The signals we give: Performance feedback, gender, and competition*

Alexander Coutts Boon Han Koh Zahra Murad

December 2025

Abstract

Feedback is vital for growth and learning, yet anecdotal evidence suggests people often hesitate to provide it, and its provision may be shaped by asymmetries and gender-related biases. We study feedback provision across variations in the nature of performance signals, their instrumental value, and the recipient's gender. We find that a surprising degree of both positive and negative feedback is withheld, with a follow-up experiment suggesting that advisors' feedback decisions are driven mainly by transparency- and duty-related considerations, or, to a lesser extent, are motivated by self-serving reasons. Additionally, when initial performance signals are vague, advisors are more likely to withhold non-instrumental negative than positive feedback—an effect we conjecture may be stemming from the lower psychological cost of lying (by omission) under uncertainty. Suggestive evidence shows the difference is more pronounced for female recipients, and exploratory analysis traces this to stronger ego-protective concern by advisors for women than for men.

Keywords: Feedback provision; gender; ego/belief utility; competitiveness; discrimination.

JEL Classification: C90, D83, D91, J16, M54.

*__Coutts__: Schulich School of Business, York University (e-mail: acoutts@schulich.yorku.ca); __Koh__: Department of Economics, University of Exeter (e-mail: b.koh@exeter.ac.uk); __Murad__: Economics and Finance, University of Portsmouth and UNEC Cognitive Economics Centre, Azerbaijan State University of Economics (e-mail:zahra.murad@port.ac.uk). We are grateful to the Associate Editor, three anonymous referees, Billur Aksoy, Pol Campos-Mercade, Lan Guo, Prachi Jain, seminar and conference participants at the Burgundy School of Business, Monash University, Nanyang Technological University, Stockholm University, University of Amsterdam, University of Leicester, University of Melbourne, University of Sydney, Discrimination and Disparities seminar series, 2021 Annual Meeting of the Canadian Economics Association, 2022 Southern Ontario Behavioural Decision Research Conference, 2022 European Meetings of the Economic Science Association, 2022 North American Meetings of the Economic Science Association, and 2023 CIRANO Workshop on Behavioral and Experimental Economics for Innovative Policy Making for their comments and feedback. This study was funded by the British Academy (SRG1920/100428), the University of Exeter Business School, and the Portuguese Foundation for Science and Technology (FCT grant: PTDC/EGE-ECO/5020/2021). The study is pre-registered on the American Economic Association's registry for randomized controlled trials (AEARCTR-0006966), and it has been approved by the Scientific Council of Nova School of Business and Economics and the Research Ethics Committees of both the University of East Anglia (0347) and the University of Exeter (8317399). Declaration of interests: none. For the purpose of open access, the authors have applied a 'Creative Commons Attribution' (CC BY) licence to any Author Accepted Manuscript version arising.

1 Introduction

Giving feedback to others is ubiquitous in today's world, and it is one of the most powerful tools that organizations and educators have to improve worker performance, increase student learning, and spark individual growth (Prue and Fairbank, 1981; Greve et al., 2003; Hattie and Timperley, 2007). Despite clear benefits, evidence suggests that feedback (both positive and negative) is often withheld (London, 2014; Solomon, 2016; Zenger and Folkman, 2017). Moreover, survey evidence reveals that women may receive vaguer (i.e., more obscure) feedback than men (Correll and Simard, 2016), even though they are no less eager to receive performance feedback (Coffman and Klinowski, 2025). Obscuring feedback or providing low-quality feedback may hinder high-ability workers from entering competitive environments, and more broadly could impede performance improvement (Drouvelis and Paiardini, 2021). Hence, any gender bias that exists in feedback provision may contribute to the observed gender gaps in labor market outcomes and competitive preferences (Niederle and Vesterlund, 2007). Against this backdrop, in this study, we use an online experiment to examine biases in feedback provision and assess the consequences of these biases with a particular focus on gender.

In the environment we consider, an *advisee* (feedback recipient) performs a Raven's matrices IQ task and is then matched to an *advisor* (feedback sender) who provides the advisee with performance feedback. Specifically, in our experiment, each advisor observes their advisee's gender and an information signal about their performance quartile, and the advisor has to decide how to convey this information to their advisee. The advisee subsequently has the option to enter their performance into a tournament.

We assume in our theoretical framework that altruistic advisors care about two factors which contribute to their advisee's welfare: (i) instrumental utility; and (ii) ego utility. The first factor is modelled as a standard utility from economic payoffs (in this case, derived from the advisee's tournament-entry decision and performance), while the second assumes that the advisee also derives utility from believing that they are of high ability. We design treatments to evaluate (gender) biases in the provision of feedback along three main dimensions of interest: (i) the precision of information that the advisor receives about the advisee's performance; (ii) the valence of this information; and (iii) the instrumentality of advisor's feedback in influencing the advisee's decision to enter into a tournament. We also assess what consequences the provided feedback has on advisees' decisions and on the gender-competition gap.

Our first treatment dimension of interest is the precision of the signal. In many situations, advisors or managers may only have incomplete information about the performance of those under their responsibility. Previous research suggests that the precision of information may contribute to biases in feedback provision (Sun and Papadokonstantaki, 2023; Serra-Garcia et al., 2011; Bol, 2011). In our context, vaguer signals of performance may provide more ambiguity to advisors about the advisee's true performance, thus providing some wiggle room for advisors to obscure feedback. To investigate how feedback depends on the precision of

signals, advisors are provided with either: (i) signals that are *precise*, i.e., an exact quartile rank from 1 to 4; (ii) signals that are *vague*, i.e., top or bottom half performer; or (iii) signals that provide *no information*, i.e. quartile rank in the range between 1 and 4. The feedback that advisors can provide to their advisees depends on the precision of the signal received. That is, those who receive precise or vague signals can decide to obscure feedback, while those who receive no information have no active decision. Advisors who receive precise signals can decide whether to: (i) send precise feedback by fully revealing to advisees their exact rank; (ii) send vague feedback by disclosing whether the rank is in the top or bottom half; or (iii) send no feedback by simply indicating that the rank is anywhere in the range between 1 and 4. Advisors who receive vague signals can only choose between the latter two options. In this way, advisors can choose to (partially) obscure information, but they cannot lie to advisees. We ask whether the advisor's propensity to obscure information about their advisee's performance depends on the precision of the information received.

Our second dimension of interest is the valence of the signals advisors receive. That is, we study whether advisors obscure positive and negative signals of performance differently. Conditional on receiving performance information, the information advisors receive in the experiment is either positive and indicates above-average performance (rank 1 or 2 or top-half), or negative and indicates below-average performance (rank 3 or 4 or bottom-half). Under the assumption that individuals receive a higher ego utility from receiving positive feedback than negative feedback, our theoretical framework predicts that advisors will be more likely to obscure negative signals from their advisees.

Finally, our third dimension of interest pertains to the instrumentality of feedback. That is, we vary whether advisors have any influence (via their feedback decisions) on the instrumental utility of their advisees. Advisors in the *instrumental* treatment provide their advisees with performance feedback *before* the advisees make their tournament-entry decision. On the other hand, advisors in the *non-instrumental* treatment provide their advisees with feedback only *after* the advisees have made their decision, thereby rendering the feedback non-instrumental in influencing the advisee's tournament-entry decision. Given that feedback provides information that can aid advisees in making optimal decisions, we predict that advisors will be less likely to obscure information when feedback is instrumental.

Overall, we find that approximately one-quarter of advisors obscure their signal when providing feedback to their advisee. Surprisingly, contrary to the theoretical framework, positive signals are obscured nearly as much as negative signals. We also find that, on aggregate, advisors obscure feedback similarly between the instrumental feedback and non-instrumental feedback treatments, and also for male and female advisees.

To investigate why a non-negligible proportion of advisors obscure positive signals from advisees, we conduct a follow-up experiment where we systematically elicit advisors' motiva-

tions for their feedback. Specifically, advisors are asked to pick from a list all possible options that explain their feedback choice. We interpret their responses as a broader reflection of the motivations that are salient to advisors and which shape their thinking when making their feedback choice. We find that advisors' feedback decisions reflect preferences for transparency and a sense of duty toward the advisee, and that advisors who obscure feedback are significantly less likely to express these preferences. These results are also consistent with analysis of advisors' free-form text explanations of their feedback choices from the main experiment. Finally, we also find suggestive evidence that some advisors obscure positive feedback either because they do not perceive the feedback to be useful to the advisees, or because they compare their own performance with that of the advisee's (e.g., due to envy, or due to a consideration of what they would have done in the advisee's shoes if they had obtained a similar level of performance), or because they are concerned with how the advisee would view them.

Our results at the aggregate level conceal key differences between advisors who receive either precise or vague signals of performance. While there are no differences in obscuring positive or negative precise signals, we find that advisors obscure negative vague signals nearly twice as frequently as they do positive vague signals. Consistent with our theoretical predictions, this difference vis-à-vis vague signals is entirely driven by advisors providing non-instrumental feedback. Moreover, we replicate this finding in our follow-up experiment.

Further analysis of advisors' indicated motives for their feedback decisions in our follow-up experiment shows that transparency-related preferences are correlated with advisors being more likely to obscure negative than positive feedback when signals are vague. On the other hand, these preferences do not differentially relate to obscuring of negative versus positive feedback for precise signals of performance. We posit that, among advisors who place relatively less weight on transparency, the uncertainty inherent in vague signals may have lowered their psychological cost of lying (by omission), thus increasing their likelihood of obscuring feedback. This is likely to be more apparent when advisors have motives to omit, such as a desire to protect the advisee's ego from negative signals of performance. When signals are precise, the psychological cost of omission is higher, leaving less scope for selective obscuring of negative feedback.

We next investigate advisors' feedback decisions by the advisee's gender. We find suggestive evidence that the higher likelihood of obscuring negative signals than positive signals by advisors in the non-instrumental treatment may be attributed to advisors who are matched with female advisees. While this gender difference is of marginal significance and holds for advisors who pass the comprehension question in our follow-up experiment, two features of our data suggest that advisors may be more inclined to shield the ego utility of female advisees from negative feedback. First, examination of advisors' second-order beliefs reveals that advisors obscure negative feedback for female advisees whom they perceive as more confident about their ability, and who would consequently stand to suffer the greatest ego utility loss from

negative feedback. Second, in our follow-up experiment, advisors who are concerned about the ego of female advisees are more likely to obscure negative feedback (and less likely to obscure positive feedback), while concern for the male advisee's ego does not affect the advisors' feedback decisions.

