Discussion Paper Series

IZA DP No. 18777

July 2026

Gender Differences in Self-Promotion and Career Advice

Rigissa Megalokonomou
Monash University, IZA@LISER
and CESifo

Juliana Silva-Goncalves
The University of Sydney,
Life Course Centre and IZA@LISER

Roel van Veldhuizen
Lund University

IZA DP No. 18777 July 2026

Gender Differences in Self-Promotion and Career Advice*

Abstract

We study the role of self-promotion and career advice in sustaining gender differences in labor market outcomes. We conduct a pre-registered experiment in which "advisers" advise "workers" to attempt either a more or a less ambitious task. We find that women promote themselves less than men and, as a result, are 12 percentage points less likely to be advised to choose the more ambitious task. This gender gap in advice persists across both quantitative and qualitative self-assessments and is robust to variation in advisers' information sets—including when advisers observe workers' actual performance—but is eliminated and even reversed when advisers are informed of the gender gap in self-promotion.

JEL classification

C91, D91, J16

Keywords

advice, gender, self-promotion, randomized experiment

Corresponding author

Rigissa Megalokonomou
rigissa.megalokonomou@monash.edu

* We thank Katie Coffman, Adeline Delavande, Kirby Nielsen, and participants at seminars at Lyon (GATE), Jönköping, Berlin (BBE), Warwick, and the UTS Business School, as well as participants at the 2025 Australia New Zealand Workshop on Experimental Economics (ANZWEE), the 2024 ESA European Meeting in Helsinki, the 2025 Nordic Conference for Behavioral and Experimental Economics in Stavanger and the 2026 Esade Workshop in Behavioral and Experimental Economics for helpful comments. The study was approved by The University of Sydney Human Research Ethics Committee (Project No. 2022/872). We pre-registered our experimental design, hypotheses, and analysis plan (AEARCTR-0013803). Rigissa Megalokonomou acknowledges financial support from the Monash Department Support and Monash Establishment Grant (No. 1755774). Roel van Veldhuizen gratefully acknowledges financial support from the Jan Wallander och Tom Hedelius foundation.

1 Introduction

Every day, people seek advice before making consequential decisions—whether to go to university, which degree to choose, which career to pursue, and whether to apply for a promotion. The quality of that advice shapes economic trajectories, and if advice is given differently to men and women, it may help explain why large gender gaps in labor market outcomes persist despite decades of convergence in educational attainment (Blau & Kahn, 2017; Goldin, 2014). Women remain underrepresented in high-paying male-dominated fields and occupations such as STEM (Goulas, Griselda, & Megalokonomou, 2024; Oreopoulos & Petronijevic, 2013) and are less likely to reach leadership positions (Bertrand, Goldin, & Katz, 2010; Eagly & Carli, 2003). While prior work has pointed to stereotypes, discrimination, childbearing responsibilities, and differences in beliefs and preferences as explanations (Blau & Kahn, 2017), a recent literature highlights a more specific mechanism: women are less likely to advocate for their own achievements—a phenomenon known as the gender gap in self-promotion (Exley & Kessler, 2022; Exley & Nielsen, 2024). How men and women advocate for themselves, however, does not only affect their own decisions—it also shapes the advice they receive from others, a channel that has received little empirical attention despite its potential to amplify gender gaps in the labor market.

Studying whether and how the gender gap in self-promotion shapes the advice men and women receive is, however, difficult for two reasons. First, advice is typically private and therefore hard to measure directly. Second, even when advice can be observed, any gender difference in advice received may simply reflect gender differences in unobserved characteristics—ability, ambition, or qualifications—rather than bias on the part of the adviser. These challenges are compounded by the fact that advice does not operate in isolation: advisers must rely on the advisee's own account of their performance, and since women are less likely to advocate for their skills and accomplishments than men (Exley & Kessler, 2022), a gender gap in advice may emerge not from adviser bias per se, but from the self-promotion gap shaping adviser judgment—explaining why, despite the scholarly and policy importance of this question, well-identified empirical evidence remains scarce. We overcome both challenges through a pre-registered online experiment on Prolific in which we directly observe the advice given to equally performing men and women, which allows us to isolate exactly this mechanism.

In our experiment, 350 participants (the "workers") complete a STEM task and then evaluate their own performance—their self-promotions. These self-promotions are then sent to 1,750 participants in the role of "advisers," who each observe five workers' self-promotions and recommend whether each worker should attempt either a more or a less ambitious task. Critically, advisers are fully informed about the tasks and incentivized to maximize workers' earnings, and workers are paid according to whichever task they choose—ensuring that advice is informed and consequential. Yet advisers observe only workers' self-promotions and not their actual performance, which allows us to isolate self-promotion as the key channel. We find a substantial gender gap in advice: women are 12 percentage points less likely to be advised to choose the more ambitious task. We show that self-promotion differences are the central driver: women provide less optimistic self-assessments, which lead advisers to steer them toward less ambitious decisions even when men and women perform identically.

Several additional treatments shed further light on the mechanism. In the first two treatments, we vary the subjectivity of advisers' information set—providing either verbal self-assessments only or full knowledge of workers' actual performance—and this leaves the gender gap in advice largely unchanged, which suggests it is not driven by adviser reliance on subjective judgment per se. We also show that the gender gap in advice remains in a replication study with advisers who have professional experience in managerial or teaching roles, as well as in a vignette study, which suggests that the mechanism generalizes beyond the framed laboratory setting. In the third treatment, we inform advisers of the gender gap in self-promotion and find that this nudge eliminates and even reverses the gender gap in advice. This suggests that simple informational interventions can meaningfully reduce gender disparities in career guidance.

This paper contributes to three strands of the literature. First, it contributes to the literature on the drivers of gender differences in the labor market (Blau & Kahn, 2017) by identifying a new channel through which gender gaps persist: the role of advice. Previous work has examined gender stereotypes, differences in earnings expectations, family trade-offs and childbearing responsibilities, and behavioral traits such as overconfidence, competitive-

ness, and risk preferences in shaping gender gaps in career decisions.[1] We also contribute to a recent branch of this literature that studies gender differences in self-promotion (Chang, Saccardo, & Gallus, 2025; Exley & Kessler, 2022; Exley & Nielsen, 2024; Römer & Schröder, 2025). We build on this work and show that the gender gap in self-promotion shapes third-party advice, which in turn shapes individuals' choices and outcomes.

Second, this paper contributes to a broader literature on advice. One branch of this literature uses laboratory experiments to study whether the advice of other participants or experimenters affects behavior in, for example, economic games (Çelen, Kariv, & Schotter, 2010; Chaudhuri, Schotter, & Sopher, 2009; Schotter & Sopher, 2003, 2007); matching markets (Guillen & Hakimov, 2018); and tournament entry (Kessel, Mollerstrom, & Van Veldhuizen, 2021). Another branch relies on exogenous information, mentoring, or role model interventions to proxy for the effect of advice in the field (Falk, Kosse, & Pinger, 2026; Hoxby & Turner, 2015; Porter & Serra, 2020). When it comes to gender, existing work suggests that advice does not close gender gaps (Brandts, Groenert, & Rott, 2015), that male and female advisers behave similarly (Brandts & Rott, 2021), and that career advisers adjust their focus depending on the gender of their advisee (Gallen & Wasserman, 2021)—collectively pointing to gender gaps in advice as a persistent and poorly understood phenomenon.[2] We contribute to this literature by directly observing both the advice given and the information on which it is based—something prior work has not done—and our answer is surprising: the gender gap in advice is not driven by adviser bias, but by the gender gap in self-promotion, and a simple informational nudge is sufficient to eliminate it.

Third, it contributes to the literature on subjective judgment and gender bias (Bohren, Imas, & Rosenberg, 2019; Krawczyk & Smyk, 2016; Mengel, Sauermann, & Zölitz, 2019; Özgümüs, Rau, Trautmann, & König-Kersting, 2020; Sarsons, Gërxhani, Reuben, & Schram, 2021). Prior work provides mixed evidence on whether subjectivity generates gender gaps in evaluations—in some settings gaps are driven by evaluators' personal beliefs or preferences, in

[1] The literature covers gender gaps in career aspirations and expectations (Bohren, Imas, & Rosenberg, 2019; Carlana, 2019); earnings expectations (Blau & Kahn, 2017; Reuben, Wiswall, & Zafar, 2017); family trade-offs and childbearing responsibilities (Angelov, Johansson, & Lindahl, 2016; Bertrand, Goldin, & Katz, 2010; Cortés & Pan, 2023); overconfidence (Bandiera, Parekh, Petrongolo, & Rao, 2022; Bordalo, Coffman, Gennaioli, & Shleifer, 2019; Coffman, Collis, & Kulkarni, 2024); competitiveness (Niederle & Vesterlund, 2007; Niederle, 2016); and risk preferences (Croson & Gneezy, 2009).

[2] Also related are Carlana (2019), Lavy and Megalokonomou (2024), and Wolter and Zöllner (2024), who investigate gender stereotypes among teachers and parents, respectively. While some of the effects they document likely operate through advice, they do not directly observe this.

others they are not. In contrast, we find no evidence that advisers have gender-biased beliefs or preferences per se—regardless of how subjective the information they receive is—which suggests that subjectivity alone is not sufficient to generate gender bias in advice.

Our findings have direct policy implications. The substantial gender gap in self-promotion indicates that women communicate their achievements less favorably than men, and this gap is important because it shapes both the advice women receive and the career decisions they ultimately make. The gender gap in advice documented in this paper is not driven by adviser bias and therefore cannot be addressed by diversity training or changing who gives advice. Instead, our results point to two more targeted levers. On the supply side, interventions that help women communicate their achievements more confidently—such as mentorship programs and role model exposure (Breda, Grenet, Monnet, & Van Effenterre, 2023; Goulas, Gunawardena, Megalokonomou, & Zenou, 2024; Porter & Serra, 2020)—may reduce the self-promotion gap and, in turn, the gender gap in advice. On the demand side, our nudge treatment shows that simply informing advisers of the gender gap in self-promotion eliminates and even reverses the gap in advice. This suggests that low-cost awareness interventions can be highly effective in equalizing career guidance across genders.

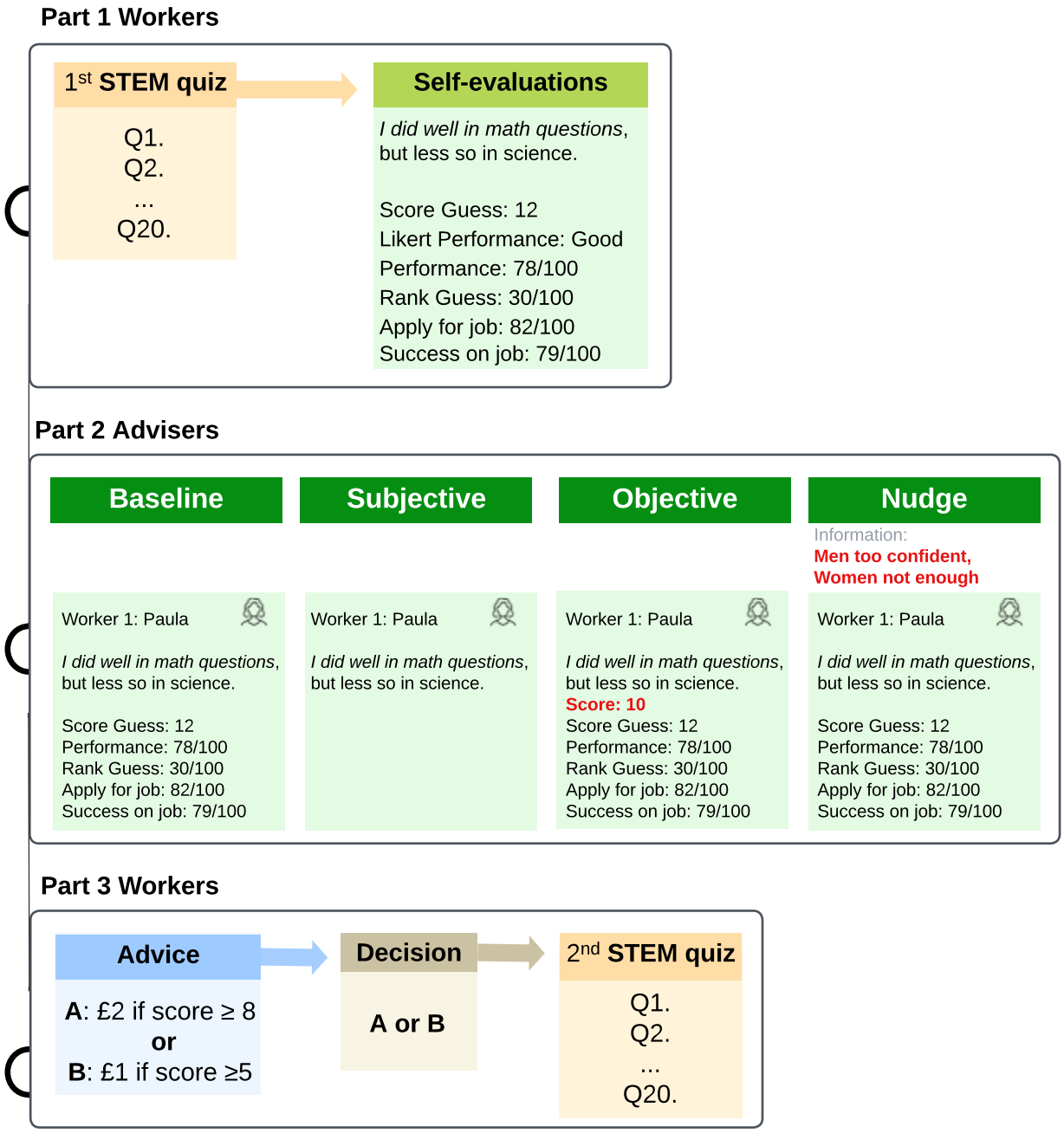
2 Experimental Design

We conducted a pre-registered online experiment in which participants randomly assigned to the role of workers received advice from other participants randomly assigned to the role of advisers. The key feature of our design is that it allows us to separately observe workers' objective performance, their self-promotion, advisers' recommendations, and workers' subsequent decisions. The experiment consisted of three parts. In Part one, workers completed a STEM quiz and subsequently evaluated their own performance. In Part two, advisers observed workers' self-evaluations and recommended whether workers should perform a more or a less ambitious follow-up STEM task. In Part three, workers received this advice and decided which of the two follow-up STEM tasks to undertake.

This section describes the experimental design. Figure 1 provides an overview of the sequence of decisions, while Appendix A.11 reports the complete experimental instructions and materials. We previously conducted a pilot lab experiment based on the same design,

the results of which are reported in Silva-Goncalves and van Veldhuizen (2020). Additional online pilot experiments are described in Appendix A.8.

Figure 1: Overview of the Experimental Design



2.1 Self-Promotion Measurement (Workers)

Part one featured only participants assigned to the role of workers, who were informed that they would participate in two separate sessions. These participants started with a brief

sociodemographic survey, which included questions on gender, age, education, employment, and income (see Appendix A.11 for the full survey). We then instructed participants to pick a name and avatar to represent themselves during the rest of the experiment. Available names and avatars were based on the person's reported gender (identifiably female for women and identifiably male for men).[3] Unbeknownst to the workers, we included this feature to give advisers a way to identify the gender of the workers. A similar approach has been used in experimental research; for example, by Gangadharan, Jain, Maitra, and Vecci (2016) and Mengel (2020).

Workers then completed a 20-question math and science quiz with multiple choice questions. They had 30 seconds for each question and were paid £0.10 per correct answer. Following Exley and Kessler (2022), we selected 20 questions from practice tests for the ASVAB (an aptitude test of the US Armed Services), and conducted several pilots to ensure that the questions were neither too easy nor too difficult. Upon completing the quiz, we asked workers to guess their performance. In order to encourage careful responses, we informed them that they would receive £0.50 for a correct guess and £0.20 if they were within 3 of their true score.[4]

As our next step, we informed workers that in the following week they would once again have to complete some kind of math and science task (Part 3 in Figure 1). We also told workers that they would be able to choose between two versions of the task (job A and job B) and that their payment would depend on their chosen task. However, we did not tell them anything about the task's specific content or incentives. Instead, we informed workers that they would receive advice from another participant—the adviser—with knowledge of the nature of the two jobs and an incentive to give good advice. This feature captures the information asymmetry between advisers and workers that makes advice invaluable in many applications.

Next, we informed workers that they would be asked to answer several self-evaluation questions. Importantly, they were told that one or more of these questions would be sent to the adviser to help them give the best possible advice. The questions we used are summarized

[3]Participants who identified as non-binary in the sociodemographic survey could have chosen a set of identifiably male or female or gender-neutral names and avatars. In our experiment, none of the workers identified as non-binary. The full list of aliases and avatars is presented in Appendix A.11.

[4]In order to reduce cheating, we presented the questions as images in order to prevent, e.g., copy-pasting questions into search engines.

in Table 1. The first five questions are taken from Exley and Kessler (2022), who find that men give more optimistic answers on each of these questions despite being worse at the task. The rank guess provides a measure of relative ability beliefs. Following Exley and Kessler (2022), we interpret these communicated self-evaluations as measures of self-promotion, because workers knew that their responses would be observed by a consequential third party.[5] Part one concluded with two questions to measure general self-confidence and risk aversion (as in Dohmen, Falk, Huffman, Sunde, Schupp, & Wagner, 2011), followed by a reminder to return for the next week's session.

Table 1: Self-Evaluation Questions

Measure	**Question**	**Scale**
Verbal Assessment	Describe how well you think you performed on the quiz.	Open-ended
Likert Performance	Indicate how good your performance was on the quiz	7-point Likert scale ("terrible" to "exceptional")
Performed Well	I performed well on the quiz	0–100 scale ("entirely disagree" to "entirely agree")
Would Apply for STEM Job	I would apply for a job that required me to perform well on the quiz.	0–100 scale ("entirely disagree" to "entirely agree")
Would Succeed in STEM Job	I would succeed in a job that required skills tested on the quiz.	0–100 scale ("entirely disagree" to "entirely agree")
Guessed Rank	How well do you think you performed compared with 100 other participants?	1 (best)–101 (worst) rank
Guessed Score	How many quiz questions do you think you answered correctly?	0-20

[5]There is a conceptual difference between sharing self-evaluations with an adviser incentivized to give good advice (as in our study) and sharing them with an employer incentivized to hire strong performers (as in Exley and Kessler, 2022). As discussed in the results section below, however, the gender gaps we observe are similar in magnitude to those documented when self-evaluations are shared with employers or only with experimenters (Exley & Kessler, 2022).

2.2 Advice Measurement (Advisers)

Part two featured only those participants assigned to the role of advisers. As a first step, we informed advisers that they would give advice to a worker about choosing between two jobs (A and B). Advisers then learned the specifics of the jobs, both of which involved a math and science quiz with twenty 30-second multiple-choice questions. The only difference between the two jobs was the payment scheme. Job A required workers to solve 8 questions to receive a bonus of £2. Job B had a lower threshold (5 questions) and lower bonus (£1). Workers who failed to reach their chosen threshold would receive no bonus. Advisers also learned that workers were unaware of these details, which rendered their advice invaluable.

For each adviser, there was a one in five chance that their advice would be delivered to a worker. When this happened, the adviser would receive an additional payment equal to half of the worker's bonus. We introduced a monetary incentive to give good advice in order to increase advisers' attention to the information provided to them when giving advice. A risk-neutral adviser should advise any worker with at least a 62.5% chance of making the higher threshold to pick the more ambitious job A.

On the next page, advisers were then presented with the name and avatar chosen by the worker. This introduced advisers to workers in a natural way and allowed advisers to clearly identify the gender of the worker, without making it too apparent that gender was the focus of our study. On the following page, advisers saw the worker's answers to the self-evaluation questions. On this page we used gendered pronouns to reinforce the signal of the worker's gender (all workers in our sample identified as either male or female). The information available to advisers differed by treatment, as we will explain later in this section.

Advisers were then given the opportunity to provide advice on the same page. The different types of advice are presented in Table 2. The first question asked the adviser for their advice on a Likert scale including "definitely job B", "most likely job B", "no advice", "most likely job A", and "definitely job A". The second question was similar, but used a binary scale (A or B). Third, we asked advisers for their confidence in the binary question on a scale from 0 to 10. If the adviser's advice was selected to be sent to a worker, one of these two pieces of advice (Likert scale advice or binary advice with confidence level) would randomly be chosen to be sent to the worker. On the next page, we then also asked advisers to imagine that instead of having to choose between two jobs, the worker would be able

to select their own threshold, and would earn a bonus of £0.20 times the threshold if they reached it. Advisers were asked which performance goal they would recommend the worker choose in this scenario. This item provides us with an additional, more continuous—but hypothetical—measure of advice.

Table 2: Advice Questions

Measure	**Question**	**Scale**
Likert Advice	Which advice do you want to send to [worker's name]?	5-point Likert scale ("definitely B" to "definitely A")
Binary Advice	Which job do you advise that [worker's name] chooses?	Binary (0-B, 1-A)
Binary Advice Confidence	On a scale from 0 to 10, how confident are you that your advice is the best for [worker's name]?	0-10
Performance Goal	Now imagine that instead of having to choose between Job A (Advanced) and Job B (Basic), [worker's name] would choose [his/her/their] performance goal as follows: • Goal 1: £0.20 for 1 or more correct answers • (...) • Goal 20: £4.00 for 20 correct answers Which goal would you advise [worker's name] to choose in this case?	1-20

After giving advice to their initial worker, advisers learned that they would be giving advice to four additional workers. We included these additional (randomly selected) workers to increase power and only revealed this to advisers at this point to allow us to treat the first worker as a one-shot observation in robustness checks. We informed advisers that each of the five workers was equally likely to receive their advice. Advisers read the information on each of the four additional workers sequentially, and gave their advice by answering the same questions as they had for the first worker.

At the end of the session, the advisers completed a sociodemographic survey. They also answered questions about their general level of confidence and risk attitudes (as in Dohmen, Falk, Huffman, Sunde, Schupp, & Wagner, 2011) and their beliefs in regard to gender differences in performance and confidence. Finally, we asked advisers to choose a name, which was used when communicating their advice to workers.

The procedures detailed so far were followed in our four adviser-level treatments. Below, we outline the procedures specific to each treatment, which differ in terms of the information the adviser receives prior to giving advice to the workers.

2.2.1 Baseline Treatment

In the baseline treatment, advisers see workers' answers to all self-evaluation questions (listed in Table 1), with the exception of the Likert performance measure, which we left out to avoid duplicate information. This treatment captures the usual situation, in which objective performance cannot be perfectly observed by advisers. This forces them to rely on subjective indicators that may be influenced by the workers' self-promotion skills and the adviser's gender attitudes.

2.2.2 Subjective Treatment

The subjective treatment removes all quantitative self-evaluation measures from the information given to advisers. Advisers only see the workers' answer to the verbal open-ended question when determining their advice. We use this treatment to study whether providing advisers with only subjective, verbal information attenuates or increases the gender gap in advice. Attenuation is possible if gender differences in self-promotion are less pronounced or less easily detected in brief self-evaluation verbal statements than in quantitative ones. Yet the gender gap in advice may also increase if any gender biases held by advisers, driven by their beliefs or preferences, become more important by increasing subjectivity in self-evaluations, as suggested by previous work on motivated reasoning and subjective judgment (e.g., Bohren, Imas, & Rosenberg, 2019).

2.2.3 Objective Treatment

The objective treatment takes the opposite approach, by decreasing subjectivity in the worker portfolio. It presents advisers with the same information as in the baseline treatment and an additional objective indicator: the worker's performance in Part one. Access to this indicator will reduce the role of the subjective indicators, which may limit both the role of gender differences in self-promotion and the role of potential underlying gender biases among advisers. Providing more objective indicators has also often been suggested as an equity-improving policy.[6]

2.2.4 Nudge Treatment

The purpose of the nudge treatment is to test whether a gender gap in advice could be mitigated by making advisers aware of prior results on gender gaps in self-confidence just before they give advice. Advisers see the same information in the worker's portfolio, as in the baseline treatment. In addition, prior to seeing the portfolio of their first worker, we presented advisers with three statements that describe situations in which a gender gap in self-confidence in a STEM setting led to unequal gender outcomes. Two of those statements summarize the results from two recent research studies, as follows:

1. Using data from more than 4,000 participants in online labor markets, researchers at Harvard University found that women rated their performance in a math and science quiz lower than men, despite similar performance levels. As a result, women who were hired performed better than men on average.

2. Using data from more than 3,600 people, researchers at University College London report that men are more confident in the self-assessment of their abilities than women. This gender gap in self-assessment contributes to the gender gap in who gets hired in a top job.

The first statement was based on Exley and Kessler (2022) and the second on Adamecz-Völgyi and Shure (2022). The third statement summarized a claim made by a major IT firm

[6]For example, Heilman and Haynes (2005) and Cecchi-Dimeglio (2017) point out a potential role for subjectivity in generating gender bias in evaluations, and propose using objective measures to address this bias. Objective information has also long been proposed as a primary method to reduce statistical discrimination (Autor & Scarborough, 2008).

as follows:

3. An internal study by the IT firm Hewlett Packard found that men who apply for a job meet only 60% of the qualifications, but women who apply meet 100% of them.[7]

We asked two comprehension questions to ensure that advisers read the statements. We also asked advisers whether they thought the information might affect their advice to encourage them to reflect on the information. Although there was no explicit recommendation that advisers adjust their advice based on this information, this recommendation was implicit. Effectively, the information provided by the nudge encouraged advisers to discount self-evaluations from male workers and inflate those of female workers. In this sense, the nudge may capture both a pure "informational" effect and an "encouragement" effect (see, e.g., Kessel, Mollerstrom, and Van Veldhuizen, 2021). Both aspects are likely to be present in similar information campaigns in the real world as well, where the encouragement effect would come from the authorities providing the information rather than experimenters.

2.2.5 Treatment Assignment

We included the baseline, subjective, and nudge treatments as between-subject treatments. The objective treatment was implemented as a within-subject treatment. For this purpose, the advisers in the baseline treatment (but not the other treatments) were given the chance to re-evaluate their five original workers while knowing the workers' actual score. We used a within-subject design for the objective treatment to increase statistical power, as detailed in our pre-analysis plan.

2.3 Second Quiz (Workers)

For Part three, all workers who completed Part one were invited to return for a second session approximately one week later. Workers received the advice from their matched adviser in binary or Likert-scale form and were then asked to make their decision (job A or B). After making their decision, workers learned both the nature of their task (twenty 30-second math and science questions for both tasks) and its incentives (A: £2 for solving at least 8 questions;

[7]Source: "Why Women Don't Apply for Jobs Unless They're 100% Qualified," Harvard Business Review (2014).

B: £1 for solving at least 5 questions). They were not told anything about the task they had not chosen. After completing the quiz, workers were given the option to provide general feedback.

2.4 Procedures

We conducted our experiment online on Prolific with 2,100 participants from the United Kingdom in July-August 2024. The experiment was programmed in oTree (Chen, Schonger, & Wickens, 2016). Throughout the experiment participants answered multiple comprehension questions. We recruited 350 participants (175 men and 175 women) to serve as workers. Our aim was to minimize the rate of worker attrition by making their bonus payments conditional on showing up a second time and by asking them to make a written commitment to complete the second part of the study (referenced as Part three above). In the end, 320 out of 350 workers (91%) completed both parts of the study. Workers only received feedback on their performance on the two quizzes at the end of their second session. Workers earned an average of £8.60, which includes a fixed payment of £6 (£3 paid out after completing each part of the study), for a median time commitment of about 25 minutes.

The remaining 1,750 participants were assigned to the role of advisers. Of these, 700 advisers were assigned to the baseline/objective treatment, 525 to the subjective treatment, and 525 to the nudge treatment. We oversampled the baseline/objective treatment to increase power, since this treatment is used for all of our statistical tests. Advisers earned a fixed payment of £3 for completing the study, for a median time commitment of 15 minutes. Three hundred and fifty advisers, whose advice was randomly selected to be sent to workers, also received an additional payment of either 0, £0.50, or £1, depending on the matched worker's performance in Part three.

