lifecoursecentre
WORKING PAPER SERIES

SUBJECTIVE JUDGMENT AND GENDER BIAS IN ADVICE: EVIDENCE FROM THE LABORATORY

Juliana Silva-Goncalves
School of Economics, University of Sydney

Roel van Veldhuizen
Department of Economics, Lund University

No. 2020-29
December 2020

Australian Government
Australian Research Council
The University of Queensland Australia
The University of Sydney
The University of Melbourne
The University of Western Australia

NON-TECHNICAL SUMMARY

Despite recent progress, there are still sizable gender gaps in labor market outcomes. Women and men choose different occupations, and women earn lower wages and are less likely to attain leadership positions even within a given profession. Therefore, better understanding and reducing gender gaps in the labor market remains an important policy goal.

In this study we test whether advice can contribute to sustain the gender gap in labour market outcomes. Many of the key decisions we make in our lifetimes are hard decisions that are made under a great deal of uncertainty. An example is the decision of which career to pursue. Many of us often rely on the advice from our family, teachers, career counselors, colleagues and managers to make that decisions. Although such advice can be helpful in many cases, we hypothesise that it may also generate or help sustain gender gaps in the labor market if advisers give advice in a gender-biased way.

We conduct an economic experiment where advisers advise workers to choose between a more ambitious and a less ambitious task based on subjective information about the quality of the worker. We expected female workers to be less confident and advisers to hold gender stereotypes, leading to a gender bias in advice. However, we find no evidence that women are less confident or that advice is gender-biased. Our results contribute to our understanding of the mechanisms driving gender differences in the labour market. They also call for caution when making general interpretations of research findings pointing to a gender bias in specific settings.

ABOUT THE AUTHORS

Juliana Silva-Goncalves is a research fellow of the Centre of Excellence for Children and Families over the Life Course at the School of Economics at the University of Sydney. Her research focuses on behavioural and experimental economics applied to educational outcomes, discrimination and inequality. Recent work has been published in the *Journal or Economic Behaviour & Organization*, *Journal of Economic Psychology*, and *Economic Record*. Email: Juliana.goncalves@sydney.edu.au

Roel van Veldhuizen is an Assistant Professor at the Department of Economics at Lund University in Sweden. His research uses behavioral and experimental economics to study gender differences, unethical behavior and trust. His research has been published in the *Journal of Public Economics*, the *Journal of the European Economic Association* and *Games and Economic Behavior*, among others. Email: rrvanveldhuizen@gmail.com

ACKNOWLEDGEMENTS

We gratefully acknowledge funding from the ARC Life Course Centre. We also thank Andy Schotter and seminar participants at the WZB Berlin for their very helpful comments

ABSTRACT

Better understanding and reducing gender gaps in the labor market remains an important policy goal. We study the role of advice in sustaining these gender gaps using a laboratory experiment. In the experiment, “advisers” advise “workers” to choose between a more ambitious and a less ambitious task based on the worker’s subjective self-assessment. We expected female workers to be less confident and advisers to hold gender stereotypes, leading to a gender bias in advice. However, we find no evidence that women are less confident or that advice is gender-biased. Our results contribute to our understanding of the mechanisms driving gender differences in the labor market. They also call for caution when making general interpretations of research findings pointing to a gender bias in specific settings.

Keywords: advice; subjective judgment, gender bias

Suggested citation: Silva-Goncalves, J. & van Veldhuizen, R. (2020). Subjective Judgment and Gender Bias in Advice: Evidence from the Laboratory'. Life Course Centre Working Paper Series, 2020-29. Institute for Social Science Research, The University of Queensland.

1. Introduction

Despite recent progress, there are still sizable gender gaps in labor market outcomes (Goldin, 2014). Women and men choose different occupations, and women earn lower wages and are less likely to attain leadership positions even within a given profession (Blau and Kahn, 2017; Ganguli et al., 2020). These results have prompted a large body of research investigating potential drivers of these gender gaps, as nicely summarized by Blau and Kahn (2017) and Bertrand (2011).

In this paper, we study the role of advice in sustaining the gender gap in labor market outcomes. Many of the key decisions we make in our lifetimes are hard decisions that are made under a great deal of uncertainty. An example is the decision of which career to pursue. According to standard economic theory, we should make this decision by comparing the lifetime expected utility of all possible options and select the most favorable one. In reality, very few of us have the information required to make this decision alone. Instead, we often rely on the advice from our family, teachers, career counselors, colleagues and managers. Although such advice can be helpful in many cases, it may also generate or help sustain gender gaps in the labor market if advisers give advice in a gender-biased way. Such biases are particularly likely to emerge since even the best advisers do not have perfect information on all available options and therefore have to rely on their subjective judgment. Recent research suggests that gender biases are particularly likely to emerge in settings characterized by subjective judgment (e.g., Barron et al., 2020; Bohren et al., 2019; Carlana, 2019; Sarsons et al., 2020). If advice is gender-biased, then this bias may help to explain the persistence of gender differences in labor market outcomes.

Studying gender differences in advice requires overcoming two key empirical challenges. First, advice is often private and qualitative, and therefore difficult to measure and quantify. Second, in studying gender differences in advice, there is also the typical identification problem that any gender differences in advice could be driven by gender differences in (unobserved) characteristics. These issues may help explain why, despite the importance and ubiquity of advice, it has traditionally received little attention in the literature on gender gaps.

In this study, we introduce a novel identification strategy that allows us to circumvent both issues. To address the measurement issue, we run a laboratory experiment in which "advisers" advise "workers" to opt for either a more ambitious job or a less ambitious job, in terms of possible earnings and difficulty. Both jobs involve performing a stereotypically-male math task, for which previous laboratory studies have typically observed no gender difference in actual performance. The advisers have an informational advantage relative to the workers about the characteristics of both jobs, making their advice valuable for the workers. The advisers also face incentives to give good advice, by virtue of receiving a bonus of 50% of the worker's earnings. Prior to giving advice, advisers see the workers' portfolio in which workers subjectively describe their general skills in math both qualitatively, by writing a motivation letter, and quantitatively by self-rating their skills. These procedures give us direct measures of both the advice itself and all the subjective information used by advisers to make their decision.

To address the fact that gender is unlikely to be independent from relevant worker characteristics, we use the following approach. We assign each adviser to a single worker, whose gender was revealed in a subtle way. The adviser sees and gives advice to that worker's portfolio, as well as to the portfolios of four other workers, without knowing which of the five portfolios belongs to their worker. Because only the advice given to their actual worker will ever be sent, it is optimal for the adviser to treat all five portfolios as-if they belong to their actual worker. This feature implies that a given portfolio is evaluated both by advisers assigned to a female worker, incentivized to judge all portfolios as if they belong to a woman, and advisers assigned to a male worker, incentivized to judge all portfolios as if they belong to a man. In other words, the same portfolio is evaluated by advisers incentivized to treat them as 'male' or 'female' portfolios respectively, providing a source of exogenous variation in the gender of a given portfolio.

We ran the laboratory experiment at the University of Sydney in 2019 with 190 participants. We pre-registered our design, identification strategy and analysis plan in the American Economic Association Randomized Controlled Trials registry and reprinted it in Appendix B for convenience. Overall, we expected advisers to be gender-biased and female workers to be less confident, leading to a gender-bias in advice. However, we find no evidence of a gender bias in advice in our sample. Instead, we find that advisers appear to rely on the workers' motivation letter and self-ratings to determine their advice, and the gender of their assigned worker does not seem to play a role. We show that this result is robust across a number of pre-registered specifications and is not the result of a lack of

power. We do find some suggestive evidence that male advisers (but not female advisers) give less ambitious advice to female workers, but this effect only appears in one specification and is not robust to adjustments for multiple testing.

Our pre-registered power calculations suggest that our result is not due a lack of statistical power and appears in a setting where a gender bias in advice might well be expected. For one, we used a task (addition problems) in which men are thought to be better than women, which could induce advisers to give different advice to female workers. For another, we study advice based on subjective judgment, creating ample scope for underlying biases to influence advice. Finding no gender difference in advice in this setting implies that such a bias may also not be found in settings that are less favorable to a gender bias.

Our study builds on a large literature studying the reasons why men and women sort into different careers.[1] Gender stereotypes (Carlana, 2019; Nürnberger et al., 2016), lower earnings expectations for women than men in top career tracks (Blau & Kahn, 2017) and greater childbearing responsibilities for women are all thought to explain why women are less likely to pursue and succeed in ambitious career tracks (Charlesworth & Banaji, 2019). Discrimination may also be a contributing factor, but here the evidence is more mixed (Bertrand and Mullainathan, 2004; Blau & Kahn, 2017; Breda and Ly, 2015; Moss-Racusin et al., 2012).[2] Gender differences in preferences and beliefs such as risk aversion, competitiveness and optimism may also explain gender differences in career choices (Baldiga, 2014; Buser et al., 2014; Ors et al., 2013; Reuben et al., 2012). Recent evidence from the field suggests that advice may also contribute to the gender gap in career progression, with women benefiting less from advice in a same-gender mentorship program compared to men (Ellis and Gershenson, 2020) and compared to mix-gender mentorship programs (AlShebli et al., 2020).