How are advisees' tournament-entry decisions influenced by their advisors' feedback? By design, feedback in the non-instrumental treatment does not influence advisees' tournament-entry decisions. However, in the instrumental treatment, feedback *exacerbates* the gender-competition gap. This is because male advisees are more likely than female advisees to disregard negative feedback that advises against entering the tournament. Despite the widened gender-competition gap, advisors' feedback leads to a greater increase in earnings for female advisees than for male advisees.

Taken together, our results have critical implications for the labor market and educational settings. Given that feedback is a pivotal tool for enhancing performance and encouraging growth, the fact that a substantial amount of feedback is obscured suggests that some individuals may be operating at a disadvantage, deprived of the insights necessary to refine their skills and optimize their potential. Of particular concern is the gender disparity in feedback provision, which could perpetuate existing inequalities in the workplace and academic arenas. If women receive less feedback or more vague feedback, then they may face challenges in identifying areas for improvement, potentially hindering their career progression and/or academic achievements. Importantly, these consequences can arise even when feedback is not perceived to be immediately instrumental, since feedback received today may shape decisions and/or performance in future tasks, or even in other domains (Huang and Murad, 2020; Banerjee et al., 2020).

Our findings also provide us with new insights on potential ways to counteract biases in feedback provision. One immediate recommendation is for institutions to combine structured and objective feedback tools backed by incentives or mandates, with deliberate norm- or culture-building that portrays feedback provision as a core duty and defining feature of good mentorship. This allows the alignment of transparency norms with advisors' professional identity, and is especially important in situations where there may be no direct financial incentives for advisors to provide transparent or honest feedback (e.g., as in the case of mentorship or peer-to-peer evaluations). More broadly, our findings underscore the need for measures which can actively discourage stakeholders from obscuring feedback.

In what follows, Section 2 discusses our contributions to the related literature, Section 3 details the experimental design, and Section 4 summarizes our theoretical framework and predictions. Section 5 presents the results for advisors' behavior, with Section 6 presenting potential mechanisms driving their behavior. Section 7 turns our attention to advisees' behavior and their response to advisors' feedback. Section 8 concludes.

2 Contributions to Related Literature

Our study contributes to the large literature on feedback provision. This literature, highlighted in recent reviews by Villeval (2020, 2023), has largely focused on the role of feedback provision on performance and behavior (e.g., Eriksson et al., 2009; Gürtler and Harbring, 2010; Gill and Prowse, 2014; Chen and Schildberg-Hörisch, 2019; Gill et al., 2019; Drouvelis and Paiardini, 2021; Benistant et al., 2022). The majority of this literature focuses on environments where it is assumed that unbiased feedback is exogenously provided to decision-makers, e.g., through an institutionally-enforced feedback provision mechanism.

A further set of studies examines contexts where other parties can lie in their provision of feedback to decision makers. Gneezy et al. (2017) find that individuals tend to lie in the feedback they give about the beauty of other individuals in order to avoid providing negative feedback. In a client-agent interaction where the agent's payoff may depend on the client's happiness level, Ho and Yeung (2014) find that agents use feedback strategically by inflating their feedback when it is (financially) beneficial for them to do so, and clients are naïve when interpreting this distorted feedback. Serra-Garcia et al. (2011), Sun and Papadokonstantaki (2023), and Alempaki et al. (2025) find that individuals prefer partial truth, vague language, and silence, to direct lying because they do not like to deceive and to be seen as deceptive.[1] Specifically examining the effects of gender in feedback decisions, Jampol and Zayas (2021) find that under-performing women tend to receive more "white lies" and more positive feedback in subjective evaluations of written essays.

However, in many contexts where performance feedback is provided, advisors have considerable latitude to offer feedback that falls between full honesty and outright deception. Hence, we depart from previous studies by considering a setup where the feedback provider can choose to either partially disclose information (i.e., by providing vague feedback) or to not disclose any information at all. Beyond this, we consider an environment where the financial incentives of the feedback sender and recipient are decoupled. This allows us to systematically evaluate whether and how feedback provision is affected by the feedback sender's concerns about the ego and instrumental utilities of the recipient, absent strategic concerns which may otherwise arise if the feedback sender's payoffs are affected by the recipient's actions.[2] This mirrors many settings and relationships that we observe in practice, such as mentoring and peer evaluations (Gibbs, 1991), as well as performance evaluation in corporate settings, where the feedback

[1] Departing from feedback provision to individual decision makers, Ertac et al. (2016, 2019) consider a team environment where a principal has superior information about the performance of two agents under their charge, and they have to decide whether or not to withhold feedback, and whether to provide feedback privately to each agent or publicly to both.

[2] Regarding such strategic behavior, prior research has examined behavior in "disclosure games", but in these environments feedback typically does not have any impact on the recipient's ego utility. Instead, the sender is modelled as a self-interested party whose payoff is directly affected by the recipient's choice. Jin et al. (2021) examine the binary decision to either reveal or not reveal the true state of the world, while Deversi et al. (2021) provide senders with the option of vague disclosure.

provider often has indirect or misaligned financial incentives to provide feedback (Prendergast and Topel, 1993). Feedback provision in such contexts may be biased by factors such as information precision, social preferences, or in-group/out-group biases (Bol, 2011).

With our examination of gender differences in feedback provision, our paper contributes to the large literature seeking to understand the factors contributing to gender gaps in labor market outcomes. Within this literature, a body of work has focused on gender gaps in competitive preferences as well as possible interventions to close this gap (Croson and Gneezy, 2009; Niederle and Vesterlund, 2011). Recent work by van Veldhuizen (2022) shows that men are significantly more overconfident than women, resulting in gender differences in self-selection into competitive payment schemes and leadership positions (Reuben et al., 2012; Alan et al., 2020). To this end, scholars have also investigated the role of information and feedback interventions in mitigating gender gaps in competitiveness (Hannan et al., 2008; Wozniak et al., 2014; Brandts et al., 2015; Berlin and Dargnies, 2016; Kessel et al., 2021; Lovász et al., 2022). The general consensus is that exogenous feedback helps to mitigate gender gaps.

In practice, feedback is often endogenously determined by decision makers (e.g., mentors, advisors, managers) and may reflect their biases. A related literature examines advice giving for future decisions, noting that this literature typically focuses on generic advice giving without knowledge of recipient characteristics or performance information (e.g., Schotter, 2003; Chaudhuri et al., 2006; Schotter and Sopher, 2007; Chaudhuri et al., 2009; Ding and Schotter, 2019).[3] Three recent exceptions, which also focus on gender, are Gallen and Wasserman (2021), Brandts and Rott (2021) and Silva-Goncalves and van Veldhuizen (2020). In a field experiment, Gallen and Wasserman (2021) find female students receive more work-life balance advice than males. Brandts and Rott (2021) find in a lab experiment that while advice-giving differs by advisor gender, it does not depend on advisee gender. On the other hand, Silva-Goncalves and van Veldhuizen (2020) find no gender difference in advice giving when the advisors have subjective information about advisees' performance (based on their motivation letters and confidence ratings), where the advice given relates to whether to undertake jobs of varying difficulty levels. Distinct from these studies which examine advice giving, we focus on an environment with asymmetric information where advisors have private information about the performance of advisees, and they have the opportunity to obscure this performance information from them.

3 Experimental Design

Participants are randomly assigned the roles of advisors and advisees, and they take part in the experiment asynchronously in three phases. Advisees participate in Phases 1 and 3, while advisors participate in Phase 2. When recruited, advisees are explicitly informed that they

[3] Brandts et al. (2015) do provide advisors (and advisees) with performance information, and find that advice improves decisions but does not eliminate gender gaps.