To ensure that the adviser-level treatment effects are not confounded with order, day of the week, time of day, or any other effects, the treatments were randomly assigned to participants. In all treatments, advisers gave advice to the same group of 350 workers, such that treatment differences in advice are not driven by random variation in workers across treatments. In each treatment, each adviser gave advice to 5 randomly selected workers. Our random assignment of workers to advisers mechanism ensured the following. In the baseline/objective treatment, each of the 350 workers was assessed by 10 advisers. Therefore,

in the baseline treatment, we have a total of 3,500 adviser-worker pairings and observations. In the subjective and nudge treatments, 175 workers were assessed by 7 advisers and the other 175 workers were assessed by 8 advisers. In these treatments, we therefore have a total of 2,625 adviser-worker pairings and observations.

2.5 Main Hypotheses

We will test three pre-registered hypotheses.

Hypothesis 1 (Gender Gap in Self-promotion) *Female workers are less confident than male workers in their self-evaluations of performance on the STEM quiz.*

This auxiliary hypothesis replicates Exley and Kessler (2022)'s study, in which men revealed greater confidence than women in five self-evaluation questions about performance on a math and science quiz. This was true both when the evaluations were sent to a third party (and thus measure self-promotion) and when they were private. We hypothesize that we will find a similar gender gap in self-promotion in our data.

Hypothesis 2 (Gender Gap in Advice) *Female workers receive less ambitious advice than male workers in the baseline treatment.*

This is our first main hypothesis. We expect women to receive less ambitious advice for two potential reasons. The first directly relates to the gender gap in self-promotion: If women promote themselves less, we also expect advisers (who likely rely heavily on workers' self-evaluations) to give less ambitious advice to female workers. Second, advisers may give less ambitious advice to women, even when controlling for self-promotion, if advisers are influenced by stereotypes.

Hypothesis 3 (Treatment Differences) *The gender gap in advice is significantly smaller in the subjective, objective, and nudge treatments than in the baseline.*

This is our second main hypothesis. We expect all three treatments to reduce the gender gap in advice. For the subjective treatment, the gap in advice could either increase, due to increased scope for gender biases with less information, or decrease, due to reduced influence of the gender gap in self-promotion. We expect the latter effect to dominate and lead to

a smaller gender gap in advice in this treatment relative to the baseline. For the objective treatment, we expect that making advice less subjective by informing advisers of workers' past performance will lead to a smaller gender gap in advice, relative to the baseline. For the nudge treatment, we expect that providing advisers with information on the general gender gap in self-confidence will encourage advisers to give more ambitious advice to women and less ambitious advice to men, relative to the baseline.

3 Results

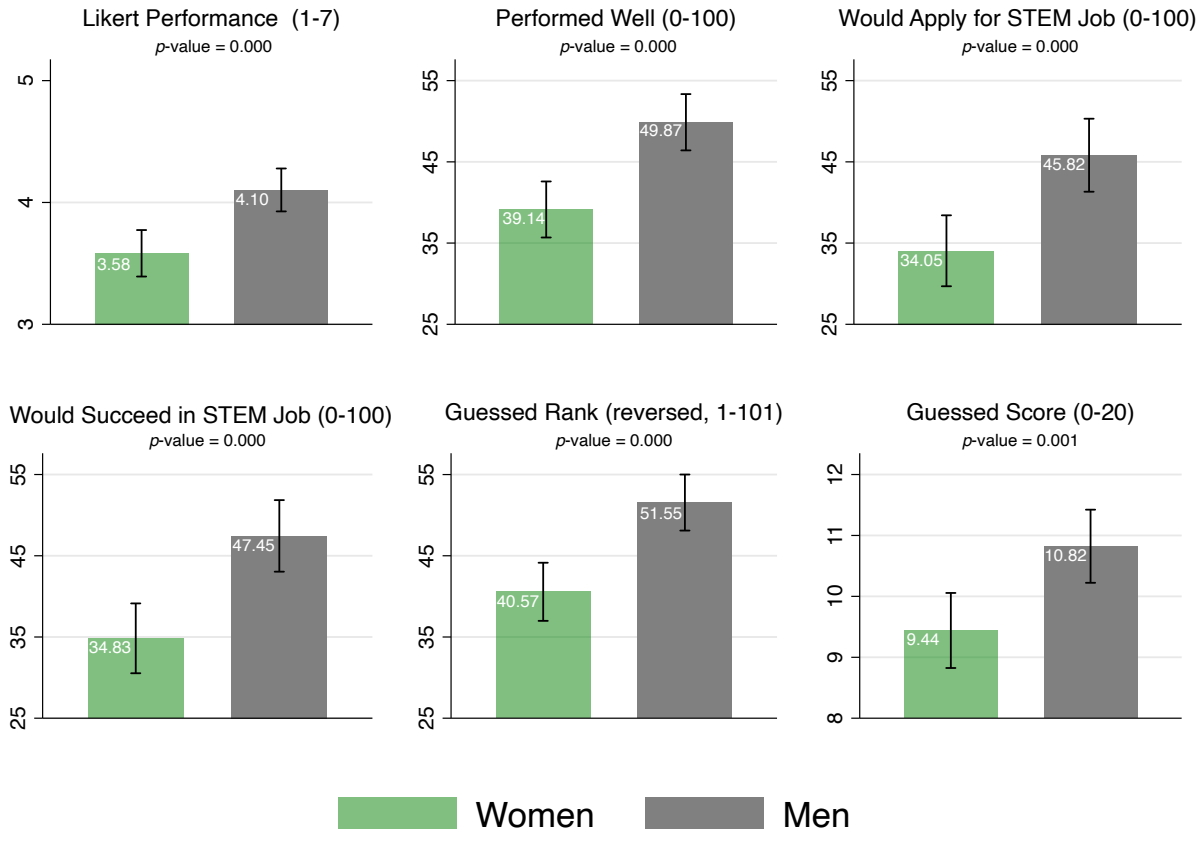
In this section we will present the pre-registered tests for our main hypotheses. We start with the gender gap in self-promotion, followed by the gender gap in advice and finally our treatment effects. We also include additional tests that serve to identify mechanisms, check robustness, or investigate other outcomes. All of our tests are pre-registered unless otherwise indicated. We present our pre-analysis plan and power calculations in Appendix A.9, alongside a brief description of where each of the pre-registered tests can be found in the paper. We use one-sided tests for every pre-registered test with a directional hypothesis, and two-sided tests in all other cases.

3.1 Gender Gap in Self-promotion

Figure 2 plots the average answer to each quantitative self-evaluation question, separately for men and women. In line with Exley and Kessler (2022) and our first hypothesis, women are less confident in all six measures (see also Table 3). Following Exley and Kessler (2022), we refer to this gender gap in self-evaluations as a gender gap in self-promotion because workers are aware that the evaluations are sent to an interested third party (in our case, the adviser).[8] In Table A.1 and Figure A.1 in Appendix A.1 we show that there is a positive correlation between a worker's score and their self-evaluation and that the correlation does not differ by gender.

[8]The exception is the score guess variable, which was elicited before workers were informed that their evaluations would be shared with an adviser.

Figure 2: Gender Gap in Self-promotion



Notes: Mean self-promotion scores for women and men are shown for each of the six measures. Error bars denote 95% confidence intervals. p-values come from the corresponding tests in Table 3.

The magnitudes of the gender gap in self-promotion in our sample are similar to those of Exley and Kessler (2022), although slightly smaller in most variables. For example, the gender gap in "Would Apply for STEM Job" in our sample is 12 percentage points, whereas in Exley and Kessler (2022) it was between 13 and 19 percentage points across their various samples. This similarity suggests that the gender gap in self-promotion is not specific to self-assessments that are shared with employers, but also arises when the recipient is an adviser whose role is to provide guidance. Similar gender differences in self-promotion are reported by Exley and Nielsen (2024) and Römer and Schröder (2025), as well as in studies with different math or science tasks (e.g., Niederle and Vesterlund, 2007).

With the notable exception of Exley and Kessler (2022), these studies also tend to find that men outperform women on math and science quizzes, which is reasonable given that men tend to have more training in STEM. We also find that men outperform women by an

average of 1 point (Table A.2). For our purpose, it is important that the gender difference in self-promotion persists when controlling for score fixed effects (Table A.3), and that advisers did not anticipate the gender gap in performance (Table A.19). In other words, the gender difference in self-evaluations captures both a gender difference in underlying performance and a gender difference in self-promotion.

Table 3: Gender Gap in Self-promotion

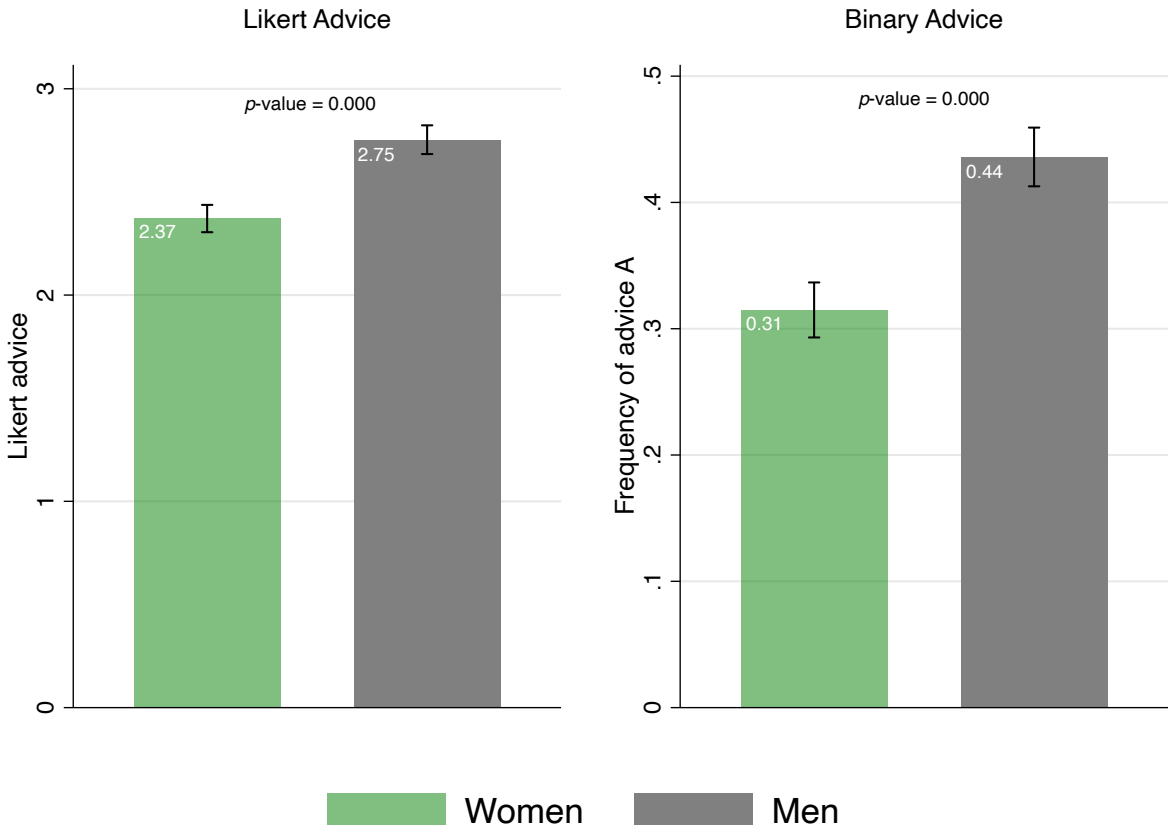
	Women		Men		Diff.	p-value
	Mean	SD	Mean	SD		
Likert Performance (1-7)	3.583	1.283	4.103	1.185	-0.520	0.000
Performed Well (0-100)	39.137	23.208	49.874	23.303	-10.737	0.000
Would Apply for STEM Job (0-100)	34.051	29.374	45.817	30.229	-11.766	0.000
Would Succeed in STEM Job (0-100)	34.834	28.901	47.446	29.629	-12.611	0.000
Guessed Rank (reversed, 1-101)	40.566	24.069	51.554	23.198	-10.989	0.000
Guessed Score (0-20)	9.440	4.143	10.823	4.037	-1.383	0.001

Notes: p-values come from pre-registered one-sided t-tests. The *Guessed Rank* variable is reversed, such that a higher value corresponds to a higher rank. The sample includes a total of 175 women and 175 men.

3.2 Gender Gap in Advice

We now turn to our main analysis and examine the extent to which there is a gender gap in advice in the baseline treatment. In line with our pre-analysis plan, we focus on the two main measures of advice (Likert and binary scale). Figure 3 plots the results. In line with our second hypothesis, women receive less ambitious advice on both the Likert scale (0.38 units on a 5-point scale) and on the binary measure (12 percentage points). Both differences are significant at the 1% level; the pre-registered test is presented in columns 1 and 3 of Table 4.

Figure 3: Gender Gap in Advice in the Baseline Treatment



Notes: 95% CIs shown. p-values derived from the analysis in columns (1) and (3) of Table 4. A higher value on the Likert scale means that advice is closer to A.

What explains the gender gap in advice? To answer this question, columns 2 and 4 in Table 4 control for all five quantitative self-evaluation questions.[9] Two results stand out. First, the self-evaluation questions are very predictive of advice received. This suggests that advisers relied heavily on workers' self-evaluations when giving advice. Second, the gender gap in advice vanishes after controlling for self-evaluations. In other words, the gender gap appears to have been driven entirely by gender differences in self-evaluations. The lack of a residual gender gap in advice also suggests that there was no intrinsic gender bias (e.g., due to stereotypes) among advisers in our population. Had such a bias existed, advisers would have given more ambitious advice to men even between men and women with similar self-evaluations.

[9]We deviate from our pre-analysis plan by controlling for all five self-evaluation questions as opposed to only the guessed score. This allows us to capture all information available to advisers. When we instead control only for the guessed score, a significant residual gender gap does remain; see Table A.5.

Table 4: Drivers of the Gender Gap in Advice

Advice Type (1-A)	Likert		Binary	
	(1)	(2)	(3)	(4)
Female	-0.382***	0.022	-0.121***	0.004
	(0.048)	(0.038)	(0.016)	(0.013)
Performed Well (0-100)		0.003*		0.000
		(0.002)		(0.001)
Guessed Rank (1-101)		0.006***		0.002***
		(0.002)		(0.001)
Would Apply for STEM Job (0-100)		0.004**		0.002***
		(0.002)		(0.001)
Would Succeed in STEM Job (0-100)		0.008***		0.002***
		(0.002)		(0.001)
Guessed Score (0-20)		0.113***		0.034***
		(0.008)		(0.003)
Constant	2.704***	0.446***	0.408***	-0.285***
	(0.058)	(0.073)	(0.020)	(0.024)
N	3,500	3,500	3,500	3,500
Clusters (Advisers)	700	700	700	700
R^2	0.019	0.432	0.017	0.382

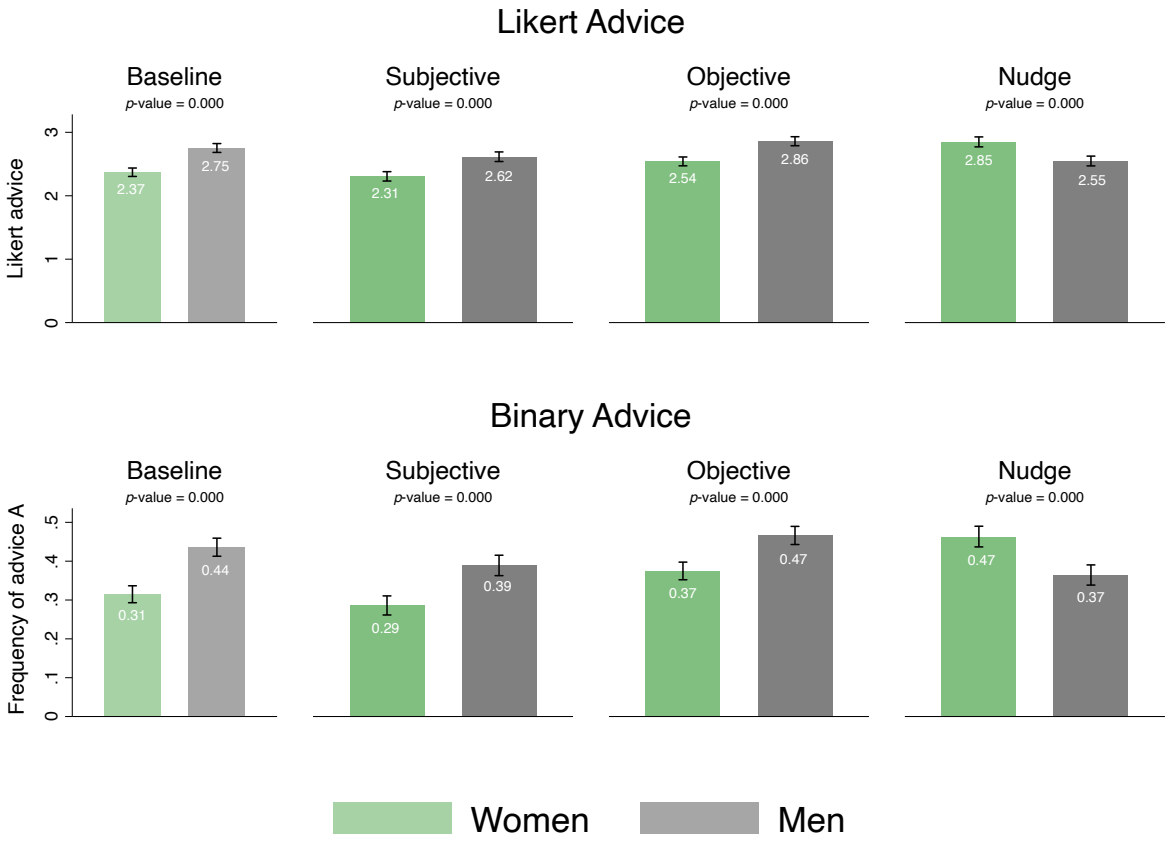
Notes: Results are from OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.3 Treatment Differences in Advice

The next question is whether this gender gap can be reduced. To answer this, we compare advice across the baseline condition and our three treatment conditions, each of which changes the information available to advisers. Figure 4 presents the advice given to men and women separately for the four treatments and the two types of advice. The biggest difference emerges in the nudge treatment: Informing advisers of prior results on gender gaps in confidence eliminates and even reverses the gender gap in advice. Indeed, women in the nudge treatment

are 10 percentage points *more likely* than men to receive ambitious advice. Columns 7 and 8 in Table 5, which formally compare the gender gap in advice across treatments, confirm that these differences are statistically significant. In other words, making advisers aware of the gender gap in self-confidence is a highly effective way to remove and even reverse the gender gap in advice in our context.

Figure 4: Gender Gap in Advice by Treatment



Notes: Mean advice for women and men is shown for each treatment, separately for the Likert (upper panel) and binary (lower panel) advice measures. A higher value on the Likert scale means that the advice is closer to A. Vertical bars indicate 95% confidence intervals. p-values correspond to gender-difference tests and are derived from the analysis in Table 5.

Results for the other two treatments are less clear-cut. Decreasing subjectivity in advice by informing advisers of the workers' true performance reduces the gender gap in advice by 0.06 units (p-value=0.06) or 3 percentage points (p-value=0.02) on the Likert and binary scale, respectively (see columns 5 and 6 in Table 5). Still, the gender gap in the objective treatment remains sizable and significant. This residual gap is driven by gender differences

in performance; when we control for performance differences, the gender gap in advice is small and no longer significant (see Table A.4 in Appendix). In other words, while advisers rely heavily on objective performance indicators, doing so does not greatly reduce the gender gap in advice due to men's superior performance. Our regression results also imply that the objective treatment would have been more effective in a setting with no gender gap in performance (or a superior performance among women, as in Exley and Kessler, 2022).

Table 5: Gender Gap in Advice, by Treatment

			Baseline vs. *Treatment*					
Treatment	Baseline		Subjective		Objective		Nudge	
Advice Type (1-A)	Likert (1)	Binary (2)	Likert (3)	Binary (4)	Likert (5)	Binary (6)	Likert (7)	Binary (8)
Female	-0.382***	-0.121***	-0.382***	-0.121***	-0.382***	-0.121***	-0.382***	-0.121***
	(0.048)	(0.016)	(0.048)	(0.016)	(0.048)	(0.016)	(0.048)	(0.016)
Subjective			-0.138**	-0.047**				
			(0.056)	(0.019)				
Objective					0.107***	0.030***		
					(0.032)	(0.011)		
Nudge							-0.205***	-0.068***
							(0.058)	(0.019)
Female×Subjective			0.074	0.019				
			(0.072)	(0.024)				
Female×Objective					0.064*	0.030**		
					(0.043)	(0.014)		
Female×Nudge							0.683***	0.221***
							(0.077)	(0.026)
Constant	2.704***	0.408***	2.710***	0.399***	2.732***	0.414***	2.722***	0.412***
	(0.058)	(0.020)	(0.050)	(0.017)	(0.056)	(0.018)	(0.050)	(0.017)
Gender Gap in *Treatment*								
Coefficient			-0.308***	-0.102***	-0.318***	-0.091***	0.301***	0.099***
Standard error			(0.053)	(0.018)	(0.049)	(0.016)	(0.060)	(0.021)
N	3,500	3,500	6,125	6,125	7,000	7,000	6,125	6,125
Clusters (Advisers)	700	700	1,225	1,225	700	700	1,225	1,225
R^2	0.019	0.017	0.018	0.018	0.017	0.015	0.018	0.017

Notes: Results are from OLS regressions with standard errors clustered at adviser level. All specifications control for worker order (1 to 5). *Female* is a binary indicator for the workers' gender (0-male, 1-female). The other variables are binary indicators for the respective treatments. Statistical significance of the coefficient estimates for the interaction terms is based on a one-sided test for a positive coefficient, in line with our pre-analysis plan. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Finally, we find no evidence that the gender gap in advice is reduced by increasing subjectivity in advice, when advisers are only provided with qualitative self-evaluation measures (Table 5, columns 3-4). To shed further light on this result, we used Microsoft Copilot (ver-

sion 365) to rate the confidence level expressed by the qualitative performance question (see Appendix A.3 for more details). Table A.7 shows that the gender gap in advice vanishes when controlling for these AI ratings. This suggests that women also promoted themselves less on the qualitative question, which in turn fully explains why women receive less ambitious advice in this treatment. Importantly, it demonstrates that gender gaps in advice driven by gender differences in self-promotion can arise even when self-promotion is limited to brief (generally, one or two sentences in our sample) self-evaluation narratives. These results also imply that increased subjectivity does not generate an increased intrinsic gender bias among our advisers.

In summary, we find that the gender gap in advice is remarkably robust to increasing subjectivity in advice, by providing advisers only with workers' brief verbal self-evaluation statements (subjective treatment). The gender gap in advice is also robust to decreasing subjectivity in advice, by providing advisers with information on workers' actual performance (objective treatment). However, providing advisers with factual information about the general gender gap in confidence affects the gender gap in advice, eliminating and even reversing the gap.

3.4 Worker Earnings

What do these differences in advice imply for worker earnings? As a first step, it is important to note that, as expected, the vast majority of workers (97%) followed the advice received (see Appendix A.5 for more details). Rather than looking at realized earnings (which depend on the adviser randomly matched with each worker), we examine *expected* earnings, based on all the advice generated for a particular worker (across all advisers), assuming that workers would have followed the advice. We compute expected earnings using both their initial (part one) and realized (part three) scores. Table 6 presents results based on workers' initial score (part one) using either (1) a dummy for whether the advice maximized the worker's expected earnings (our pre-registered test, columns 1 and 2); (2) expected earnings when following the advice (columns 3 and 4, not pre-registered); and (3) the gap between expected earnings based on following the advice and what a risk-neutral worker would have done based on their part one score guess, that is, choose job A if their score guess in part one was 8 or greater (columns 5 and 6, not pre-registered).

Table 6: Expected Earnings Based on Part 1 Score

	Advice Maximized Expected Earnings		Expected Earnings Based on Advice		Gap in Expected Earnings With vs. Without Advice	
	(1)	(2)	(3)	(4)	(5)	(6)
Female		-0.027*		-0.161***		0.034*
		(0.017)		(0.018)		(0.018)
Subjective	-0.032**	-0.046**	-0.032**	-0.045**	-0.035**	-0.047**
	(0.013)	(0.019)	(0.015)	(0.020)	(0.016)	(0.022)
Objective	0.087***	0.073***	0.083***	0.073***	0.083***	0.073***
	(0.008)	(0.011)	(0.008)	(0.011)	(0.008)	(0.011)
Nudge	0.022	-0.038**	0.024*	-0.040**	0.027*	-0.035
	(0.013)	(0.019)	(0.015)	(0.020)	(0.016)	(0.022)
Female×Subjective		0.028		0.025		0.025
		(0.026)		(0.027)		(0.029)
Female×Objective		0.029**		0.022		0.022
		(0.014)		(0.014)		(0.014)
Female×Nudge		0.118***		0.129***		0.123***
		(0.026)		(0.029)		(0.028)
Constant	0.469***	0.482***	1.262***	1.343***	-0.270***	-0.287***
	(0.014)	(0.016)	(0.015)	(0.018)	(0.016)	(0.018)
N	12,250	12,250	12,250	12,250	12,250	12,250
Clusters (Advisers)	1,750	1,750	1,750	1,750	1,750	1,750
R^2	0.009	0.011	0.007	0.021	0.007	0.012

Notes: Results are from OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). The other variables are binary indicators for the respective treatments. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Our results show that in all three measures, expected earnings were highest when advisers knew the worker's part one score (objective treatment) and lowest when they only had access to brief open-ended self-assessments of performance (subjective treatment). In other words, the more information advisers had, the better their advice, which is intuitive. Columns 2, 4, and 6 show that these effects are quite similar for men and women. These columns also show that the nudge treatment reduces the expected earnings of men (although not statistically significant in column 6) while increasing those of women. This result is driven by a tendency for advisers to recommend job A at suboptimally low rates (see Appendix A.8). The nudge offsets this tendency for female workers, which increases their expected earnings and hence

the quality of advice. These results are similar when we calculate expected earnings based on the part three score instead (Table A.9, not pre-registered).

4 Discussion

4.1 Additional Results

We report pre-registered robustness checks and additional analyses in Appendix A.6. Our main findings are robust to controlling for worker socio-demographic characteristics (Table A.10) and adviser socio-demographic characteristics, confidence, and risk attitudes (Table A.11). Treatment differences in the gender gap in advice also persist when (1) using alternative advice measures (Table A.12) and (2) restricting the analysis to the first worker advised (Table A.13). Relative to the main results, the main difference is that, in both cases, the gender gap in the objective treatment is no longer statistically different from the baseline gap. In the baseline treatment, advisers who recommend job A are less confident in their advice than advisers who recommend job B, especially when advising women, driven by women's lower confidence (Tables A.14 and A.15). On average, advisers in the baseline, objective, and subjective treatments believe that women and men will perform equally well, but that women will be less confident in their performance. These beliefs, however, do not lead to adjustments in advice (Table A.16). Moreover, the gender gap in advice is largest among advisers who believe men perform better than women, and is not statistically different from zero among those who believe women perform better than men (Table A.17). The nudge treatment shifted these beliefs in the expected direction, increasing the share of advisers who believe that men are more confident and that women perform better (Tables A.18 and A.19). Finally, male and female advisers generally give similar advice, except in the nudge treatment, where female advisers are more likely to give women more ambitious advice. This suggests that the nudge was particularly effective in changing female advisers' recommendations (Table A.20, not pre-registered).

4.2 Additional Treatment: Experienced Advisers and Vignettes

As a next step, we report the results of an additional treatment that examines the robustness of our results along two dimensions. First, we test whether our main results hold up in a sample of participants who regularly give advice in real-world settings (managers and teachers). Second, we investigate whether our results hold up in a set of vignette tasks that capture more concrete career decision contexts.

For this purpose, we replicated our baseline treatment on Prolific in April 2026 with two main changes. First, we required our advisers (N=525) to have either management or teaching experience.[10] Second, after the incentivized advice task, advisers completed a vignette task in which they provided hypothetical advice in realistic career scenarios. Each vignette featured a common career decision, such as whether to apply for a promotion, switch to a riskier form of employment, or relocate for professional reasons (see Table A.24 for the full list). We included six different scenarios and randomized the advisee's gender (conveyed through names and pronouns) and confidence level (conveyed through the text) within each scenario. Each adviser evaluated 12 vignettes; we present additional design details in Appendix A.7 and the pre-analysis plan in Appendix A.10.