Our study also builds on previous experimental research on advice. One branch of this literature examines how exogenous variation in advice affects, e.g., college enrolment (Hoxby and Turner, 2015), truth-telling in matching markets (Guillen and Hakimov, 2018)

---

[1] See Blau and Kahn (2017) for a review of some of the mechanisms. Charlesworth & Banaji (2019) provides a recent survey of the reasons for the persistent gender gap in STEM.

[2] Some studies find no evidence of gender discrimination (Bertrand and Mullainathan, 2004; Booth et al., 2012; Edo et al., 2019), including among high skilled job candidates (Duguet et al., 2012). Breda and Ly (2015) even find evidence of positive discrimination of women among high achievers. In the context of a French selective higher education institution, they report that women are favored by examiners and benefit from a grade premium in male-related fields (math and physics).

and tournament entry (Kessel, Mollerstrom and Van Veldhuizen, 2017). Another branch of the literature examines how institutional features affect advice sent by participants to subsequent generations of participants (e.g., Çelen, Kariv and Schotter, 2010; Chaudhuri, Schotter and Sopher 2009; Schotter and Sopher 2003, 2007). While most of this work does not look explicitly at gender, notable exceptions are Brandts, Groenert and Rott (2015) and Brandts and Rott (2020), who study advice in the context of tournament entry and, like in our study, show that advisers give similar advice to men and women and advice therefore has no impact on the gender gap in tournament entry. A key feature that is distinct from our study is that advisers know about the performance of the advisee. In our study, advisers have to rely on their own and the advisee's subjective judgment, potentially increasing the scope for gender-biased advice.

Our study also relates to previous work on the role of subjectivity in generating biased judgments. Previous research shows that unethical and biased behavior becomes more prevalent in settings characterized by subjective judgment (Babcock et al., 1995; Dana, Weber and Kuang, 2007; Exley, 2016; Gneezy et al., 2019, 2020). When it comes to gender, Bohren et al. (2019) find that questions asked by inexperienced women on online STEM-related forums receive lower (subjective) ratings than those asked by inexperienced men, whereas questions asked by experienced women receive higher ratings than those asked by experienced men. Krawczyk and Smyk (2016) find that academic papers with a female author are evaluated less favorably (by non-experts) than academic papers with a male author. Similarly, Özgümüs et al. (2020) find that the subjective evaluation of teaching materials is worse when male evaluators believe that the materials belong to a female lecturer instead of a male lecturer. Relatedly, Sarsons et al. (2020) show that women receive less credit for group work when objective contributions cannot be easily established.

Our study makes several contributions to the economics literature. Firstly, it extends the literature on gender gaps by exploring the role of advice in sustaining them. Secondly, it contributes to the literature on the role of subjectivity in generating biased judgment by examining the role of subjective judgment in generating a gender gap in advice. Finally, our study makes a methodological contribution to the experimental literature on gender

differences by introducing a novel method that allows us to generate exogenous variation in personal attributes like gender without the use of deception.[3]

The paper proceeds as follows. In Section II we detail our experimental design and proceedings, in Section III we describe our identification strategy and hypothesis, in Section IV we present our results and in Section V we discuss the implications of our findings.

2. Experimental Design

In this section, we describe our design in a sequential order. Figure 1 provides an overview of all parts of our design. We also briefly describe the procedures we followed for each session.

Part 1: Addition Task

We randomly assigned all participants to a computer terminal upon entering the laboratory. Participants were informed that half of them would be asked to work on an addition task, and the other half would have to wait for the first group to finish. The addition task consisted of adding five two-digit numbers for five minutes; participants received 1 dollar per correct answer.[4] The group of participants working on the task would go on to play the role of worker in the remainder of the experiment; the other participants would play the role of adviser. The allocation of the adviser and worker roles was only revealed to participants at the start of part 2. This first part was aimed at giving us an objective measure of workers' skill at the task and giving workers some idea of what task to expect in part 2.

We chose the addition task because it is thought to be a stereotypically male task. In line with this idea, previous studies report that despite the absence of a gender gap in performance, men are typically more optimistic than women about their performance in this task (e.g., Gillen, Snowberg and Yariv, 2019; Niederle & Vesterlund, 2007; Van Veldhuizen, 2020). If our participants share these stereotypes, this could induce advisers to give different advice to male and female workers.

---

[3] The most common method to exogenously assign personal attributes is the one followed in audit studies which consists of creating fictitious individual profiles (see, for example, Bertrand and Mullainathan, 2004; Bohren et al., 2019; Özgümüs et al., 2020). Other studies use vignettes which also involve creating fictitious individual profiles but do not involve deception because study subjects are aware that they are making hypothetical decisions based on fictitious profiles (see, for example, Kübler et al., 2018; Tobback et al., 2020). Krawczyk and Smyk (2016) exogenously assign gender to academic papers, without the use of deception, by letting judges know about the gender of one of the authors which could either be male or female since all papers were co-authored by at least one man and one woman. Some judges evaluated the paper knowing that it has been co-authored by a woman, whereas other judges evaluated the same paper knowing that it had been co-authored by a man.

[4] All earnings in our experiment are expressed in Australian dollars. At the time of the experiment, 1 AUD = 0.6 EUR/0.7 USD.

Figure 1: Overview of the Experimental Design

	Worker	Adviser	Duration
Part 1	Works on addition task	Waits	5 min.
Part 2a	Read instructions		~5 min.
Part 2b	Writes the motivation letter and rates own math skills	Waits and reads additional instructions	10 min.
Part 2c	Confirm presence in the lab		~10 min.
Part 2d	Waits	Sees the worker portfolios and determines advice	~10 min.
Part 2e	Sees advice and decides on which job to take	Waits	~3 min.
Part 2f	Works on addition task	Rates worker' chance of success	10 min.
Earnings	Receive information on earnings		~2 min.
Questionnaire	Answer a questionnaire		~5 min.
IAT	Complete an Implicit Association Test		~5 min.

Part 2a: Instructions

Following part 1, all participants received a new set of instructions that informed them about their role in the second part of the experiment (worker or adviser) and their tasks. In particular, we informed all participants that workers would need to opt between working on an 'advanced' (A) and a 'basic' (B) job. We also informed them that each adviser would send a binary signal ('advanced' or 'basic') to one worker and had an incentive to give good advice by virtue of receiving a bonus of 50% of the worker's earnings. Workers only learned that both jobs were related to the task they performed in part 1, and would last for 10 minutes. By contrast, all participants were told that advisers would learn the exact content of each job. We introduced this design feature because it represents an extreme version of the information asymmetry between advisers and workers in the real world, and it implied that workers would be very likely to follow the adviser's recommendation. The latter helps rule out strategic considerations (where advisers expect workers not to follow their advice and adjust their advice accordingly).

Part 2b: Motivation Letter and Self-rating

After all participants had finished reading the instructions, workers were asked to rate their ability at the task (on a scale from 0 to 10) and to write a motivation letter. In the

letter, we asked them to describe their math experience and math skills in general terms. They had ten minutes to write the letter and no restrictions were imposed on what they could write. In the instructions we emphasized that each letter and self-rating (their "worker portfolio") would be sent to their adviser. This implied that the more informative their letter, the better the advice they were likely to receive. The reason we provided advisers with these two subjective indicators (and not, for example, the worker's actual score on the task) was to ensure that advice would be based on subjective factors. This introduces scope for biased advice if advisers have gender stereotypes, by allowing them to interpret the subjective information they receive in a way that fits their underlying stereotypes.[5]

While workers wrote their motivation letter, advisers received detailed information about what the advanced and the basic jobs were. They were told that workers would perform the same task as in the first part of the experiment, but the payment for performance would be different. For the basic job, workers would earn $13 if they got at least 13 correct answers and earned nothing if they fell short of this performance target. For the advanced job, workers would earn $26 if they got at least 26 correct answers and nothing otherwise. These thresholds were calibrated using data from Buser, Ranehill and Van Veldhuizen (2017) to ensure that about half of the participants could be expected to meet the advanced job's threshold. The advisers received a bonus equal to 50% of their worker's earnings. Therefore, advisers had an incentive to advise workers to choose the job that best matched their expected performance.