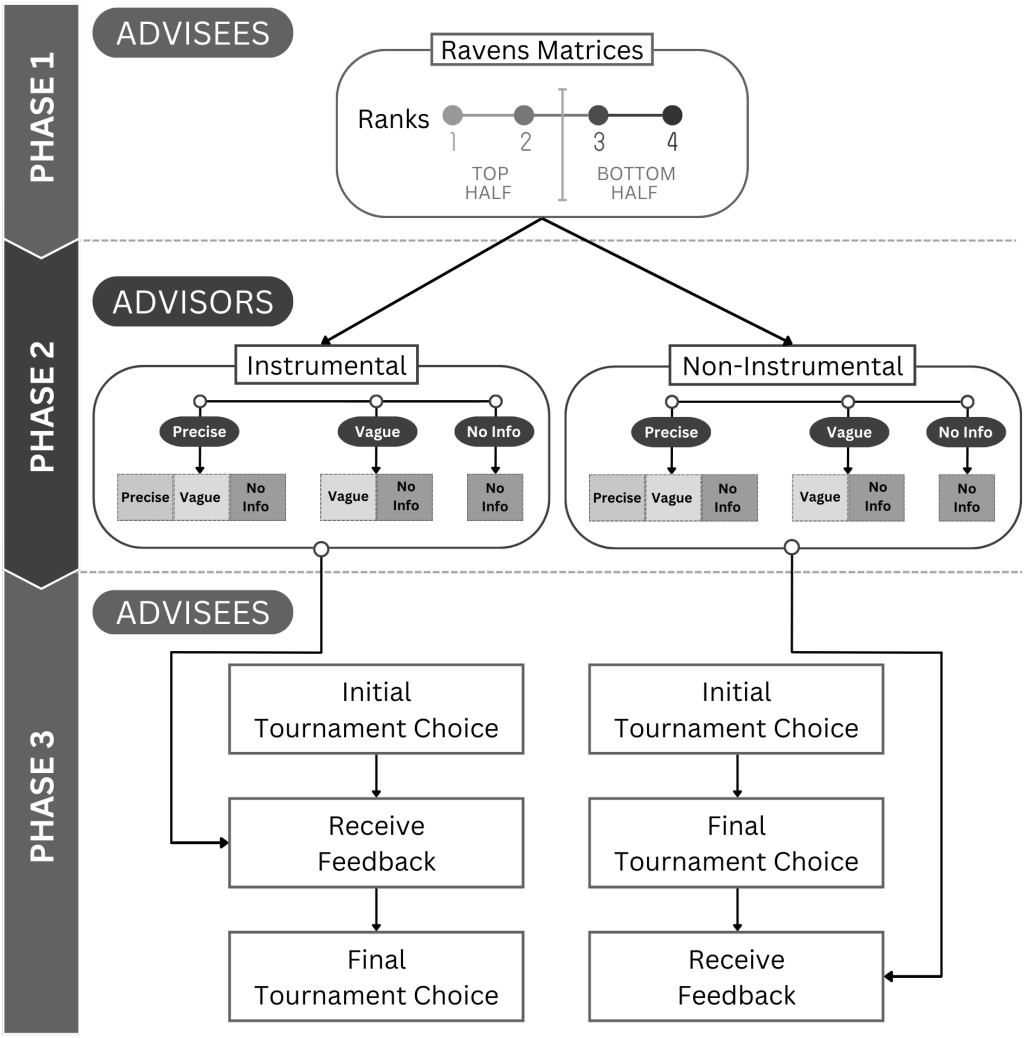


Figure 1: Timeline of the Experiment

are participating in a two-part study, and that they will be invited to complete the second part of the study within 7 days of completing Phase 1. Figure 1 summarizes the timeline of the experiment, with screenshots of the interface available in Appendix E.[4] We detail each phase of the experiment below.

3.1 Timeline of Experiment

Phase 1: Advisees complete task and demographic questionnaire. In Phase 1, participants are recruited as advisees. Advisees first complete a demographic questionnaire. Their responses are used to create a short profile that will be presented to their matched advisors in Phase 2.

Next, advisees are asked to solve 20 Raven's Progressive IQ task questions. Each advisee

[4]To enhance participant comprehension in the experiment, we framed the decision task using a manager/worker context, with advisors labeled as "Managers" and advisees labeled as "Workers".

is ranked against 19 other advisees and paid based on their quartile rank in this group of 20. Advisees ranked in the top quartile (ranks 1-5) receive £0.20 for each correct answer, while those in the remaining three quartiles (ranks 6-20) receive £0. Hence, the tournament incentives faced by advisees are similar to that in Niederle and Vesterlund (2007), where only the top 25% of performers receive a positive piece-rate payment for their task performance. To increase the ego-relevance of the task, we explain to participants that the task is based on an IQ test that is commonly used to measure people's intelligence levels, and performance in similar tasks have been found to be correlated with individuals' salaries, job positions, and life satisfaction (a similar ego-relevance manipulation was used in Drobner and Goerg, 2024).

After advisees complete the task, we elicit their prior beliefs about their relative performance. Specifically, they assign 10 tokens across 4 quartile ranks to indicate their belief of falling within each performance quartile. We incentivize advisees' beliefs using the Binarized Scoring Rule (Hossain and Okui, 2013; Erkal et al., 2020), where advisees may receive a fixed reward (£0.10 in this case) with a probability that is determined by the accuracy of their guesses. Instead of providing full details of the scoring rule, we state to participants that the payment scheme is designed such that they can secure the largest chance of receiving this payment by reporting their most accurate guess (Danz et al., 2022).

At the end of Phase 1, advisees receive no information about their task performance. Moreover, since the advisee's bonus payment is determined by their responses to several questions throughout the experiment in addition to their task performance (from either Phase 1 or Phase 3 of the experiment), it is unlikely that they will be able to infer their performance based on their earnings.

Phase 2: Advisors receive information about advisees and choose feedback. In Phase 2, a different sample of participants are recruited as advisors. The Phase 2 experiment is divided into two parts which advisors complete in one sitting. In Part 1, to allow advisors to understand the task and incentives faced by the advisees, advisors complete the exact same set of tasks as the advisees have done in Phase 1.

In Part 2, each advisor is randomly matched to one advisee and shown a profile that is constructed based on the advisee's responses to the demographic questionnaire. Specifically, the profile displays the advisee's gender, quarter of birth, favorite hot beverage, and favorite color. To increase the salience of the advisee's gender, gendered pronouns (e.g., "he/him" or "she/her") are used whenever the instructions refer to the advisee. Crucially, our between-subjects design, where each advisor is matched to only one advisee with the advisee's gender manipulated across different advisors, allows us to reduce the salience of the gender treatment manipulation to our participants, thus mitigating potential experimenter demand effect concerns.

After advisors are shown their advisee's profile, we elicit their prior beliefs about the advisee's performance in the Raven's IQ task. These beliefs are also incentivized using the Binarized Scoring Rule (with the fixed reward being £0.20). Next, advisors receive a signal of their advisee's quartile rank in the task. There are four possible quartile ranks (1, 2, 3, and 4). The advisor could receive either a *precise* signal (i.e., they are told the advisee's precise quartile rank), a *vague* signal (i.e., they are told whether the advisee is in the top half or bottom half quartiles) or *no* signal (i.e., they are simply told that the advisee's quartile is between 1 and 4). The probabilities of receiving a *precise*, a *vague*, or *no* signal are 0.65, 0.30, and 0.05, respectively, so that there is roughly an equal distribution of advisors receiving each signal type (i.e., an equal distribution of rank 1, 2, 3, 4, top-half, and bottom-half signals). These probabilities are common knowledge.[5] Advisors are then asked to choose how they would like to convey this signal to their advisees. Specifically, they can choose to send either precise, vague, or no feedback. Each advisor's options for feedback are limited based on the nature of the signal they received (e.g., an advisor is only able to send either vague or no feedback if they observe a vague signal).[6] Moreover, feedback is restricted to be true, i.e., advisors cannot convey feedback that is not contained in the original signal they have received.

When deciding how to convey the feedback to advisees, the advisors are informed that: (i) their advisee will be re-matched to a new group of 20 advisees and assigned a new rank within this group based on their absolute performance in Phase 1, and (ii) their advisee can choose whether to enter their performance into a tournament within their new group or be paid a piece rate. Under the tournament scheme, advisees will receive £0.20 for each correct answer if they are ranked in the top quartile within the new group, while under the piece-rate scheme, advisees will receive £0.05 for each correct answer in Phase 1 regardless of their rank in the new group.

Advisors (and their matched advisees) are randomly assigned to one of two treatments which vary the instrumentality of the feedback that the advisors send to their advisees. In the *Instrumental-Feedback* treatment, advisors are informed that their advisee will receive the feedback *before* they make their tournament-entry decision in Phase 3, while in the *Non-Instrumental Feedback* treatment, advisees will receive the advisor's feedback only *after* they have made their choice in Phase 3. Hence, in the Instrumental-Feedback treatment, advisors' feedback can potentially play an instrumental role in shaping the advisee's tournament-entry decision, while such an instrumental role is absent in the Non-Instrumental Feedback treatment. Advisors are explicitly informed that, aside from their feedback, advisees will not receive any additional information about their task performance. This design element is essential for emphasizing both the instrumentality of the advisor's feedback and the impact of feedback on the

[5] As explained in our theoretical framework (Appendix A), it is important that a non-zero proportion of advisors receive no signals. This implies that when advisees receive no feedback, they are unable to determine whether their advisor has received no signal or has chosen to obscure a precise/vague signal.

[6] Since an advisor who receives no signal has no option other than to send their advisee no feedback, we exclude these advisors from our analysis.

advisee's ego utility.

After the advisors have made their feedback decisions, we elicit their beliefs about the advisee's *new* quartile rank given the signal they received earlier. Note that, since advisees will be matched to a new group of 20 in Phase 3 and be given a new rank in this new group, the previous signal advisors receive about their advisee's performance may be different to the new rank the advisee receives. At the end of the experiment, advisors complete a short survey including questions about the decisions they have made during the experiment, incentivized manipulation check questions, questions eliciting their beliefs about their matched advisee, (second-order) beliefs about the advisee's reported confidence in their task performance, as well as a modified Implicit Association Test (IAT) relating to gender bias (Greenwald et al., 1998).

Phase 3: Advisees receive advisor's feedback and make tournament-entry decisions. In Phase 3, we invite advisees back for the second part of the study. Advisees are first reminded of the tasks they had to perform in Phase 1, as well as how their quartile rank and incentives were determined. Then, they are informed that they will be matched to a new group of 20 advisees and given a *new* quartile rank based on how their Phase 1 performance compares within the new group. Advisees are also informed that they can choose between receiving a piece rate based on their Phase 1 performance or entering their performance into a tournament within this new group.

In both the Instrumental-Feedback and Non-Instrumental Feedback treatments, we first ask advisees to decide whether or not they would like to enter into the tournament before receiving any feedback. This provides us with a baseline measure of the advisees' competitive preferences across both treatments. It also allows us to examine whether, in the Instrumental-Feedback treatment, advisees revise their entry decisions given their advisor's feedback.

Next, advisees are informed that they have been matched with an advisor and given detailed information about the decisions that their advisor faced. This includes a description of the distribution of possible signals that their advisor could have received, as well as the feedback decisions the advisors are asked to make. In both treatments, advisees are asked to make a tournament-entry decision *again*, and they are informed that the latter decision will overwrite the earlier decision they have made and be used to determine their payoff in Phase 3. The timing in which advisees receive their advisor's feedback, relative to when they make their second tournament-entry decision, depends on which of the two treatments (Instrumental-Feedback or Non-Instrumental Feedback) they and their matched advisor have been assigned to.

Finally, in both treatments, after receiving the advisor's feedback, we elicit advisees' beliefs about their new quartile rank. They then complete a short survey about the decisions they have made during the experiment, including an open-ended question on the reasons behind their

tournament-entry choice, as well as a one-shot risk task (Gneezy and Potters, 1997).

3.2 Procedures

The experiment was conducted online using oTree (Chen et al., 2016). We recruited a total of 2,620 participants residing in the United States (1,305 advisees and 1,315 advisors) via Prolific, with a balanced sample of male and female participants for both the advisor and advisee sessions. Moreover, the advisor-advisee matches were implemented such that there was a balanced sample of male-to-male, female-to-female, male-to-female, and female-to-male matches, thus allowing us to examine whether advisors' feedback decisions depend on both the advisee's and their own gender.[7] The experiment was pre-registered at AEA RCT Registry (AEARCTR-0006966) prior to the collection of data.