The results closely support the mechanisms identified in the main experiment. In the original incentivized task, experienced advisers provide less ambitious advice to female workers than to male workers, and this gap is driven by gender differences in self-promotion (Table A.25). Indeed, neither the gender gap in advice nor the dominant role of self-promotion differs significantly by advisers' real-life advising experience (Table A.26). The vignette evidence points in the same direction: in hypothetical career scenarios, experienced advisers give more ambitious advice to confident individuals, while giving similar advice to men and women holding confidence fixed (Table A.27). These results also closely resemble those in our main experiment (Table A.28), suggesting that advisers' responses to confidence signals also operate in the more typical career decision contexts captured by the vignette task. The results are robust to several pre-registered alternative specifications (see Appendix A.7). These findings suggest that our conclusions extend beyond the main experimental setting:

[10] Participants were required to be currently employed and either (i) have at least three years of management experience, supervise at least two subordinates, and regularly interact with employees, or (ii) work as teachers, either in educational institutions or workplace settings. Advice provision is frequent in both groups: 84 percent of those with management experience and 87 percent of teachers report giving advice at least weekly (see Tables A.21 and A.23).

they hold among advisers with real-world advising experience and in more concrete career decision contexts.

5 Conclusion

Why do women receive different career advice than men? A common explanation is that advisers evaluate men and women differently. In this paper, we provide evidence for a different mechanism. We conducted an online experiment in which advisers recommend whether workers should pursue a more or a less ambitious opportunity based on workers' self-promotion. We find that women systematically self-promote less than equally performing men and, as a result, receive less ambitious career advice. This gender gap arises because advisers respond to workers' expressed confidence rather than because they treat otherwise identical men and women differently. We find these results with both experienced and inexperienced advisers and in both incentivized and vignette tasks.

Our findings suggest that gender differences in self-promotion are a key mechanism underlying gender differences in career advice. This perspective also helps reconcile our results with previous studies that find little or no gender gap in advice (Brandts, Groenert, & Rott, 2015; Gallen & Wasserman, 2021). In those settings, advisers relied on objective performance measures or standardized information provided by the experimenter, leaving little scope for gender differences in self-promotion to influence advice. By contrast, advisers in our study observed workers' own self-evaluations, which differed systematically by gender even after accounting for workers' actual performance. Consistent with this interpretation, the gender gap in advice disappears once we account for gender differences in self-promotion.

Our findings suggest that persistent gender inequalities need not arise because decision makers evaluate men and women differently. Instead, they can emerge because decision makers respond similarly to information that is itself systematically gendered. Advisers in our study recognize that men are, on average, more confident than women, yet they nevertheless interpret confidence signals similarly across genders, consistent with Exley and Nielsen (2024). Likewise, changing the objectivity of the information available to advisers has little effect on the gender gap in advice. Together, these findings point to gender differences in self-promotion—not intrinsic adviser bias—as the primary mechanism generating unequal advice.

Consequently, even unbiased and well-intentioned advisers may inadvertently reinforce gender differences in educational and career trajectories by responding rationally to systematically gendered information.

Unlike Exley and Kessler (2022), men in our experiment outperform women on average, although this pattern is consistent with Exley and Nielsen (2024), Römer and Schröder (2025), and Chang, Saccardo, and Gallus (2025). Importantly, however, gender differences in self-promotion remain substantial after controlling for actual performance. Advisers did not anticipate the gender performance gap, and therefore evaluated workers "as-if" there were no gender gap in performance. This implies that our setting was functionally similar to a setting with no gap in performance, as might be expected in most real-world applications. Our conclusions regarding the role of self-promotion are unlikely to be driven by gender differences in underlying performance.

Our results have important policy implications. The substantial gender gap in self-promotion suggests that equally qualified women are less likely than men to communicate confidence in their abilities. Because advisers respond to these signals, such differences shape the advice women receive and may ultimately influence their career choices. We show that, in our setting, these disparities can be eliminated by informing advisers about prior evidence on gender differences in self-confidence and self-promotion. Complementary policies could instead target the supply of self-promotion through mentorship, coaching, and communication-skills programs that help women more effectively convey their abilities.

References

Adamecz-Völgyi, A., & Shure, N. (2022). The Gender Gap in Top Jobs–The Role of Overconfidence. *Labour Economics*, *79*, 102283.

Angelov, N., Johansson, P., & Lindahl, E. (2016). Parenthood and the Gender Gap in Pay. *Journal of Labor Economics*, *34*(3), 545–579. https://doi.org/10.1086/684851

Autor, D. H., & Scarborough, D. (2008). Does Job Testing Harm Minority Workers? Evidence from Retail Establishments. *The Quarterly Journal of Economics*, *123*(1), 219–277.

Bandiera, O., Parekh, N., Petrongolo, B., & Rao, M. (2022). Men are from Mars, and Women too: A Bayesian meta-analysis of Overconfidence Experiments. *Economica*. https://doi.org/10.1111/ecca.12407

Bertrand, M., Goldin, C., & Katz, L. F. (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics*, *2*(3), 228–255.

Blau, F. D., & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, *55*(3), 789–865.

Bohren, J. A., Imas, A., & Rosenberg, M. (2019). The Dynamics of Discrimination: Theory and Evidence. *American Economic Review*, *109*(10), 3395–3436.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019). Beliefs About Gender. *American Economic Review*, *109*(3), 739–73.

Brandts, J., Groenert, V., & Rott, C. (2015). The Impact of Advice on Women's and Men's Selection Into Competition. *Management Science*, *61*(5), 1018–1035.

Brandts, J., & Rott, C. (2021). Advice from Women and Men and Selection Into Competition. *Journal of Economic Psychology*, *82*, 102333. https://doi.org/10.1016/j.joep.2020.102333

Breda, T., Grenet, J., Monnet, M., & Van Effenterre, C. (2023). How Effective are Female Role Models in Steering Girls Towards STEM? Evidence from French High Schools. *The Economic Journal*, *133*(653), 1773–1809. https://doi.org/10.1093/ej/uead019

Carlana, M. (2019). Implicit Stereotypes: Evidence from Teachers' Gender Bias. *The Quarterly Journal of Economics*, *134*(3), 1163–1224.

Cecchi-Dimeglio, P. (2017). How Gender Bias Corrupts Performance Reviews, and What to Do About it. *Harvard Business Review*.

Çelen, B., Kariv, S., & Schotter, A. (2010). An Experimental Test of Advice and Social Learning. *Management Science*, *56*(10), 1687–1701.

Chang, J., Saccardo, S., & Gallus, J. (2025). Leaning in Softly: on Gender and Self-Promotion. *Available at SSRN 5240652*.

Chaudhuri, A., Schotter, A., & Sopher, B. (2009). Talking Ourselves to Efficiency: Coordination in Inter-Generational Minimum Effort Games with Private, Almost Common and Common Knowledge of Advice. *The Economic Journal*, *119*(534), 91–122.

Chen, D. L., Schonger, M., & Wickens, C. (2016). OTree—An Open-source Platform for Laboratory, Online, and Field Experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Coffman, K., Collis, M. R., & Kulkarni, L. (2024). Stereotypes and Belief Updating. *Journal of the European Economic Association*, *22*(3), 1011–1054. https://doi.org/10.1093/jeea/jvad063

Cortés, P., & Pan, J. (2023). Children and the Remaining Gender Gaps in the Labor Market. *Journal of Economic Literature*, *61*(4), 1359–1409.

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, *47*(2), 448–474.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, *9*(3), 522–550.

Eagly, A. H., & Carli, L. L. (2003). The Female Leadership Advantage: an Evaluation of the Evidence. *The leadership quarterly*, *14*(6), 807–834.

Exley, C. L., & Kessler, J. B. (2022). The Gender Gap in Self-promotion. *The Quarterly Journal of Economics*, *137*(3), 1345–1381.

Exley, C. L., & Nielsen, K. (2024). The Gender Gap in Confidence: Expected but Not Accounted for. *American Economic Review*, *114*(3), 851–885.

Falk, A., Kosse, F., & Pinger, P. (2026). Mentoring and Schooling Decisions: Causal Evidence. *Journal of Political Economy*, *134*(1), 366–396. https://doi.org/10.1086/738484

Gallen, Y., & Wasserman, M. (2021). Informed Choices: Gender Gaps in Career Advice. *CEPR discussion paper No. DP15728*.

Gangadharan, L., Jain, T., Maitra, P., & Vecci, J. (2016). Social Identity and Governance: the Behavioral Response to Female Leaders. *European Economic Review*, *90*, 302–325.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, *104*(4), 1091–1119.

Goulas, S., Griselda, S., & Megalokonomou, R. (2024). Comparative Advantage and Gender Gap in STEM. *Journal of Human Resources*, *59*(6), 1937–1980.

Goulas, S., Gunawardena, B., Megalokonomou, R., & Zenou, Y. (2024). Gender Role Models in Education. *CESifo Working Paper No. 11330 CESifo Munich*.

Guillen, P., & Hakimov, R. (2018). The Effectiveness of Top-down Advice in Strategy-proof Mechanisms: a Field Experiment. *European Economic Review*, *101*, 505–511.

Heilman, M. E., & Haynes, M. C. (2005). No Credit Where Credit Is Due: Attributional Rationalization of Women's Success in Male-female Teams. *Journal of Applied Psychology*, *90*(5), 905.

Hoxby, C. M., & Turner, S. (2015). What High-Achieving Low-Income Students Know About College. *American Economic Review*, *105*(5), 514–517.

Kessel, D., Mollerstrom, J., & Van Veldhuizen, R. (2021). Can Simple Advice Eliminate the Gender Gap in Willingness to Compete? *European Economic Review*, *138*, 10377.

Krawczyk, M., & Smyk, M. (2016). Author's Gender Affects Rating of Academic Articles: Evidence from an Incentivized, Deception-free Laboratory Experiment. *European Economic Review*, *90*, 326–335.

Lavy, V., & Megalokonomou, R. (2024). The short- and the long-run impact of gender-biased teachers. *American Economic Journal: Applied Economics*, *16*(2), 176–218. https://doi.org/10.1257/app.20210052

Mengel, F., Sauermann, J., & Zölitz, U. (2019). Gender Bias in Teaching Evaluations. *Journal of the European Economic Association*, *17*(2), 535–566.

Mengel, F. (2020). Gender Differences in Networking. *The Economic Journal*, *130*(630), 1842–1873.

Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away from Competition? Do men compete too much? *The Quarterly Journal of Economics*, *122*(3), 1067–1101.

Niederle, M. (2016). Gender. In J. Kagel & A. E. Roth (Eds.), *Handbook of Experimental Economics, 2nd ed.* (pp. 481–553). Princeton University Press.

Oreopoulos, P., & Petronijevic, U. (2013). Making college worth it: A review of the returns to higher education. *The Future of Children*, *23*(1), 41–65.

Özgümüs, A., Rau, H. A., Trautmann, S. T., & König-Kersting, C. (2020). Gender Bias in the Evaluation of Teaching Materials. *Frontiers in Psychology*, *11*, 1074.

Porter, C., & Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, *12*(3), 226–54. https://doi.org/10.1257/app.20180426

Reuben, E., Wiswall, M., & Zafar, B. (2017). Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender. *The Economic Journal*, *127*(604), 2153–2186.

Römer, N., & Schröder, M. (2025). *Effects of Written Self-promotion on Gender Bias and Decision Quality* (tech. rep.). Hannover Economic Papers (HEP).

Sarsons, H., Gërxhani, K., Reuben, E., & Schram, A. (2021). Gender Differences in Recognition for Group Work. *Journal of Political Economy*, *129*(1), 101–147.

Schotter, A., & Sopher, B. (2003). Social Learning and Coordination Conventions in Intergenerational Games: an Experimental Study. *Journal of Political Economy*, *111*(3), 498–529.

Schotter, A., & Sopher, B. (2007). Advice and Behavior in Intergenerational Ultimatum Games: an Experimental Approach. *Games and Economic Behavior*, *58*(2), 365–393. https://doi.org/10.1016/j.geb.2006.03.005

Silva-Goncalves, J., & van Veldhuizen, R. (2020). Subjective Judgment and Gender Bias in Advice: Evidence from the Laboratory. *Life Course Centre Working Paper*, (2020-29).

Wolter, S. C., & Zöllner, T. (2024). *Are Parents an Obstacle to Gender-atypical Occupational Choices?* (Tech. rep.). IZA Discussion Papers.

A Appendix

A.1 Gender Gap in Performance

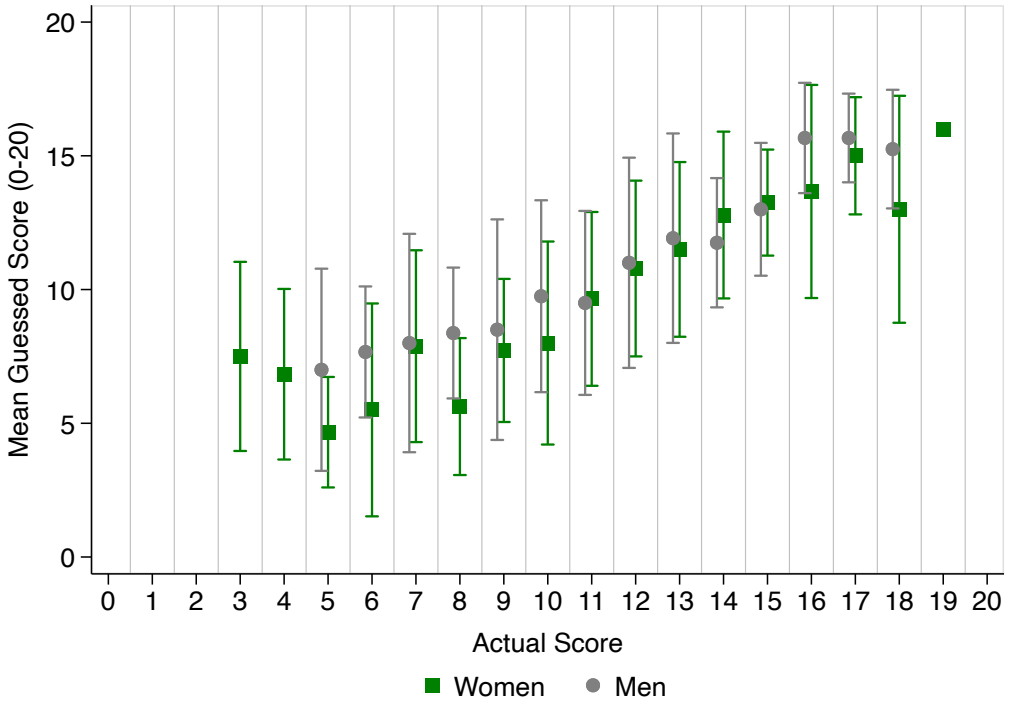
Table A.1 and Figure A.1 (not pre-registered) show that the relationship between the actual score and self-promotion does not differ by gender. The estimated coefficients of the interaction terms between gender and (actual) score for the different outcomes are small and statistically indistinguishable from 0. Table A.2 presents the means, standard deviations, differences of the means, and P-values of differences in worker scores on the 1st and 2nd quizzes for men and women. The scores range from 0 to 20. Men outperform women in both cases by about 1 point (out of 20). Table A.3 (pre-registered) shows that the gender gap in self-promotion remains significant when we include score fixed effects. This is shown by the negative and statistical significant coefficients of the variable "Female" across the different outcomes. Table A.4 (not pre-registered) shows that the gender gap in advice in the baseline treatment remains significant when controlling for score fixed effects, and that the gender gap in advice in the objective treatment vanishes when controlling for workers' performance in the STEM quiz in part 1.

Table A.1: Relationship Between Actual Score and Self-promotion

	Likert Performance (1-7)	Performed Well (0-100)	Would Apply for STEM Job (0-100)	Would Succeed in STEM Job (0-100)	Guessed Rank (1-101)	Guessed Score (0-20)
	(1)	(2)	(3)	(4)	(5)	(6)
Score	0.164***	3.352***	3.346***	3.073***	2.845***	0.713***
	(0.024)	(0.460)	(0.629)	(0.610)	(0.479)	(0.075)
Female	-0.801**	-10.646	-12.532	-21.014**	-17.204**	-0.860
	(0.388)	(7.310)	(9.991)	(9.696)	(7.609)	(1.198)
Female×Score	0.043	0.327	0.408	1.106	0.876	0.022
	(0.033)	(0.631)	(0.862)	(0.836)	(0.656)	(0.103)
Constant	2.208***	11.106**	7.119	11.903	18.654***	2.573***
	(0.293)	(5.530)	(7.559)	(7.336)	(5.756)	(0.906)
N	350	350	350	350	350	350
R^2	0.300	0.304	0.197	0.222	0.271	0.381

Notes: Estimates are based on OLS regressions. *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Score* is a continuous variable in the range [0,20] and captures the worker's actual score. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.1: Mean and Standard Deviation of Guessed Score by Actual Score by Gender



Notes: Markers show mean guessed scores by actual score, with vertical lines indicating the corresponding standard deviations. Circles correspond to means for men, and squares correspond to means for women.

Table A.2: Gender Gap in Performance

	Women		Men		Diff.	p-value
	Mean	SD	Mean	SD		
Score in 1st Quiz, *Before Advice*	10.514	3.510	11.566	3.291	-1.051	0.004
Score in 2nd Quiz, *After Advice*	12.062	3.592	13.044	3.654	-0.981	0.016

Notes: The sample *"Before Advice"* is based on part 1, and includes 175 women and 175 men. The sample *"After Advice"* is based on part three, and includes 160 women and 160 men.

Table A.3: Gender Gap in Self-promotion Controlling for Performance

	Likert Performance (1-7)	Performed Well (0-100)	Would Apply for STEM Job (0-100)	Would Succeed in STEM Job (0-100)	Guessed Rank (1-101)	Guessed Score (0-20)
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.359***	-7.357***	-8.361***	-8.942***	-7.725***	-0.733**
	(0.118)	(2.202)	(2.977)	(2.903)	(2.298)	(0.361)
Constant	8.233***	3.359***	46.857***	48.775***	56.861***	56.942***
	(2.341)	(0.763)	(14.280)	(14.899)	(19.304)	(18.824)
Score FE	✓	✓	✓	✓	✓	✓
N	350	350	350	350	350	350
R^2	0.319	0.333	0.248	0.264	0.298	0.407

Notes: Estimates are based on OLS regressions. *Score FE* (fixed effects) controls for actual performance. *Female* is a binary indicator for the worker's gender (0-male, 1-female). Statistical significance of the coefficient estimate for the binary indicator *Female* is based on a one-sided test. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Gender Gap in Advice Controlling for Performance

Treatment	Baseline		Subjective		Objective		Nudge	
Advice Type (1-A)	Likert	Binary	Likert	Binary	Likert	Binary	Likert	Binary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.225***	-0.075***	-0.192***	-0.063***	-0.025	-0.005	0.433***	0.140***
	(0.047)	(0.016)	(0.052)	(0.018)	(0.040)	(0.014)	(0.055)	(0.019)
Constant	2.226***	0.297***	1.700***	0.088	1.131***	0.088	2.366***	0.309**
	(0.306)	(0.100)	(0.243)	(0.072)	(0.083)	(0.069)	(0.385)	(0.132)
Score FE	✓	✓	✓	✓	✓	✓	✓	✓
N	3,500	3,500	2,625	2,625	3,500	3,500	2,625	2,625
Clusters (Advisers)	700	700	525	525	700	700	525	525
R^2	0.203	0.186	0.125	0.115	0.457	0.388	0.160	0.142

Notes: Results are from OLS regressions with standard errors clustered at adviser level. All specifications control for worker order (1 to 5). *Female* is a binary indicator for the workers' gender (0-male, 1-female). *Score FE* (fixed effects) controls for actual performance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.2 Gender Gap in Advice in Baseline Treatment

In line with the pre-analysis plan, we report the analysis of the drivers of the gender gap in advice in the baseline treatment, controlling for only one self-evaluation measure—namely, participants' guessed score. The results show that the gender gap in advice is strongly attenuated, with a small gender gap persisting. Since this gap is less than one-third of the

original gap and since it vanishes completely when we also control for the other self-evaluation variables, we decided to report the results that control for all variables available to advisers in the main text.

Table A.5: Drivers of the Gender Gap in Advice

Advice Type (1-A)	Likert	Binary
	(1)	(2)
Female	-0.082**	-0.029**
	(0.041)	(0.014)
Guessed Score (0-20)	0.217***	0.067***
	(0.005)	(0.002)
Constant	0.356***	-0.316***
	(0.075)	(0.024)
N	3,500	3,500
Clusters (Advisers)	700	700
R^2	0.384	0.334

Notes: Results are from OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Guessed Score (0-20)* is the worker's guessed score in part 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.3 Gender Gap in Qualitative Self-promotion

In line with the pre-analysis plan, we use AI to quantify responses to the open-ended self-evaluation question. Most workers answered this question with one or two sentences. We explore the extent to which there are gender differences in the qualitative measure of self-promotion using AI tools (Copilot for Microsoft 365), which allow us to quantify the expressed level of confidence in the statements. Specifically, we asked Copilot to evaluate the statements according to how confident the worker is about their performance (on a scale from 0 to 10), to guess the likelihood that the worker gave more than 8 correct answers (the threshold for success in job A), to guess the actual score of the worker, and to rate the statement based on the quality of the writing (on a scale from 0 to 10). The results show that (a) women were also significantly less confident in the open-ended question (Table A.6, pre-registered) and (b) these gender differences can fully explain the gender gap in advice in the subjective treatment (Table A.7, not pre-registered). Note that we did not pre-register the exact question, or the specific AI used to quantify the qualitative self-promotions.

Table A.6: Gender Gap in Confidence Based on AI Ratings of the Qualitative Performance Statements

	AI Confidence in Performance (0-10)	AI More Than 8 Correct	AI Score Guess (0-20)	AI Writing Quality (0-10)
	(1)	(2)	(3)	(4)
Female	-0.367***	-0.421***	-0.888***	-0.226***
	(0.134)	(0.145)	(0.266)	(0.084)
Constant	3.367***	3.421***	6.888***	5.226***
	(0.870)	(0.939)	(1.727)	(0.545)
Score FE	✓	✓	✓	✓
N	350	350	350	350
R^2	0.296	0.247	0.285	0.281

Notes: Estimates are based on OLS regressions. *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Score FE* (fixed effects) controls for actual performance. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Drivers of the Gender Gap in Advice in Subjective Treatment

Advice A Measure:	Likert		Binary	
	(1)	(2)	(3)	(4)
Female	-0.308***	-0.020	-0.102***	-0.014
	(0.053)	(0.047)	(0.018)	(0.016)
AI Confidence in Performance		0.566***		0.197***
		(0.059)		(0.021)
AI More Than 8 Correct		0.028		0.011
		(0.044)		(0.016)
AI Score Guess		-0.043		-0.027*
		(0.041)		(0.015)
AI Writing Quality		0.067		0.030*
		(0.049)		(0.017)
Constant	2.579***	-0.005	0.339***	-0.496***
	(0.065)	(0.199)	(0.022)	(0.066)
N	2,625	2,625	2,625	2,625
Clusters (Advisers)	525	525	525	525
R^2	0.015	0.307	0.016	0.269

Notes: Estimates are based on OLS regressions with standard errors clustered at the adviser level. All specifications control for worker order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.4 Advice and Worker Guessed Score

Table A.8 (pre-registered) re-examines the role of self-evaluations more flexibly, which complements Table 4. The estimates reveal a monotonically increasing relationship between workers' guessed scores and the likelihood of receiving advice A: coefficients are negative and significant for guesses below the performance threshold of 8, cross zero in its neighbourhood, and grow increasingly large and positive above it. Advisers thus recommend job A predominantly to workers whose self-assessed score comfortably exceeds the threshold—a pattern that is consistent across both the Likert and binary measures and economically meaningful in magnitude.

Table A.8: Effect of Worker Score Guess on Advice in Baseline Treatment

Advice A measure:	Likert (1)	Binary (2)
Guessed Score=2	-0.800***	-0.118***
	(0.127)	(0.044)
Guessed Score=3	-0.768***	-0.139***
	(0.102)	(0.029)
Guessed Score=4	-0.556***	-0.112***
	(0.106)	(0.029)
Guessed Score=5	-0.568***	-0.091***
	(0.096)	(0.030)
Guessed Score=6	-0.450***	-0.073**
	(0.114)	(0.035)
Guessed Score=7	-0.200**	-0.029
	(0.095)	(0.031)
Guessed Score=9	0.000	0.007
	(0.110)	(0.036)
Guessed Score=10	0.235**	0.083**
	(0.100)	(0.034)
Guessed Score=11	0.700***	0.236***
	(0.111)	(0.039)
Guessed Score=12	0.798***	0.305***
	(0.108)	(0.038)
Guessed Score=13	0.742***	0.282***
	(0.143)	(0.052)
Guessed Score=14	1.461***	0.514***
	(0.107)	(0.037)
Guessed Score=15	1.689***	0.597***
	(0.119)	(0.039)
Guessed Score=16	1.983***	0.713***
	(0.104)	(0.036)
Guessed Score=17	2.323***	0.765***
	(0.109)	(0.033)
Guessed Score=18	2.375***	0.757***
	(0.154)	(0.049)
Guessed Score=19	2.650***	0.832***
	(0.220)	(0.026)
Constant	2.001***	0.140***
	(0.091)	(0.030)
N	3,500	3,500
Clusters (Advisers)	700	700
R^2	0.400	0.363

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. The omitted category is Guessed Score=8. All specifications control for worker order (1 to 5). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.5 Worker Behavior and Expected Earnings

Of the 320 workers who completed the second session, 121 were advised to choose job A on either the binary or Likert scale, 198 were advised to choose job B, and 1 worker received no advice (the middle point on the Likert scale). Of the 319 workers who received advice,

308 chose to follow it (96.6%), whereas 11 workers (5 men and 6 women) did not. The high rate of advice-following is understandable, as workers had minimal information beyond the adviser's recommendation. In total, 121 workers were advised to choose job A and 198 were advised to choose job B.

How did advice impact earnings and how do earnings differ by gender and treatment? We can answer this question in two ways. First, we can look at expected earnings based on all the advice generated for a particular worker (across all advisers). The results of this approach are discussed in Table 6 in Section 3.4 in the main text based on part 1 score. Results are similar if we re-do this analysis using part three score instead (Table A.9). Second, we can examine realized earnings in part three. This has the benefit of not requiring us to assume that workers follow advice. However, it comes at the cost of a much smaller sample size (320 workers rather than 8,750 pieces of advice), which reduces power and increases the scope for random adviser draws to impact the results. Across all treatments, women are 6.3 percentage points less likely to choose job A (N=320, p=0.122, one-sided t-test). The gender gap is also insignificant in the baseline treatment (10.8 pp, N=88, p=0.150, one-sided t-test); nudge treatment (12.1 pp, N=62, p=0.166, one-sided t-test); subjective treatment (5.2 pp, N=71, p=0.328, one-sided t-test); and objective treatment (-0.7 pp, N=99, p=0.527, one-sided t-test) individually. Due to the small sample size, these estimates are much noisier than the adviser-level estimates presented in Section 3.4 in the main text. The fact that women also make less ambitious choices in the nudge treatment is driven by female workers randomly receiving much less ambitious advice from their assigned advisers (28% job A) than the average advice sent by all advisers in this treatment (47% job A for women).

Table A.9: Expected Earnings Based on Part 3 Score

	Advice Maximized Expected Earnings		Expected Earnings Based on Advice		Gap in Expected Earnings With vs. Without Advice	
	(1)	(2)	(3)	(4)	(5)	(6)
Female		-0.047***		-0.102***		0.013
		(0.017)		(0.019)		(0.019)
Subjective	-0.036***	-0.048**	-0.034**	-0.048**	-0.040**	-0.050**
	(0.014)	(0.019)	(0.015)	(0.021)	(0.017)	(0.022)
Objective	0.068***	0.066***	0.065***	0.061***	0.060***	0.054***
	(0.008)	(0.011)	(0.008)	(0.011)	(0.008)	(0.011)
Nudge	0.028**	-0.054***	0.029*	-0.061***	0.034**	-0.067***
	(0.014)	(0.019)	(0.015)	(0.021)	(0.017)	(0.022)
Female×Subjective		0.023		0.026		0.020
		(0.026)		(0.028)		(0.029)
Female×Objective		0.002		0.008		0.011
		(0.015)		(0.015)		(0.015)
Female×Nudge		0.162***		0.179***		0.202***
		(0.027)		(0.030)		(0.030)
Constant	0.403***	0.426***	1.287***	1.338***	-0.341***	-0.348***
	(0.014)	(0.016)	(0.015)	(0.018)	(0.016)	(0.018)
N	11,197	11,197	11,197	11,197	10,531	10,531
Clusters (Advisers)	1,750	1,750	1,750	1,750	1,750	1,750
R^2	0.007	0.011	0.006	0.012	0.006	0.016

Notes: Estimated results are generated by OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). The other variables are binary indicators for the respective treatments. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.6 Additional Results

Table A.10 replicates Table 3 while also controlling for worker sociodemographics. We find that less educated participants are less confident in all six dimensions, as are participants who report low income satisfaction. Table A.11 replicates Table 4 while controlling for adviser-level demographics. Older and more risk-averse participants give less ambitious advice. Both the gender gap in self-promotion and the gender gap in advice remain similar when controlling for demographics.