At this time, we also informed advisers that they would receive five worker portfolios instead of one. We informed them that one of the five portfolios belonged to their assigned worker and would determine their payment, and the other portfolios belonged to other workers in the session. They were also told that they would only be informed about which portfolio belonged to their worker at the end of the experiment. It was, therefore, optimal for them to treat each portfolio as if it belonged to their worker.

Part 2c: Confirming Presence in the Lab

After all workers had finished their motivation letter and self-rating, we informed all participants that we would start calling up each worker-adviser pair sequentially to ask

[5] Previous studies have used a similar approach to engender biased judgments in the context of bargaining (Babcock et al., 1995), bribery (Gneezy et al., 2019) and financial advice (Gneezy et al., 2020).

them to confirm their presence in the laboratory. We called out each participant's number and asked them to confirm their presence by loudly saying "I am worker / adviser (their desk number) and I am here", so that all other participants could hear them. This procedure, introduced by Bordalo et al. (2019), allows participants to hear the voice of the person they are matched with and, therefore, form an accurate guess on whether that person is a man or a woman. This allows us to subtly reveal participants' gender without conveying additional information (such as physical appearance or personality) or alerting participants to the purpose of the study. This helps alleviate the concerns regarding a possible gender bias in advice being driven by an experimenter demand effect. We set up the participant desks in the laboratory so that workers and advisers could not see each other but could hear each other clearly.[6]

Part 2d: Giving Advice

After all participants had confirmed their presence in the laboratory, advisers received the worker portfolios (consisting of the motivation letter and self-rating) of their worker and four other workers. They were then asked, for each portfolio, to read through the portfolio and advise its worker to choose the basic or advanced job. We presented the five worker portfolios sequentially; advisers were unable to change their advice after moving on to the next worker portfolio.[7] We also reminded advisers that only the performance of their worker would actually determine their payment, and that it was in their best interest to judge each worker portfolio as if it belonged to their worker. This means that if advisers have a different advice strategy for male and female workers (e.g., because of a gender bias), they are incentivized to use the strategy corresponding to the gender of their assigned worker. Having each adviser evaluate five different worker portfolios also greatly increases our statistical power; see the power calculation in Appendix B for more details.

---

[6] As reported by participants in the post-experimental questionnaire, advisers correctly identified the gender of the workers in 84% of the cases, showing that our method was generally successful in revealing gender. Our success rate is slightly lower than, but nonetheless similar, to the one found in Bordalo et al. (2019) (92% and 95%, depending on the laboratory where the sessions were conducted).

[7] In the experiment, we randomly (and unbeknownst to participants) formed matching groups of five advisers and five workers. Each adviser sees the worker portfolio of each of the five workers in their group, in a random order. This way, we ensure that each worker portfolio is judged by five advisers. We aimed to have 20 participants taking part in each experimental session, giving us two matching groups, each including five workers and five advisers. In the cases where we had more or fewer than 20 participants in a session, each adviser who was not part of a group of 10 saw the worker portfolio from their own worker and four other randomly chosen worker portfolios.

Part 2e: Communication of Advice and Job Choice

After all advisers had given their advice to all five worker portfolios, workers received the advice from their adviser, and chose between the advanced and the basic job. After making their choice, they received detailed instructions about the task they would perform and its associated payment (a payment of $26 for obtaining at least 26 correct answers if they chose the advanced job, and a payment of $13 for obtaining at least 13 correct answers if they chose the basic job) and then proceeded to the task. In the meantime, advisers were waiting for workers to make their decision.

Part 2f: Addition Task and Adviser Ratings

Workers had 10 minutes to work on the addition task. While workers performed the task, advisers saw the five worker portfolios once again, and indicated how confident they were, on a scale from 0 to 10, that each worker would achieve the performance targets of the basic and the advanced job. This gives us a more continuous measure of the advisers' beliefs about the ability of the worker at the task. The advisers' answers were not incentivized.

Earnings, Questionnaire, and Implicit Association Test

At the end of part 2, all participants received a summary of their earnings. Workers were told whether they achieved their threshold and final earnings, including their earnings from part 1 and part 2 as well as a $5 participation fee. Advisers were also informed about their final earnings, including their earnings in part 2 (50% of their matched worker's earnings plus a $10 flat fee) and a $5 participation fee. Advisers were also informed about which of the five worker portfolios belonged to their worker.

All participants were then given a short questionnaire that asked them to explain how they made their decisions in the experiment, to indicate the gender of their partner in the experiment and to answer sociodemographic questions. We also asked advisers about the importance of several factors (on a scale between 0 and 10) for the advice they had given to the workers, including the workers' self-confidence, willingness to take risk, skill at the task, and motivation letter. They also had the possibility to describe other factors that influenced their advice. Meanwhile, the workers were asked to explain how they chose between the advanced and the basic job. All participants were asked about their willingness to take risks in general (Dohmen et al., 2011), their general self-confidence

and the extent to which they believed that men and women differed in their ability at the additions task.

Finally, participants performed an implicit association test (IAT) to measure their implicit attitudes with regards to personality traits of men and women.[8] Participants had to associate personality traits – which either were or were not related to professional success and ambition – with male and female roles as fast as they could. In one stage, the association rule was consistent with the stereotypical view, with male roles being associated with traits that relate to professional success and ambition (such as confidence, competitiveness, leadership) and female roles being associated with other traits (such as kindness, caution and generosity). In another stage, the association rule was reversed. The extent to which a participant held a stereotypical view is measured by the gap in response time between the first and the second case, with a quicker response in the first case indicative of a stereotypical belief that men are more competitive and confident and less generous and kind. More detailed information on the IAT procedures can be found in Appendix A1.

Experimental Procedure

We conducted our experiment with students at the University of Sydney. Our participants were recruited through ORSEE (Greiner, 2015), and each participant took part in one session only. The experiment was computerized using the experimental software oTree (Chen et al, 2016). A session lasted about 1 hour, and participants earned on average AU$22 (including a $5 participation fee), paid out in cash at the end of the session. One hundred and ninety participants took part in the experiment across 10 experimental sessions.

Participants read all instructions on the computer screen. The experimenter also gave oral instructions to the participants during the experimental proceedings aimed at revealing the participants' gender. During the experiment, participants could ask questions, which were answered quietly at their desk. Throughout the instructions, they also answered comprehension questions displayed on the computer screen.

[8] The conceptualization of our IAT follows Greenwald et al. (1998). The IAT has been widely used in social psychology and more recently in economics, to measure implicit stereotypical attitudes (e.g. Beaman et al., 2009; Glover et al., 2017; Carlana, 2019).

3. Hypothesis and Identification Strategy

In this section we will first set up a simple general model that illustrates the various channels that may generate gender differences in advice in our experiment. We will then outline our identification strategy and introduce our main hypothesis.

Let us first consider the general problem faced by an agent i (the 'worker' in our experiment) choosing from a discrete set of possible careers c (or, equivalently, jobs, tasks, or other uncertain prospects) among a set of possible alternatives. Assuming that the agent's preferences $\succsim_i$ over outcomes satisfy the axioms of Savage Expected Utility, we can represent the agent's maximization problem as follows:

$$\max_{c \in C} EU_i = \sum_{s=1}^{S} p_{ic}^s u_i^s$$

where, p_{ic}^s is the agent's subjective probability that she will end up in state $s \in S$ for a given career c, and u_i^s is the utility obtained in a given state.[9] Gender differences in career choices can emerge from gender differences in beliefs p_{ic}^s, for example if men expect greater success in traditionally male careers and women expect greater success in traditionally female careers. They can also emerge from gender differences in preferences u_i^s, for example if women and men prefer different professions or differ in their risk aversion or competitiveness (Niederle and Vesterlund, 2007). Both beliefs and preferences may be affected by advice.

Now let us consider the problem faced by an adviser k. We assume that the adviser may care not just about how the agent's choice affects their own compensation but may also care about the agent's utility and may have some personal preferences over the agent's choice or the advice given (e.g., due to stereotypes). Assuming that the adviser's references $\succsim_k$ satisfy Savage's axioms, we can represent the adviser's problem of giving advice to agent i as follows:

$$\max_{a \in A} EU_k = U(EU_i(c^*), \pi_k(a, g_i), c^*, g_i)$$

[9] States can be thought of as job outcomes, such that u_i^s could be interpreted as sum of discounted utilities if a given job outcome is reached, i.e., $u_i^s = \sum_{t=0}^{\infty} \delta(t) u_{it}^s$.

where $\pi_k(a)$ is adviser k's compensation for giving advice a, g_i is the agent's gender, and c^* is the career the agent chooses after observing the advice. In order to help determine $EU_i(c^*)$, the adviser forms beliefs about the likelihood that agent i ends up in a given state s after selecting a career c, i.e., $p_{ik} = (p_{ik1}, \dots, p_{ikc})$, where $p_{ikc} = p_{ikc}^1, \dots, p_{ikc}^S$ based on a prior $\overline{p_{ik}}$ and a signal σ_i. Based on his beliefs, the adviser then sends a simple signal a (advice) to the agent recommending that she chooses a specific action.