To minimize attrition, we conducted the experiments in eight waves, with each wave lasting about a week. Within each wave, we recruited advisors for Phase 2 immediately after all the advisees had completed Phase 1. Once Phase 2 was completed, advisees were invited back to complete Phase 3 within the same week. Of the 1,305 advisees who completed Phase 1, 1,199 (92%) returned to complete Phase 3 of the experiment. Of the 106 advisees who did not return to complete Phase 3, 53 are male, 52 are female, and 1 identifies as non-binary. There are no gender differences in attrition rates (Fisher's exact test: p-value = 0.359).[8] Phases 1 and 3 lasted about 10 minutes each, while Phase 2 lasted about 20 minutes in total.

Before making their decisions, participants had to complete comprehension and attention check questions to test their understanding of the decision-making environment and to measure their attention to the on-screen instructions. Participants were only paid if they did not fail any attention checks. Advisees received a fixed reward of £2 and a bonus payment based on their decisions in either Phase 1 or Phase 3 of the experiment, with payments made only if they completed Phase 3. In Phase 2, advisors received a fixed reward of £2.25, as well as a bonus payment based on either: (i) their own performance in the Raven IQ task (Part 1), or (ii) the accuracy of their beliefs about the advisee's performance (Part 2) and for completing the IAT. On average, the bonus payments paid to advisees and advisors were £1.21 and £1.00, respectively. This translates to a total average payment of approximately £9.70 per hour.

[7] Table B.1 provides a break down of the sample by treatment assignment. Due to a recruitment error, two advisors were mistakenly matched with the same advisee in 10 advisor-advisee pairs. In these instances, we randomly picked one of the two advisors' decisions to implement.

[8] Table B.2 presents OLS estimates of attrition by advisees against all other advisees' characteristics. We find that the attrition rate is lower for advisees who perform better in the Raven's Matrices task (p-value = 0.027), as well as those who are older (p-value = 0.062), while the attrition rate is higher for advisees who are Black or African American (p-value = 0.049). Importantly, column (2) shows that the interaction between gender and performance in the Raven's task is not statistically significant (p-value = 0.287). Hence, even though low-performing advisees are less likely to return to Phase 3, there is no gender difference in the impact of task performance on attrition rate. Nonetheless, we include controls for these characteristics in our analysis.

4 Theoretical Framework

We present a simple theoretical framework in Appendix A to help guide the interpretation of our results. Here, we summarize the key observations resulting from our theoretical framework.

We consider an advisor who is altruistic and cares about both the ego and instrumental utilities of the advisee. Regardless of the instrumentality of feedback, sending positive feedback will increase the advisee's ego utility while sending negative feedback will decrease the advisee's ego utility. On the other hand, sending either type of feedback will increase the advisee's instrumental utility, since the advisee will be able to make a more informed tournament-entry decision when provided with feedback about their performance. Note that this latter effect only applies in the Instrumental-Feedback treatment.

Against this backdrop, we make the following observations. First, independent of the instrumentality of feedback, an advisor who receives a positive signal will always send the received signal as feedback. This is because sending feedback will increase both the ego- and instrumental utility of the advisee. Second, when feedback is non-instrumental and an advisor receives a negative signal, the advisor will obscure feedback. This is driven by the advisor's concern about the advisee's ego utility, as sending negative feedback will decrease the advisee's ego utility.

Finally, when feedback is instrumental and an advisor receives a negative signal, whether they obscure feedback or send it depends on their belief about the advisee's confidence. The more (over)confident the advisor believes their advisee to be, the less likely they are to obscure feedback. This is because advisors expect more confident advisees to mistakenly enter the tournament, a loss in instrumental utility that would outweigh the ego-utility benefit of obscuring negative feedback.

Taken together, the likelihood of obscuring signal decreases if: (i) signal is positive, (ii) feedback is instrumental, and (iii) the advisor holds higher second-order beliefs about the advisee's confidence. In terms of gender differences in feedback provision, our theoretical framework is agnostic as to whether advisors treat male and female advisees differently, and this is an empirical question that we will evaluate in our data. One channel through which a gender bias could manifest is if advisors have different second-order beliefs about the advisee's confidence (e.g., Exley and Nielsen, 2024). Additionally, advisors may care differently about the ego and/or instrumental utilities of the advisee, e.g., if some advisors exhibit benevolent sexism or engage in paternalistic discrimination (e.g., Glick and Fiske, 2001; Viki et al., 2003; Buchmann et al., 2023; Erkal et al., 2023).

5 Advisors' Feedback Decisions

5.1 Do advisors obscure feedback to advisees and does this depend on the instrumentality of feedback?

Figure 2 presents advisors' feedback choice based on the signal that they receive, both overall (panel a) and separately by the instrumentality of feedback (panels b and c).[9] We consider the analysis separately for advisors who receive precise signals of performance and those who receive vague signals of performance. Previous research suggests that the precision of information may contribute to biases in feedback provision (Bol, 2011; Serra-Garcia et al., 2011; Sun and Papadokonstantaki, 2023). In our context, vaguer signals of performance may provide more ambiguity to advisors about the true performance of their advisee, thus providing some wiggle room for advisors to obscure feedback. As robustness, Appendix Figure B.1 pools advisors' feedback decisions across both vague and precise signals separately for each treatment, where we find advisors' behavior to be generally in line with our conclusions from Figure 2.

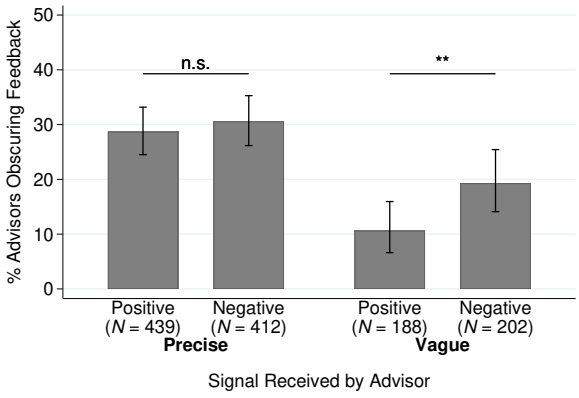
In all the figures, we report the proportion of advisors who obscure feedback, that is, they provide feedback that is vaguer than the signals that they received. This is our key variable of interest. Note that this includes advisors who provide vague or no feedback when they receive precise signals, or those who provide no feedback when they receive vague signals. In all figures, vertical error bars represent 95% confidence intervals.

We observe from Figure 2 that a significant amount of feedback is obscured. Overall, advisors obscure feedback to their advisees about 25% of the time. This is the case even when they receive a positive signal about their advisee's performance. Panel (a) reveals that, across both treatments, of the advisors who are given a precise or vague positive signal, 29% and 11%, respectively, choose to obscure this feedback from their advisees.

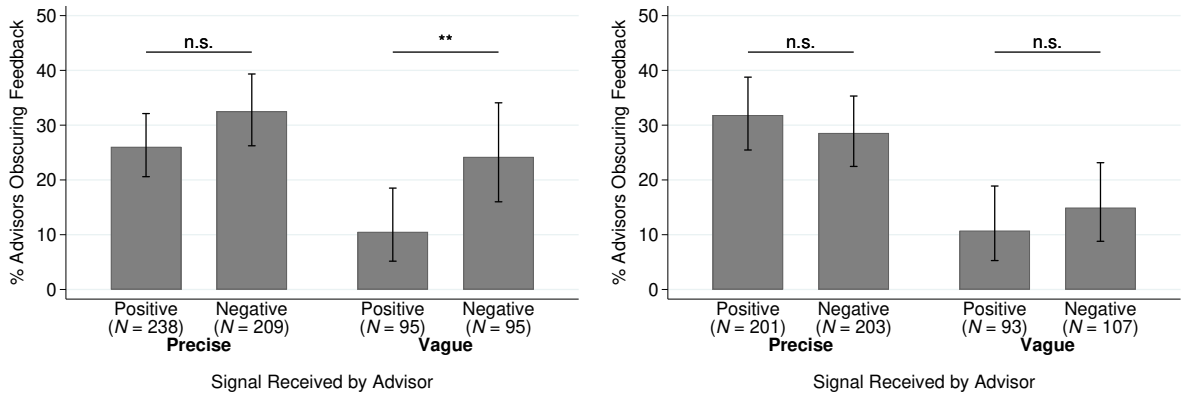
Moreover, panel (a) reveals that advisors are more likely to obscure feedback when they observe a vague negative signal than a vague positive signal of their advisee's performance. 19% of advisors choose to obscure their feedback (by sending no feedback) when they observe a vague negative signal (that the advisee is in the bottom half of the distribution), as compared to 11% of those who observe a vague positive signal (that the advisee is in the top half of the distribution) (Fisher's exact test: p-value = 0.023). The difference in feedback provision between positive and negative signals is not statistically significant with precise signals (Fisher's exact test: p-value = 0.549).

We next investigate whether advisors' decisions depend on the instrumentality of their feed-

[9]For our main analysis, we consider "positive" signals to be ranks 1, 2, or top half, and "negative" signals to be ranks 3, 4, or bottom half. Given that entering the tournament only pays off for advisees if they are rank 1, advisors may treat signals of ranks 1 and 2 differently. To account for this, as robustness, we recode "positive" signals to be ranks 1 or top half only, and "negative" signals to be ranks 2, 3, 4, or bottom half. The results using the recoded definition of positive and negative signals, reported in Figures B.2 and B.3, and Tables B.3 and B.4, are consistent with the conclusions from our main analysis.



(a) Overall



(b) Non-Instrumental Feedback

(c) Instrumental Feedback

Note: Error bars represent 95% confidence intervals. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, n.s. $p > 0.10$.