Table A.10: Gender Gap in Self-promotion Controlling for Worker Performance and Sociodemographics

	Likert Performance (1-7)	Performed Well (0-100)	Guessed Rank (rev)	Would Apply for STEM job (0-100)	Would Succeed in STEM (0-100)	Score Guess (0-20)
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.374***	-7.549***	-7.680***	-9.244***	-9.751***	-0.703**
	(0.118)	(2.193)	(2.323)	(2.986)	(2.871)	(0.367)
Age	-0.011*	-0.211*	0.042	-0.142	-0.147	0.003
	(0.006)	(0.110)	(0.116)	(0.150)	(0.144)	(0.018)
Education (ref: Postgraduate Degree)						
High School	-0.550**	-10.379**	-12.024**	-18.324***	-20.118***	-1.173
	(0.237)	(4.393)	(4.652)	(5.982)	(5.751)	(0.735)
Vocational Training	-0.465***	-9.151***	-8.952***	-15.144***	-19.600***	-1.132**
	(0.175)	(3.242)	(3.433)	(4.415)	(4.244)	(0.543)
Undergraduate Degree	-0.248	-6.557**	-4.464	-6.304	-9.493**	-0.427
	(0.153)	(2.845)	(3.013)	(3.874)	(3.725)	(0.476)
Occupation (ref: Employed)						
Studying	0.069	3.015	2.787	1.843	2.088	0.819
	(0.194)	(3.595)	(3.807)	(4.895)	(4.706)	(0.602)
Retired	0.445	4.810	0.741	0.698	3.795	1.094
	(0.299)	(5.539)	(5.866)	(7.542)	(7.251)	(0.927)
Unemployed	0.265	2.983	2.692	8.336	2.806	0.331
	(0.268)	(4.981)	(5.275)	(6.782)	(6.520)	(0.834)
Not in Labor Force	0.179	-0.385	-1.307	2.963	1.728	-1.134
	(0.323)	(5.991)	(6.344)	(8.158)	(7.843)	(1.003)
Income Satisfaction (ref: Live Comfortably)						
Find it difficult	-0.380**	-8.084**	-7.499**	-7.279	-7.189*	-0.620
	(0.175)	(3.243)	(3.434)	(4.416)	(4.246)	(0.543)
Coping	-0.114	-2.580	-2.111	-1.729	-1.401	-0.469
	(0.133)	(2.470)	(2.616)	(3.364)	(3.234)	(0.414)
Constant	4.488***	68.009***	64.596***	83.039***	86.020***	9.370***
	(0.816)	(15.140)	(16.032)	(20.615)	(19.819)	(2.534)
Score FE	✓	✓	✓	✓	✓	✓
N	350	350	350	350	350	350
R^2	0.370	0.394	0.342	0.306	0.339	0.437

Notes: Results are from OLS regressions. *Score FE* (fixed effects) controls for actual performance. Statistical significance of the coefficient estimates for the variable *Female* is based on a one-sided test. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11: Gender Gap in Advice With Adviser Sociodemographics

Advice A Measure:	Likert		Binary	
	(1)	(2)	(3)	(4)
Worker Characteristics:				
Female	-0.387***	0.020	-0.123***	0.004
	(0.047)	(0.038)	(0.016)	(0.013)
Performed Well (0-100)		0.003*		0.000
		(0.002)		(0.001)
Guessed Rank (1-101)		0.006***		0.002***
		(0.002)		(0.001)
Would Apply for STEM job (0-100)		0.004**		0.002***
		(0.002)		(0.001)
Would Succeed in STEM job (0-100)		0.008***		0.002***
		(0.002)		(0.001)
Guessed Score (0-20)		0.114***		0.034***
		(0.009)		(0.003)
Adviser Characteristics:				
Female	0.028	0.090*	0.009	0.029*
	(0.057)	(0.054)	(0.018)	(0.017)
Age	-0.003	-0.005*	-0.001	-0.002**
	(0.003)	(0.003)	(0.001)	(0.001)
Education (ref: Postgraduate Degree)				
High School	0.065	0.162	0.021	0.051
	(0.122)	(0.122)	(0.039)	(0.039)
Vocational Training	0.077	0.131	0.029	0.045*
	(0.088)	(0.086)	(0.027)	(0.026)
Undergraduate Degree	-0.025	-0.001	-0.006	0.000
	(0.074)	(0.072)	(0.023)	(0.022)
Occupation (ref: Employed)				
Studying	0.108	0.019	0.023	-0.004
	(0.099)	(0.100)	(0.032)	(0.033)
Retired	0.159	0.102	0.061	0.042
	(0.158)	(0.148)	(0.050)	(0.046)
Unemployed	-0.064	-0.091	-0.024	-0.034
	(0.101)	(0.086)	(0.035)	(0.030)
Not in Labor Force	-0.108	-0.024	-0.044	-0.016
	(0.103)	(0.107)	(0.037)	(0.035)
Income Satisfaction (ref: Live Comfortably)				
Find it difficult	-0.059	-0.033	-0.056**	-0.048*
	(0.087)	(0.084)	(0.026)	(0.025)
Coping	-0.040	-0.016	-0.017	-0.010
	(0.064)	(0.063)	(0.020)	(0.020)
Self-confidence	-0.005	-0.003	-0.004	-0.003
	(0.017)	(0.016)	(0.005)	(0.005)
Risk Attitude	0.032**	0.036**	0.009**	0.010**
	(0.016)	(0.015)	(0.004)	(0.004)
Constant	2.657***	0.386**	0.425***	-0.272***
	(0.166)	(0.171)	(0.052)	(0.056)
N	3,500	3,500	3,500	3,500
Clusters (Advisers)	700	700	700	700
R^2	0.026	0.441	0.024	0.390

Notes: Results are from OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We next investigate whether our results on advice are robust to using two alternative measures of advice: the performance goal and a version of binary advice that adjusts for adviser confidence. In the latter case, in line with our pre-analysis plan, we combine the binary variable with the confidence variable in the following way: We code advice given with confidence 1-3 as "no advice," confidence 4-8 as "probably A/B," and 9+ as "definitely A/B," and encode this as a 5-point scale from 1-definitely B to 5-definitely A, to approximate the distribution of responses on the Likert scale. Our results, reported in Table A.12, (a) replicate the gender gap in advice in the baseline, (b) replicate the effect of the nudge, and (c) replicate the null effect of the subjective treatment on the gender gap in advice. The main difference from our main results is that we no longer find a significant effect of the objective treatment on the gender gap in advice.

Table A.12: Treatment Differences in Gender Gap in Secondary Advice Measures

			Baseline vs. *Treatment*					
Treatment	Baseline		Subjective		Objective		Nudge	
Advice Measure:	Goal	Confidence	Goal	Confidence	Goal	Confidence	Goal	Confidence
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-1.062***	-0.342***	-1.062***	-0.342***	-1.062***	-0.342***	-1.062***	-0.342***
	(0.134)	(0.046)	(0.134)	(0.046)	(0.134)	(0.046)	(0.134)	(0.046)
Subjective			-0.307	-0.136**				
			(0.191)	(0.054)				
Objective					0.306***	0.118***		
					(0.083)	(0.031)		
Nudge							-0.515***	-0.199***
							(0.187)	(0.055)
Female×Subjective			0.092	0.033				
			(0.204)	(0.069)				
Female×Objective					0.122	0.033		
					(0.105)	(0.040)		
Female×Nudge							1.547***	0.668***
							(0.216)	(0.073)
Constant	9.890***	2.648***	10.186***	2.640***	10.129***	2.662***	10.257***	2.670***
	(0.120)	(0.056)	(0.140)	(0.048)	(0.152)	(0.054)	(0.143)	(0.048)
Gender Gap in *Treatment*								
Coefficient			-0.970***	-0.308***	-0.940***	-0.309***	0.485***	0.327***
Standard error			(0.154)	(0.052)	(0.120)	(0.049)	(0.170)	(0.057)
N	3,500	3,500	6,125	6,125	7,000	7,000	6,125	6,125
Clusters (Advisers)	700	700	1,225	1,225	700	700	1,225	1,225
R^2	0.015	0.018	0.017	0.019	0.018	0.017	0.014	0.020

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). The other variables are binary indicators for the respective treatments. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We next examine whether our results are robust to restricting the analysis to the first worker assessed by each adviser (Table A.13). Advisers received information about workers 2–5 only after completing their evaluation of worker 1, allowing us to treat the first worker as a one-shot observation. Focusing on the first worker also eliminates any potential spillovers from evaluations of previous workers. The results are broadly consistent with our main findings. Female workers continue to receive less ambitious advice than male workers (columns 1 and 2), the nudge treatment reverses the gender gap (columns 7 and 8), and the gender gap in the subjective treatment does not differ from that in the baseline treatment (columns 3 and 4). The main difference is that the gender gap in the objective treatment is no longer statistically different from the baseline gender gap (columns 5 and 6).

Table A.13: Treatment Differences in Gender Gap in Advice, First Worker Only

			Baseline vs. *Treatment*					
Treatment	Baseline		Subjective		Objective		Nudge	
Advice A measure:	Likert	Binary	Likert	Binary	Likert	Binary	Likert	Binary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.289***	-0.106***	-0.289***	-0.106***	-0.289***	-0.106***	-0.289***	-0.106***
	(0.106)	(0.036)	(0.106)	(0.036)	(0.106)	(0.036)	(0.106)	(0.036)
Subjective			-0.124	-0.087**				
			(0.113)	(0.039)				
Objective					0.203***	0.046**		
					(0.062)	(0.022)		
Nudge							-0.288***	-0.094**
							(0.111)	(0.039)
Female×Subjective			0.072	0.056				
			(0.159)	(0.053)				
Female×Objective					-0.017	0.023		
					(0.085)	(0.031)		
Female×Nudge							0.920***	0.287***
							(0.158)	(0.055)
Constant	2.657***	0.400***	2.657***	0.400***	2.657***	0.400***	2.657***	0.400***
	(0.076)	(0.026)	(0.076)	(0.026)	(0.076)	(0.026)	(0.076)	(0.026)
Gender gap in *Treatment*								
Coefficient			-0.217**	-0.050	-0.306***	-0.083***	0.631***	0.182***
Standard error			(0.118)	(0.040)	(0.112)	(0.037)	(0.117)	(0.042)
N	700	700	1,225	1,225	1,400	1,400	1,225	1,225
Clusters (Advisers)	700	700	1,225	1,225	700	700	1,225	1,225
R^2	0.010	0.012	0.010	0.013	0.015	0.013	0.032	0.025

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. *Female* is a binary indicator for the worker's gender (0-male, 1-female). The other variables are binary indicators for the respective treatments (i.e., subjective, objective, and nudge). The sample is restricted to the first worker advised by each adviser. Statistical significance of the coefficient estimates for the variable Female× *Treatment* is based on a one-sided test. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.14 reports the stated confidence of advisers giving binary advice. The key results are that advisers are less confident when advising job A, and they are less confident when giving ambitious advice (job A) to female workers than when giving this advice to male workers. Table A.15 repeats this analysis while controlling for worker guessed score fixed effects. The results are similar, though the gender gap for advice A falls from 0.209 to 0.131 and is no longer significant (p=0.125, two-sided Wald test). Table A.16 shows that the gender gap in advice in the baseline treatment is similar for advisers who anticipate that men are more confident than women. Table A.17 shows that the gender gap in advice is largest when advisers believe that men perform better than women and non significantly different from zero when advisers believe that women perform better than men. Table A.18 shows that the nudge treatment increases the share of advisers who expect men to be more confident relative to the baseline/objective treatment, whereas the Subjective treatment reduces it. The nudge treatment also increases the share of advisers who expect women to outperform men. Table A.19 presents the full distribution of answers to the expected gender performance gap question in each treatment. Finally, Table A.20 looks at heterogeneity by the gender of the adviser. There are no differences in the baseline and subjective treatments. Female advisers appear to give less ambitious advice to female workers relative to male workers in the objective treatment, though this is only significant when looking at the Likert scale. In the nudge treatment, by contrast, female advisers are significantly more likely to advise female workers to go for the ambitious task (job A). This suggests that female advisers internalized the nudge to a greater extent than their male counterparts.

Table A.14: Gender Gap in Adviser Confidence

	Advice		Worker gender		Diff.	p-value
	Job A	Job B	Female	Male		
Confidence in Binary Advice	7.467	7.986			-0.520	0.000
Confidence in Binary Advice			7.786	7.796	-0.010	0.870
Confidence when Advising A			7.345	7.554	-0.210	0.019
Confidence when Advising B			7.989	7.983	0.006	0.935

Notes: The sample includes 700 advisers in the baseline treatment.

Table A.15: Gender Gap in Adviser Confidence With Worker Guessed Score Fixed Effects

	Confidence in Advice		
	(1)	(2)	(3)
Female	-0.065		-0.058
	(0.057)		(0.074)
Job A		-0.572***	-0.542***
		(0.094)	(0.111)
Female×Job A			-0.073
			(0.112)
Constant	9.001***	8.948***	9.009***
	(0.280)	(0.263)	(0.285)
Guessed Score FE	✓	✓	✓
N	3,500	3,500	3,500
Clusters (Advisers)	700	700	700
R^2	0.048	0.064	0.065

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Job A* is a binary indicator that takes the value 1 if the adviser advised the worker to choose Job A and 0 otherwise. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.16: Gender Gap in Advice by Advisers Beliefs about the Gender Gap in Confidence in Baseline

Advice A measure:	Men are more confident				Men and women equally confident				Women more confident			
	Likert		Binary		Likert		Binary		Likert		Binary	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	-0.456***	-0.010	-0.147***	-0.008	-0.261***	0.093	-0.091***	0.016	-0.230	-0.003	-0.018	0.055
	(0.058)	(0.048)	(0.020)	(0.017)	(0.091)	(0.067)	(0.030)	(0.023)	(0.196)	(0.173)	(0.063)	(0.052)
Performed Well (0-100)		0.005**		0.001		-0.002		-0.001		0.009		0.000
		(0.002)		(0.001)		(0.004)		(0.001)		(0.007)		(0.002)
Guessed Rank (1-101)		0.006***		0.002***		0.005		0.001		0.012**		0.006***
		(0.002)		(0.001)		(0.003)		(0.001)		(0.006)		(0.002)
Would Apply for STEM Job (0-100)		0.005**		0.002***		0.002		0.002		-0.004		0.002
		(0.002)		(0.001)		(0.003)		(0.001)		(0.007)		(0.002)
Would Succeed in STEM Job (0-100)		0.007***		0.002***		0.010***		0.003**		0.011		0.001
		(0.002)		(0.001)		(0.004)		(0.001)		(0.007)		(0.002)
Guessed Score (0-20)		0.105***		0.029***		0.148***		0.048***		0.066*		0.028**
		(0.010)		(0.004)		(0.015)		(0.005)		(0.034)		(0.011)
Constant	2.723***	0.461***	0.417***	-0.263***	2.604***	0.268**	0.381***	-0.335***	2.984***	0.964***	0.438***	-0.278***
	(0.072)	(0.093)	(0.025)	(0.031)	(0.106)	(0.118)	(0.035)	(0.041)	(0.246)	(0.334)	(0.084)	(0.081)
N	2,240	2,240	2,240	2,240	1,015	1,015	1,015	1,015	245	245	245	245
Clusters (Advisers)	448	448	448	448	203	203	203	203	49	49	49	49
R^2	0.025	0.442	0.023	0.383	0.008	0.435	0.009	0.384	0.006	0.401	0.000	0.442

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.17: Gender Gap in Advice by Advisers Beliefs about the Gender Gap in Performance in Baseline

Advice A measure:	Men perform better				Men and women perform equally well				Women perform better			
	Likert		Binary		Likert		Binary		Likert		Binary	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Female	-0.458***	-0.010	-0.147***	-0.008	-0.257***	0.093	-0.090***	0.016	-0.237	-0.003	-0.019	0.055
	(0.058)	(0.048)	(0.020)	(0.017)	(0.090)	(0.067)	(0.030)	(0.023)	(0.189)	(0.173)	(0.060)	(0.052)
Performed Well (0-100)		0.005**		0.001		-0.002		-0.001		0.009		0.000
		(0.002)		(0.001)		(0.004)		(0.001)		(0.007)		(0.002)
Guessed Rank (1-101)		0.006***		0.002***		0.005		0.001		0.012**		0.006***
		(0.002)		(0.001)		(0.003)		(0.001)		(0.006)		(0.002)
Would Apply for STEM Job (0-100)		0.005**		0.002***		0.002		0.002		-0.004		0.002
		(0.002)		(0.001)		(0.003)		(0.001)		(0.007)		(0.002)
Would Succeed in STEM Job (0-100)		0.007***		0.002***		0.010***		0.003**		0.011		0.001
		(0.002)		(0.001)		(0.004)		(0.001)		(0.007)		(0.002)
Guessed Score (0-20)		0.105***		0.029***		0.148***		0.048***		0.066*		0.028**
		(0.010)		(0.004)		(0.015)		(0.005)		(0.034)		(0.011)
Constant	2.806***	0.461***	0.455***	-0.263***	2.627***	0.268**	0.398***	-0.335***	2.810***	0.964***	0.422***	-0.278***
	(0.046)	(0.093)	(0.015)	(0.031)	(0.069)	(0.118)	(0.022)	(0.041)	(0.146)	(0.334)	(0.043)	(0.081)
N	670	670	670	670	2,330	2,330	2,330	2,330	500	500	500	500
Clusters (Advisers)	134	134	134	134	466	466	466	466	100	100	100	100
R^2	0.029	0.419	0.031	0.355	0.016	0.441	0.013	0.391	0.010	0.427	0.012	0.413

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.18: Treatment Effects on Adviser's Opinion About Gender Gap in Confidence and Performance

	Share answering that:	
	Men more confident	Women perform better
	(1)	(2)
Ref: Objective		
Nudge	0.225***	0.253***
	(0.026)	(0.023)
Subjective	-0.053**	-0.010
	(0.026)	(0.023)
Constant	0.640***	0.143***
	(0.017)	(0.015)
N	1,750	1,750

Notes: Estimates are based on OLS regressions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.19: Advisers Stated Beliefs About Gender Gaps (Final Questionnaire)

	% of sample in each treatment		
	Objective ($N = 700$)	Nudge ($N = 525$)	Subjective ($N = 525$)
Gender performance gap			
Women perform better / a lot better	4.71	14.10	5.33
Women perform slightly better	9.57	25.52	8.00
Women and men perform equally well	66.57	46.29	67.62
Men perform slightly better	14.71	11.24	14.86
Men perform better / a lot better	4.43	2.86	4.19

Notes: Since the question was asked at the end of the experiment and we use a within-subject design for the baseline and objective treatments, we do not have advisers' stated beliefs for the baseline treatment.

Table A.20: Heterogeneity in Advice by Adviser Gender

Treatment	Baseline		Subjective		Objective		Nudge	
Advice A measure:	Likert	Binary	Likert	Binary	Likert	Binary	Likert	Binary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WFemale	0.043	0.004	0.004	-0.002	0.088	0.021	0.509***	0.161***
	(0.057)	(0.019)	(0.066)	(0.023)	(0.063)	(0.022)	(0.067)	(0.023)
Afemale	0.051	0.012	0.069	0.020	0.169**	0.039	-0.030	-0.024
	(0.065)	(0.022)	(0.080)	(0.026)	(0.075)	(0.025)	(0.079)	(0.026)
Wfemale×**A**female	-0.040	0.000	0.031	0.008	-0.187**	-0.038	0.336***	0.112***
	(0.076)	(0.026)	(0.095)	(0.032)	(0.085)	(0.029)	(0.098)	(0.034)
Performed Well (0-100)	0.003*	0.000	0.017***	0.005***	0.003	0.001	0.005**	0.002*
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Guessed Rank (1-101)	0.006***	0.002***	0.002	0.001	0.007***	0.002***	0.008***	0.002***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Would Apply for STEM Job (0-100)	0.004**	0.002***	-0.004*	-0.001*	-0.000	0.001	0.002	0.001
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Would Succeed in STEM Job (0-100)	0.008***	0.002***	0.011***	0.004***	0.000	-0.000	0.006***	0.002**
	(0.002)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Guessed Score (0-20)	0.113***	0.034***	0.029***	0.007*	0.142***	0.044***	0.104***	0.034***
	(0.009)	(0.003)	(0.010)	(0.004)	(0.010)	(0.003)	(0.011)	(0.004)
Constant	0.418***	-0.291***	0.983***	-0.152***	0.707***	-0.196***	0.407***	-0.318***
	(0.081)	(0.027)	(0.090)	(0.030)	(0.091)	(0.030)	(0.089)	(0.031)
N	3,500	3,500	2,625	2,625	3,500	3,500	2,625	2,625
Clusters (Advisers)	700	700	525	525	700	700	525	525
R^2	0.432	0.383	0.271	0.238	0.280	0.241	0.382	0.330

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). **W** means worker; **A** means adviser. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.7 Additional Treatment: Experienced Advisers and Vignettes

A.7.1 Experimental Design

We conducted our additional treatment with experienced advisers and vignettes on Prolific in April 2026. We recruited "experienced" advisers who were required to be currently employed and either (i) have at least three years of management experience, supervise at least two subordinates, and regularly interact with employees, or (ii) work as teachers, either in educational institutions or workplace settings. Among these advisers, 434 report managerial or supervisory experience and 326 report teaching experience. Advice provision is frequent in both groups: 84 percent of those with management experience and 87 percent of teachers report giving advice at least weekly (see Tables A.21 and A.23). Tables A.21–A.23 describe the characteristics and professional backgrounds of the experienced-adviser sample in greater detail.

Experienced advisers completed the same task as advisers in the baseline treatment, with one modification: they advised six workers rather than five. Five workers were drawn from the original baseline sample, allowing us to hold worker characteristics fixed and rule out compositional differences as an explanation for any treatment differences. The sixth worker was drawn from a new wave of data collection, comprising 105 additional workers, and was used to incentivize advisers' decisions. Our main analysis focuses on the data of the first five workers in order to maximize comparability with the baseline treatment.

After completing this initial incentivized advice task, our advisers took part in a vignette task to elicit advice in a more natural context. Table A.24 presents the exact vignettes we used in greater detail. We used a total of 24 vignettes that varied the scenario (one of six options) as well as the gender (male or female) and confidence (modest vs confident) of the advisee. Each adviser went through twelve vignettes in total.

A.7.2 Results

We discuss the main results in section 4.2 in the main text. Overall, we find similar results among experienced advisers as in our main sample (Tables A.25 and A.26) and find similar results in our vignette task as in our main incentivized advice task (Tables A.27 and A.28). A series of robustness exercises confirms the robustness of these findings. Tables A.29, A.30,

and A.31 include the sixth worker drawn from the new worker sample and yield results comparable to the main specifications. Table A.32 documents significant gender gaps in self-promotion in both the new worker sample ($N = 105$) and the full worker sample ($N = 455$). Tables A.33 and A.34 further show that these gender gaps persist even after controlling for performance, although estimates in the new worker sample are generally imprecise because of limited statistical power.

Tables A.35 and A.36 restrict the sample to the first time each adviser encountered a given vignette scenario, while Tables A.37, A.38, and A.39 restrict the analysis to the first worker evaluated by each adviser. Both sets of exercises produce results that are highly consistent with the main findings. Finally, Tables A.40 and A.41 show that the results are robust to alternative measures of advice. Table A.42 indicates that experienced advisers do not provide higher-quality advice; if anything, there is suggestive evidence that their advice quality is slightly lower, with no meaningful differences by advisee gender.

Table A.21: Experienced Advisers: Experience in Managing / Supervising Subordinates and Teaching

	Manager / Superviser		
Teacher	Yes	No	Total
Yes	257	69	326
No	177	22	199
Total	434	91	525

Notes: The table reports the overlap between teaching experience and managerial/supervisory experience among experienced advisers. Categories are not mutually exclusive. Advisers were classified as managers/supervisors if they reported at least three years of managerial experience and supervision of at least two subordinates.

Table A.22: Experienced Advisers: Manager / Supervisor Experience

	N	Percent
Years of experience		
Less than 6 months	3	0.69
6-12 months	6	1.38
1–2 years	21	4.84
3–5 years	90	20.74
6–10 years	115	26.50
More than 10 years	199	45.85
Number of subordinates		
1	11	2.53
2-3	104	23.96
4-6	121	27.88
7-10	85	19.59
11-20	68	15.67
More than 20	45	10.37
Hours spent managing / supervising others per week		
Less than 1 hour	16	3.69
2-3 hours	48	11.06
4-7 hours	83	19.12
8-14 hours	94	21.66
15-25 hours	71	16.36
More than 25 hours	122	28.11
Frequency of advising others		
Rarely	9	2.07
Monthly or less	59	13.59
Weekly	135	31.11
Several times per week	135	31.11
Daily	96	22.12

Notes: The table reports descriptive statistics for advisers who indicated that they had prior managerial or supervisory experience. The table shows the distribution of respondents by years of managerial experience, number of subordinates supervised, hours spent managing or supervising others per week, and frequency of providing advice to others. Percentages are calculated using the sample of experienced advisers.

Table A.23: Experienced Advisers: Teacher Experience

	N	Percent
Teaching experience		
School	77	23.62
Higher Education	57	17.48
Workplace	224	68.71
Other	14	4.29
Years of experience		
Less than 6 months	3	0.92
6-12 months	5	1.53
1-2 years	18	5.52
3-5 years	62	19.02
6-10 years	77	23.62
More than 10 years	161	49.39
Typical number of students		
1-2	34	10.43
3-10	104	31.90
11-30	84	25.77
31-100	55	16.87
More than 100	49	15.03
Hours spent teaching per week		
Less than 1 hour	19	5.83
2-3 hours	81	24.85
4-7 hours	73	22.39
8-14 hours	54	16.56
15-25 hours	40	12.27
More than 25 hours	59	18.10
Frequency of advising others		
Rarely	8	2.45
Monthly or less	33	10.12
Weekly	100	30.67
Several times per week	90	27.61
Daily	95	29.14

Notes: The table reports descriptive statistics for advisers who indicated that they had prior teaching experience. The table shows the distribution of respondents by teaching setting, years of teaching experience, typical number of students taught, hours spent teaching per week, and frequency of advising others. Percentages are calculated using the sample of experienced teacher advisers and may not sum to 100 due to rounding.

Table A.24: Vignettes

Scenario	Confidence		Gender	
	High confidence	Low confidence	Female	Male
Job risk	[name] has just received two job offers: one at a large company with a stable salary, and another at a start-up with more uncertainty but potentially higher rewards.		Alice	Andrew
	The start-up sounds exciting, and [name] feels confident in [pronoun] abilities and optimistic about succeeding in a fast-paced environment.	Although the start-up sounds exciting, [name] feels unsure about its future and doubts success in such an unpredictable environment.		
STEM vs. non-STEM	[name] is a high school student strong in both mathematics and literature. Teachers encourage [pronoun] to pursue engineering due to strong job prospects, but [name] prefers writing and considers English or philosophy.		Sarah	Simon
	[pronoun] feels confident in math and science skills and believes [pronoun] would succeed in a demanding STEM program.	Although competent, [name] doubts [pronoun] ability to succeed in STEM and worries about keeping up.		
Promotion	[name] works in a medium-sized company. A managerial position has opened up. The promotion offers higher pay and responsibility but also longer hours and pressure.		Natalie	Nathan
	[name] feels confident in leadership abilities and expects to manage challenges successfully.	[name] is unsure about managing a team and worries about performance in a leadership role.		
Relocation	[name] has been offered a job in another city with 25% higher pay but requiring relocation away from friends and family.		Maria	Mark
	[name] feels able to adapt and build a fulfilling life in the new city.	[name] is uncertain about moving and worries about adjusting and building a new social circle.		
Starting own business	[name] has worked for several years in a stable corporate job. Recently, [pronoun] has been considering starting [pronoun] own business based on an idea [pronoun]'s passionate about. Starting the business would mean quitting [pronoun] job and taking on significant financial risk.		Josephine	Joseph
	[name] feels quite confident in [pronoun] abilities and is convinced that [pronoun] skills and motivation would help [pronoun] make the business succeed.	Although [pronoun] likes the idea, [name] is not fully convinced that [pronoun] has the skills and perseverance to make the business succeed.		