This simple framework illustrates that there are several reasons why advisers may give less ambitious advice to female agents. Firstly, male and female agents may send different *signals* σ_i to their adviser about their aptitudes in different careers. For example, it may be that men are more confident and that this makes advisers give them more ambitious advice. Secondly, even conditional on receiving the same signal, *beliefs* p_{ik} may differ by gender if advisers interpret signals differently based on the agent's gender or if they have different priors about male and female agents. The beliefs channel is likely to play an important role in settings like in our experiment where advice is exclusively based on subjective signals about the quality of the advisee. Thirdly, advisers may have gender-specific *preferences* when it comes to giving advice (as reflected by the fourth argument of the utility function). For example, they may prefer that agents follow a career that is congruent with existing gender norms. Fourthly, even conditional on the worker's expected aptitude, advisers may have a *financial incentive* $\pi_k(a, g_i)$ to give different advice to men and women with equal aptitude, e.g., if they are rewarded for conforming with existing stereotypes. Finally, there may be *strategic reasons* to give different advice to men and women, e.g., when the adviser has to give different advice a to each gender to induce them to choose the same action c^* (as could be true if women are more risk averse).

Our design allows us to rule out the fourth channel (incentives) by making the advisers' financial incentives independent of the agent's gender (conditional on expected aptitude). It also allows us to rule out the fifth channel (strategic reasons) because workers have so little information that they should be expected to always follow the stated advice (and indeed do so also in practice).

Our design allows us to distinguish between the remaining channels (signals, beliefs, and preferences) given the way we elicit advice. To see this, first recall that advisers learn the gender of their assigned worker by hearing their voice. They then give advice to their worker portfolio plus four other randomly selected worker portfolios without knowing

which portfolio belongs to their worker and are incentivized to judge each of them as-if they belonged to their worker. This implies that if advisers are gender-biased in either their preferences or their beliefs, they should give less ambitious advice if their worker had a female voice than if their worker had a male voice. The signal received by the adviser (i.e., the worker's portfolio), however, does not depend on the gender of their assigned worker. Instead, it depends on the gender of the person to whom the portfolio actually belongs.

Table 1 gives an overview of our identification strategy, by presenting the different possible cases in our experiment. By comparing across columns, we can study how a given portfolio is evaluated depending on whether it belongs to a man or a woman. This allows us to identify the importance of the signaling channel. By comparing across rows, we can identify the importance of the preference and belief channels. We can separate between these two channels by comparing the gender gap in advice (which captures both channels) to the gender gap in elicited beliefs about the likelihood that the worker will be successful in the advanced and in the basic job (which should only capture the belief channel).

Table 1: Identification Strategy

	Within-Advisor:	
	Sees a Male Portfolio	Sees a Female Portfolio
Between-Advisor:		
Hears Male Voice	Male Signal Male-based Belief Male-based Preference	Female Signal Male-based Belief Male-based Preference
Hears Female Voice	Male Signal Female-based Belief Female-based Preference	Female Signal Female-based Belief Female-based Preference

Previous research suggests that the beliefs and preferences channels may lead advisers to give less ambitious advice when matched to a worker with a female voice. In particular, there is evidence suggesting that advisers may have different prior beliefs about the ability of men and women in our task and that beliefs may be updated differently depending on

the agent's gender.[10] Jointly referring to this evidence as a 'gender bias', we hypothesize that advisers are less likely to advice female workers than male workers to opt for the advanced job.

5. Results

In this section, we start by presenting descriptive statistics of our sample. We then investigate a possible gender bias in advice along the lines of the pre-analysis plan, followed by a brief analysis of workers' job choices. Finally, we present exploratory (i.e., not pre-registered) results seeking to better understand the processes driving advice in our experiment.

Descriptive Statistics

Table 2 provides some descriptive statistics on our sample by gender. Notably, male workers scored higher in both parts of the experiment, though this difference is only significant in part 2. Male workers were also more likely to actually reach the threshold score of 26 in part 2 than female workers (50% versus 27%, p=0.021, Fisher's exact test). Though advisers were unaware of their workers' performance, the gender difference in performance may affect advice if it makes male workers appear more self-confident in their portfolios relative to female workers. We will take this into account in our analysis in the next section.

[10] For example, Geraldes (2020) shows that both men and women prefer to compete against women in this task, and Reuben, Sapiena and Zingales (2014) show that both men and women prefer to hire a man to do this task on their behalf. Coffman, Collis and Kulkarni (2019) show that belief updating may depend on gender and gender stereotypes, and Carlana (2019) demonstrates that female students perform worse when their teachers are gender-biased, possibly due to receiving negative feedback from their teacher.

Table 2: Descriptive Statistics

	(1) Male	(2) Female	(3) Difference
Workers			
Score in Part 1	12.152	10.939	1.213
	(5.086)	(4.059)	(0.941)
Score in Part 2	26.304	21.878	4.427**
	(10.581)	(6.582)	(1.796)
Reached Threshold Part 2	0.500	0.265	0.235**
	(0.506)	(0.446)	(0.098)
All			
Age	22.408	22.207	0.202
	(3.404)	(3.714)	(0.516)
International Student	0.378	0.293	0.084
	(0.487)	(0.458)	(0.069)
Economics	0.337	0.337	-0.000
	(0.475)	(0.475)	(0.069)
Engineering	0.255	0.076	0.179***
	(0.438)	(0.267)	(0.053)
Science	0.194	0.207	-0.013
	(0.397)	(0.407)	(0.058)
Social Science	0.071	0.109	-0.037
	(0.259)	(0.313)	(0.042)
Other	0.143	0.272	-0.129**
	(0.352)	(0.447)	(0.058)
Observations	98	92	190

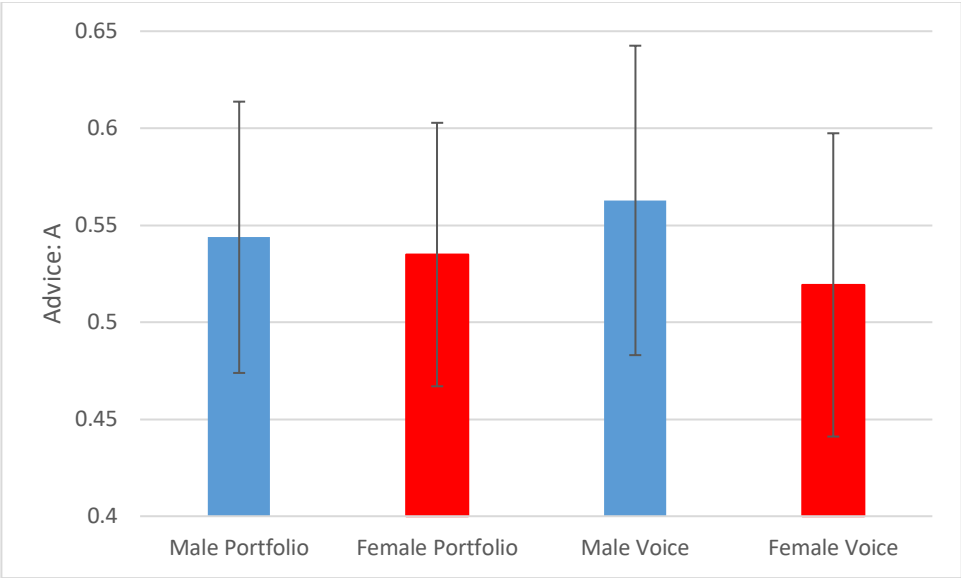
Notes: Columns 1 and 2 show the average outcome for each gender (standard deviation in parentheses. Column 3 shows the difference in the average outcome between genders (standard error in parentheses). 'Score in part 1' (part 2) is the number of additions correctly solved in part 1 (part 2). 'Reached Threshold Part 2' is a dummy indicating whether a participant reached the target score of 26 in part 2. 'Age' is a participant's self-reported age. 'International Student' is a dummy variable for international students (1-international, 0-domestic). Economics, engineering, science and social science are dummy variables for the respective study majors. Other consists of students who did not classify themselves in any of the previous categories. *p<0.1, **p<0.05, ***p<0.01.

It is also useful to note that there appears to have been sufficient scope for a gender bias to affect the adviser's judgment. Recall that we expect a gender bias to play a larger role when judgment is subjective, that is, when it is not obvious to the adviser what advice they should give. In our sample, in 41 out of 85 portfolios evaluated by at least two advisers (48%), there was substantial disagreement among advisers in the sense that less than 70% gave the majority advice. If we treat these cases as situations with subjective judgment, more than 95% of advisers ($1 - 0.52^5$) received at least one portfolio with scope for a gender bias to play in role in expectation.

