply for a job that required me to perform well on the test." What would you like to do?

Decision 13 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **27, indicating disagreement** with the following statement: "I would apply for a job that required me to perform well on the test I took." What would you like to do?

Decision 14 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **46, indicating neither much disagreement nor agreement** with the following statement: "I would apply for a job that required me to perform well on the test I took." What would you like to do?

Decision 15 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **64, indicating agreement** with the following statement: "I would apply for a job that required me to perform well on the test I took." What would you like to do?

Decision 16 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **91, indicating strong agreement** with the following statement: "I would apply for a job that required me to perform well on the test." What would you like to do?

The instructions for, and examples of, decisions relating to the *Success Assessments* are displayed in Figures B.26 and B.27, respectively.

Figure B.26: Instructions for *Success Assessments* Decisions

Instructions for Decisions 17 - 21

In each decision below, you will learn how the worker in that decision answered a question in which they indicated the extent to which they agreed, on a scale from 0 (entirely disagree) to 100 (entirely agree), with the following statement: "I would succeed in a job that required me to perform well on the test I took."

Recall that, in each decision-that-counts, the additional payment given to the worker in that decision and to you will be determined as follows:

- If you choose NOT to hire the worker, that worker's additional payment will equal 25 cents and your additional payment will equal 100 cents.

- If you choose to hire the worker, that worker's additional payment will equal how much you choose to pay them, and your additional payment will equal (i) 100 cents minus how much you choose to pay them, plus (ii) 5 cents times the number of questions that worker answered correctly on the test. You can choose to pay that worker any amount between 25 cents and 100 cents.

Figure B.27: *Success Assessments* Decisions

Decision 17 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **6, indicating strong disagreement** with the following statement: "I would succeed in a job that required me to perform well on the test I took." What would you like to do?

Decision 18 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **33, indicating disagreement** with the following statement: "I would succeed in a job that required me to perform well on the test I took." What would you like to do?

Decision 19 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **44, indicating neither much disagreement nor agreement** with the following statement: "I would succeed in a job that required me to perform well on the test I took." What would you like to do?

Decision 20 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **76, indicating agreement** with the following statement: "I would succeed in a job that required me to perform well on the test I took." What would you like to do?

Decision 21 out of 21: On a scale from 0 (entirely disagree) to 100 (entirely agree), the worker in this decision chose **96, indicating strong agreement** with the following statement: "I would succeed in a job that required me to perform well on the test I took." What would you like to do?