Table A.24 continued

Scenario	Confidence		Gender	
	High confidence	Low confidence	Female	Male
Work-life balance	[name] works at a growing company where [pronoun] is well respected. Recently, [pronoun] manager offered [pronoun] a chance to take on an important project that could lead to a promotion. However, the project would require frequent overtime and weekend work for several months.		Sophie	James
	[name] feels quite confident in [pronoun] ability to manage the workload effectively while still maintaining a healthy balance between [pronoun] job and personal life.	[name] is not 100% sure that [pronoun] could handle the additional workload without burning out or letting [pronoun] personal life suffer.		

Notes: In each vignette, advisers report the option they would recommend to the decision-maker.

Table A.25: Experienced Advisers: Drivers of the Gender Gap in Advice

Advice type (1-A)	Likert		Binary	
	(1)	(2)	(3)	(4)
Female	-0.415***	0.016	-0.133***	-0.002
	(0.056)	(0.041)	(0.018)	(0.014)
Performed Well (0-100)		0.001		-0.000
		(0.002)		(0.001)
Guessed Rank (1-101)		0.012***		0.004***
		(0.002)		(0.001)
Would Apply for STEM Job (0-100)		0.009***		0.003***
		(0.002)		(0.001)
Would Succeed in STEM Job (0-100)		0.003		0.001
		(0.002)		(0.001)
Guessed Score (0-20)		0.112***		0.035***
		(0.010)		(0.003)
Constant	2.563***	0.152**	0.376***	-0.360***
	(0.069)	(0.075)	(0.022)	(0.023)
N	2,625	2,625	2,625	2,625
Clusters (Advisers)	525	525	525	525
R^2	0.024	0.476	0.023	0.434

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.26: Baseline vs. Experienced Advisers: Drivers of the Gender Gap in Advice

Advice type (1-A)	Likert			Binary		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.382***	0.033	0.022	-0.121***	0.007	0.004
	(0.048)	(0.038)	(0.038)	(0.016)	(0.013)	(0.013)
Experience	-0.108*	-0.107**	-0.233**	-0.030	-0.030*	-0.065**
	(0.056)	(0.046)	(0.093)	(0.018)	(0.015)	(0.028)
Female×Experience	-0.033	-0.031	-0.006	-0.012	-0.011	-0.006
	(0.074)	(0.055)	(0.056)	(0.024)	(0.018)	(0.019)
Performed Well (0-100)[1]		0.002	0.003*		-0.000	0.000
		(0.001)	(0.002)		(0.000)	(0.001)
Guessed Rank (1-101)[2]		0.009***	0.006***		0.003***	0.002***
		(0.001)	(0.002)		(0.000)	(0.001)
Would Apply for STEM Job (0-100)[3]		0.006***	0.004**		0.002***	0.002***
		(0.001)	(0.002)		(0.000)	(0.001)
Would Succeed in STEM Job (0-100)[4]		0.006***	0.008***		0.002***	0.002***
		(0.002)	(0.002)		(0.000)	(0.001)
Guessed Score (0-20)[5]		0.113***	0.113***		0.035***	0.034***
		(0.007)	(0.008)		(0.002)	(0.003)
Experience×[1]			-0.002			-0.001
			(0.003)			(0.001)
Experience×[2]			0.005**			0.001*
			(0.002)			(0.001)
Experience×[3]			0.005*			0.001
			(0.003)			(0.001)
Experience×[4]			-0.005*			-0.002
			(0.003)			(0.001)
Experience×[5]			-0.001			0.001
			(0.013)			(0.005)
Constant	2.690***	0.367***	0.420***	0.407***	-0.304***	-0.289***
	(0.051)	(0.059)	(0.069)	(0.017)	(0.019)	(0.022)
N	6,125	6,125	6,125	6,125	6,125	6,125
Clusters (Advisers)	1,225	1,225	1,225	1,225	1,225	1,225
R^2	0.023	0.451	0.452	0.021	0.404	0.405

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Experience* is a binary indicator for the sample of experienced advisers (managers and teachers) (the reference group is the general sample of advisers). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.27: Experienced Advisers: Gender Gap in Advice in Vignettes

Advice Type (1-A)		By Vignette Scenario						
		Job risk	Field of study	Apply for promotion	Take job in other city	Start own business	Work-life balance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	-0.001	0.001	-0.004	-0.003	-0.014	0.006	0.004	-0.002
	(0.006)	(0.013)	(0.013)	(0.014)	(0.011)	(0.016)	(0.015)	(0.009)
Confident	0.477***	0.637***	0.326***	0.453***	0.457***	0.454***	0.536***	0.477***
	(0.015)	(0.030)	(0.037)	(0.031)	(0.033)	(0.035)	(0.030)	(0.016)
Female×Confident								0.001
								(0.011)
Constant	-0.327***	-0.512***	-0.202***	0.035	-0.030	-0.400***	-0.147**	0.149***
	(0.032)	(0.061)	(0.068)	(0.076)	(0.073)	(0.067)	(0.070)	(0.023)
N	6,300	1,050	1,050	1,050	1,050	1,050	1,050	6,300
Clusters (Advisers)	525	525	525	525	525	525	525	525
R^2	0.320	0.412	0.123	0.259	0.257	0.226	0.334	0.320

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a vignette order (1 to 12). Specifications (1) and (8) control for vignette scenario (6 in total). *Female* is a binary indicator for the vignette's gender (0-male, 1-female). *Confident* is a binary indicator for the level confidence displayed in the vignette (0-low confidence, 1-high confidence). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.28: Experienced Advisers: Advice in Vignettes vs. Incentivized Task

Advice Type (1-A)	(1)	(2)
Female	-0.133***	-0.002
	(0.018)	(0.014)
Vignette	0.118***	0.135***
	(0.026)	(0.038)
Female×Vignette	0.132***	0.000
	(0.019)	(0.015)
Guessed Score (0-20)		0.035***
		(0.003)
Performed Well (0-100)		-0.000
		(0.001)
Guessed Rank (1-101)		0.004***
		(0.001)
Would Apply for STEM Job (0-100)		0.003***
		(0.001)
Would Succeed in STEM Job (0-100)		0.001
		(0.001)
Vignette Confidence		0.477***
		(0.016)
Constant	0.406***	-0.320***
	(0.014)	(0.019)
N	8,925	8,925
Clusters (Advisers)	525	525
R^2	0.043	0.312

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for vignette scenario (1 to 12). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Vignette* is a binary indicator for whether advice was given in the vignette or incentivized task (0-incentivized, 1-vignette). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. *Vignette Confidence* is a binary indicator for the level confidence displayed in the vignette (0-low confidence, 1-high confidence). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.29: Experienced Advisers: Drivers of the Gender Gap in Advice With Full Worker Sample (6 Workers per Adviser)

Advice type (1-A)	Likert		Binary	
	(1)	(2)	(3)	(4)
Female	-0.415***	0.021	-0.132***	-0.001
	(0.050)	(0.037)	(0.016)	(0.012)
Performed Well (0-100)		0.002		-0.000
		(0.002)		(0.001)
Guessed Rank (1-101)		0.011***		0.003***
		(0.002)		(0.001)
Would Apply for STEM Job (0-100)		0.008***		0.003***
		(0.002)		(0.001)
Would Succeed in STEM Job (0-100)		0.005**		0.001**
		(0.002)		(0.001)
Guessed Score (0-20)		0.106***		0.034***
		(0.010)		(0.003)
Constant	2.563***	0.135*	0.376***	-0.358***
	(0.068)	(0.072)	(0.022)	(0.022)
N	3,150	3,150	3,150	3,150
Clusters (Advisers)	525	525	525	525
R^2	0.027	0.501	0.025	0.452

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 6). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.30: Baseline vs. Experienced Advisers: Drivers of the Gender Gap in Advice With Full Worker Sample (6 Workers per Adviser in Experienced Adviser Treatment)

Advice type (1-A)	Likert			Binary		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.382***	0.041	0.022	-0.121***	0.008	0.004
	(0.048)	(0.038)	(0.038)	(0.016)	(0.013)	(0.013)
Experience	-0.107*	-0.103**	-0.250***	-0.030*	-0.029*	-0.063**
	(0.055)	(0.046)	(0.090)	(0.018)	(0.015)	(0.027)
Female×Experience	-0.033	-0.039	-0.001	-0.011	-0.013	-0.006
	(0.069)	(0.052)	(0.053)	(0.023)	(0.018)	(0.018)
Performed Well (0-100)[1]		0.003*	0.003*		0.000	0.000
		(0.001)	(0.002)		(0.000)	(0.001)
Guessed Rank (1-101)[2]		0.008***	0.006***		0.003***	0.002***
		(0.001)	(0.002)		(0.000)	(0.001)
Would Apply for STEM Job (0-100)[3]		0.006***	0.004**		0.002***	0.002***
		(0.001)	(0.002)		(0.000)	(0.001)
Would Succeed in STEM Job (0-100)[4]		0.007***	0.008***		0.002***	0.002***
		(0.001)	(0.002)		(0.000)	(0.001)
Guessed Score (0-20)[5]		0.110***	0.113***		0.034***	0.034***
		(0.006)	(0.008)		(0.002)	(0.003)
Experience×[1]			-0.001			-0.000
			(0.003)			(0.001)
Experience×[2]			0.005**			0.001
			(0.002)			(0.001)
Experience×[3]			0.004*			0.001
			(0.003)			(0.001)
Experience×[4]			-0.003			-0.001
			(0.003)			(0.001)
Experience×[5]			-0.008			-0.001
			(0.013)			(0.004)
Constant	2.690***	0.349***	0.420***	0.407***	-0.306***	-0.289***
	(0.051)	(0.058)	(0.069)	(0.017)	(0.019)	(0.022)
N	6,650	6,650	6,650	6,650	6,650	6,650
Clusters (Advisers)	1,225	1,225	1,225	1,225	1,225	1,225
R^2	0.023	0.464	0.465	0.021	0.415	0.415

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 6). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Experience* is a binary indicator for the sample of experienced advisers (managers and teachers) (the reference group is the general sample of advisers). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.31: Experienced Advisers: Advice in Vignettes vs. Incentivized Task With Full Worker Sample (6 Workers per Adviser in Experienced Adviser Treatment)

Advice type (1-A)	(1)	(2)
Female	-0.132***	-0.001
	(0.016)	(0.012)
Vignette	0.106***	0.127***
	(0.026)	(0.037)
Female×Vignette	0.132***	-0.000
	(0.017)	(0.014)
Guessed Score (0-20)		0.034***
		(0.003)
Performed Well (0-100)		-0.000
		(0.001)
Guessed Rank (1-101)		0.003***
		(0.001)
Would Apply for STEM Job (0-100)		0.003***
		(0.001)
Would Succeed in STEM Job (0-100)		0.001*
		(0.001)
Vignette Confidence		0.477***
		(0.016)
Constant	0.418***	-0.311***
	(0.013)	(0.018)
N	9,450	9,450
Clusters (Advisers)	525	525
R^2	0.042	0.324

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for vignette scenario (1 to 12). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Vignette* is a binary indicator for whether advice was given in the vignette or incentivized task (0-incentivized, 1-vignette). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. *Vignette Confidence* is a binary indicator for the level confidence displayed in the vignette (0-low confidence, 1-high confidence). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.32: Gender Gap in Self-promotion in Additional and Full Worker Sample

	Women		Men		Diff.	p-value
	Mean	SD	Mean	SD		
Additional Worker Sample ($N = 105$)						
Likert Performance (1-7)	3.623	1.333	4.173	1.216	-0.550	0.015
Performed Well (0-100)	41.245	25.122	49.731	24.658	-8.485	0.042
Would Apply for STEM Job (0-100)	34.340	31.048	42.365	32.918	-8.026	0.101
Would Succeed in STEM Job (0-100)	34.792	30.087	44.731	32.022	-9.938	0.052
Guessed Rank (reversed, 1-101)	42.679	25.932	53.135	25.207	-10.455	0.019
Guessed Score (0-20)	9.698	4.668	11.115	3.833	-1.417	0.046
Full Worker Sample ($N = 455$)						
Likert Performance (1-7)	3.592	1.292	4.119	1.190	-0.527	0.000
Performed Well (0-100)	39.627	23.626	49.841	23.565	-10.214	0.000
Would Apply for STEM Job (0-100)	34.118	29.702	45.026	30.825	-10.908	0.000
Would Succeed in STEM Job (0-100)	34.825	29.114	46.824	30.143	-11.999	0.000
Guessed Rank (reversed, 1-101)	41.057	24.473	51.916	23.625	-10.859	0.000
Guessed Score (0-20)	9.500	4.262	10.890	3.985	-1.390	0.000

Notes: p-values come from pre-registered one-sided t-tests. The *Guessed Rank* variable is reversed, such that a higher value corresponds to a higher rank. The sample includes a total of 175 women and 175 men.

Table A.33: Gender Gap in Self-promotion Controlling for Performance in Additional Worker Sample

	Likert Performance (1-7)	Performed Well (0-100)	Would Apply for STEM Job (0-100)	Would Succeed in STEM Job (0-100)	Guessed Rank (1-101)	Guessed Score (0-20)
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.394**	-4.893	-4.263	-6.290	-7.450*	-0.876
	(0.232)	(4.468)	(5.847)	(5.402)	(4.631)	(0.725)
Constant	3.697***	45.946***	48.631**	54.145***	48.225***	9.438***
	(0.804)	(15.459)	(20.229)	(18.692)	(16.023)	(2.507)
Score FE	✓	✓	✓	✓	✓	✓
N	105	105	105	105	105	105
R^2	0.374	0.380	0.349	0.416	0.377	0.446

Notes: Estimates are based on OLS regressions. *Score FE* denotes score fixed effects that control for workers' actual performance. *Female* is a binary indicator for the worker's gender (0-male, 1-female). Statistical significance of the coefficient estimate for the binary indicator *Female* is based on a one-sided test. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.34: Gender Gap in Self-promotion Controlling for Performance in Full Worker Sample

	Likert Performance (1-7)	Performed Well (0-100)	Would Apply for STEM Job (0-100)	Would Succeed in STEM Job (0-100)	Guessed Rank (1-101)	Guessed Score (0-20)
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.371***	-6.948***	-7.761***	-8.575***	-7.773***	-0.795***
	(0.103)	(1.941)	(2.611)	(2.531)	(2.026)	(0.317)
Constant	3.371***	46.448***	56.261***	56.575***	52.273***	8.295***
	(0.764)	(14.384)	(19.344)	(18.750)	(15.010)	(2.351)
Score FE	✓	✓	✓	✓	✓	✓
N	455	455	455	455	455	455
R^2	0.315	0.329	0.252	0.273	0.300	0.404

Notes: Estimates are based on OLS regressions. *Score FE* denotes score fixed effects that control for workers' actual performance. *Female* is a binary indicator for the worker's gender (0-male, 1-female). Statistical significance of the coefficient estimate for the binary indicator *Female* is based on a one-sided test. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.35: Experienced Advisers: Gender Gap in Advice in Vignettes Including First 6 Vignettes Only

Advice type (1-A)		By vignette scenario						
		Job risk	Field of study	Apply for promotion	Take job in other city	Start own business	Work-life balance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female	0.018	0.034	0.006	-0.001	0.001	0.035	0.021	0.013
	(0.015)	(0.035)	(0.040)	(0.034)	(0.036)	(0.039)	(0.034)	(0.024)
Confident	0.467***	0.618***	0.324***	0.434***	0.450***	0.422***	0.548***	0.462***
	(0.016)	(0.034)	(0.039)	(0.035)	(0.035)	(0.039)	(0.034)	(0.023)
Female×Confident								0.009
								(0.031)
Constant	-0.332***	-0.532***	-0.214**	0.059	-0.040	-0.396***	-0.194**	0.155***
	(0.039)	(0.081)	(0.089)	(0.088)	(0.093)	(0.085)	(0.086)	(0.026)
N	3,150	525	525	525	525	525	525	3,150
Clusters (Advisers)	525	525	525	525	525	525	525	525
R^2	0.311	0.388	0.126	0.251	0.241	0.207	0.340	0.311

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a vignette order (1 to 6). Specifications (1) and (8) control for vignette scenario (6 in total). *Female* is a binary indicator for the vignette's gender (0-male, 1-female). *Confident* is a binary indicator for the level confidence displayed in the vignette (0-low confidence, 1-high confidence). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.36: Experienced Advisers: Advice in Vignettes vs. Incentivized Task Including First 6 Vignettes Only

Advice type (1-A)	(1)	(2)
Female	-0.133***	-0.002
	(0.018)	(0.014)
Vignette	0.092***	0.115**
	(0.035)	(0.045)
Female×Vignette	0.158***	0.023
	(0.026)	(0.021)
Guessed Score (0-20)		0.035***
		(0.003)
Performed Well (0-100)		-0.000
		(0.001)
Guessed Rank (1-101)		0.004***
		(0.001)
Would Apply for STEM Job (0-100)		0.003***
		(0.001)
Would Succeed in STEM Job (0-100)		0.001
		(0.001)
Vignette Confidence		0.466***
		(0.017)
Constant	0.406***	-0.320***
	(0.014)	(0.019)
N	5,775	5,775
Clusters (Advisers)	525	525
R^2	0.053	0.341

Notes: Estimates are based on OLS regressions. with standard errors clustered at adviser level. All specifications control for vignette scenario (1 to 6). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Vignette* is a binary indicator for whether advice was given in the vignette or incentivized task (0-incentivized, 1-vignette). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. *Vignette Confidence* is a binary indicator for the level confidence displayed in the vignette (0-low confidence, 1-high confidence). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.37: Experienced Advisers: Drivers of the Gender Gap in Advice Among, 1st Worker Only

Advice type (1-A)	Likert		Binary	
	(1)	(2)	(3)	(4)
Female	-0.288**	0.120	-0.064	0.066**
	(0.126)	(0.098)	(0.040)	(0.031)
Performed Well (0-100)		0.000		-0.001
		(0.005)		(0.001)
Guessed Rank (1-101)		0.011***		0.005***
		(0.004)		(0.001)
Would Apply for STEM Job (0-100)		0.008		0.003*
		(0.006)		(0.001)
Would Succeed in STEM Job (0-100)		0.005		0.001
		(0.006)		(0.002)
Guessed Score (0-20)		0.098***		0.031***
		(0.022)		(0.007)
Constant	2.500***	0.250*	0.342***	-0.375***
	(0.092)	(0.143)	(0.029)	(0.041)
N	525	525	525	525
Clusters (Advisers)	525	525	525	525
R^2	0.010	0.434	0.005	0.424

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.38: Baseline vs. Experienced Advisers: Drivers of the Gender Gap in Advice Among, 1st Worker Only

Advice type (1-A)	Likert			Binary		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.289***	0.077	0.042	-0.106***	0.013	0.003
	(0.106)	(0.092)	(0.092)	(0.036)	(0.031)	(0.032)
Experience	-0.157	-0.172*	-0.747***	-0.058	-0.063**	-0.239***
	(0.119)	(0.094)	(0.204)	(0.039)	(0.031)	(0.060)
Female×Experience	0.001	-0.005	0.078	0.042	0.040	0.062
	(0.165)	(0.132)	(0.135)	(0.054)	(0.044)	(0.044)
Performed Well (0-100)[1]		0.003	0.005		0.001	0.002
		(0.003)	(0.005)		(0.001)	(0.002)
Guessed Rank (1-101)[2]		0.007**	0.004		0.003***	0.001
		(0.003)	(0.004)		(0.001)	(0.001)
Would Apply for STEM Job (0-100)[3]		0.007**	0.006		0.002**	0.002
		(0.003)	(0.004)		(0.001)	(0.001)
Would Succeed in STEM Job (0-100)[4]		0.006*	0.007*		0.002*	0.002
		(0.004)	(0.004)		(0.001)	(0.001)
Guessed Score (0-20)[5]		0.073***	0.057***		0.022***	0.017**
		(0.015)	(0.021)		(0.005)	(0.007)
Experience×[1]			-0.005			-0.003
			(0.007)			(0.002)
Experience×[2]			0.007			0.003**
			(0.005)			(0.002)
Experience×[3]			0.002			0.001
			(0.007)			(0.002)
Experience×[4]			-0.002			-0.001
			(0.008)			(0.002)
Experience×[5]			0.041			0.014
			(0.031)			(0.010)
Constant	2.657***	0.758***	0.996***	0.400***	-0.209***	-0.136***
	(0.076)	(0.114)	(0.145)	(0.026)	(0.035)	(0.044)
N	1,225	1,225	1,225	1,225	1,225	1,225
Clusters (Advisers)	1,225	1,225	1,225	1,225	1,225	1,225
R^2	0.013	0.349	0.355	0.011	0.331	0.337

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Experience* is a binary indicator for the sample of experienced advisers (managers and teachers) (the reference group is the general sample of advisers). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.39: Experienced Advisers: Advice in Vignettes vs. Incentivized Task, 1st Worker / Vignette Only

Advice Type (1-A)	(1)	(2)
Female	-0.064	0.066**
	(0.040)	(0.031)
Vignette	0.166**	0.043
	(0.075)	(0.094)
Female×Vignette	0.067	-0.047
	(0.062)	(0.049)
Guessed Score (0-20)		0.031***
		(0.007)
Performed Well (0-100)		-0.001
		(0.001)
Guessed Rank (1-101)		0.005***
		(0.001)
Would Apply for STEM Job (0-100)		0.003*
		(0.001)
Would Succeed in STEM Job (0-100)		0.001
		(0.002)
Vignette Confidence		0.533***
		(0.037)
Constant	0.342***	-0.375***
	(0.029)	(0.041)
N	1,050	1,050
Clusters (Advisers)	525	525
R^2	0.044	0.375

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Vignette* is a binary indicator for whether advice was given in the vignette or incentivized task (0-incentivized, 1-vignette). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. *Vignette Confidence* is a binary indicator for the level confidence displayed in the vignette (0-low confidence, 1-high confidence). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.40: Experienced Advisers: Drivers of the Gender Gap in Secondary Advice Measures

Advice type (1-A)	Goal		Confidence	
	(1)	(2)	(3)	(4)
Female	-1.210***	-0.032	-0.387***	0.023
	(0.167)	(0.126)	(0.054)	(0.038)
Performed Well (0-100)		0.021***		0.002
		(0.006)		(0.002)
Guessed Rank (1-101)		0.022***		0.011***
		(0.005)		(0.002)
Would Apply for STEM Job (0-100)		0.000		0.008***
		(0.006)		(0.002)
Would Succeed in STEM Job (0-100)		0.011*		0.002
		(0.006)		(0.002)
Guessed Score (0-20)		0.406***		0.102***
		(0.029)		(0.009)
Constant	9.606***	2.413***	2.530***	0.250***
	(0.186)	(0.240)	(0.066)	(0.068)
N	2,625	2,625	2,625	2,625
Clusters (Advisers)	525	525	525	525
R^2	0.023	0.472	0.025	0.481

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.41: Baseline vs. Experienced Advisers: Drivers of the Gender Gap in Secondary Advice Measures

Advice type (1-A)	Goal			Confidence		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-1.062***	0.089	0.073	-0.342***	0.047	0.032
	(0.134)	(0.106)	(0.106)	(0.046)	(0.037)	(0.037)
Experience	-0.435**	-0.431***	-0.616**	-0.103*	-0.102**	-0.274***
	(0.185)	(0.153)	(0.291)	(0.054)	(0.045)	(0.088)
Female×Experience	-0.147	-0.141	-0.103	-0.045	-0.044	-0.009
	(0.214)	(0.158)	(0.165)	(0.070)	(0.051)	(0.053)
Performed Well (0-100)[1]		0.021***	0.021***		0.002*	0.002
		(0.004)	(0.006)		(0.001)	(0.002)
Guessed Rank (1-101)[2]		0.018***	0.014***		0.008***	0.006***
		(0.003)	(0.005)		(0.001)	(0.001)
Would Apply for STEM Job (0-100)[3		0.006*	0.011**		0.006***	0.004***
		(0.004)	(0.005)		(0.001)	(0.002)
Would Succeed in STEM Job (0-100)[4]		0.007*	0.005		0.005***	0.007***
		(0.004)	(0.005)		(0.001)	(0.002)
Guessed Score (0-20)[5]		0.409***	0.412***		0.104***	0.105***
		(0.020)	(0.026)		(0.006)	(0.008)
Experience×[1]			0.001			0.000
			(0.008)			(0.003)
Experience×[2]			0.008			0.004**
			(0.007)			(0.002)
Experience×[3]			-0.011			0.004
			(0.008)			(0.002)
Experience×[4]			0.006			-0.004
			(0.008)			(0.003)
Experience×[5]			-0.007			-0.002
			(0.039)			(0.013)
Constant	10.114***	3.066***	3.145***	2.642***	0.474***	0.547***
	(0.142)	(0.174)	(0.205)	(0.049)	(0.058)	(0.069)
N	6,125	6,125	6,125	6,125	6,125	6,125
Clusters (Advisers)	1,225	1,225	1,225	1,225	1,225	1,225
R^2	0.022	0.447	0.448	0.023	0.440	0.441

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Experience* is a binary indicator for the sample of experienced advisers (managers and teachers) (the reference group is the general sample of advisers). *Performed Well* indicates how much the worker agreed with having performed well in part 1 (from 0 to 100%). *Would Apply for STEM Job* expresses willingness to apply for jobs that utilize similar skills as the STEM quiz (from 0 to 100%). *Would Succeed in STEM Job* expresses agreement with expecting to perform well on jobs featuring similar tasks (from 0 to 100%). *Guessed Rank* indicates the expected rank compared with 100 other individuals, reverse coded such that 101 is the highest rank. *Guessed Score* is the worker's guessed score in part 1. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.42: Baseline vs. Experienced Advisers: Expected Earnings Based on Part 1 Score

	Advice Maximized Expected Earnings		Expected Earnings Based on Advice		Gap in Expected Earnings With vs. Without Advice	
	(1)	(2)	(3)	(4)	(5)	(6)
Experience	-0.020*	-0.020	-0.019*	-0.018	-0.021*	-0.020
	(0.013)	(0.019)	(0.013)	(0.020)	(0.015)	(0.020)
Female×Experience		0.001		-0.002		-0.001
		(0.025)		(0.027)		(0.028)
Constant	0.471***	0.484***	1.266***	1.346***	-0.276***	-0.293***
	(0.015)	(0.018)	(0.017)	(0.019)	(0.018)	(0.020)
N	6,125	6,125	6,125	6,125	6,125	6,125
Clusters (Advisers)	1,225	1,225	1,225	1,225	1,225	1,225
R^2	0.002	0.003	0.002	0.024	0.002	0.003

Notes: Estimates are based on OLS regressions with standard errors clustered at adviser level. All specifications control for a worker's order (1 to 5). *Female* is a binary indicator for the worker's gender (0-male, 1-female). *Experience* is binary indicators for the experienced adviser treatments (0-baseline, 1-baseline with experienced advisers). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.8 Pilot Experiments

We conducted several pilots for our online experiment. We conducted our first pilot in April 2023 with 24 workers. Each worker answered 30 questions randomly selected from a set of 46. Based on the results, we dropped two easy questions (accuracy rates greater than 94%) and two hard questions (accuracy rates lower than 20%). To obtain set of 30 questions for our second pilot, we also dropped two questions in which women strongly outperformed men and nine questions in which men strongly outperformed women. We conducted our second pilot in May 2023 with 23 participants. Men and women performed equally well, but men were more confident in the quantitative measures. We tested two qualitative questions in this pilot, one rating their performance on the task and one on general math and science skills. We each evaluated the responses on these questions in a gender-blind way and found a larger gap in the performance question. We therefore decided to include this question in the main experiment.