Preregistered Analysis for Advisers

We present the results of our main analysis, pre-registered in our pre-analysis plan, in Figure 2 and Table 3. We test our main hypothesis that advisers are less likely to advise women to opt for the advanced job than men by investigating two potential determinants of advice. First, we consider the content of the worker's portfolio (motivation letter and self-rating). Any effect of the portfolio's content would capture the signaling channel, as discussed in the previous section. The second determinant is the voice of the worker assigned to the adviser. If advisers have gender-biased beliefs or preferences, they may give less ambitious advice when matched to a worker with a female voice.

Figure 2. Advice for Male and Female Portfolios and Voices



Notes: Advice: A is a equal to one if the advisor advised a worker to choose job A. The error bars are 95% confidence intervals calculated using standard errors that are clustered at the advisor level.

The first result is that advisers who are assigned a worker with a female voice are 4 percentage points less likely to advise the worker to opt for the advanced job (Figure 2 and column 1 of Table 3). However, this effect is small and not statistically significant. In other words, we do not find evidence that advisers have gender-biased beliefs or preferences that lead them to systematically give different advice to female and male workers. Second, the coefficient estimate of 'female portfolio' is not statistically different from zero either (column 1 of Table 3). This implies that there are no systematic gender differences in motivation letter content and self-ratings. In other words, there is no

evidence that male and female workers send different signals to advisers, and that this in turn generates a gender difference in advice.

Table 3: Main Analysis for Advisers

	(1) Advice:A	(2) Rating	(3) Advice:A	(4) Rating	(5) Advice:A	(6) Rating
Female Voice	-0.043	-0.021	-0.024	-0.009	-0.058	-0.105
	(0.055)	(0.367)	(0.062)	(0.420)	(0.059)	(0.382)
Female Portfolio	-0.004	-0.193	0.004	-0.129	0.012	-0.120
	(0.041)	(0.227)	(0.045)	(0.251)	(0.043)	(0.225)
Male Bias					-0.040**	-0.158
					(0.020)	(0.158)
Confidence					0.011	0.053
					(0.017)	(0.109)
Risk Prefs					0.042**	0.312**
					(0.017)	(0.119)
Constant	0.565***	6.121***	0.533***	6.071***	0.353*	3.939***
	(0.044)	(0.259)	(0.048)	(0.289)	(0.187)	(1.352)
Sample	All	All	Remember	Remember	All	All
Observations	475	475	400	400	475	475
Demographics	No	No	No	No	Yes	Yes
Cluster	95	95	80	80	95	95
R-squared	0.002	0.001	0.001	0.001	0.045	0.061

Notes: Robust standard errors are in parentheses. Advice:A is equal to 1 if the advisor advised the worker to choose job A. Rating is the advisor's subjective rating of the worker's ability. Female voice is equal to 1 for advisers matched to a female worker. 'Female portfolio' is equal to 1 for portfolios (motivation letter and self-rating) created by female workers. Male Bias is a question that asked advisers whether they believed that men are better at the task than women. Confidence is the adviser's general confidence, 'Risk Prefs' is their general attitude towards risk. The demographic controls in column 5 and 6 are field-of-study dummies and a dummy for international students; none of these variables are significant. *p<0.1, **p<0.05, ***p<0.01.

In column 2 of Table 3, we show that we obtain similar results when we consider the adviser's subjective rating of the worker's ability as the dependent variable. The non-significant coefficient for female voice indicates that the belief channel by itself does not generate a gender difference in worker ratings. In columns 3 and 4, we show that the results are also similar if we restrict our sample to the 80 advisers (out of 95) who correctly identified the gender of their worker in the post-experimental questionnaire. In columns 5 and 6, we show that the results are robust to controlling for field-of-study dummies, a dummy for international students and the risk attitudes, self-confidence and explicit gender bias variables elicited in the questionnaire. These results also demonstrate that advisers who are personally more willing to take risks are significantly more likely to advise workers to opt for the advanced job and give higher ratings to their workers,

whereas advisers who believe that men are better at the task are less likely to advise workers of either gender to opt for the advanced job. Note, however, that none of these results remain statistically significant at conventional levels when we adjust our *p*-values for multiple testing (Holm-Bonferroni correction).[11]

In Table 4, we present additional tests that investigate whether the effects presented in Table 3 vary depending on the gender of the adviser (columns 1-3 of Table 4) or the adviser's score on the implicit association test (IAT, columns 4-6 of Table 4). The results suggest that male advisers are less likely to advise a worker with a female voice to opt for the advanced job and that this effect vanishes for female advisers. In other words, male advisers appear to be gender-biased, whereas female advisers are not. However, these effects are statistically significant only when we control for demographics (in column 3 of Table 4) and are no longer significant when adjusting for multiple testing. In addition, the adviser's IAT score does not appear to significantly determine their advice.

Overall, there is no evidence that advisers consistently give different advice to men and women. In other words, there is no evidence that gender differences in the quality of the portfolio (the signaling channel) or a gender bias among advisers (the preferences and beliefs channels) generate differences in advice received by men and women in our experiment. Further, our power calculation suggests that this is not a power issue: with 95 advisers evaluating five letters each, we expected to have a power of 0.80 to detect even a small effect size (Cohen's D of 0.28), see the pre-registered power calculation in Appendix B1 for more details. We do observe some suggestive evidence that male advisers appear to be more gender-biased than their female counterparts, though these results appear in only one specification and are not robust to adjusting for multiple testing.

Table 4: Main Analysis for Advisers by Adviser Gender and IAT score

	(1)	(2)	(3)	(4)	(5)	(6)

[11] To test for order effects, we also re-ran regression (1) separately for the first, second, third, fourth, and fifth motivation letter read by a given adviser. We do not find a significant effect for either voice gender or portfolio gender in any of these regressions. The 'Female Voice' ('Female Portfolio') coefficients also remain small and insignificant if we replace the 'Female Portfolio' ('Female Voice') variable by portfolio (adviser) fixed effects respectively.

Dependent Variable: Advice A						
Female Voice	-0.112	-0.085	-0.154**	-0.041	-0.017	-0.058
	(0.075)	(0.080)	(0.073)	(0.059)	(0.065)	(0.062)
Female Portfolio	0.016	0.017	0.038	-0.009	-0.004	0.006
	(0.057)	(0.058)	(0.057)	(0.044)	(0.048)	(0.045)
Female Adviser	-0.100	-0.093	-0.177**			
	(0.086)	(0.094)	(0.076)			
Female Adviser * Female Voice	0.153	0.137	0.212**			
	(0.109)	(0.123)	(0.101)			
Female Adviser * Female	-0.048	-0.027	-0.069			
	(0.085)	(0.094)	(0.086)			
IAT				-0.032	-0.031	-0.034
				(0.103)	(0.090)	(0.112)
IAT * Female Voice				-0.083	-0.178	-0.020
				(0.149)	(0.154)	(0.154)
IAT * Female Portfolio				0.068	0.149	0.069
				(0.126)	(0.123)	(0.126)
Constant	0.608***	0.573***	0.359*	0.566***	0.535***	0.243
	(0.065)	(0.067)	(0.191)	(0.044)	(0.049)	(0.205)
Sample	All	Remember	All	All	Remember	All
Controls	no	no	yes	no	no	yes
Observations	475	400	475	445	380	445
Cluster	95	80	95	89	76	89
R-squared	0.010	0.006	0.063	0.003	0.006	0.049

Notes: Robust standard errors are in parentheses. 'Female Adviser' is equal to 1 for female advisors. IAT is the adviser's normalized score on the Implicit Association Test. For remaining variable definitions, see the notes to Table 3. *p<0.1, **p<0.05, ***p<0.01.

Workers Choices

Table 5 presents the results of the pre-registered analysis for workers. Column 1 shows that women are 12 percentage points less likely to choose the advanced job than men. This effect is driven by the fact that women received slightly less (4 percentage points) ambitious advice (Table 3) plus the fact that women, randomly, appear to have been matched with less ambitious advisers (the remaining 8 percentage points). Note, however, that this effect is not significant, which is because our study was designed (and is) adequately powered to identify a gender gap in advice, but not in job choices of the workers. Column 2 shows that a worker's chosen job is driven almost entirely by the advice they received. Indeed, after controlling for advice, the gender effect falls to 0.8 percentage points. Taken together, these results imply that the lack of a statistically significant gender difference in job choices in column 1 is due to a combination of the fact

that, by design, advice was a key determinant of workers' choices, and the fact that, as we have shown previously, men and women received similar advice.