Figure 2: Proportion of advisors sending feedback that is vaguer than received signal (by signal valence, signal precision, and treatment) [Main Experiment]

back. Comparing behavior across treatments (panels b and c), we observe no statistically significant treatment differences in advisors' propensity to obscure either positive or negative signals. That is, advisors are equally likely to obscure non-instrumental and instrumental feedback, regardless of whether they receive a positive or a negative signal of performance. This is true for both vague and precise signals.[10]

We now examine how advisors respond to positive and negative signals of performance within each treatment. Panel (b) of Figure 2 reveals that, when advisors' feedback is non-instrumental in influencing advisee's tournament-entry decisions, their propensity to obscure feedback is in line with what we observe in panel (a). Specifically, with non-instrumental feedback, advisors are more likely to obscure vague negative signals than vague positive signals (Fisher's exact test: p-value = 0.021). While advisors are also slightly more likely to obscure

[10]Fisher's exact test of the choice to obscure feedback comparing non-instrumental to instrumental treatments in the case of: (i) precise positive signal: p-value = 0.204; (ii) precise negative signal: p-value = 0.394; (iii) vague positive signal: p-value = 1.000; and (iv) vague negative signal: p-value = 0.110.

Table 1: OLS regression results of advisors' feedback decisions (overall and by treatment) [Main Experiment]

	Overall			Non-Instrumental		Instrumental	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable: Advisor obscures feedback to advisee							
Negative signal	0.087**	0.097***	0.144***	0.137**	0.131**	0.042	0.079
	(0.036)	(0.036)	(0.054)	(0.054)	(0.054)	(0.047)	(0.049)
Precise signal	0.180***	0.184***	0.163***	0.155***	0.156***	0.211***	0.229***
	(0.031)	(0.031)	(0.042)	(0.043)	(0.043)	(0.046)	(0.047)
Negative signal × Precise signal	−0.068	−0.083*	−0.087	−0.072	−0.077	−0.075	−0.117*
	(0.048)	(0.048)	(0.070)	(0.069)	(0.070)	(0.066)	(0.068)
Instrumental feedback	−0.008	0.002	0.016				
	(0.024)	(0.025)	(0.044)				
Negative signal × Instrumental			−0.089				
			(0.073)				
Precise signal × Instrumental			0.046				
			(0.063)				
Negative signal × Precise signal			−0.003				
× Instrumental feedback			(0.097)				
Constant	0.110***	0.187**	0.184*	0.105***	0.109	0.108***	0.314**
	(0.026)	(0.093)	(0.095)	(0.032)	(0.132)	(0.032)	(0.138)
Negative signal	0.019	0.014	0.057	0.065	0.054	−0.033	−0.038
+ Negative signal × Precise signal	(0.031)	(0.032)	(0.044)	(0.043)	(0.044)	(0.046)	(0.047)
Controls	N	Y	Y	N	Y	N	Y
Observations	1,241	1,241	1,241	637	637	604	604
R^2	0.028	0.049	0.053	0.026	0.064	0.037	0.085

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Robust standard errors in parentheses.
Other controls include the advisor's gender, age, education level, political views on social issues, and U.S. political party affiliation, as well as advisee's gender, quarter of birth, favorite color, and favorite beverage.
The baseline group in all specifications is advisors who observe a positive signal of "Rank 1", "Rank 2", or "Top Half".

precise negative signals than precise positive signals, this difference is not statistically significant (Fisher's exact test: p-value = 0.144). On the other hand, when advisors' feedback is instrumental (panel c), advisors do not differ in their propensity to obscure negative versus positive signals, regardless of whether the signal is precise or vague (Fisher's exact tests: p-values = 0.516 and 0.407, respectively).

Table 1 presents estimates of OLS regressions of advisors' decision to obscure feedback to their advisees, both overall (columns 1, 2, and 3), and separately for non-instrumental feedback (columns 4 and 5) and instrumental feedback (columns 6 and 7). The baseline comparison group in all columns is advisors who observe a vague positive signal. Hence, the coefficients in the first row provides a test of whether advisors are more or less likely to obscure negative vague signals than positive vague signals. In the last row, we evaluate whether advisors differ in their propensity to obscure negative precise signals relative to positive precise signals. Coefficient estimates of probit regressions presented in Appendix Table B.5 lead to similar conclusions.

The estimates in Table 1 are consistent with our observations in Figure 2. Overall, there is no statistically significant difference in the propensity of advisors to obscure feedback between the instrumental feedback and non-instrumental feedback treatments. The interaction terms in column (3) suggest that the absence of this treatment difference holds regardless of the

signal valence or signal precision. Nonetheless, on average, advisors are more likely to obscure negative vague signals than positive vague signals, and this is driven by behavior when feedback is non-instrumental. Regardless of the instrumentality of feedback, there is no statistically significant difference in advisors' propensity to obscure negative versus positive signals when advisors receive precise signals.

We summarize as follows.

Result 1. *(a) A non-trivial proportion of advisors obscure feedback to their advisees, even when they receive positive signals about the advisee's performance.*
(b) Advisors do not differ in their overall propensity to obscure performance signals between the instrumental and non-instrumental feedback treatments. This holds regardless of signal precision or signal valence.
(c) Advisors are no different in their propensity to obscure positive and negative signals from their advisee when feedback is instrumental.
(d) When feedback is non-instrumental, advisors are more likely to obscure feedback when they receive a vague negative signal about their advisee's performance than when they receive a vague positive signal. Nonetheless, advisors are no different in their propensity to obscure precise positive and precise negative signals.

5.2 Do advisors obscure feedback to male and female advisees differently?

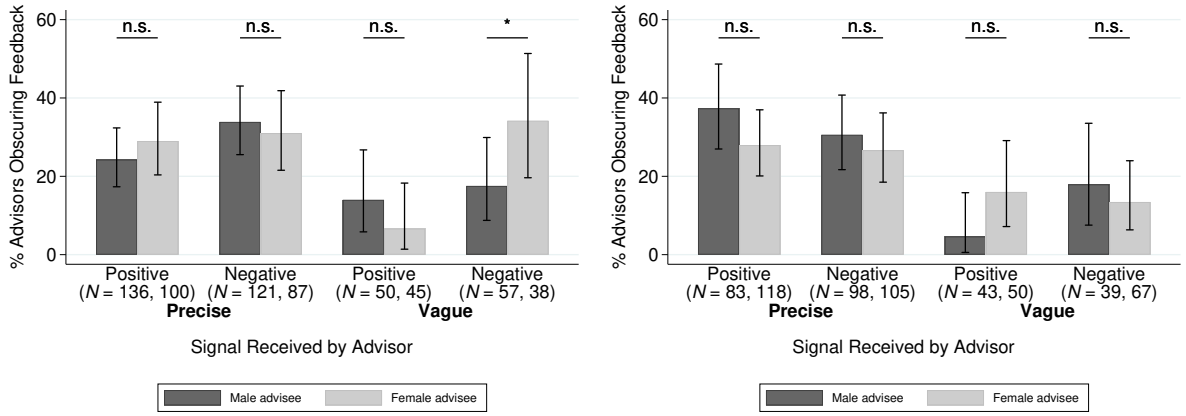
We next examine whether advisors' decisions differ based on the advisee's gender. We find that our design successfully manipulates the gender of the advisee, whereby 1,215 of 1,241 advisors (97.9%) correctly recall the gender of their advisee in the incentivized post-experimental questionnaire.[11]

Appendix Figure B.4 presents advisors' propensity to obscure feedback based on the signal valence and instrumentality of feedback, where we pool data for both precise and vague signals. The figure reveals that, in aggregate, advisors are equally likely to obscure feedback from male and female advisees, regardless of the signal valence and the instrumentality of feedback.[12]

Next, we consider whether this holds regardless of whether advisors receive vague or precise signals of their advisee's performance (Figure 3). We first focus on the case where advisors provide non-instrumental feedback (panel a). There is suggestive evidence of gender differences in advisors' propensity to obscure feedback when they receive a vague signal. Specifically, when feedback is non-instrumental, advisors are nearly twice as likely to obscure negative

[11] Specifically, 26 advisors (2% of the sample) are unable to correctly recall their advisee's gender. Of these, 13 of them are matched with a male advisee while 13 are matched with a female advisee, and this difference is not statistically significant (Fisher's exact test: p-value = 1.000).

[12] Fisher's exact tests comparing advisors' choice to obscure feedback from male versus female advisees: (i) non-instrumental feedback, positive signals: p-value = 1.000; (ii) non-instrumental feedback, negative signals: p-value = 0.611; (iii) instrumental feedback, positive signals: p-value = 0.786; and (iv) instrumental feedback, negative signals: p-value = 0.285.



(a) Non-Instrumental Feedback

(b) Instrumental Feedback

Note: Error bars represent 95% confidence intervals. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, n.s. $p > 0.10$.

Figure 3: Proportion of advisors sending feedback that is vaguer than received signal (by signal valence, signal precision, treatment, and advisee's gender) [Main Experiment]

vague signals to female advisees (34.2%) than to male advisees (17.5%), although this difference is only marginally statistically significant (Fisher's exact test: p-value = 0.087). On the other hand, there is no gender difference in the propensity to obscure positive signals (6.7% for female advisees and 14.0% for male advisees) (p-value = 0.324). When advisors receive a precise signal about their advisee's performance, there are no statistically significant differences in the proportion of advisors who obscure feedback between male and female advisees (positive signals: p-value = 0.456; negative signals: p-value = 0.765). Hence, we find suggestive evidence that Result 1(d)—that advisors are more likely to obscure vague negative than vague positive feedback in the non-instrumental treatment—may be driven by advisors matched with female advisees.

When feedback is instrumental (panel b), we observe no statistically significant gender differences in the proportion of advisors who obscure feedback when they receive a precise signal (positive signals: p-value = 0.170; negative signals: p-value = 0.539). This is also the case when advisors receive vague signals. Specifically, there is no gender difference in the propensity to obscure negative signals (p-value = 0.580). Moreover, advisors are more likely to obscure positive signals from female advisees than to male advisees. However, this difference is (marginally) not statistically significant (p-value = 0.100).

Table 2 presents OLS estimates separately for non-instrumental feedback (columns 1 and 2) and instrumental feedback (columns 3 and 4).[13] The baseline comparison group in all specifications is advisors who are matched with a male advisee and who observe a vague positive signal. For ease of interpretation, we focus on the last four rows, which present the combined effect on the propensity to obscure feedback when advisors are matched with a female advisee

[13]Probit coefficient estimates presented in Appendix Table B.6 yield similar conclusions.