Our third pilot took place in September 2023 and involved 20 workers and 20 advisers. Each worker answered 25 questions. We piloted the baseline treatment and found a gender gap in self-promotion and a gender gap in advice. However, we saw a reverse gender gap (with

women more confident) in the guessed rank question, which we suspected was driven by this question's being reverse-coded to all other questions (most confident being on the left rather than right of the screen). We therefore decided to flip the scale for this question. We also included a second rank question in which participants were asked to compare themselves with similar individuals. However, the answers to the two rank questions were almost perfectly correlated, so we decided to drop this second (more complicated) question.

Our fourth pilot took place in early October 2023 with a further 20 workers and 20 advisers. This time we found no gender gap in self-evaluations, which led us to conduct a fifth pilot in late October 2023 with 100 workers, and we once again did find a gender gap in self-evaluations. We also introduced a qualitative question asking participants to reflect on their math and science skills, for which they needed to spend at least 1 minute and received 50 pence for answering. We also saw a gender gap in performance. To reduce the gender gap, we decided to remove the five questions with the strongest male advantage in performance. We tested the set of questions in a sixth pilot in early December 2023 with 100 workers, and found no gender gap in performance. We then decided to use this set of 20 questions for our main experiment. We conducted a seventh pilot in late December 2023 with 100 workers to test the new qualitative question from the fifth pilot. After evaluating the answers using ChatGPT and finding no gender difference and little correlation with the quantitative question on our AI, we decided to remove this question.

The purpose of our eighth pilot (conducted in February 2024 with 100 workers and 100 advisers) was to test the final design of the baseline treatment and be used as the basis for our power calculations. Upon observing the results, we realized that the performance thresholds for task A might be a little too ambitious. In our ninth and final pilot (conducted in March 2024), we therefore lowered the threshold for job A from 10 to 8 correct answers. We used the data from pilots six to nine as the basis for our power calculations. These, in turn, convinced us to move away from a one-to-one matching of workers and advisers to the one-to-many matching we used in our main experiment. The time between the final pilot and the main sessions was spent on the pre-analysis plan and finalizing the design of the three other treatments (which we did not pilot).

A.9 Pre-Analysis Plan

In this section, we reprint our pre-analysis plan. We also added notes explaining where in the paper the respective analysis can be found. For any analysis not included in the paper, we will both (a) summarize its results and (b) explain why we did not include it in the paper. We also include the power calculation. The pre-analysis plan is also available from our pre-registration page at https://www.socialscienceregistry.org/trials/13803.

A Description of the Experiment

We conduct a three-part experiment. In session 1, participants in the role of workers work on a math and science quiz. Upon completing the quiz, they indicate their confidence in several ways, forming a worker "portfolio". In session 2, a different set of participants in the role of advisers sees the portfolio of five workers and advises each worker whether they should go for an advanced (A) or a basic (B) version of the task. In session 3, each worker receives the advice from one adviser and decides which of the two versions of the task to pursue.

We consider three between-subject treatments that vary the information the adviser receives in session 2. In the "baseline" treatment, the adviser receives the full portfolio of worker confidence measures from session 1. In the "nudge" treatment, advisers also receive a brief message informing them about the gender gap in self-assessment. In the "subjective" treatment, the advisers instead only receive the worker's verbal self-evaluation (a qualitative and subjective self-assessment). In addition, participants in the baseline treatment (but not the other treatments) are also asked to evaluate the same five workers a second time while also having access to those workers' true performances on the task. We refer to this second set of advice as the (within-subject) "objective" treatment.

B Main Outcomes

1. Worker Confidence (Session 1)
    a) Guessed number of correct answers
    b) I performed well on the quiz (0-100)
    c) I would apply for a job that would require me to perform well on the quiz (0-100)
    d) I would succeed in a job requiring me to perform well on the quiz (0-100)
    e) Relative performance rank (compared to 100 other participants)
    f) Subjective performance Likert scale (1-terrible to 5-great).
    g) Standardized sum of the previous measures
    h) Verbal self-evaluation
2. Advice (Session 2)
    a) Binary advice (A or B)
    b) Five-point scale advice (Definitely A to Definitely B)

C Secondary Outcomes

3. Worker Task Choice (Session 3)
4. Other measures of advice (Session 2)

D Main Explanatory Variable

5. Worker Gender: dummy variable comparing males to females.

E Other Variables

6. Workers (session 1): general confidence/risk aversion measures, score, demographics
7. Advisors: gender, other demographics, confidence/risk aversion
8. Advisor beliefs about gender differences:
   a) Are men or women better at the task?
   b) Are men or women more confident in their self-assessment?
9. Worker score in session 3.

Demographics include household income, income satisfaction, age, occupation, job title, and education.

F Main Analysis

Our experiment aims to test the following hypotheses:

1. Gender Gap in Self-Promotion: Female workers are less confident than male workers.
   (a) Main tests: one-sided t-test for each of the quantitative worker confidence variables listed in part B1 above (a-g).
   Authors' note: We included these tests in Figure 2
2. Gender Gap in Advice: Female workers receive less ambitious advice than male workers in the baseline treatment.
   (a) Main tests: linear regression of the (a) binary advice measure and (b) Likert scale advice measure on a dummy for the worker's gender. We use a one-sided test on the gender dummy.
   Authors' note: We included these tests in Table 4 (columns 1 and 3) and reprint the results in Table 5 (columns 1 and 2).
3. Treatment Differences: The gender gap in advice is significantly smaller in the objective, nudge and subjective treatments than in the baseline.
   (a) Main tests: a linear regression of advice on gender, treatment and a gender*treatment interaction. The key test is a one-sided test on the significance of the gender*treatment interaction. We do three regressions comparing each of nudge, subjective and objective treatment to the baseline.
   Authors' note: We included these tests in Table 5 (columns 3, 5 and 7).

We will use robust standard errors clustered at the referee level for the regression analysis (F2- F3). In addition to testing these three hypotheses, we will also investigate the drivers of the gender gap in advice:

4. Drivers of the Gender Gap in Advice: Is the gender gap in advice (F2) driven by gender differences in self-promotion or by a gender bias among advisers?

   (a) Main tests: linear regression of the (a) binary advice measure and (b) Likert scale advice measure on (i) a dummy for the worker's gender and (ii) the worker's guessed score (B1a). We use a one-sided test for both variables (less ambitious advice for women, positive effect of the guessed score).
   Authors' note: The pre-registered specification can be found in Table A.5. In our main analysis in Table 4 we decided to control for all six self-evaluation questions instead (as per analysis G11 below). The main difference between the two tables is that the residual gender gap is no longer significant in Table A.5. Upon reflection, Table A.5 seemed to us like a fairer reflection of our results since it controls for all information available to advisers.

G Additional Analysis

Here is an overview of additional tests we intend to include in our paper to shed further light on our results, but that do not directly address our main hypotheses. We start with some further analysis of worker confidence.

5. Is there a gender difference in performance?

   a) Two-sided t-test for a gender difference in worker score (session 1).
   b) Two-sided t-test for a gender difference in worker score (session 3), after receiving advice.
   Authors' note: We present these results in Table A.2.

6. Are any gender differences in confidence significant after controlling for performance and demographics?

   a) Regression of confidence on the worker gender dummy (1-female, 0-male) while controlling for (a) worker score from session 1 and (b) also controlling for worker demographics. The relevant test is a one-sided test on the regression coefficient on the female dummy (as per main analysis F1).
   Authors' Note: We present the results (a) in Table A.3 and (b) in Table A.10.

7. Are there any gender differences in the subjective self-evaluations?

   a) For this purpose, we intend to explore the use of AI methods to quantify the expressed levels of confidence.
   Authors' Note: We present these results in A.6.

Next we plan to present robustness checks for our analysis on gender differences in advice, as well as some additional outcomes.

8. Is the gender difference in advice F2 robust to only looking at the first worker evaluated?

   a) Replication of main analysis F2 while including only the first worker evaluated.
   Authors' note: We present this analysis in Table A.13.

9. Do gender differences and treatment differences in advice (F2 and F3) also appear in our secondary measures (C4)?

   a) Replication of main analysis F2 and F3 for the performance target C4a.

b) Replication of main analysis F2 and F3 using the confidence in advice variable (C4b). For this purpose, we combine the binary variable with the confidence variable as follows: confidence 1–3 → "no advice", confidence 4–8 → "probably A/B", confidence 9+ → "definitely A/B", and encode this as a 5-point scale from 1 = definitely B to 5 = definitely A.
Authors' note: This analysis is present in Table A.12

10. Does the adviser's confidence (C4b) depend on the worker's gender and advice sent?
    a) Two-sided t-test investigating whether confidence differs significantly by (i) worker gender and (ii) binary advice (A or B). Also two t-tests studying whether confidence differs by worker gender *within* each advice category.
    b) A linear regression of adviser confidence on (i) worker gender, (ii) binary advice (A or B), or (iii) gender within the sample of a given type of advice, always controlling for the worker's guessed score.
    Authors' note: This analysis can be found in Tables A.14 and A.15.
11. What are the main drivers of advice?
    a) Linear regressions of advice (B2a and B2b) in the baseline on (i) all quantitative measures of confidence available to advisers (B1a–e) and (ii) referee demographics.
    b) Regression exploring whether the effect of confidence is non-linear by including guessed-score fixed effects.
    Authors' note: We present the analysis controlling for all measures of confidence in Table 4, and the one also controlling for demographics in Table A.11. Results with guessed score fixed effects are in Table A.8.
12. What is the quality of advice and how does it differ by treatment and gender?
    a) We create a binary indicator for optimal advice—whether advice would maximize a worker's expected earnings based on session-1 performance. Regress this indicator on treatment dummies. Also replicate including a gender dummy and interaction terms.
    Authors' note: This analysis is presented in Table 6.
13. Do we observe heterogeneous effects based on adviser bias?
    a) Replicating main analysis F2 while splitting the sample based on adviser bias E8.
    b) We split the sample into three groups: no difference, men at least slightly better/more confident and women at least slightly better/more confident. We only analyze a given group if at least 5% of our sample are in this group.
    c) We analyze variables E8a and E8b separately.
    Authors' note: This analysis is presented in Table A.16 and Table A.17.
14. Do any of our treatments reduce any gender bias that exists in variables E8a and E8b?
    a) Regression of each variable on treatment dummies (two-sided tests).
    b) Conduct this test only if the baseline shows significant bias (one-sided test: women perceived as worse/less confident).
    Authors' note: This analysis is presented in Table A.18.

We also consider several tests of behavior in Session 3. These are of lesser importance because we (a) expect nearly all workers to follow advice and (b) we have less power due to the smaller sample size per treatment.

15. Do female workers choose less ambitious jobs in session 3?
    a) One-sided t-test for worker choice in session 3 by gender.
    Authors' note: This test is presented in Appendix A.5.
16. An analysis of the efficiency consequences of advice.
    a) Create a binary indicator of the optimal choice (maximizing expected earnings based on session-1 performance). Regress on treatment dummies, and replicate including gender and interaction terms.
    b) Perform a similar analysis for whether advisers' advice is itself optimal given session-3 performance.
    Authors' note: This analysis is presented in Table 6 and Table A.9.
17. Are gender differences in worker choice in Session 3 driven by advice, and do they differ by treatment?
    a) Split the sample by whether workers received binary or Likert advice.
    b) For binary advice: regress choice on gender and advice. Also use an ordinal confidence-weighted scale (1–3 no advice, 4–8 probably A/B, 9+ definitely A/B).
    c) For Likert advice: regress choice on gender and the 1–5 Likert recommendation.
    d) Test whether the gender gap in choices differs by treatment using regressions with gender, treatment dummies, and gender×treatment interactions.
    Authors' note: Due to the small sample size for these tests, we do not report the full pre-registered tables on the influence of advice on worker decisions. Instead, we analyze the quality of advice (Table 6 and Table A.9).

9.8 Power Calculations

Authors' note: In what follows we reproduce the power calculation from our pre-analysis plan.

We conducted power calculations to check our power to detect (a) a gender gap in worker confidence (self-promotions), (b) a gender gap in advice and (c) treatment differences in the gender gap in advice. Based on these power calculations, we aim to collect data from 350 workers and 1750 advisers (700 in the baseline, and 525 in the Nudge and Subjective treatments). With the intended sample size, we will have a power greater than 0.90 to detect (a) gender differences in worker confidence in line with the pilot and previous literature, (b) a gender gap in advice of half the size of the gap implied by our pilots and (c) treatment effects that eliminate 70% or more of the baseline gender gap in advice. In the remainder of this document, we will present the power calculations in greater detail.

H1: Gender Differences in Self-Promotion

Our main interest in this experiment is in testing three sets of hypotheses. Our first hypothesis relates to the gender gap in self-confidence (self-promotion) among workers. We can do power calculations using the confidence data from three pilot sessions (N=301 participants). Table 1 presents the pilot results for our quantitative confidence variables (B1a-g). The gender gap was significant at the 1% level for all variables, and the point estimates for the first five variables are consistent with Exley and Kessler (2022) albeit slightly smaller.[11] We use

[11]The gender gaps in Exley and Kessler (2022) for the first five variables are 2.28 (p. 1359), 0.67, 13.83, 17.28 and 16.12 (Table II). They do not elicit the rank guess variable.

the averages and standard deviations for men and women to compute the power associated with a one-sided t-test testing whether men are more confident than women. The results suggests that a sample size of 210 workers (105 men and 105 women) is sufficient to achieve a power of 0.80 or greater for all six of our confidence variables, as well as when summing over all variables. A sample size of 290 workers would be sufficient for achieving a power of 0.90.

Table 1: Power Calculations for the Gender Gap in Self-Promotions

	Score Guess	Perform (Likert)	Perform (0-100)	Apply	Perf Job	Rank Guess	Sum
Men	10.14 (3.95)	3.06 (1.02)	44.82 (22.4)	38.63 (28.1)	39.96 (27.3)	54.44 (24.3)	0.29 (0.98)
Women	8.81 (3.75)	2.64 (1.02)	35.89 (22.5)	25.95 (25.1)	27.66 (26.1)	62.67 (22.8)	-0.29 (0.94)
Diff	1.33 (3.91)	0.42 (1.04)	8.93 (22.9)	12.68 (27.3)	12.30 (27.3)	-8.23 (23.9)	0.58 (1.0)
Samp (0.80)	210	150	158	112	118	206	70
Samp (0.90)	290	206	220	154	164	284	96

Notes: The cells report the means of the relevant variables and standard deviations are reported in parentheses. "Score Guess" asks participants to guess the number of questions they solved correctly. "Perform" elicits subjective performance on a five-point Likert scale and a 0-100 scale respectively. The next two questions ask, on a 0-100 scale, whether they would apply for a job that requires them to perform well on the quiz or be successful in such a job respectively. "Rank Guess" variable asks participants to guess their rank compared to 100 other participants (1-best, 101-worst). The final variable takes the sum of the standardized first six variables (reverse coding the rank guess variable) and then standardizes this sum. Required sample sizes ("Samp") are based on "power twomeans mu_men mu_women, sd1(sd_men) sd2(sd_women) onesided", where the four parameters are the ones presented in the first two rows of the table (means and sd). We present results for both power calculation with a power of 0.80 ("Samp (0.80)") and for a power of 0.90.

H2: Gender Differences in Advice

Our second hypothesis relates to the gender gap in advice. We conducted pilot sessions with a total of 200 advisers who each evaluated 5 worker portfolios for a total of 1000 adviser-level observations. The results revealed that (a) the various confidence measures were highly correlated with each other ($r>0.6$ in all cases) and (b) advisers relied heavily on the confidence variables in determining their advice. For example, a regression of binary advice on the worker's guessed quiz score yielded a coefficient of 0.067 (se: 0.003, $p<0.0001$). The resulting gender gap in advice received was also significant at 5.6pp ($p=0.015$, one-sided t-test). This gap might understate the true expected gender gap in the main sessions, however, because the sample of workers the advisers evaluated had a lower gender gap in confidence (0.42 points on the score guess variable) than the one we saw when combining worker data from all pilots (1.33 points).

Hence, we compute power in the following way. First, we randomly pick a sample of workers from the pilot sessions with replacement. We then simulate advice data using the pilot coefficient estimates of a regression of advice on workers' guessed quiz score. We also add noise to the simulated advice to the point that the guessed quiz score explains the same fraction of the total variance as in the pilot. We ensure that the Likert scale advice has five values by winsorizing at 1 and 5 and rounding to the nearest integer. Similarly, we binarize the binary advice measure by converting, for example, a predicted advice of 0.75 into either

advice A (with 75% chance) or advice B (with 25% chance). Power is then equal to the fraction of simulated samples in which women receive less ambitious advice with $p<0.05$ (one-sided).

Table 2 presents the results. The power calculations predict an approximately 8.6pp gender gap in the binary measure and a 0.25-point gap in the Likert scale measure. Noting that each adviser evaluates five workers, a sample size of 120-150 advisers would already yield power above the 0.80 threshold, where power is slightly higher for the Likert scale measure. Note that this assumes that each piece of advice is an independent observation, which seems reasonable in the present context (and given that the data were simulated from data in an identical environment). Note also that we conservatively assume that advisers are not intrinsically biased against women; adding such a bias would raise the predicted gender gap and thereby further increase our power.

Table 2: Power Calculations for a Gender Difference in Advice

	Gender Gap (Binary)	Power	Gender Gap (Likert)	Power
600 evals / 120 adv	-0.085	0.724	-0.249	0.816
800 evals / 150 adv	-0.088	0.876	-0.263	0.930
1000 evals / 200 adv	-0.087	0.929	-0.254	0.964
2000 evals / 400 adv	-0.086	0.993	-0.251	1.000

Notes: The table presents results from four sets of simulations that vary the sample size from 600 to 2000 portfolio evaluations (evenly split across the two genders). Binary advice is a dummy variable for whether advice A is given, and Likert advice is a five-point scale ranging from "definitely B" to "definitely A". Each row is based on 1000 simulated samples. "Gender Gap" is the difference between men and women on the relevant outcome averaged across all simulation samples. Power is the fraction of simulated samples in which an OLS regression of the outcome on gender produced a significant result at the 5% level. Since we predict that women receive less ambitious advice, we use a one-sided test.

H3a: Treatment Effects in Advice in the Subjective Treatment

Our third set of hypotheses relates to treatment differences in the gender gap in advice. A first between-subject treatment examines what happens if referees only have access to a subjective verbal self-evaluation (and no longer see the worker's guessed score or other quantitative self-evaluations). For our power calculations, we assume that advice in this treatment is no longer connected to (guessed) performance. This implicitly assumes that subjective self-evaluations are not informative or, if they are, that male and female workers write similar subjective self-evaluations. We also assume that advisers are not directly influenced by the gender of the worker, i.e., they are not intrinsically gender-biased. In practice, this implies that we randomly generate advice in such a way that its distribution resembles the distribution of advice given in the pilot in terms of advice, while removing any connection to performance and gender. We then test whether the gender difference in advice is significantly smaller in this treatment (where it is zero in expectation) than in the baseline treatment (where it is approximately 8.6pp). Table 3 presents the results: having 400 advisers in each treatment gives us a power greater than 0.90.

Table 3: Power Calculations for the Subjective Treatment

Evaluations	Advisers	Power (Binary)	Power (Likert)
1000	200	0.682	0.706
2000	400	0.910	0.923
3000	600	0.977	0.990
3500	700	0.989	0.994
4000	800	0.995	0.996

Notes: The table presents results from simulations. The first column presents the number of evaluations included in each simulation sample for the baseline and (in parentheses) the subjective treatment. The second column presents the corresponding number of advisers (one per five evaluations). The two subsequent rows summarize the results of 1000 simulations, where power is calculated as the proportion of simulated samples in which the gender gap in the subjective treatment is significantly smaller than in the baseline as per a difference-in-difference test (5% level). Since we predict a smaller gender gap in the Subjective treatment, we use a one-sided test.

Two further remarks are in order. First, since we use the baseline treatment in multiple comparisons, it may be prudent to oversample this treatment. Table 4 shows that shifting observations from the Subjective to the baseline treatment only starts reducing power when more than 30% of observations are shifted this way. Second, the true treatment effect may be larger than assumed here (e.g., if advisers are gender-biased) or smaller (e.g., if men write more optimistic subjective self-evaluations). We will discuss minimum detectable effect sizes at the end of this document.

Table 4: Power Calculations for Overrepresented Baseline in the Subjective Treatment

Per Treatment		Binary Advice	Likert Advice
#Evaluations	#Advisers	Power	Power
2000 (2000)	400 (400)	0.902	0.917
2200 (1800)	440 (360)	0.913	0.912
2400 (1600)	480 (320)	0.904	0.917
2600 (1400)	520 (280)	0.883	0.911
2800 (1200)	560 (240)	0.860	0.842
3000 (1000)	600 (200)	0.837	0.819
3200 (800)	640 (160)	0.773	0.782
3400 (600)	680 (120)	0.710	0.686
3600 (400)	720 (80)	0.578	0.555
3800 (200)	760 (40)	0.377	0.375

Notes: The table presents simulation-based power calculations under alternative allocations of observations between the baseline and subjective treatments. The first column reports the number of evaluations in the baseline treatment and, in parentheses, the number of evaluations in the subjective treatment. The second column reports the corresponding numbers of advisers (one adviser per five evaluations). Power is calculated from 1,000 simulations as the proportion of samples in which the gender gap in the subjective treatment is significantly smaller than in the baseline treatment in a difference-in-differences test at the 5% significance level. One-sided tests are used.

H3b: Treatment Effects in Advice in the Nudge Treatment

Our second between-subject treatment manipulation concerns a nudging/information intervention that attempts to lower the gender gap in advice by nudging the advisers to take the existing gender gap in self-promotion into account in their advice. For our power calculations, we assume that the intervention completely eliminates the gender gap in self-confidence/self-promotion (which is equivalent to assuming that the advisers remove the gender bias in worker confidence). We implement this in the power calculation by randomly drawing observations (i.e., confidence elicitations) from both male and female workers in the pilot, and then assigning a gender randomly. This removes the gender gap in worker self-promotions by construction, while ensuring that the confidence distribution is the same as in the pilot. Table 5 presents the results, which are very similar to the Subjective treatment (Table 3).

Table 5: Power Calculations for the Nudge Treatment

Evaluations	Advisers	Power (Binary)	Power (Likert)
1000	200	0.682	0.759
2000	400	0.909	0.951
3000	600	0.910	0.993
3500	700	0.993	0.996
4000	800	0.994	0.999

Notes: The table reports power calculations based on 1,000 simulations. Power is defined as the proportion of simulated samples in which the gender gap in advice is significantly smaller in the nudge treatment than in the baseline treatment in a difference-in-differences test at the 5% significance level. The simulations assume that the nudge eliminates the gender gap in self-promotion. One-sided tests are used.

H3c: Treatment Effects in Advice in the Objective Treatment

As a final step, we include a within-subject treatment testing whether providing objective information (the worker's actual score) decreases the gender gap. For our power calculations, we simulate the data for this treatment in the same manner as in the baseline treatment above (see the section H2), except that we replace the "guessed score" with the "actual score" in the process generating simulated advice. Since we use a within-subject design, we assume that the noise term used in the simulations is identical for the baseline and Objective treatment. This essentially assumes that the adviser's reading of the worker's portfolio does not change other than through replacing the worker's subjective performance with their true performance. Table 6 shows that in this treatment the gender gap reduces by about two thirds (approximately 5.3pp on the binary scale and 0.15 points on the Likert scale), which is because the gender gap in performance (while not zero) is smaller than the gender gap in confidence/self-promotions.

Table 6 also suggests that a relatively small sample size may be sufficient to detect these effects. This is because we use a within-subject design and because we assume that all advisers update in line with the new information (e.g., less ambitious advice if the true objective performance was less than the subjective performance). In practice, power may be lower if advisers choose not to incorporate the objective information (e.g., because they are

Table 6: Power Calculations for the Objective Treatment

Evaluations	Advisers	Treat Effect (Binary)	Power	Treat Effect (Likert)	Power
600	120	-0.054	0.505	-0.148	0.785
800	150	-0.054	0.584	-0.153	0.882
1000	200	-0.051	0.641	-0.151	0.939
2000	400	-0.052	0.901	-0.150	0.993
3000	600	-0.053	0.971	-0.153	1.000

Notes: The table presents results from five sets of simulations that vary the sample size from 600 to 3000 portfolio evaluations. Binary advice is a dummy variable for whether advice A is given, and Likert advice is the five-point scale ranging from "definitely B" to "definitely A". Each row presents the averages from 1000 simulated samples. "Treat Effect" is the estimated difference-in-difference between the gender gap in the baseline and the gender gap in the objective treatment averaged across all simulation samples. Power is the fraction of simulated samples in which an OLS of the difference between the outcome in the two treatments on gender produced a significant result at the 5% level (this is equivalent to a difference-in-difference test). Since we predict a smaller gender gap in the objective treatment, we use a one-sided test.

happy with their original advice), or if the noise term is not perfectly correlated across the baseline and objective evaluations (e.g., if advisers change their mind about the interpretation of the portfolio from the baseline to objective choice, or if they are making mistakes or are just noisy). Nevertheless, even in such cases we can expect our power to be high for the intended sample size of our experiment, as we discuss in the next section.

Final Sample Size and Cost Calculations

The previous sections suggest that a sample of around 300 workers would be sufficient to detect a gender gap in confidence with a power 0.90 (Table 1). To detect a gender gap in advice with a power of 0.90 a sample of 200 advisers would be sufficient (Table 2). And for studying treatment effects in the gender gap in advice, approximately 400 advisers per treatment would also achieve a power of 0.90. A larger sample size may still be helpful, however, in particular if the treatment effects are smaller than assumed in the power calculations. In addition, it is useful to oversample the baseline treatment, since it will be used in all of our tests (in contrast to the other treatments, which are used in a single comparison only).

Also keeping in mind budget constraints, we therefore aim to collect data from 350 workers and 1750 advisers (700 in the baseline, and 525 in the Nudge and Subjective treatments). The total costs are estimated to be around 9 AUD per adviser and 21.5 AUD per worker for a total of 23,275 AUD. This also implies that each adviser has a 20% chance of being matched to a worker.

With the intended sample size, we have a power greater than 0.90 to detect a gender difference in confidence among workers based on the pilot data (H1). For the gender gap in advice (H2) the present sample size would give us a power greater than 0.90 even if the gender gap in confidence is half the size in the pilot (or equivalently: the advisers put only half as much weight on worker confidence in their final decision as they did in the pilot). For our between-subject treatments, we expect to still have a power of 0.84 on the Likert scale variable to detect a decrease in the gender gap by 60%, assuming that the gender gap in the

baseline is in line with the pilot (and a power greater than 0.90 if the gender gap decreases by 70% or more).

A.10 Pre-Analysis Plan: New Treatment with Experienced Advisers

In this section, we reprint our pre-analysis plan for the new experimental treatment involving experienced advisers and a vignette task. We also added notes explaining where in the paper the respective analysis can be found. The pre-analysis plan is also available from our pre-registration page at https://www.socialscienceregistry.org/trials/17915/history/297617.

A Introduction

This pre-analysis plan concerns a second data collection wave that builds upon a first wave of data collected in summer 2024. The purpose of the wave two data collection effort is two-fold. First, we want to replicate our original baseline treatment with a novel sample of advisers with teaching or management experience to test whether our results hold up in this sample. Second, we append several vignettes to the end of the survey to test for gender differences in advice received in those vignettes.

The main structure of the experiment remains the same as in our original data collection. In particular, we will conduct a three-part experiment. In session 1, participants in the role of workers work on a math and science quiz. Upon completing the quiz, they indicate their confidence in several ways, forming a worker "portfolio". In session 2, a different set of participants in the role of advisers sees the portfolio of six workers and advises each worker whether they should go for an advanced (A) or a basic (B) version of a future task. The first five pieces of advice pertain to workers from the first wave, with the last worker being a worker from wave 2. The advisers then go through twelve vignette advice tasks with variation in (a) the confidence of the advisee (high or low), (b) the advisee's gender (male or female) and (c) the context (six different scenarios). In session 3, each worker receives the advice from one adviser and decides which of the two versions of the task to pursue. We will conduct only our "baseline" treatment from wave 1 in which the adviser in session 2 receives the full portfolio of worker confidence measures from session 1.