In columns 3 and 4 we look at gender differences in self-confidence among workers. Even though women seem to be less confident in their math ability (they have lower self-ratings), and less overconfident about the number of problems they expect to solve than men, these gender differences are not statistically significant. These results are different from previous work, in which men are typically more confident than women even in samples that are smaller than ours (e.g., Niederle and Vesterlund, 2007; Van Veldhuizen, 2020). One possible reason is that, in contrast to these previous studies, we ask about self-confidence in absolute ability instead of relative ability.

Table 5: Workers' Job Choices

Dependent Variable:	(1) Choice:A	(2) Choice:A	(3) Self-Rating	(4) Overconfidence
Female	-0.123	-0.008	-0.476	-1.204
	(0.100)	(0.030)	(0.364)	(1.529)
Advice: A		0.948***		
		(0.037)		
Constant	0.674***	0.056	7.435***	-0.674
	(0.070)	(0.043)	(0.244)	(1.030)
Observations	95	95	95	95
Cluster	95	95	95	95
R-squared	0.016	0.916	0.018	0.007

Notes: Robust standard errors are in parentheses. Choice:A is equal to 1 if the worker chose job A. Self-rating is a worker's self-rating sills in math in the portfolio. Overconfidence is the score a participant expects minus their actual score in the addition task. Female is the worker's actual gender. Advice:A is equal to 1 if the advisor advised the worker to choose job A. *p<0.1, **p<0.05, ***p<0.01.

We find no systematic evidence that men and women differ in their ability to select the optimal job, both when we look at the payment-maximizing job based on participants' actual scores in part 2 (ex post optimality: 65% for men and 55% for women; p=0.32, t-test) and when we compute their expected score in part 2 based on part 1 performance (ex ante optimality: 54% for men and 55% for women; p=0.95, t-test). Even though women were 12 percentage points less likely to choose the advanced job, their overall lower performance in part 2 meant that they were as likely to select the optimal payment scheme overall. Finally, there was no tendency for workers to be more or less responsive to advice sent by male or female advisors, because nearly all workers (93 out of 95) followed the advice.

What Explains Advice?

The results in the previous sections demonstrate that male and female workers receive very similar advice and, as a result, make very similar decisions. In this section, we present some exploratory (i.e., non-preregistered) analyses to help understand what factors advisers took into account when giving advice, and whether these factors can help explain the absence of a gender bias in our study.

Table 6 follows the approach of Table 3, but now includes the worker's self-rating as a third control variable. The results show that workers' self-ratings were a key driver of advice in our experiment. In particular, a one-point increase in self-rating (on a scale from 0 to 10) increased the probability of being advised to choose the advanced job by approximately 15 percentage points, and increased the adviser's subjective rating of a worker by 1.1 points on a scale from 0 to 10.

However, the motivation letter likely played a key role as well. In particular, there is still substantial unexplained variance even after we control for worker self-ratings (in column 1 of Table 6); this remains true even after we include adviser fixed effects (in column 2 of Table 6). More importantly, when we asked advisers about the factors determining their advice in the questionnaire, the motivation letter (8.6 out of 10) was deemed to be more important than the worker's self-rating (8.3 out of 10; $p=0.085$, t-test). It was also deemed to be more important than the worker's risk attitudes (6.6 out of 10; $p<0.001$, t-test), and as important as the worker's skill level (8.7 out of 10; $p=0.782$, t-test).[12] Moreover, after controlling for confidence, motivation letters written by women are actually *more* likely to receive the advice to choose the advanced job. One possible interpretation of these results is that women may have written better motivation letters encouraging advisers to advise them to choose the advanced job for a given level of stated confidence. However, this result must be interpreted with caution since we did not preregister this test and the effect in question is not particularly strong.

[12] Advisers matched to workers with female voices reported that they put less weight on the worker's overall skill level ($p=0.023$, t-test). However, this difference is not reflected by the actual advice given, and does not survive an adjustment for multiple testing. The 'Female Voice' ('Female Portfolio') coefficients also remain small and insignificant if we replace the 'Female Portfolio' ('Female Voice') variable by portfolio (adviser) fixed effects respectively. They also remain small and insignificant if we restrict our analysis to the 41 portfolios for which advisers were split 30/70 or more evenly, and hence are most likely to require subjective judgment.

Table 6: Determinants of Advice

Dependent Variable:	(1) Advice:A	(2) Advice:A	(3) Rating	(4) Rating
Worker's Self-Rating	0.147***	0.150***	1.093***	1.023***
	(0.009)	(0.011)	(0.051)	(0.054)
Female Voice	-0.022		0.133	
	(0.050)		(0.296)	
Female Letter	0.083*	0.082+	0.449*	0.266
	(0.039)	(0.045)	(0.179)	(0.170)
Constant	-0.550***	-0.584***	-2.162***	-1.492***
	(0.086)	(0.088)	(0.471)	(0.399)
Advisor FE	no	yes	no	yes
Observations	475	475	475	475
Cluster	95	95	95	95
R-squared	0.266	0.507	0.484	0.767

Notes: Robust standard errors are in parentheses. For variable definitions, please see the notes of Table 2. Advisor fixed-effects are included in columns 2 and 4. *p<0.1, **p<0.05, ***p<0.01.

In Appendix A2, we test whether our main results in Table 3 are heterogeneous by worker confidence or by worker performance. The results suggest that female workers in the third confidence quartile are more likely to be advised to choose the advanced job based on their portfolio than their male counterparts, whereas workers in the second performance quartile are less likely to be advised to choose the advanced job if they have a female voice. However, these results do not survive adjustments for multiple testing.

Overall, we do not find support for our main hypothesis that there is a gender bias in advice. In addition, advisers do not appear to display a gender bias outside the context of the experiment. Firstly, there is no overall gender bias in the IAT among advisers in our sample (p=0.736; t-test). Secondly, 65 out of 95 advisers (68%) think that men and women are equally good at the task in the experiment. Out of the remaining 30 advisers, 20 do expect men to be (slightly) better than women and 10 expect the reverse (p=0.090, Wilcoxon). However, the overall gender bias is not significant (p=0.489, t-test). Taken together, these results suggest that the lack of a gender bias in advice appear to be driven by the fact that advisers in our experiment are not gender-biased in general.

6. Conclusion

We conducted a laboratory experiment to study if gender-biased advice could help explain the sustained gender gap in labor market outcomes. Overall, we found no systematic

evidence that advisers display a gender bias in their advice. There is some suggestive evidence that male advisers give less ambitious advice to female workers compared to male workers, but this result vanishes when adjusting for multiple testing. The lack of supportive evidence for a gender bias in advice cannot be explained by insufficient statistical power and appears in a setting where a gender bias could be expected because the advice related to choosing between a more ambitious and a less ambitious stereotypically male job and the advice was based on subjective judgement. Finding no gender bias in advice in this setting implies that a gender bias is also unlikely to be found in other settings that are less favorable to a gender bias.

Our results suggest that the main reason why we found no gender bias in advice is that advisers in our experiment were not gender-biased. Our participants were university students in a high-income country, like participants in previous experiments that document a bias against women (Krawczyk and Smyk, 2016; Özgümüs et al., 2020), as well as participants in experiments that find no bias against women or that find a bias against men (Breda and Ly, 2015; Beugnot and Peterle, 2020). These mixed results show that we need further research to identify the situations in which biases against women or men are likely to occur. In the context of advice specifically, we need to better understand the situations in which gender-biased advice is likely to occur and the extent to which it can influence education, career or financial choices, and across the life course.

Another potential explanation for our results is that we deliberately did not make gender particularly salient in the experiment. In particular, we sought to reduce concerns related to experimenter demand effects by subtly revealing the advisee's gender through their voice. While gender was revealed just prior to giving advice and most advisers recalled the gender of their advisee, it still seems possible that gender was not a major factor in determining advice. To the extent that gender is also not particularly salient in other settings, our results may therefore be more predictive of behaviour in such settings than comparable experiments in which gender was made more salient. Nevertheless, it could be of interest to replicate our study in a setting where gender is made more salient by revealing it in a less subtle way.

Our experimental design may also have contributed to our results. When judging the five worker portfolios, all the advisers were certain about was that at least one of the portfolios belonged to a man or a woman, depending on whether they heard a male or a female voice when introduced to their worker. Even though we told advisers that it was in

their best interest to judge all portfolios as if they belonged to their worker, it is possible that they did not follow our advice. We used this approach because it allowed us to exogenously assign gender to the worker portfolios and because it substantially increases our statistical power to detect a gender bias in advice. Our design is analogous to the strategy-method in which participants make a series of binding choices conditional on different possible future scenarios, which are commonly used in experiments to elicit individual preferences and behaviour. Nevertheless, it would be interesting to see if a larger gender gap emerges in a high-powered study that does not rely on the strategy method.