Table 2: OLS regression results of advisors' feedback decisions (by treatment, interacting with advisee's gender) [Main Experiment]

	Non-Instrumental		Instrumental	
	(1)	(2)	(3)	(4)
Dependent Variable: Advisor obscures feedback to advisee				
Female advisee	−0.073	−0.069	0.113*	0.090
	(0.062)	(0.064)	(0.061)	(0.064)
Negative signal	0.035	0.028	0.133*	0.176**
	(0.071)	(0.068)	(0.070)	(0.074)
Precise signal	0.103*	0.113*	0.327***	0.328***
	(0.062)	(0.061)	(0.062)	(0.064)
Female advisee × Negative signal	0.240**	0.244**	−0.159	−0.168*
	(0.111)	(0.112)	(0.097)	(0.098)
Female advisee × Precise signal	0.121	0.098	−0.207**	−0.180**
	(0.085)	(0.085)	(0.091)	(0.091)
Negative signal × Precise signal	0.061	0.041	−0.200**	−0.239**
	(0.091)	(0.090)	(0.100)	(0.102)
Female advisee	−0.316**	−0.282**	0.213	0.210
× Negative signal × Precise signal	(0.142)	(0.143)	(0.134)	(0.134)
Constant	0.140***	0.171	0.047	0.236*
	(0.049)	(0.138)	(0.032)	(0.140)
Female Advisee Relative to Male Advisee by Signal Type				
Positive Vague	−0.073	−0.069	0.113*	0.090
	(0.062)	(0.064)	(0.061)	(0.064)
Negative Vague	0.167*	0.175*	−0.045	−0.078
	(0.093)	(0.094)	(0.075)	(0.077)
Positive Precise	0.047	0.029	−0.094	−0.090
	(0.059)	(0.060)	(0.068)	(0.067)
Negative Precise	−0.028	−0.008	−0.039	−0.048
	(0.066)	(0.068)	(0.064)	(0.066)
Controls	N	Y	N	Y
Observations	634	634	603	603
R^2	0.034	0.071	0.045	0.090

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Robust standard errors in parentheses.
Other controls include the advisor's gender, age, education level, political views on social issues, and U.S. political party affiliation, as well as advisee's quarter of birth, favorite color, and favorite beverage.
The baseline group in all specifications is advisors who observe a vague positive signal for a male advisee.

(as compared to with a male advisee) separately given each signal type.[14]

The estimates in the last four rows of Table 2 are broadly in line with the conclusions of

[14]In other words, conditional on signal type, it shows the impact of the advisee being female. Since the vague positive signal type is the baseline group, the estimate for *Female advisee* in the first part of the table, is therefore identical to the impact of *Positive Vague* in the second part of the table.

our non-parametric tests above. Gender differences emerge when advisors observe vague signals, but not when they receive precise signals about their advisees' performance. Specifically, columns (1) and (2) reveal that advisors are about 17 percentage points (p.p.) more likely to obscure vague negative signals to female advisees than to male advisees when feedback is non-instrumental, although these differences are only of marginal statistical significance (p-values = 0.072 and 0.063, respectively).

We summarize our key result as follows.

Result 2. *(a) When feedback is non-instrumental, there is suggestive evidence that advisors are more likely to obscure vague negative signals from female advisees than from male advisees. There are no statistically significant gender differences in feedback provision given precise signals.*
(b) When feedback is instrumental, there are no statistically significant gender differences in feedback provision.

Finally, we note that the estimates showcased in the bottom rows of Table 2 reveal that advisors are more likely to obscure positive vague signals from female advisees when feedback is instrumental (column 3: p-value = 0.065), but this difference is no longer statistically significant with the inclusion of controls in column (4) (p-value = 0.160). Investigating this further, we consider the inclusion of controls individually and observe that the decrease in significance is due to the inclusion of advisor's gender in column (4). To explore this finding further, Table B.7 breaks down the analysis by the advisor's gender. The estimates reveal that the observed gender biases are driven by *male advisors*. Specifically, male advisors are more likely to obscure vague negative signals from female advisees when feedback is non-instrumental (p-value = 0.021). They are also more likely to obscure vague positive signals from female advisees when feedback is instrumental (p-value = 0.016).

Taken together, both Result 1 and Result 2 imply that advisors are more likely to obscure negative signals than positive signals, and this is largely driven by advisors giving feedback that is non-instrumental in driving advisees' tournament-entry decisions. We find suggestive evidence of a gender bias in feedback provision, where advisors are more likely to obscure negative non-instrumental and, to some extent, positive instrumental feedback from female advisees than from male advisees.

6 What are the potential drivers of advisors' feedback decisions?

In this section, we explore potential mechanisms behind advisors' feedback decisions. We focus on three aspects of our results. First, we investigate why a non-trivial proportion of advisors choose to obscure positive feedback. Second, we ask why the difference in the provision of non-instrumental feedback by signal valence exists for vague signals but not for precise signals.

Finally, we explore potential mechanisms driving gender differences in feedback provision.[15]

6.1 Why do advisors obscure positive signals?

A non-negligible proportion of advisors obscure positive signals from advisees (Result 1a), which is a surprising result that is not consistent with our theoretical predictions. To better understand the mechanisms behind this finding, we design and run a follow-up experiment with the primary goal of capturing the potential motives that drive advisors' feedback decisions in a systematic manner. Specifically, we replicate the original non-instrumental feedback treatment and include a structured survey question in the post-experimental questionnaire to elicit advisors' reasons for their feedback decisions. Each advisor is shown: (i) the signal they received about the advisee's performance; (ii) the feedback they chose to send to their advisee; and (iii) a list of potential statements that may motivate their feedback choice, presented in a random order (see Appendix Figure B.5). They are asked to indicate which statement(s) explain their feedback choice.

The statements that are shown to advisors in Appendix Figure B.5 can be organized into several themes. A first set of explanations appeals to broad moral motives, such as desire for transparency or discretion (Abeler et al., 2019), and a sense of duty toward the advisee.[16] Second, advisors may be driven by self-serving motives, such as feeling envious of high-performing advisees and subsequently obscuring positive feedback. Relatedly, advisors may be concerned about how they will be perceived by the advisee based on the feedback they send. Third, some advisors may want to avoid the responsibility of providing detailed feedback, or they may worry that providing feedback could reveal to advisees that they have made a sub-optimal payment choice, thereby causing regret. Fourth, some advisors may be concerned about how the feedback will influence their advisee, either in terms of its usefulness, or its impact on the advisee's ego, motivation, or feelings, which is most closely aligned with the motives posited in our theoretical framework. Finally, some advisors may obscure feedback simply due to confusion. In total, we explore 14 different potential reasons, along with an open-ended "other" option.[17]

The follow-up experiment was conducted online using a similar protocol as described in Section 3.2. We collected data from 1,023 advisors who were all assigned to the non-instrumental feedback treatment (of whom 27, or 2.6%, are dropped in the analysis as they received no signal of their advisee's performance).[18] Appendix Figure B.6 presents advisors'

[15] We thank the associate editor and three anonymous reviewers for their excellent and constructive suggestions pertaining to the various potential mechanisms driving advisors' behavior.

[16] We use the term duty as an accessible and intuitive label to capture advisors' moral commitment to take a certain action (Sen, 1977).

[17] We can also examine some of these motives using data from the main experiment. The analysis, summarized in Appendix C, does not yield systematic evidence to suggest that advisors' envy of the advisee's performance, advisors' competence, or a lack of comprehension by advisors about the decision-making environment can explain our result. The follow-up experiment allows us to examine the motives in a more systematic manner.

[18] We focus only on the non-instrumental treatment for two key reasons. First, since we observe in our main experiment that advisors are equally likely to obscure positive feedback between treatments, we concentrate our lim-

feedback choice based on the signal that they receive in this follow-up experiment. A visual comparison with panel (b) of Figure 2 shows that we are able to replicate the core of our Result 1. Moreover, advisors in the follow-up experiment are even more likely to obscure signals (both positive and negative) as compared to those in our main experiment.[19]

The central variable of interest in this follow-up experiment is the reason(s) indicated by advisors for their feedback choice. Before presenting this data, it is worth noting that while this exercise allows us to study multiple potential motives, it may come at a cost of precise identification. The question is phrased to elicit all considerations that shaped participants' thinking, although some respondents may have focused on the consideration that ultimately tipped the balance in their feedback choice. In addition, some motives may be more socially desirable to endorse than others, even though the data collection is anonymous. Our main analysis relies on comparisons across feedback choices and signal conditions. All advisors are presented with the same set of statements regardless of the signal received or the feedback sent, thus allowing consistent comparisons across treatments. We also further complement this analysis with a large-language-model (LLM) analysis of free-form text responses from the main experiment, presented in Appendix D.[20]

Focusing exclusively on advisors who receive positive performance signals, Appendix Table B.8 summarizes the proportion of advisors who indicate each reason as a motivation for their feedback decision. It distinguishes between advisors who choose to send feedback (columns 1-3) and those who opt to obscure feedback (columns 4-6), and within each sub-group, between those who receive precise signals and vague signals of their advisee's performance. The last column in the table reports p-values from Fisher's exact tests of differences in the proportion of advisors motivated by a given reason between those who send feedback versus those who obscure feedback.

We make a number of key observations from Appendix Table B.8. First, transparency/ discretion and sense of duty stand out as the primary considerations in advisors' self-reported explanations of their feedback decisions. Those who send positive feedback are more likely to report a preference for transparency as well as a sense of duty than those who obscure feedback. Conversely, those who obscure positive feedback are more likely to indicate a preference for

ited budget on one treatment to boost statistical power. Second, since Result 1(d) is driven by the non-instrumental treatment, our follow-up experiment enables us to test the replicability of this finding. To further increase the sample size of advisors receiving vague signals, advisors receive a precise signal with a 49% chance, a vague signal with a 49% chance, and no signal with a 2% chance (these percentages are 65%, 30%, and 5%, respectively, in our main experiment).