B Main Outcomes

1. Incentivized Advice (Session 2)
    a) Binary advice (A or B)
    b) Five-point scale advice (Definitely A to Definitely B)
2. Vignette Advice (Session 2)
    a) Binary advice (A or B)

C Secondary Outcomes

3. Worker Confidence (Session 1)
    a) Guessed number of correct answers
    b) I performed well on the quiz (0-100)
    c) I would apply for a job that would require me to perform well on the quiz (0-100)
    d) I would succeed in a job requiring me to perform well on the quiz (0-100)
    e) Relative performance rank (compared to 100 other participants)

f) Subjective performance Likert scale (1-terrible to 7-great).
g) Verbal self-evaluation

4. Other measures of incentivized advice (Session 2)
   a) Performance target for workers
   b) Confidence in binary advice

D Main Explanatory Variables

5. Worker Gender: dummy variable comparing males to females.
6. Confident Vignette: dummy variable comparing vignettes with a confident worker (1) to vignettes with a non-confident worker (0).
7. Vignette scenarios: six dummy variables equal to 1 for choices made under one of the six vignette scenarios.

E Other Variables

8. Advisors: gender, other demographics, confidence/risk aversion
9. Advisor beliefs about gender differences:
   a) Are men or women better at the task?
   b) Are men or women more confident in their self-assessment?

Demographics include personal income, household income satisfaction, age, occupation, job title, education, experience in STEM fields, experience in teaching and / or management, experience in giving advice to others in the context of their teaching and / or management experience.

F Main analysis

We will include in our main analysis only the first five workers evaluated by the advisers. These workers are the same exact workers our advisers encountered in wave 1, allowing us to attribute any differences we observe between waves to advisers rather than workers. We will conduct robustness checks including the sixth worker as well (G1 below). We use one-sided tests for gender gaps in advice received (F1 and F3), since we expect women to receive less ambitious advice based on both previous research and our results from wave 1. We will use robust standard errors clustered at the adviser level for any regression analysis unless otherwise specified.

Our experiment aims to answer the following research questions:

1. Gender Gap in Advice (Incentivized Task): Do female workers receive less ambitious advice than male workers?
   (a) Main test: two linear regressions of the (a) binary advice measure and (b) Likert scale advice measure respectively on a dummy for the worker's gender. We use a one-sided test on the gender dummy.
   (b) With controls: two linear regressions of the (a) binary advice measure and (b) Likert scale advice measure respectively on a dummy for the worker's gender. We control linearly for each of the worker confidence variables C3a-e. We use a one-sided test on the gender dummy.

2. Experience Effect on Advice: Is the gender gap similar in the wave 1 sample and the wave 2 sample with advisors with management/teaching experience?
   (a) Main test: a linear regression of advice on gender, experience, and a gender*experienced interaction. The key test is a two-sided test on the significance of the gender*experienced interaction. Here, "experienced" is a dummy equal to one for participants from the second wave (those with management/teaching experience).
   (b) With controls: same as the main test but now controlling linearly for each of the worker confidence variables C3a-e.
   (c) With interacted controls: same as (b) except we now interact all confidence variables C3a-e with the experienced dummy.

   In addition to investigating the incentivized measure, we will also examine the following hypothesis related to our vignettes:

3. Gender Gap in Advice (Vignette Task): Do women receive less ambitious advice than men?
   (a) Overall: linear regression of the vignette advice measure on a dummy for the worker's gender. We use a one-sided test on the gender dummy. For this analysis, we will pool data from all vignettes and include scenario-fixed effects (one for each of the six possible scenarios D7) and control for worker confidence D6 (1-confident, 0-not confident).
   (b) By vignette: we repeat the previous test separately for each specific scenario. This gives us the gender gap in advice separately for each scenario.
   (c) By confidence: we repeat test (a) but include a confidence*gender interaction effect. We use a one-sided test for both the gender dummy (the gender gap among non-confident workers) and the sum of the gender dummy and interaction term (the gap among confident workers). For the interaction term, we will use a two-sided test.
4. Comparison between Vignettes and Incentivized Task: is the gender gap similar in the two tasks?
   (a) Main test: Regression comparison: linear regression of binary advice on a dummy for the worker's gender, a dummy for vignettes (1-vignette, 0-incentivized) and their interaction. We use the data from all vignettes and include scenario-fixed effects for the vignettes. The key test is a two-sided test on the interaction term.
   (b) With controls: same as the main test but now controlling linearly for the confidence variables C3a-e (incentivized data) and for the confident vignette dummy D6 (vignette data).

G Robustness Checks

We plan to include the following robustness checks.

1. Main results with all workers: do our main results F1, F2 and F4 hold up when also including the sixth worker (who only received advice in wave 2)?

(a) Replication of analysis F1, F2 and F4 while now also including the data for the sixth worker that receives advice.

2. Gender Gap in Self-Promotion: Are female workers are less confident than male workers?
   (a) New sample: one-sided t-test for each of the quantitative worker confidence variables listed in part C3 above (a-g) for the 105 workers in wave 2.
   (b) New sample with controls: Regression of confidence on the worker gender dummy (1-female, 0-male) while controlling for worker score fixed effects from session 1. The relevant test is a one-sided test on the regression coefficient on the female dummy. Separate test for each variable listed in part C3 (a-g) above. 105 workers from wave 2 only.
   (c) Full sample: same as (a) but now including all workers from wave 1 and wave 2.
   (d) Full sample with controls: same as (b) but now including all workers from wave 1 and wave 2.
3. Are the gender gap in vignette advice F3 and the difference between vignettes and the incentivized task F4 robust to only looking at the first six vignettes, i.e., to only encountering each scenario once?
   (a) Replication of main analysis F3 and F4 while including only the first six vignettes evaluated.
4. Main results with first worker: do our main results F1, F2 and F4 hold up when only analyzing the first worker?
   (a) Replication of analysis F1, F2 and F4 while only including the first piece of advice.
5. Do gender differences and experience effects in advice (F1 and F2) also appear in our secondary measures (C4)?
   (a) Replication of main analysis F1 and F2 for the performance target C4a.
   (b) Replication of main analysis F1 and F2 using the confidence in advice variable (C4b). For this purpose, we combine the binary variable with the confidence variable in the following way: we will code advice given with confidence 1-3 as "no advice", confidence 4-8 as "probably A/B" and 9+ as "definitely A/B" and encode this as a five-point scale from 1-definitely B to 5-definitely A, so as to approximate the distribution of responses on the Likert scale.
6. Do experienced advisers give better advice?
   (a) a. We create a binary indicator for optimal advice based on whether advice, if followed, would maximize the worker's expected earnings based on their actual score from session 1. We then regress this advice on the "experienced" dummy (see F2 above). Our test is a one-sided test on this dummy.
   (b) We regress expected earnings based on following the advice and part 1 score on the experienced dummy. Our test is a one-sided test on this dummy.
   (c) We create a new variable to compare the gap in expected earnings based on advice received and what would have maximized expected earnings based on part 1 score. We regress this variable on the experienced dummy. Our test is a one-sided test on this dummy.

(d) We will also repeat regressions a-c while controlling for gender and interacting gender with the experienced dummy. This allows us to test whether the gender gap in advice quality is different for experienced advisers. Here the key test is a two-sided test on the gender*experienced dummy.

H Sample, Recruitment and Randomization

We aim to recruit 525 advisers and 105 workers from the sample of UK participants on Prolific. We will require adviser participants to have experience in either management or teaching roles. Should we fail to reach our target sample of advisers within one week, we will supplement our sample with Americans using the same criteria. If more than 20% of our adviser sample ends up being US-based, we will conduct a robustness check for our main analysis (section F) using only our UK advisers.

Advisers first receive portfolios from five workers from our first data collection effort (wave 1). Each of the 350 worker portfolios from wave 1 gets randomly assigned to either 7 or 8 advisers. Each adviser then receives the portfolio from one of the workers from wave 2, so that each worker's portfolio is randomly assigned to five unique advisers.

For the vignette task, the first six vignettes the advisers see correspond to the six scenarios, presented in random order. For each of these six vignettes, we randomly vary the worker's confidence level (high or low) and gender (male or female). They will then see the same six scenarios a second time, changed so that either confidence or gender is different from the first set of six scenarios.

10.9 Power Calculations

Summary

We conducted power calculations to check our power for the four components of our main analysis, that is, to detect (a) a gender gap in incentivized advice, (b) the experience effect on incentivized advice, (c) a gender gap in vignette advice and (d) comparisons between vignette and incentivized advice. Based on these power calculations, we aim to collect data from 525 advisers and (for procedural reasons) 105 workers. With the intended sample size, we will have (a) a power greater than 0.80 to detect a gender gap in incentivized or vignette advice of 40% the size of the gap in our first wave and (b) a power greater than 0.80 to detect a 60% difference in advice due to experience or due to the task used (vignettes vs incentivized tasks). In the remainder of this document, we will present the power calculations in greater detail.

H1: Gender Difference in Incentivized Advice

Our first hypothesis relates to the gender gap in advice. In our first wave we collected data from 350 workers and 700 baseline treatment advisers who each evaluated 5 worker portfolios for a total of 3500 adviser-level observations. We compute power for the binary advice measure in wave 2 by using simulations to randomly draw A-advice using the first wave averages for men (43.6%) and women (31.5%). We do this for our intended sample size (525 advisers and 2625 evaluations) and calculate power for effect sizes equal to or smaller than in our first wave. For the Likert scale we use a similar approach where we draw advice equal to the average response for men (2.75 points) and women (2.37 points) respectively

and then add noise to achieve a similar R2 as in the first wave data set. The table below presents the results. The main takeaway is that we still have good power ($> .80$) even for a gender gap only 40% of the size of our first wave (-0.048 in the binary measure).

Table A.43: Power Calculations for a Gender Difference in Advice

	Binary Advice		Likert Advice	
% Evaluations	Gender Gap	Power	Gender Gap	Power
100%	-0.121	1.00	-0.38	1.00
80%	-0.097	1.00	-0.30	1.00
60%	-0.072	0.98	-0.22	0.99
40%	-0.048	0.81	-0.15	0.85
N evaluations = 3,150; N advisers = 525				

Notes: The table reports results from four sets of simulations that vary the effect size from 100% to 40% of the first-wave effect in steps of 20 percentage points. Binary advice is an indicator for whether advice A is given, and Likert advice is measured on a five-point scale ranging from "definitely B" to "definitely A." Each row is based on 1,000 simulated samples. The gender gap is the average difference between men and women across simulations. Power is the fraction of simulations in which an OLS regression of the outcome on gender yields a statistically significant effect at the 5% level. Because we hypothesize that women receive less ambitious advice, tests are one-sided.

H2: Effect of Experience

Our second hypothesis relates to the effect of management/teaching experience. To calculate power, we use the same simulations as for hypothesis 1 and compare the simulated gender gaps to the observed gender gap in our first wave data. Here the key test is a two-sided t-test on the interaction term of a regression of advice on gender, a dummy for experienced (i.e., wave 2) advisers and their interaction. The results (presented in the table below) suggest that with our intended sample size we have a power of at least 0.83 to detect 60% increases or 60% decreases in the gender gap, with a larger power for larger changes.

Table A.44: Power Calculations for Experience Effect

	Binary Advice			Likert Advice		
	Baseline	Experience	Power	Baseline	Experience	Power
+100%	-0.121	-0.242	1.00	-0.38	-0.72	1.00
+60%	-0.121	-0.193	0.92	-0.38	-0.58	0.83
+50%	-0.121	-0.182	0.76	-0.38	-0.55	0.65
-50%	-0.121	-0.059	0.72	-0.38	-0.19	0.81
-60%	-0.121	-0.048	0.91	-0.38	-0.15	0.92
-100%	-0.121	0.000	1.00	-0.38	0.000	1.00

Notes: The table reports results from six sets of simulations that vary the gender gap among experienced advisers. Power is the fraction of simulated samples in which an OLS regression of the outcome on gender, an indicator for experienced advisers, and their interaction yields a statistically significant coefficient on the interaction term at the 5% level. See Table **??** for further details. Because we do not have an ex ante prediction regarding the direction of the difference in gender gaps, tests are two-sided.

H3: Gender Gap in Vignette Advice

Our third hypothesis posits that women receive less ambitious advice in the vignette tasks. In our pilot, 55% of men received ambitious advice across all vignettes. With our intended sample size of 525 evaluators and 6300 evaluations, we can calculate minimum effect size using the power twoproportions command in Stata. If we assume that each piece of advice is an independent observation, we obtain a minimum detectable effect size of 3.6pp to detect a gender gap across all vignettes with a power of 0.90 (power twoproportions 0.55, n(6300) power(0.9) alpha(0.05) onesided). The minimum detectable effect sizes are 5.2pp when split by confidence level (n=3150 evaluations), 8.9pp for each individual vignette scenario (N=1050) and 12.4pp for each vignette/confidence combination (N=525). If we instead assume that pieces of advice are not independent but correlated for each adviser, our minimum detectable effect size would lie somewhere between 12.4pp and the respective lower values mentioned above.

H4: Comparison between Vignettes and Incentivized Advice

Our fourth hypothesis compares the gender gap between vignettes and incentivized advice. For this purpose, we can use simulations similar to H2 above. For the incentivized data, we simulate new data assuming the same rates of ambitious advice for men and women as in wave 1 in the baseline (2625 evaluations). For the vignette task, we use the same process, where we vary the size of the gender gap to be smaller or larger than in the incentivized task (6300 evaluations). Power is then calculated as the fraction of simulated samples in which we observe a significant interaction term in a regression of advice-A on gender, a vignette dummy and their interaction. The table below shows that we have good power to detect a 50-60% increase or decrease in the gender gap between the vignettes and the incentivized task.

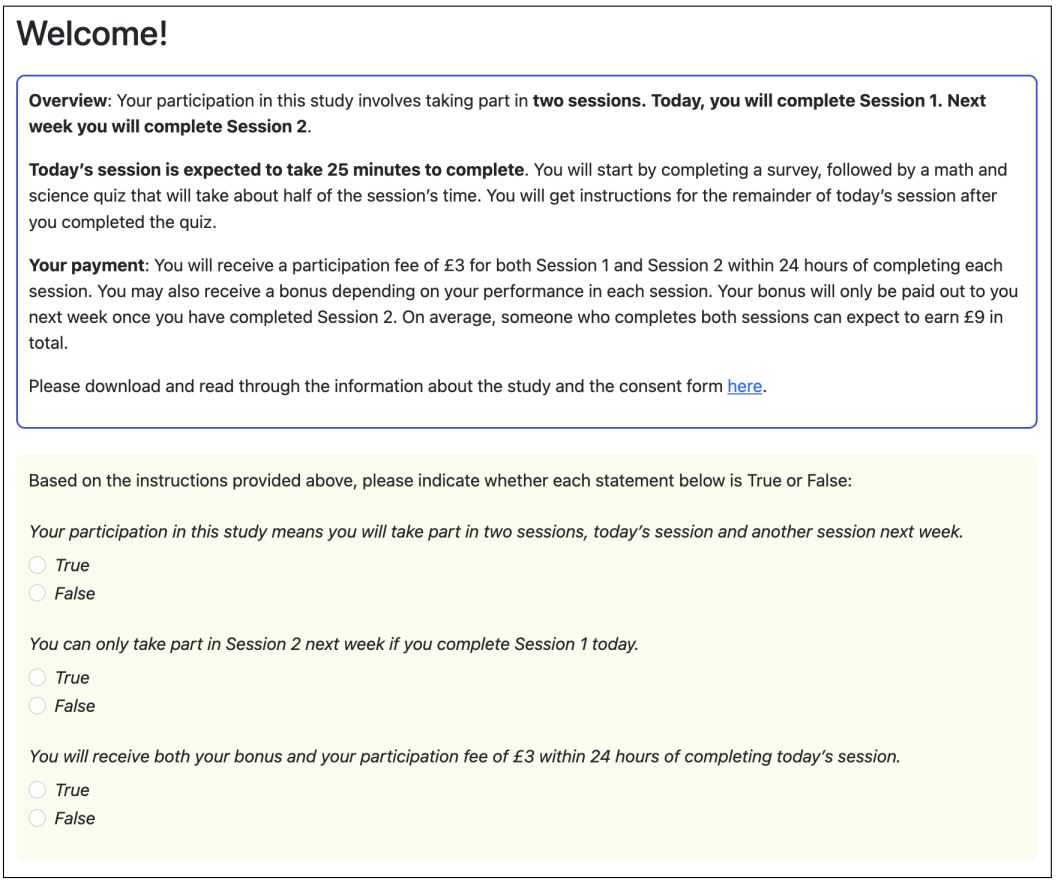
Table A.45: Power Calculations for Vignette vs. Incentivized Task

	Binary Advice		
	Incentivized	Difference	Power
-100%	-0.121	-0.121	1.00
-60%	-0.121	-0.073	0.91
-50%	-0.121	-0.060	0.78
+50%	-0.121	0.059	0.77
+60%	-0.121	0.071	0.91
+100%	-0.121	0.121	1.00

Notes: The table reports results from six sets of simulations that vary the gender gap in the vignette responses while holding the gap in incentivized choices constant. Each simulated sample includes 2,525 incentivized evaluations and 6,300 vignette evaluations, each split equally by gender. "Incentivized" denotes the gender gap in incentivized evaluations. "Difference" is the difference between the incentivized and vignette gender gaps. Power is the fraction of simulated samples in which the difference in gender gaps between vignette and incentivized responses is statistically significant at the 5% level (two-sided test).

A.11 Experimental Instructions/Screenshots

11.1 Part 1 Worker Session



Welcome!

Overview: Your participation in this study involves taking part in **two sessions. Today, you will complete Session 1. Next week you will complete Session 2**.

Today's session is expected to take 25 minutes to complete. You will start by completing a survey, followed by a math and science quiz that will take about half of the session's time. You will get instructions for the remainder of today's session after you completed the quiz.

Your payment: You will receive a participation fee of £3 for both Session 1 and Session 2 within 24 hours of completing each session. You may also receive a bonus depending on your performance in each session. Your bonus will only be paid out to you next week once you have completed Session 2. On average, someone who completes both sessions can expect to earn £9 in total.

Please download and read through the information about the study and the consent form here.

Based on the instructions provided above, please indicate whether each statement below is True or False:

Your participation in this study means you will take part in two sessions, today's session and another session next week.

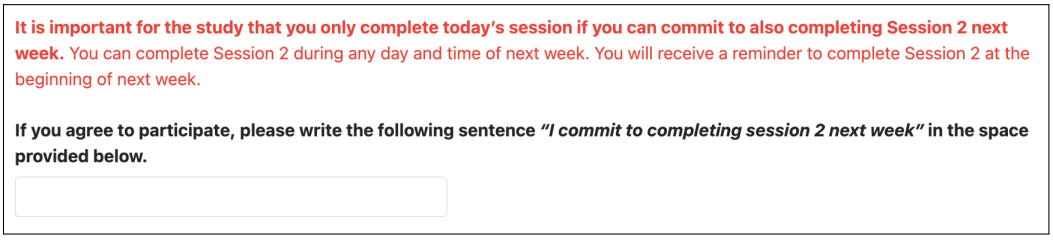
- *True*
- *False*

You can only take part in Session 2 next week if you complete Session 1 today.

- *True*
- *False*

You will receive both your bonus and your participation fee of £3 within 24 hours of completing today's session.

- *True*
- *False*



It is important for the study that you only complete today's session if you can commit to also completing Session 2 next week. You can complete Session 2 during any day and time of next week. You will receive a reminder to complete Session 2 at the beginning of next week.

If you agree to participate, please write the following sentence *"I commit to completing session 2 next week"* in the space provided below.

Initial survey

1) Please indicate your age:

----

2) Are you:

Female Male Other

If you chose *Other*, please specify:

3) What is the highest degree or level of education you have completed?

----

4) What is your current occupation?

- [ ] Studying
- [ ] Employed (full-time or part-time)
- [ ] Currently looking for work
- [ ] Not in the labour force
- [ ] Retired
- [ ] Other
- [ ] Prefer not to say

5) What is (or was) your full job title? *Examples: Retail assistant, Civil engineer, Cleaner, Nurse, Primary School Teacher*

6) How much do you get paid for your job, before tax and other deductions are taken out? If your pay before tax changes from week to week, think about what you earn on average.

----

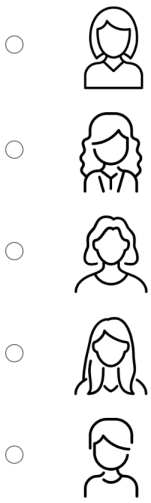
7) Now please consider your **household income** *(your household includes everyone who lives with you regularly and with whom you share money)*. How do you feel about your **household income** nowadays?

During this study, you will be represented by an "in-study" name and avatar. Please choose your name and avatar now.

Name:

- Gabriella
- Erica
- Felicia
- Caroline
- Stephanie
- Philippa
- Alana
- Georgia
- Olivia
- Louisa
- Victoria
- Jennifer
- Paula
- Alexandra
- Lucy
- Christina
- Samantha
- Julia
- Marcia
- Danielle

Avatar:

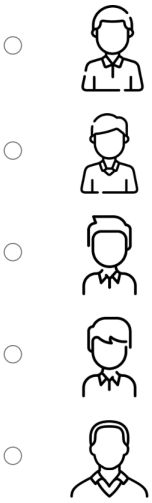


During this study, you will be represented by an "in-study" name and avatar. Please choose your name and avatar now.

Name:

- Alexander
- Charles
- Eric
- Alan
- Philip
- George
- Stephen
- Marcus
- Felix
- Victor
- John
- Gabriel
- Christopher
- Louis
- Oliver
- Daniel
- Julian
- Paul
- Luke
- Samuel

Avatar:

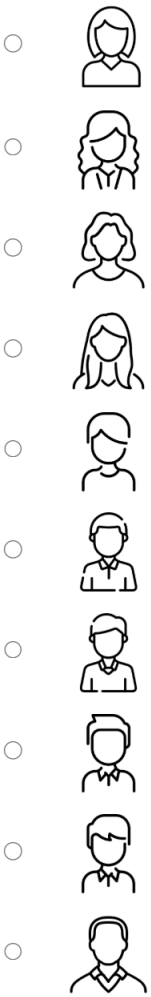


During this study, you will be represented by an "in-study" name and avatar. Please choose your name and avatar now.

Name:

- Freddie
- Alexander
- Charlie
- Louis
- Felicia
- Felix
- Sam
- Philippa
- Julian
- Stephen
- Philip
- Alex
- Louisa
- Leo
- Alexandra
- Julia
- Theo
- Stephanie

Avatar:



Math and science quiz

You will complete a **20 multiple choice questions quiz** that will test your skills in **math and science**.

Here is an example of a question:

If $5^5 = 5^2 \times 5^m$, what is the value of m?

A. 3
B. 5.5
C. 9
D. 12.5

You will see each question on a separate page, and you will have **30 seconds to answer each question**.

You will earn a bonus of **£0.10 per correct answer**. (There is no penalty for wrong answers.)

Based on the instructions provided above, please indicate whether each statement below is True or False:

The quiz will test your math and science skills.

- ○ *True*
- ○ *False*

You will earn a bonus of £0.10 per correct answer on the quiz.

- ○ *True*
- ○ *False*

Math and science quiz

When you are ready to start the quiz, please press Start.

Quiz (1/20)

Time left to complete this page: **0:25**

If n is a positive integer divisible by 7, and if $n < 70$, what is the greatest possible value of n ?

A. 49
B. 56
C. 63
D. 69

Quiz (2/20)

Time left to complete this page: **0:25**

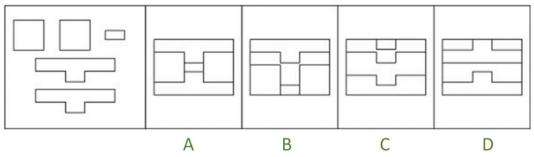
On the pH scale, any substance below what number is considered acid?

A. 0
B. 5
C. 7
D. 15

Quiz (3/20)

Time left to complete this page: **0:25**

Which of the images below *best* solves the problem in the first picture?



Quiz (4/20)

Time left to complete this page: **0:25**

There are 4 red marbles and 8 green marbles in a box. If Emma randomly selects a marble from the box, what is the probability of her selecting a red marble?

A. $\frac{1}{3}$
B. $\frac{1}{2}$
C. $\frac{2}{3}$
D. $\frac{4}{5}$

Quiz (5/20)

Time left to complete this page: **0:25**

In the field of genetics, an organism's appearance is referred to as its?

A. Genes
B. Alleles
C. Genotype
D. Phenotype

Quiz (6/20)

Time left to complete this page: **0:24**

Air is less dense than water because

A. It is lighter.
B. Its molecules are further apart.
C. Its molecules are closer together.
D. It moves quicker.

Quiz (7/20)

Time left to complete this page: **0:25**

If $5^{11} = 5^2 \times 5^m$, what is the value of m?

A. 3
B. 5.5
C. 9
D. 12.5

Quiz (8/20)

Time left to complete this page: **0:24**

In vertebrates, which substance gives red blood cell their colour?

A. carotene
B. dopamine
C. melanin
D. hemoglobin

Quiz (9/20)

Time left to complete this page: **0:26**

How much greater is the value of $3x + 5$ than the value of $3x - 7$?

A. 8
B. 10
C. 12
D. 14

Quiz (10/20)

Time left to complete this page: **0:25**

Which of the following is **not** a blood type?

A. A
B. AB
C. O
D. X

Quiz (11/20)

Time left to complete this page: **0:26**

If $(3.2 + 3.3 + 3.5)w = w$, what is the value of w ?

A. 0
B. $\frac{1}{10}$
C. 1
D. 10

Quiz (12/20)

Time left to complete this page: **0:22**

Which of the following is a characteristic of the automatic nervous system?

A. It regulates the voluntary control of body movements through the skeletal muscles.
B. It regulates involuntary activity in the heart, stomach, lungs, and intestines.
C. It is thought to be the centre of intelligence.
D. It is part of the nervous system consisting of the brain and spinal cord.

Quiz (13/20)

Time left to complete this page: **0:22**

DNA has which of the following structures?

A. Double helix
B. Helix
C. Parabola
D. Hexagon

Quiz (14/20)

Time left to complete this page: **0:25**

What is the latitude of any point on earth's equator?

A. 0
B. 90
C. 180
D. 270

Quiz (15/20)

Time left to complete this page: **0:25**

Which of the following planets is located farthest from the sun?

A. Neptune
B. Saturn
C. Jupiter
D. Uranus

Quiz (16/20)

Time left to complete this page: **0:26**

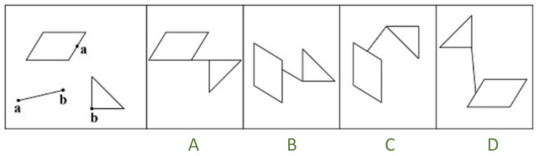
Which layer of the atmosphere is closest to the earth?

A. Mesosphere
B. Stratosphere
C. Troposphere
D. Thermosphere

Quiz (17/20)

Time left to complete this page: **0:21**

Which figure best shows how the objects in the left box will touch if the letters for each object are matched?



Quiz (18/20)

Time left to complete this page: **0:25**

Which of the following is **not** a factor of 90?

A. 5
B. 6
C. 12
D. 15

Quiz (19/20)

Time left to complete this page: **0:25**

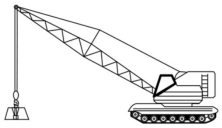
If b does not equal zero and $ab = \frac{b}{4}$, what is the value of a?

A. $\frac{1}{8}$
B. $\frac{1}{4}$
C. $\frac{1}{3}$
D. $\frac{1}{2}$

Quiz (20/20)

Time left to complete this page: **0:23**

Extending the reach of this crane will shift its:



A. Total weight
B. Allowable speed
C. Centre of gravity
D. Centre of buoyancy

End of the Quiz

Thank you for completing the quiz.

Before proceeding to the next part of today's session, you will have to **guess how many questions (out of 20) you answered correctly** in the quiz.

You will receive a bonus of £0.50 if your guess is exactly correct, and a bonus of £0.20 if your guess lies within 3 of the correct number.

How many quiz questions do you think you answered correctly (out of 20)?

What happens next week

Before proceeding with today's session tasks, you will get information about what will happen in Session 2 (next week).

In Session 2, **you will have to decide whether to work on one of two jobs: Job A or Job B**. Both jobs will last for approximately 10 minutes and involve another math and science task. Apart from this, you will not be told anything else about the two jobs until you have to decide which job you will work on.

Prior to making your decision (in Session 2), **you will receive advice from an Adviser regarding which job you should take**. The Adviser will be another Prolific participant. **Unlike you, the Adviser knows what Job A and Job B are.** Hence, **their advice may help you make an informed decision**.

Your bonus in Session 2 will depend on your performance on your chosen job. The Adviser will earn a bonus equal to 50% of your earnings on the job. This means that it is in the Adviser's best interest to advise you to take the job that will maximize your earnings, because this will maximize their earnings too.