While we focused on gender, our design could be adapted to study biases based on other sociodemographic variables like ethnicity, nationality or political affiliation. If such biases are found, it could also be worth studying whether they can be reduced by providing objective measures of quality (such as the worker's past performance). This would allow us to contrast the influence of stereotypes on advice when advice is based on subjective measures of quality compared to when it is supplemented by objective measures.

References

AlShebli, B., Makovi, K. & Rahwan, T. (2020) The association between early career informal mentorship in academic collaborations and junior author performance. *Nature Communications 11*, 5855.

Babcock, L., Loewenstein, G., Issacharoff, S., & Camerer, C. F. (1995). Biased Judgments of Fairnes in Bargaining. *American Economic Review*, *85*(5), 1337–1343.

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science, 60*(2), 434-448.

Baron, K., Ditlmann, R., Gehrig, S., Schweighofer-Kodritsch, S. (2020). Explicit and implicit belief-based gender discrimination: A hiring experiment. WZB Discussion paper 2020-306.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R., & Topalova, P. (2009). Powerful women: does exposure reduce bias?. *The Quarterly Journal of Economics, 124*(4), 1497-1540.

Bertrand, M. (2011). New Perspectives on Gender. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics Volume 4B* (pp. 1543–1590). Amsterdam: Elsevier.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review, 94*(4), 991-1013.

Beugnot, J., & Peterlé, E. (2020). Gender bias in job referrals: An experimental test. *Journal of Economic Psychology, 76*, 102209.

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.

Booth, A. L., Leigh, A., & Varganova, E. (2012). Does ethnic discrimination vary across minority groups? Evidence from a field experiment. *Oxford Bulletin of Economics and Statistics, 74*(4), 547-573.

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. (2020). Advice from women and men and selection into competition. Barcelona GSE Working Papers Series, n. 1007.

Breda, T., & Ly, S. T. (2015). Professors in core science fields are not always biased against women: Evidence from France. *American Economic Journal: Applied Economics, 7*(4), 53-75.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, *129*(3), 1409-1447.

Buser, T., Ranehill, E., & van Veldhuizen, R. (2017). *Gender differences in willingness to compete: The role of public observability*. University of Zurich, Department of Economics, Working Paper No. 257.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics, 134*(3), 1163-1224.

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

Charlesworth, T. E., & Banaji, M. R. (2019). Gender in Science, Technology, Engineering, and Mathematics: Issues, Causes, Solutions. *Journal of Neuroscience*, *39*(37), 7228-7243.

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. B., Collis, M., & Kulkarni, L. (2019). *Stereotypes and Belief Updating*. Harvard Business School Working Paper 19-068. Available at: https://www.hbs.edu/faculty/Publication%20Files/19-068_aac3e9db-c1df-48c6-b501-8026fb556d72.pdf

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.

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.

Duguet, E., Du Parquet, L., L'Horty, Y., & Petit, P. (2012). *First Order Stochastic Dominance and the Measurement of Hiring Discrimination: A ranking extension of correspondence testings with an application to gender and origin*. Sciences

de l'Homme et de la Société Working Paper No. halshs-00731005f. Available at: https://halshs.archives-ouvertes.fr/halshs-00731005

Edo, A., Jacquemet, N., & Yannelis, C. (2019). Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications. *Review of Economics of the Household*, *17*(1), 349-376.

Ellis, J. R., & Gershenson, S. (2020). Gender, peer advising, and college success. Labour Economics, 62, 101775.

Exley, C. L. (2016). Excusing Selfishness in Charitable Giving: The Role of Risk. *The Review of Economic Studies*, *83*(2), 587–628.

Ganguli, I., Hausmann, R., & Viarengo, M. (2020). Gender Differences in Professional Career Dynamics: New Evidence from a Global Law Firm. *Economica (forthcoming)*.

Geraldes, D. (2020). *Women Dislike Competing Against Men*. Unpublished Manuscript. Available at: https://drive.google.com/file/d/0B_nE1lM3ktKNdksyYU5pQzJBODQ/view

Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, *127*(4), 1826-1863.

Glover, D., Pallais, A., & Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores. *The Quarterly Journal of Economics*, *132*(3), 1219-1260.

Gneezy, U., Saccardo, S., & Van Veldhuizen, R (2019). Bribery: Behavioral drivers of distorted decisions. *Journal of the European Economic Association*, *17*(3), 917–946.

Gneezy, U., Saccardo, S., Serra-Garcia, M., & Van Veldhuizen, R. (2020). Bribing the Self. *Games and Economic Behavior*, *120*, 311–324.

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

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114-125.

Guillen, P., & Hakimov, R. (2018). The effectiveness of top-down advice in strategy-proof mechanisms: A field experiment. *European Economic Review*, *101*, 505–511.

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. (2020). *Can Simple Advice Eliminate the Gender Gap in Willingness to Compete?* Mimeo.

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.

Kübler, D., Schmid, J., & Stüber, R. (2018). Gender discrimination in hiring across occupations: a nationally-representative vignette study. *Labour Economics, 55*, 215-229.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*(41), 16474-16479.

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.

Nürnberger, M., Nerb, J., Schmitz, F., Keller, J., & Sütterlin, S. (2016). Implicit gender stereotypes and essentialist beliefs predict preservice teachers' tracking recommendations. *The Journal of Experimental Education*, *84*(1), 152-174.

Ors, E., Palomino, F., & Peyrache, E. (2013). Performance gender gap: does competition matter?. *Journal of Labor Economics, 31*(3), 443-499.

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.

Reuben, E., Rey-Biel, P., Sapienza, P., & Zingales, L. (2012). The emergence of male leadership in competitive environments. *Journal of Economic Behavior & Organization, 83*(1), 111-117.

Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, *111*(12), 4403–4408.

Sarsons, H., Schram, A., Gerxhani, K., & Reuben, E. (2020). Gender Differences in Recognition for Group Work. *Journal of Political Economy (forthcoming).*

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.

Tobback, I., Verhaest, D., & Baert, S. (2020). Student access to apprenticeships: Evidence from a vignette experiment. *Industrial Relations: A Journal of Economy and Society, 59*(3), 435–465.

Van Veldhuizen, R. (2020). *Gender differences in tournament choices: Risk preferences, overconfidence or competitiveness?* Mimeo available from the author's website.

Appendix A1: Description of the Implicit Association Test

In the IAT participants have to associate personality traits – which either are or are not typically related to professional success and ambition – with male and female roles. The personality traits that relate to professional success and ambition include assertive, competent, confident, ambitious, leader, competitive, and efficient. The personality traits which do not relate to professional success and ambition include passive, kind, cautious, timid, generous, cooperative and emotional. The female roles include woman, aunt, daughter, wife, mother, grandma, and girl. The male roles include man, uncle, son, husband, grandpa, and boy. Our IAT followed the standard procedure; it included five stages and took about 5 minutes to complete.

The first two stages aim at making participants familiar with the task. In the first stage, they sort the male and female roles. Specifically, they see each word one by one and in random other, in the middle of the computer screen and need to press a key on the right-hand side of the keyboard if the word corresponds to a female role and a key on the left-hand side of the keyboard if the word corresponds to a male role. In the second stage, they see each personality trait in the middle of the screen, also one at a time and in random order, and need to press a key on the right-hand side of the keyboard if the personality trait belongs to the category of professional success and a key on the left-hand side, if it does not.

In the third stage, participants need to associate the male roles with the personality traits which do not relate to professional success and ambition, and the female roles with personality traits that relate to professional success. As in the previous stages, they see each word belonging to one the four categories, one by one and in random order. They need to press a key on the right-hand side of the keyboard if the word is either in the category of female roles or of personality traits for professional success. They need to press a key on the left-hand side of the keyboard if the word belongs to one of the other two categories.

In the fourth stage, participants repeat the association they have made in stage one, but this time with the reverse sorting rule (they assign words for female roles to the left-hand side, and words for male roles to the right-hand side).

In the fifth stage, participants see again words belonging to all four categories, as in stage 3, but now they have to associate the male roles with the personality traits related with professional success and ambition, and the female roles with personality traits which are do relate with professional success. The extent to which a participant holds gender

stereotypes in the domain of personality traits is given by the difference in response time between stage 3 and stage 5 of the task.

Appendix A2: Heterogeneity Analysis

In this section, we investigate whether the gender effects we observe are heterogeneous along various dimensions. Column 1 of Table A1 repeats the analysis of Table 6 while splitting the confidence variable in 4 quartiles and separately estimating the effect of letter gender and voice gender in a given performance quartile. Of the estimated interaction effects, only the gender of the letter writer in the third confidence quartile is significant. Note, however, that this result does not survive any reasonable adjustment for multiple testing.