[19] Appendix Figure B.7 shows that these findings are robust to the exclusion of advisors who fail the comprehension check question.

[20] We thank an anonymous referee for highlighting these limitations. Some concerns, such as social desirability, may also influence free-form responses, so the LLM analysis is not immune to them. However, because the two approaches rely on different modes of elicitation, the ways in which such concerns manifest need not be the same. Considering both therefore provides an additional check on the robustness of the patterns we document.

discretion.[21] Second, there is suggestive evidence that advisors care about their advisee's ego-utility, as well as whether the advisee will regret their choice of payment. Advisors who send positive feedback are more likely to indicate a concern for their advisee's ego, and less likely to indicate a concern that their advisee will regret their choice, than those who obscure feedback. Third, we find patterns consistent with self-image concerns on the advisors' part. Advisors who obscure positive feedback are more likely to indicate concerns about the advisee's opinion of them and how their performance compares to their advisee's.[22] Fourth, advisors who send positive feedback are more likely to indicate the usefulness of such feedback to the advisee as compared to those who obscure feedback. This is despite feedback being non-instrumental within the context of the experiment. Finally, idiosyncratic reasons (such as random decisions or mistakes) can explain a minority of advisors' decisions to obscure positive feedback (less than 5% of advisors).[23]

Appendix Table B.9 reports OLS regression estimates of advisors' feedback decision against the reasons stated by advisors as motivating their feedback choice. Columns (1) to (15) show whether each individual reason is correlated with advisors' decisions to obscure positive feedback. Focusing on column (16) which considers all the reasons within the same specification, we find that transparency- and discretion-related considerations, as well as a sense of duty, remain as significant predictors of advisors' decisions to obscure positive performance signals. Advisors' concern for the advisee's ego is a weak predictor (only marginally significant) of their feedback choice. Moreover, both self-serving motives (such as their concern about how the advisee will view them, as well as how their performance compares with the advisee's performance) and the usefulness of feedback remain as significant predictors of the advisors' feedback choice.

Appendix Table B.10 reports results from a dominance analysis of the regression reported in column (16) of Appendix Table B.9. Interpreting these along with the regression estimates, we conclude that there are three core sets of reasons that can explain why advisors tend to obscure positive feedback. First, advisors who do not self-report a preference for transparency or a sense of duty (or conversely, those who report a preference for discretion) are more likely

[21] A non-trivial share of advisors who obscure feedback nonetheless select transparency-related reasons (40%) or a sense of duty (36%). As discussed above, the question allows respondents to indicate considerations that enter their decision process, not only the factor(s) that ultimately determine the message sent. We interpret these selections as indicating that transparency-related concerns are present but outweighed by other considerations.

[22] Research in psychology shows that many individuals are socially avoidant and prefer to avoid situations in which others form opinions about them. Weeks et al. (2008) show that there exists a "fear of positive evaluation", whereby sending positive feedback could still trigger a sense of being evaluated. In our context, the advisee may judge the advisor's motives, and this can be uncomfortable for individuals who are sensitive to how they are being perceived by others.

[23] We opted to provide a structured list to capture nuances in the factors driving behavior (such as envy, uncertainty/ambiguity in the information about the advisee's performance, and other-image concerns) that may be challenging to capture through analysis of free-form text alone. Appendix D presents complementary evidence from the LLM analysis of free-form responses, showing a similarly important role for transparency- and discretion-related considerations in motivating advisors' feedback decisions.

to obscure positive signals of performance. This set of reasons is the most dominant (ranks 1 to 3 in Appendix Table B.10). Second, advisors who do not perceive the positive feedback to be useful are more likely to obscure positive performance. This comes across as the next most dominant reason (rank 4). Finally, advisors who compare their own performance with that of the advisee, as well as those who are concerned with how the advisee would view them, are more likely to obscure positive signals of performance. These self-serving reasons are the next most important in explaining advisors' feedback decisions (ranks 5 and 6).

In summary, advisors' decisions to obscure positive feedback are best explained by heterogeneity in decision-specific preferences, reflected in the relatively lower importance some advisors attach to transparency or a sense of duty, rather than direct altruism toward their advisees. While our theory assumes that advisors are altruistic toward advisees, the evidence from our experiments does not support this (with the exception of advisors who send feedback because they perceive the feedback to be useful for the advisee). Instead, our findings point to advisors' motives for sending/obscuring feedback as stemming primarily from their personal preference and, to a lesser extent, self-interest.

6.2 Why does the difference in obscuring positive and negative feedback appear for vague but not precise signals?

Another finding in our main experiment is that the tendency for advisors to hide negative feedback more than positive feedback arises only when they receive vague, but not precise signals of performance (Result 1d). Our follow-up experiment allows us to explore possible mechanisms driving this result. First, as noted in Appendix Figure B.6, we replicate our main finding that, with non-instrumental feedback, advisors are more likely to obscure negative vague feedback than positive vague feedback from advisees, but they are no different in their propensity to obscure positive precise and negative precise feedback. Second, we find this result to be robust to the exclusion of participants who fail the comprehension check question (Appendix Figure B.7), suggesting that confusion on the advisors' part cannot explain this result.[24]

Using data from our follow-up experiment, we investigate whether advisors state different motives for their feedback decisions when receiving vague versus precise signals of their advisee's performance. Specifically, as vague signals of performance come with more noise (or uncertainty), we conjecture that this may have provided advisors with some form of "wiggle room" to obscure the feedback that they send to their advisees. This may arise, for instance,

[24]One possibility is that when signals are precise, advisors have the option to provide partial feedback, which lowers the perceived cost of obscuring. This flexibility may lead more advisors to obscure feedback (be it partially or fully) when signals are precise, potentially masking the valence effect found for vague signals. In contrast, when signals are vague, advisors can only fully obscure feedback, as partial feedback is not an available option. Separately considering whether advisors fully obscure (no message) or partially obscure (vague message) feedback given precise signals, we find a higher propensity by advisors to fully obscure negative non-instrumental feedback, but this pattern does not replicate in the follow-up study (see Appendix Figure B.8). We therefore treat these results as suggestive rather than conclusive.

if advisors use the higher ambiguity (with vague signals) as an excuse to not be transparent with advisees when they receive signals of negative performance. To this end, the statement "*I was uncertain about my worker's ability based on the information I received*" in the post-experimental questionnaire aims to investigate this conjecture.

Appendix Table B.11 presents OLS regression estimates of advisors' feedback choice in the follow-up experiment, against whether they receive a negative signal or a vague signal, as well as whether they indicate uncertainty about their advisee's performance as a motive for their feedback choice. We consider the analysis separately based on whether advisors receive a precise signal (columns 1 and 2) or a vague signal (columns 3 and 4). Column (5) pools all data and contains interaction terms of the "ambiguity" motive with signal precision and signal valence.

The outcome variable of interest in the first four columns is the interaction term between "ambiguity in signal" and "negative signal" (in columns 2 and 4). This interaction term tells us whether, where faced with a negative signal of performance, advisors who indicate signal ambiguity as a motive for their feedback choice are more likely to obscure feedback, as opposed to the case where they receive a positive signal. We observe that, in both columns (2) and (4), the interaction term is positive, but not statistically significant (p-values = 0.450 and 0.282, respectively). Moreover, the triple interaction term in column (5) (Ambiguity in signal $\times$ Negative signal $\times$ Vague signal) is not statistically significant (p-value = 0.832), suggesting that the marginal impact of signal ambiguity (as a self-reported motive) on whether advisors choose to obscure negative signals is no different between the case of precise versus vague signals.[25]

As further exploratory analysis, we examine whether any of the other reasons provided in Appendix Figure B.5 can explain advisors' differential behavior between precise versus vague signals. Of the set of motives investigated, selecting the transparency-related item ("*I preferred to be transparent/honest*") is significantly correlated with the difference in advisors' propensity to obscure vague versus precise signals. As column (4) in Appendix Table B.12 shows, given vague signals, advisors who select transparency are less likely to obscure negative feedback than positive feedback (p-value = 0.008). However, this is not the case for precise signals. Column (2) reveals that advisors who select this reason are no more or less likely to obscure negative feedback as compared to positive feedback (p-value = 0.116). Column (5) of the table shows that the triple interaction is statistically significant (p-value = 0.003). This suggests that the difference in obscuring positive versus negative feedback under vague signals is systematically related to transparency-related motives, but not under precise signals.[26]

[25]Note that even when advisors receive a precise signal, there remains some residual uncertainty about the advisee's future performance since they will be matched with a new group of participants. This means that advisors may still feel uncertain about the advisee's relative performance in the new group they are matched with, which may explain why some advisors select this option even when they have received a precise signal of performance.

[26]We also find that concern about their advisee's motivation to improve ("*I considered how the message may*

Taken together, we observe that the tendency for advisors to obscure negative (versus positive) vague feedback is driven by those who do not indicate a preference for transparency as a motive for their feedback choice. We note an interesting parallel with the literature on moral costs of lying, where evidence suggests that uncertainty can facilitate self-serving lies (Dugar et al., 2019; Villeval, 2025). While we only have suggestive evidence to this effect, we posit that, for advisors with weaker preferences for transparency, the uncertainty inherent in vague signals may have lowered their psychological cost of lying (by omission), thus making the obscuring of feedback more common. This may be more apparent when advisors have motives to lie, e.g., due to the desire to protect the advisee's ego from negative signals of performance. Conversely, the moral costs of lying (by omission) may be higher for precise signals, thus leaving less room for selective obscuring of feedback. This may explain why we do not see a significant difference in the propensity to obscure positive versus negative precise feedback.[27]

Overall, our analysis reveals that: (i) we are able to replicate the main finding that advisors obscure negative feedback more than positive feedback only when feedback is vague but not when feedback is precise; (ii) this result cannot be explained by systematic differences in full versus partial obscuring of precise signals of performance; (iii) this result is not driven by a lack of comprehension or understanding by advisors with regards to the decision-making environment; (iv) this result is not driven by higher ambiguity in the signal structure of the vague signals per se; but (v) this result may potentially be driven by the interaction between preferences for transparency and a lower moral cost of lying (by omission) with vague signal structures.

affect my worker's motivation to improve") is correlated with the difference in advisors' propensity to obscure vague versus precise signals (see Appendix Table B.13). Advisors who indicate that they care about the advisee's motivation to improve are more likely to obscure negative signals than positive signals if the signals are precise (p-value = 0.024), but not if the signals are vague (p-value = 0.989). The triple interaction term is marginally statistically significant (p-value = 0.091). This suggests that when feedback signals are precise, advisors who are concerned about their advisee's motivation may obscure negative feedback to avoid discouraging the advisee. However, this consideration becomes less relevant when signals are vague, possibly because the inherent ambiguity of vague signals reduces the perceived impact on motivation.