Based on the instructions provided above, please indicate whether each statement below is True or False:

The Adviser will decide whether you work on Job A or Job B.

- ○ *True*
- ○ *False*

The Adviser will advise you on whether to opt for Job A or Job B.

- ○ *True*
- ○ *False*

The Adviser will have more information than you about what Job A and Job B are when giving you advice.

- ○ *True*
- ○ *False*

Self-assessment questionnaire

You will now answer a **self-assessment questionnaire**, which will include several questions **about how well you think you did on the math and science quiz** today.

To help the Adviser give you the best possible advice about which job to choose in Session 2, the **Adviser will see one or more of your answers** in this questionnaire.

Based on the instructions provided above, please indicate whether the statement below is True or False:

The Adviser will see one or more of your answers in the self-assessment questionnaire.

- ○ *True*
- ○ *False*

Self-assessment questionnaire (1/4)

In the space provided below, please describe (using 2 to 3 sentences) how well you think you performed today on the math and science quiz.

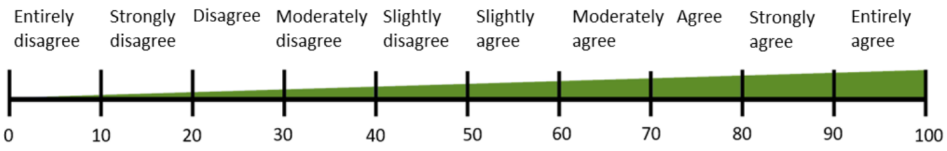
Self-assessment questionnaire (2/4)

Please indicate how good you think your performance was on the quiz.

Terrible	Very poor	Poor	So-so	Good	Very good	Exceptional
○	○	○	○	○	○	○

Self-assessment questionnaire (3/4)

Please indicate the extent to which you agree with each of the following statements, on a scale from 0 (entirely disagree) to 100 (entirely agree).



I performed well on the quiz.

0 10 20 30 40 50 60 70 80 90 100

Your answer:

I would apply for a job that required me to perform well on the quiz.

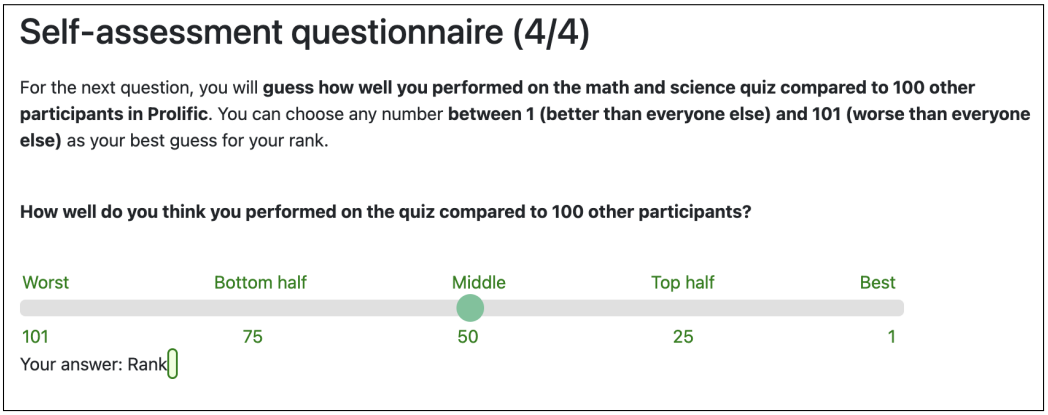
0 10 20 30 40 50 60 70 80 90 100

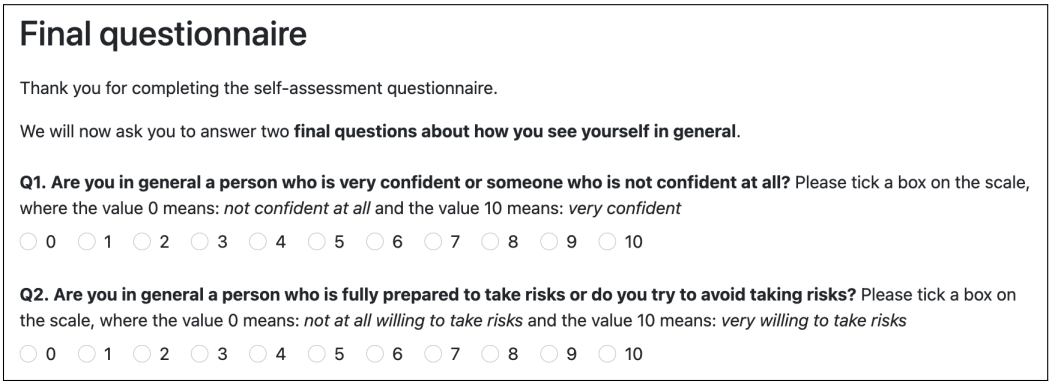
Your answer:

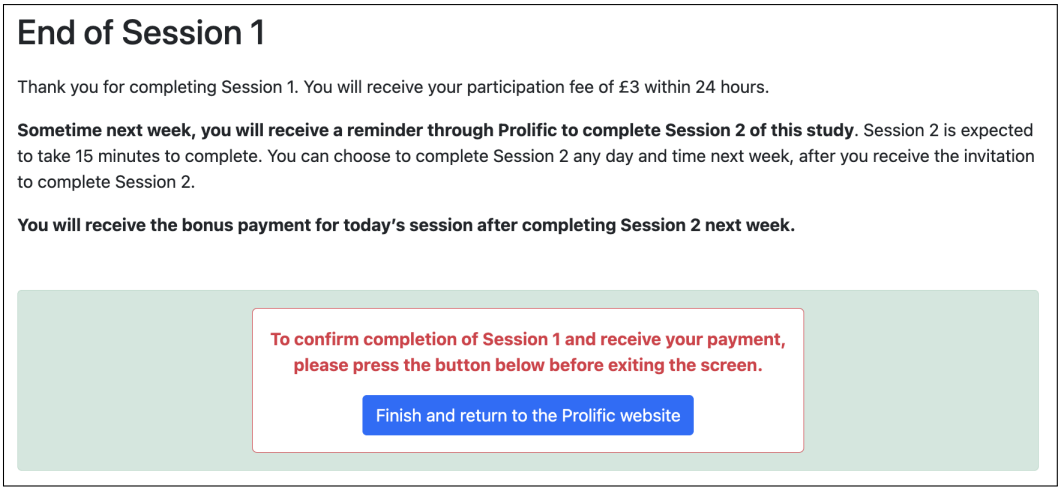
I would succeed in a job that required me to perform well on the quiz.

0 10 20 30 40 50 60 70 80 90 100

Your answer:







11.2 Part 2 Adviser Session

Baseline treatment

Welcome

Overview: In this study, you will be taking the role of an Adviser. Your main task will be to advise a Worker (another Prolific participant), on which job to complete in this study. You will also be asked to answer a short survey. The study is expected to take 20 minutes to complete.

Your payment: You will receive a participation fee of £3 within 24 hours of completing the study. In addition, there is a 20% chance that your advice will be sent to the Worker. If so, you will receive a bonus, depending on the performance of the Worker who received your advice. Your bonus will be paid out to you after the Worker has completed their job (within two weeks from today).

Please download and read through information about the study and the consent form here.

Based on the instructions provided above, please indicate whether each statement below is True or False:

You will advise the Worker on which job to choose.

- ○ *True*
- ○ *False*

You will receive any bonus payment within 24 hours of completing the study.

- ○ *True*
- ○ *False*

You will receive any bonus payment with a delay, after the Worker has performed their job.

- ○ *True*
- ○ *False*

The Worker's Job

In this study, **you will be matched to a Worker** who is another participant on Prolific. The **Worker will have to choose between two jobs: Job A (Advanced) and Job B (Basic)**. Both jobs require the Worker to answer **20 multiple choice math and science questions**, with a limited time of 30 seconds per question. You can see example questions below.

Example 1

If $5^5 = 5^2 \times 5^m$, what is the value of m?

A. 3
B. 5.5
C. 9
D. 12.5

Example 2

Which layer of the atmosphere is closest to the earth?

A. Mesosphere
B. Stratosphere
C. Troposphere
D. Thermosphere

Job A (Advanced) and Job B (Basic) require the Worker to answer the same 20 questions within the same time limit but **their bonuses differ**, as follows:

- **Job A (Advanced)**: Worker receives a bonus of **£2** if they get **8 or more correct answers**, and no bonus otherwise.
- **Job B (Basic)**: Worker receives a bonus of **£1** if they get **5 or more correct answers**, and no bonus otherwise.

Advice

Your task is to give advice to the Worker regarding which job to take. When the Worker has to decide whether to take Job A (Advanced) or Job B (Basic), they do not know what task each job entails nor the associated bonus. Hence, your advice may help the Worker make an informed decision. If your advice is sent to the Worker, they will see it before making their decision.

A few days ago, the Worker took part in another study in which they completed a math and science quiz that is similar to the job they will have to work on. They **then answered self-assessment questions regarding their performance** in that quiz, without having received any feedback about how well they actually did.

To help you give the Worker the best possible advice, **you will receive the Worker's answers to the self-assessment questions**, where the Worker describes how well they think did in the math and science quiz.

Your Bonus

There is a 20% chance that your advice is sent to the Worker. If so, you will receive a bonus equal to 50% of the earnings of the Worker. Here are the possible outcomes:

- Worker chooses **Job A (Advanced)**: **you earn a bonus of £1** if they have **8 or more correct answers**, and no bonus otherwise.
- Worker chooses **Job B (Basic)**: **you earn a bonus of £0.50** if they have **5 or more correct answers**, and no bonus otherwise.

It is in your best interest to advise the Worker to take the job that will maximize their earnings, because this will maximize your earnings too.

Based on the instructions on the previous pages, please indicate whether each statement below is True or False:

You decide whether the Worker works on Job A (Advanced) or Job B (Basic).

- ○ *True*
- ○ *False*

You give advice to the Worker on whether to opt for Job A (Advanced) or Job B (Basic).

- ○ *True*
- ○ *False*

Your bonus depends on the Worker's performance on their chosen job.

- ○ *True*
- ○ *False*

Nudge Treatment Screen

Information

On the next page, you will be able to give advice to the Worker. Below are some facts that you may want to keep in mind when giving advice:

1. *Using data from more than 4,000 participants in online labour markets, researchers at Harvard University found that women rated their performance in a math & science quiz lower than men, despite similar performance levels. As a result, women who were hired performed better than men on average.*

2. *Using data from more than 3,600 people, researchers at the University College London report that men are more confident in the self-assessment of their abilities compared to women. This gender gap in self-assessment contributes to the gender gap in who ends up in a top job.*

3. *An internal study by the IT firm Hewlett Packard found that men who apply for a job meet only 60% of the qualifications, but women who apply meet 100% of them.*

Based on the information provided above, please answer the following questions:

Men were more confident in their self-assessment in the previous studies.

- ○ True
- ○ False

Suppose you are comparing a male and female worker with the same self-assessment of their performance. Based on previous research, which of the two workers is likely to have performed better?

- ○ Male
- ○ Female
- ○ Neither

Do you think the information you received will affect your advice?

- ○ Yes
- ○ No

In one or two sentences, please briefly justify your answer to the previous question.

Advice

You will now advise the following Worker:

Name: **Lucy**

Advice

Below are Lucy's answers to the self-assessment questions. Please read her answers carefully before giving your advice.

Name: **Lucy**

Q1. Please describe how well you think you performed on the math and science quiz.

Lucy's answer:

I don't think I did very well. I didn't have time to think for very long what the answers are so i was a bit under pressure and out of time on quite a few questions

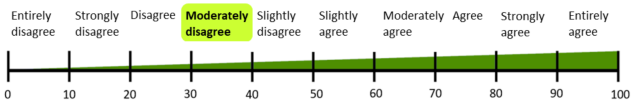
Q2. How many quiz questions do you think you answered correctly (out of 20)?

Lucy's answer: **15**

Q3. To which extent do you agree with each of the following statements? Please indicate your answer on a scale from 0 (entirely disagree) to 100 (entirely agree).

I performed well on the quiz.

Lucy's answer: **30**

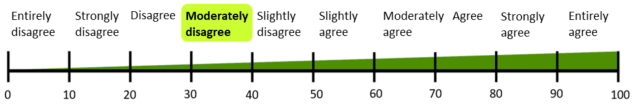


I would apply for a job that required me to perform well on the quiz.

Lucy's answer: **40**

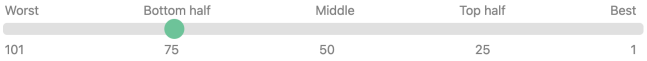
I would succeed in a job that required me to perform well on the quiz.

Lucy's answer: **30**



Q4. How well do you think you performed on the quiz compared to 100 other participants? Please give an answer **between 1 (better than everyone else) and 101 (worse than everyone else).**

Lucy's answer: Rank **75**



Please submit your advice to Lucy by answering the questions below. Your answer to one of these questions will be sent to Lucy and help her decide between **Job A (Advanced)** and **Job B (Basic)**.

Q1. Which advice do you want to send to Lucy?	**Definitely Job B**	**Most likely Job B**	**No advice**	**Most likely Job A**	**Definitely Job A**
Please indicate your answer:	○	○	○	○	○

Q2. Which job do you advise that Lucy chooses?

○ Job A (Advanced) ○ Job B (Basic)

On a scale from 0 to 10, how confident are you that your advice in **Q2** is the best for Lucy?

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

Advice

Now imagine that instead of having to choose between Job A (Advanced) and Job B (Basic), Lucy would choose her performance goal as follows:

- **Goal 1**: **£0.20** for **1** or more correct answers
- **Goal 2**: **£0.40** for **2** or more correct answers
- **Goal 3**: **£0.60** for **3** or more correct answers
- (...)
- **Goal 9**: **£1.80** for **9** or more correct answers
- **Goal 10**: **£2.00** for **10** or more correct answers
- **Goal 11**: **£2.20** for **11** or more correct answers
- (...)
- **Goal 18**: **£3.60** for **18** or more correct answers
- **Goal 19**: **£3.80** for **19** or more correct answers
- **Goal 20**: **£4.00** for **20** correct answers

This means that Lucy would earn a bonus of **£0.20 times the selected performance goal if her performance met or exceeded her goal, and no bonus otherwise**. You would always earn a bonus equivalent to 50% of Lucy's earnings.

Which goal would you advise Lucy to choose in this case?

See Lucy's answers to self-assessment questions again

Advice

Thank you for submitting your advice.

Next, you will be asked to give advice to four additional Workers. Just like for your previous Worker (Lucy), there is **a chance that these workers will receive your advice**. Overall, there is a 20% chance that your advice is sent to a Worker. Each of the five Workers (including the previous Worker, Lucy) is equally likely to be selected.

[Adviser sees 4 additional worker portfolios, sequentially]

Objective Treatment

Advice

Thank you for submitting your advice.

Next, you will see the portfolios of each of the Workers again. This time, in addition to their answers to the self-assessment questions, you will also **learn about their actual score in the previous math and science quiz**, and you will have the possibility to **change your advice** about which job they should choose.

Based on the instructions provided above, please indicate whether each statement below is True or False:

You will see the actual score of the Workers in the math and science quiz.

- ○ *True*
- ○ *False*

You will be able to change your advice.

- ○ *True*
- ○ *False*

Advice to 1st Worker

Below are Lucy's answers to the self-assessment questions and **new information about her score** in the previous math and science quiz.

Name: **Lucy**

NEW: Lucy got a **score of 6/20** in the previous math and science quiz.

Q1. Please describe how well you think you performed on the math and science quiz.

Lucy's answer:

I don't think I did very well. I didn't have time to think for very long what the answers are so i was a bit under pressure and out of time on quite a few questions

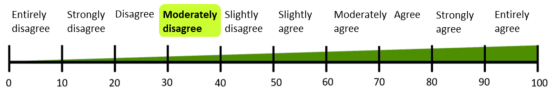
Q2. How many quiz questions do you think you answered correctly (out of 20)?

Lucy's answer: **15**

Q3. To which extent do you agree with each of the following statements? Please indicate your answer on a scale from 0 (entirely disagree) to 100 (entirely agree).

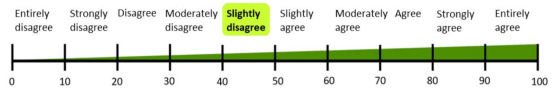
I performed well on the quiz.

Lucy's answer: **30**



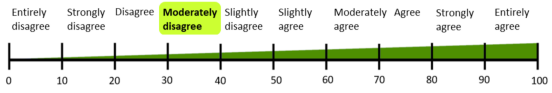
I would apply for a job that required me to perform well on the quiz.

Lucy's answer: **40**



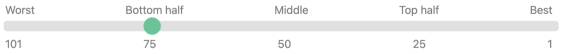
I would succeed in a job that required me to perform well on the quiz.

Lucy's answer: **30**



Q4. How well do you think you performed on the quiz compared to 100 other participants? Please give an answer **between 1 (better than everyone else) and 101 (worse than everyone else).**

Lucy's answer: Rank 75



Below is the advice you previously gave to Lucy. You can either change or maintain your previous advice. The only new information you received is that Lucy's actual score on the previous math and science quiz was 6/20. If you wish to change your advice, please select your new advice, and then press Next. If you do not wish to change your advice, please press Next.

Q1. Which advice do you want to send to Lucy?	**Definitely Job B**	**Most likely Job B**	**No advice**	**Most likely Job A**	**Definitely Job A**
Please indicate your answer:	◉	○	○	○	○

Q2. Which job do you advise that Lucy chooses?

○ Job A (Advanced) ◉ Job B (Basic)

On a scale from 0 to 10, how confident are you that your advice in **Q2** is the best for Lucy?

○ 0 ○ 1 ○ 2 ◉ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

Q3. Which performance goal would you advise Lucy to choose?

Recall that Lucy would earn a bonus of £0.20 times the selected performance goal if her performance met or exceeded her goal, and no bonus otherwise.

1 ⌄

Final questionnaire (1/2)

Thank you for submitting your advice. We will now ask you to answer a **final questionnaire**.

1) Please indicate your age:

----

2) Are you:

○ Female ○ Male ○ Other

If you chose *Other*, please specify:

3) What is the highest degree or level of education you have completed?

----

4) What is your current occupation?

☐ Studying
☐ Employed (full-time or part-time)
☐ Currently looking for work
☐ Not in the labour force
☐ Retired
☐ Other
☐ Prefer not to say

5) What is (or was) your full job title? *Examples: Retail assistant, Civil engineer, Cleaner, Nurse, Primary School Teacher*

6) How much do you get paid for your job, before tax and other deductions are taken out? If your pay before tax changes from week to week, think about what you earn on average.

----

7) Now please consider your **household income** *(your household includes everyone who lives with you regularly and with whom you share money)*. How do you feel about your **household income** nowadays?

----

8) Are you in general a person who is very confident or someone who is not confident at all? Please tick a box on the scale, where the value 0 means: *not confident at all* and the value 10 means: *very confident*

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

9) Are you in general a person who is fully prepared to take risks or do you try to avoid taking risks? Please tick a box on the scale, where the value 0 means: *not at all willing to take risks* and the value 10 means: *very willing to take risks*

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10

10) Who do you think, if anybody, would perform better on the math and science quiz?

Women a lot better	**Women better**	**Women slightly better**	**Both equally well**	**Men slightly better**	**Men better**	**Men a lot better**
○	○	○	○	○	○	○

11) Who do you think, if anybody, would be more confident in their self-assessment?

Women a lot more confident	**Women more confident**	**Women slightly more confident**	**Both equally confident**	**Men slightly more confident**	**Men more confident**	**Men a lot more confident**
○	○	○	○	○	○	○

12) Is there any feedback you would like to give us about the study?

Please **pick a name** for yourself from the list below. (They will be used when communicating your advice to the Worker.)

- ○ Stephanie
- ○ Stephen
- ○ Felicia
- ○ Sam
- ○ Louis
- ○ Louisa
- ○ Alexander
- ○ Philip
- ○ Julian
- ○ Julia
- ○ Freddie
- ○ Leo
- ○ Charlie
- ○ Alexandra
- ○ Alex
- ○ Philippa
- ○ Theo
- ○ Felix

End of the Study

Thank you for completing this study. Your payment is as follows:

Participation fee: £3. You will receive this payment within 24 hours.

In addition to the participation fee, if your advice is sent to a Worker, your may receive a bonus, depending on your worker's performance in their chosen job, which will be paid out to you within 2 weeks.

Thank you for your participation.

To confirm completion and receive your payment, please press the button below before exiting the screen.

Finish and return to the Prolific website

11.3 Part 3 Worker Session

Welcome back!

Thank you for returning to the study!

Last week you did a math and science quiz and answered questions about your performance in the quiz.

Today you will have to decide whether to work on one of two jobs: Job A (the advanced job) and Job B (the basic job). Both jobs will last for approximately 10 minutes and require you to work on a math and science task. Apart from this, you will not be told anything else about the two jobs until after you have decided which job you will work on.

Prior to making your decision, you will receive advice from an Adviser regarding which job you should take. The Adviser is another Prolific participant. The Adviser has seen some of your answers to last week's self-assessment questions and knows what the two jobs are. Because the Adviser earns a bonus equal to 50% of your earnings in the job, it is in the Adviser's best interest to help you make an informed decision.

Your payment: You will receive a participation fee of £3, your bonus for Session 1 and your bonus for today's session, within 24 hours of completing today's session.

Based on the instructions provided above, please indicate whether each statement below is True or False:

The Adviser will advise you on whether to take Job A (Advanced) or Job B (Basic).

- ○ *True*
- ○ *False*

The Adviser has more information than you about what Job A (Advanced) and Job B (Basic) are.

- ○ *True*
- ○ *False*

You will learn what your chosen job (A or B) entails after you have made your choice and before you start your chosen job.

- ○ *True*
- ○ *False*

Advice & Decision

Your Adviser, who chose the name John, sends you the following advice:

The best job for you is definitely Job A.

Please indicate **your decision** about which job you will work on:

- ○ Job A (Advanced)
- ○ Job B (Basic)

Advice & Decision

Your Adviser, who chose the name John, sends you the following advice:

You should choose Job A.

He rated his **confidence** level in his advice as **9 out of 10**.

Please indicate **your decision** about which job you will work on:

- ○ Job A (Advanced)
- ○ Job B (Basic)

Job A

You decided to work on Job A (Advanced). The job consists of a **quiz with 20 multiple choice questions testing your skills in math and science**. You will have up to 30 seconds to answer each question.

For this job, you receive a bonus of **£2 if you solve at least 8 questions correctly**. If your score is lower than 8, you will not receive any bonus.

When you are ready, please press the Next button below to start the task.

Job B

You decided to work on Job B (Basic). The job consists of a **quiz with 20 multiple choice questions testing your skills in math and science**. You will have up to 30 seconds to answer each question.

For this job, you receive a bonus of **£1 if you solve at least 5 questions correctly**. If your score is lower than 5, you will not receive any bonus.

When you are ready, please press the Next button below to start the task.

Quiz (1/20)

Time left to complete this page: **0:25**

If $(3.2 + 3.3 + 3.5)w = w$, what is the value of w ?

A. 0

B. $\frac{1}{10}$

C. 1

D. 10

Quiz (2/20)

Time left to complete this page: **0:25**

In the field of genetics, an organism's appearance is referred to as its?

A. Genes

B. Alleles

C. Genotype

D. Phenotype

Quiz (3/20)

Time left to complete this page: **0:25**

Which layer of the atmosphere is closest to the earth?

A. Mesosphere
B. Stratosphere
C. Troposphere
D. Thermosphere

Quiz (4/20)

Time left to complete this page: **0:24**

Air is less dense than water because

A. It is lighter.
B. Its molecules are further apart.
C. Its molecules are closer together.
D. It moves quicker.

Quiz (5/20)

Time left to complete this page: **0:24**

What is the latitude of any point on earth's equator?

A. 0
B. 90
C. 180
D. 270

Quiz (6/20)

Time left to complete this page: **0:25**

DNA has which of the following structures?

A. Double helix
B. Helix
C. Parabola
D. Hexagon

Quiz (7/20)

Time left to complete this page: **0:19**

If b does not equal zero and $ab = \frac{b}{4}$, what is the value of a?

A. $\frac{1}{8}$

B. $\frac{1}{4}$

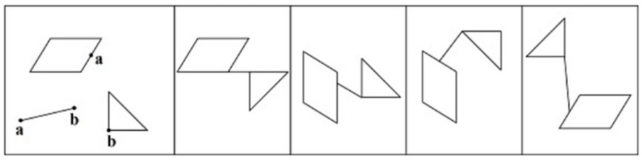
C. $\frac{1}{3}$

D. $\frac{1}{2}$

Quiz (8/20)

Time left to complete this page: **0:25**

Which figure best shows how the objects in the left box will touch if the letters for each object are matched?



Quiz (9/20)

Time left to complete this page: **0:24**

If n is a positive integer divisible by 7, and if $n < 70$, what is the greatest possible value of n ?

A. 49

B. 56

C. 63

D. 69

Quiz (10/20)

Time left to complete this page: **0:25**

There are 4 red marbles and 8 green marbles in a box. If Emma randomly selects a marble from the box, what is the probability of her selecting a red marble?

A. $\frac{1}{3}$

B. $\frac{1}{2}$

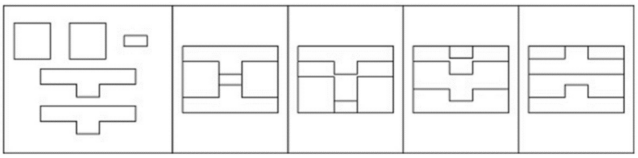
C. $\frac{2}{3}$

D. $\frac{4}{5}$

Quiz (11/20)

Time left to complete this page: **0:25**

Which of the images below *best* solves the problem in the first picture?



Quiz (12/20)

Time left to complete this page: **0:20**

Which of the following is **not** a factor of 90?

A. 5
B. 6
C. 12
D. 15

Quiz (13/20)

Time left to complete this page: **0:26**

How much greater is the value of $3x + 5$ than the value of $3x - 7$?

A. 8
B. 10
C. 12
D. 14

Quiz (14/20)

Time left to complete this page: **0:26**

On the pH scale, any substance below what number is considered acid?

A. 0
B. 5
C. 7
D. 15

Quiz (15/20)

Time left to complete this page: **0:26**

Which of the following is a characteristic of the automatic nervous system?

A. It regulates the voluntary control of body movements through the skeletal muscles.
B. It regulates involuntary activity in the heart, stomach, lungs, and intestines.
C. It is thought to be the centre of intelligence.
D. It is part of the nervous system consisting of the brain and spinal cord.

Quiz (16/20)

Time left to complete this page: **0:24**

Which of the following is **not** a blood type?

A. A
B. AB
C. O
D. X

Quiz (17/20)

Time left to complete this page: **0:24**

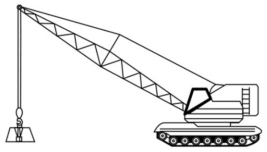
If $5^{11} = 5^2 \times 5^m$, what is the value of m?

A. 3
B. 5.5
C. 9
D. 12.5

Quiz (18/20)

Time left to complete this page: **0:24**

Extending the reach of this crane will shift its:



A. Total weight
B. Allowable speed
C. Centre of gravity
D. Centre of buoyancy

Quiz (19/20)

Time left to complete this page: **0:23**

In vertebrates, which substance gives red blood cell their colour?

A. carotene
B. dopamine
C. melanin
D. hemoglobin

Quiz (20/20)

Time left to complete this page: **0:25**

Which of the following planets is located farthest from the sun?

A. Neptune
B. Saturn
C. Jupiter
D. Uranus

End of the quiz

Thank you for completing the quiz.

Before ending today's session, is there any feedback you would like to give us about the study?

End of the study!

Thank you for completing this study. Your payment for today's session is as follows:

Bonus from Session 1: You answered 18 correct answers in the quiz in Session 1, for a bonus of £1.80. In addition, you will receive £0.00 based on the accuracy of your guess regarding your performance in Session 1.

Bonus from Session 2: You answered 5 correct answers in the quiz, for a bonus of £1.00.

Including your £3 participation fee, your total payment for today's session is: £5.80. You will receive this payment within 24 hours.

To confirm completion of the study and receive your payment, please press the button below before exiting the screen.

Finish and return to the Prolific website