Column 2 mimics the analysis of column 1, but now splits the sample in quartiles based on worker's actual performance in part 1 of the experiment. In this specification, workers with a performance in the 2nd quartile are less likely to be advised to take job A if their adviser was matched with a worker with a female voice. We also see an interesting pattern that low-ability women are more likely to be advised to choose job A, whereas high-ability women are less likely to be advised to do so compared to their male counterparts. However, this result is likely driven by the fact that stated confidence is a key determinant of advice, and it happens to be less predictive of actual ability for women in our sample than it is for men.

Table A1: Heterogeneity Analysis for Advice

	(1)	(2)
Dependent Variable:	Advice:A	Advice:A
	Worker Confidence	Worker Ability
Quartile 2	0.309**	0.289*
	(0.095)	(0.115)
Quartile 3	0.449***	0.297**
	(0.112)	(0.109)
Quartile 4	0.731***	0.365**
	(0.087)	(0.113)
Female Voice * Q1	0.013	-0.025
	(0.086)	(0.101)
Female Voice * Q2	-0.117	-0.202*
	(0.106)	(0.088)
Female Voice * Q3	-0.011	0.045
	(0.086)	(0.090)
Female Voice * Q4	0.000	0.024
	(0.063)	(0.103)
Female Portfolio * Q1	0.069	0.270**
	(0.065)	(0.088)
Female Portfolio * Q2	0.021	0.035
	(0.087)	(0.094)
Female Portfolio * Q3	0.189*	-0.033
	(0.074)	(0.088)
Female Portfolio * Q4	-0.003	-0.142+
	(0.068)	(0.078)
Constant	0.164*	0.294**
	(0.075)	(0.090)
Observations	475	475
Cluster	95	95
R-squared	0.278	0.050

Notes: Robust standard errors are in parentheses. Quartile are dummies for being in the respective quartile for worker confidence (column 1) and worker ability (column 2). Here, worker ability is the worker's score in part 1 of the experiment. For the remaining variable definitions, please see the notes of Table 2. +p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Appendix B1: Pre-Registration Documents[13]

This project will consist of two stages. First, we will investigate whether a significant gender gap in advice emerges in baseline sessions where the advisors know the worker's gender but do not have objective information about the worker's quality. Second, provided that we find a significant gender gap in advice in the first stage, we will then also see whether the gap is significantly reduced in two additional treatments that either (a) remove the gender information or (b) give the advisor additional objective information when giving advice.14 In this document, we will present the power calculation and pre-analysis plan pertaining to the first stage. The power calculation and pre-analysis plan for the second stage will be uploaded before we run these additional treatments.

Power Calculation

Our interest in these initial sessions lies in investigating whether a gender difference in advice emerges in the baseline treatment. Since each advisor evaluates multiple workers, we compute power $P(p < 0.05|H_1)$ using simulations. We use the following data generating process:

$$A_{ij}^* = Q_i^* + \left(2 * female_{ij} - 1\right) * B_j^* + \varepsilon_{ij}$$

$$A_{ij} = I(A_{ij}^* > \bar{A}_j)$$

Here, A_{ij} is the advice given by advisor j to worker i, where $A_{ij} = 1$ indicates that the advisor advised the worker to choose job A. We assume that actual advice is determined by a latent variable (A_{ij}^*) that reflects advisor j's opinion of worker i. We assume that the advisor's opinion A_{ij}^* depends on worker i's true quality $Q_i^* \sim N(0,1)$, an idiosyncratic error term $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon)$ and an advisor-specific gender bias $B_j^* \sim N(B, \sigma_B)$. Each advisor evaluates five workers and each of these five workers receives advice from one of these five advisors, forming a matching group of 10 participants. Following our design, we assume that advisors base their evaluation of the worker not on the worker's actual gender but rather on the gender that advisor j perceives worker i to be, which is the gender of the person whose voice the advisors hears. We further assume that $\sigma_\varepsilon = 0.5$, which implies that, absent a bias, two-thirds of the variation in A_{ij}^* would result from variation in quality, and one third from random noise.

[13] This section is reprinted from the pre-analysis plan that is available at https://www.socialscienceregistry.org/trials/4244

[14] Since we did not find a significant gender gap in our study, we decided not to run these additional treatments.

We then simulate 1000 samples using this DGP for a given sample size and bias B. For each sample, we regress advice on perceived gender, clustering standard errors at the advisor level. Power is then equal to the fraction of simulated samples that have a significant gender gap in the right direction (against women whenever $B > 0$). For our initial sessions, we focus on a sample size of 160 (80 workers and 80 advisors) due to budget constraints, and compute power for various effect sizes.

Table B1: Power Calculation

Bias Term (B)	Gender Gap in A	Cohen's d	Power
0.5	0.33	0.67	1
0.4	0.27	0.54	0.998
0.3	0.20	0.41	0.961
0.2	0.14	0.28	0.727

Table B1 presents the results for different values of the average bias B. Column 2 quantifies the effect size in terms of the gender gap in advice, column 3 quantifies it using Cohen's d, a standardized measure of effect size commonly used in the social sciences. The results in column 4 indicate that we should have a high power to pick up moderate or large effects, and an acceptable power of 0.727 to detect even a fairly small gender gap in advice of 14 percentage points. By comparison, 21 out of 26 studies on willingness to compete on a similar real effort task reviewed in Van Veldhuizen (2020) find a gender gap of at least 15 percentage points. We would also like to note that power is one of the main reasons why we asked advisors to evaluate five workers instead of a single one. For example, with a single worker per advisor our power to detect an effect of 14 percentage points would fall to a mere 0.260 (instead of 0.727). More generally, we would have required a roughly twice as large sample to detect the same effect size as in our current design.

Pre-Analysis Plan

We specify our pre-analysis plan for testing whether there is a significant difference in the advice received by male and female workers in our sample. We also describe additional tests we will conduct to investigate the mechanism driving any gender difference in advice and investigate gender differences in workers' decisions and outcomes. Finally, we also specify robustness checks and additional exploratory analysis.

Main Analysis

1. A regression of advice on the perceived gender of the worker and the worker's actual gender.

a. Advice is a dummy variable (1-A).

b. The worker's actual gender is a dummy variable (1-female, 0-male) that is based on the gender the worker indicates in the questionnaire. This coefficient tells us whether men and women receive different advice based on the content of their motivation letter.

c. Perceived gender is a dummy variable (1-female, 0-male) that is based on the gender the advisor believes his/her worker to be (as measured in the questionnaire). This coefficient allows us to test whether advisors who believe their worker to be male give different advice compared to advisors who believe their matched worker to be female.

d. Our main hypothesis is that workers perceived by their advisor to be female are significantly less likely to receive advanced advice (tested by the coefficient of perceived worker gender).

e. We will cluster the standard errors in the regression at the advisor level.

Mechanism Analysis

2. We are interested in distinguishing between two mechanisms: gender differences in motivation letters and gender bias in the perceived quality of these letters. We will investigate these mechanisms in the following way:

   a. To check whether men and women received different advice by virtue of the content of their letter, we check whether the coefficient for gender in regression (1) is significant. In addition, we test whether men and women differ in their stated confidence in the letter using a t-test. To control for potential gender differences in performance, we also check whether participants differ in their stated overconfidence, defined as the number of problems they expect to solve minus the number of problems they actually end up solving.

   b. To check for a gender bias in the perception of male and female letters, we regress the advisor ratings of worker letters on the worker's actual gender and perceived gender (standard errors clustered at the matching group level). The coefficient for perceived gender tells us whether perceiving a worker's gender as

female changes the way advisors evaluate a worker's quality based on the motivation letter.

Other Analysis

3. We will also investigate worker outcomes in two ways:
   a. We test whether male and female workers differ in their tendency to choose task A over task B using a t-test.
   b. We test whether male and female workers differ in their tendency to select their ex-ante and ex-post payoff maximizing task. We define the former to be equal to the advanced task for any worker who scored at least 13 in part 1. We define the latter to be equal to the advanced task for every worker who scored at least 26 in part 2.

Robustness Checks and Exploratory Analysis

4. We will do several additional analyses to address the robustness of our results.
   a. We will check whether male and female workers differ in their performance on the task using a t-test. If we find a difference in 4a, we re-run regression (1) while also controlling for the worker's score in part 1.
   b. We run a regression of advice on perceived gender, gender, the implicit association test (IAT) score, the IAT*gender interaction and the IAT*perceived gender interaction.
   c. We will include the advisor's gender as an additional control variable in regression (1), and interact this variable with the gender and perceived gender variables.
   d. We will re-run our analysis on the advisors that correctly remembered their worker's gender.
   e. We will replace perceived gender in regression (1) with the actual gender of the worker the advisor is matched with.
   f. We also run regression (1) while controlling for advisor characteristics such as their risk attitudes and socio-cultural background.

Reference

Van Veldhuizen, R., (2018). Gender Differences in Tournament Choices: Risk Preferences, Overconfidence, or Competitiveness? Working paper