[27] Precise signals may make it easier for advisors to partially withhold information by partially obscuring feedback. The availability of this intermediate option lowers the cost of withholding information, particularly for advisors who place relatively less weight on transparency, and can increase obscuring overall. This mechanism is expected to be largely valence-invariant, since partially obscuring negative feedback does little to mitigate harm to the advisee's ego. At the same time, increasing signal vagueness can lower the psychological cost of omission itself, making selective obscuring of negative feedback more common, particularly when advisors seek to protect the advisee's ego. These two effects may operate simultaneously as we move from precise to vague signals. This pattern can also explain why the aggregate rate of obscuring precise signals is greater than the rate of obscuring vague signals, as seen in Figure 2.

6.3 What drives advisors' feedback decisions to male and female advisees when feedback is non-instrumental?

In this section, we further examine factors that may affect differential behavior by advisors toward male and female advisees when feedback is non-instrumental.[28]

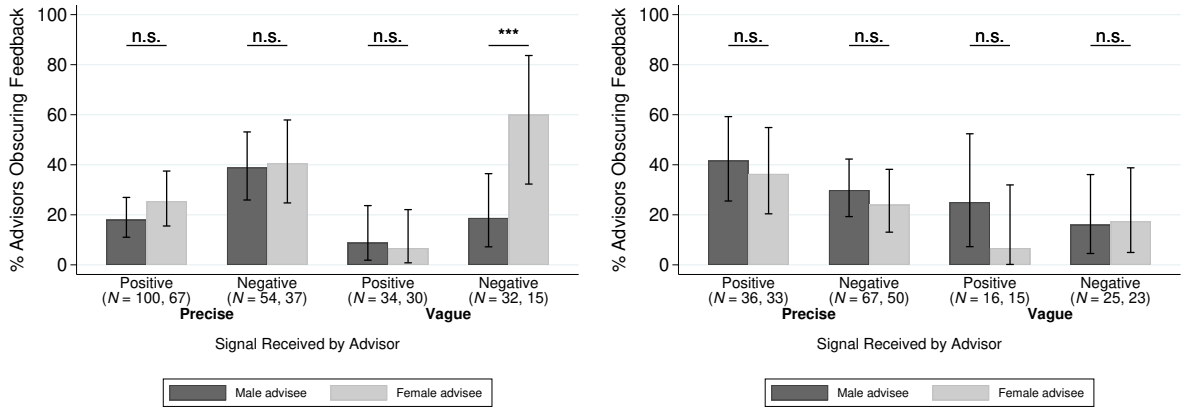
Incomprehension by advisors. We first investigate whether advisors' tendency to obscure negative signals more from women than men can be explained by a lack of comprehension on the advisors' part. Appendix Figure B.10 reproduces panel (a) of Figure 3 by excluding advisors who fail the comprehension check question on instrumentality of feedback in Part 2 of the experiment (panel a, 19.9% of advisors). We find qualitatively similar results, although we note the insignificant gender difference due to the loss of statistical power where a larger proportion of advisors are dropped.

Advisors' beliefs about advisees. We next investigate the role that advisors' beliefs play in driving their feedback decisions. All else equal, our theoretical framework suggests that advisors would be less inclined to obscure negative feedback for overconfident advisees, as they are most likely to mistakenly enter the tournament. When we reproduce the analysis in Table 2 by controlling for advisors' second-order beliefs (reported in columns 1 and 4 of Appendix Table B.14), we note that variations in second-order beliefs cannot fully explain the gender difference in obscuring non-instrumental feedback.[29]

We next split the analysis based on whether advisors have above- or below-median second-order beliefs. Advisors who hold above-median (below-median) second-order beliefs are defined as those who think that their advisees have a higher or equal confidence (lower confidence) about their performance than the median advisor. Figure 4 presents advisors' feedback decisions when feedback is non-instrumental, separately for those who hold above-median second-order beliefs (panel a) and those who hold below-median second-order beliefs (panel b). Panel (a) reveals that advisors who hold above-median second-order beliefs about their advisee's performance are more likely to obscure a vague negative signal of performance from female advisees than from male advisees (Fisher's exact test: p-value = 0.008). On the other hand, those who hold below-median second-order beliefs do not exhibit any gender difference

[28] We also investigate whether the gender result holds in our follow-up experiment (Appendix Figure B.9). The gender difference observed in our follow-up experiment is of smaller magnitude and does not reach statistical significance. The difference is marginally significant when we remove advisors who fail the comprehension check question from the analysis (panel b). Since the gender difference is only observed for these more attentive participants, this can be noted as a qualification of Result 2(a).

[29] A caveat to this analysis is that the gender difference in second-order beliefs may potentially be endogenous to our manipulation on the instrumentality of feedback. In fact, we do find this to be the case. Overall, we find that advisors believe that male advisees are more confident about their performance than female advisees (Wilcoxon rank-sum test: p-value = 0.006), but this gender difference is only present in the instrumental feedback treatment and not the non-instrumental feedback treatment (Wilcoxon rank-sum test: p-values = 0.008 and 0.318, respectively).



(a) $\geq$ Median Confidence

(b) $<$ Median Confidence

Note: Error bars represent 95% confidence intervals. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, n.s. $p > 0.10$. "$\geq$ Median" ("$<$ Median") implies that advisors believe that their advisees have a higher or equal confidence (lower confidence) about their performance than the median advisor.

Figure 4: Proportion of advisors sending feedback that is vaguer than received signal when feedback is non-instrumental (by advisor's second-order belief) [Main Experiment]

in the propensity to obscure vague negative signals of performance (Fisher's exact test: p-value = 1.000).[30]

Hence, when feedback is non-instrumental, advisors obscure vague negative signals from female advisees they perceive as more confident than average. On the other hand, confidence appears to play no role in feedback decisions for male advisees. Although not directly predicted by our theoretical framework, this finding suggests that advisors may be assigning a higher weight to the ego utility of female advisees. This is because the ego utility of the most confident advisees will be most impacted by negative feedback. When feedback is non-instrumental, advisors appear to behave as if they care more about the ego utility of confident female advisees, and therefore are more likely to obscure signals of negative performance from them.[31]

We further draw on our follow-up experiment to investigate whether advisors' concerns for their advisee's ego can explain whether they obscure negative feedback differently from female and male advisees. Appendix Table B.15 presents OLS regression estimates for advisors' feed-

[30] Our conclusions in Figure 4 are supported by regression analysis reported in Appendix Table B.14, where we reproduce the analysis in Table 2 separately for advisors who hold above-median beliefs (columns 2 and 3) and below-median beliefs (columns 5 and 6).

[31] We also investigate whether advisors' feedback decisions are influenced by their beliefs about the advisee's tournament-entry decisions. Panel (a) of Appendix Figure B.11 reveals that advisors do not believe that, prior to receiving any feedback, male advisees are significantly more likely than female advisees to choose to enter the tournament (Fisher's exact test: p-value = 0.207). This is the case for both the non-instrumental and instrumental feedback treatments (Fisher's exact tests: p-values = 0.374 and 0.408, respectively). Hence, it is not the case that advisors expect male advisees to be more likely to enter tournament than female advisees. The lack of a gender difference in advisors' beliefs about advisees' competition choice may explain why we do not observe advisors obscuring feedback from (overconfident) male advisees as a way to deter them from entering the tournament with instrumental feedback.

back choice separately for female advisees (columns 1-3) and male advisees (columns 4-6), where we additionally control for whether advisors indicate a concern for their advisee's ego as a motive for their feedback choice. We find that concern for the ego of female advisees leads to advisors being more likely, overall, to obscure feedback from them (column 1: p-value = 0.020). Importantly, the interaction of this motive with the signal valence in columns (2) and (3) reveals that advisors who are concerned about their female advisee's ego are less likely to obscure positive feedback (p-values = 0.027 and 0.046, respectively) and more likely to obscure negative feedback (p-values of interaction term < 0.001 in both columns).[32] Columns (4)-(6) reveal that advisors' concern for the ego of male advisees does not explain their feedback choice in any way.[33]

Taken together, our results from both the main and follow-up experiments provide suggestive evidence that advisors' concerns for the ego utility of their advisees have differential impact on the feedback that they give to female versus male advisees.

Advisors' implicit biases. Next, we investigate whether advisors' feedback decisions are influenced by their implicit biases (women/men being emotional/rational), as captured by the modified Implicit Association Test (IAT) (Greenwald et al., 1998).[34] Panel (b) of Appendix Figure B.11 presents the average IAT score of advisors elicited in the main experiment. A positive (negative) score indicates an implicit bias against women (men). The figure reveals that on average both male and female advisors exhibit an implicit bias against women (i.e., positive IAT score) and there is no statistically significant difference in the IAT score between male and female advisors (Wilcoxon rank-sum test: p-value = 0.958).

Appendix Table B.16 reproduces the analysis in Table 2 controlling for advisors' IAT scores (columns 1 and 4). Overall, we find that advisors' IAT scores do not have any explanatory power. When we split the analysis based on whether advisors' IAT scores are above or below median (columns 2-3 and 5-6), we also do not observe any evidence that either subgroups exhibit a different behaviour from the one reported in Result 2. Hence, we do not find that advisors' implicit biases (as captured by their IAT scores) provide any meaningful explanation

[32]The overall positive effect of concern about advisee's ego on advisors' propensity to obscure negative signals from female advisees is statistically significant in both columns. That is, test of "Concern for advisee's ego + Concern for advisee's ego $\times$ Negative signal" = 0: p-values < 0.001 in both columns.

[33]We also further examine if the envy motive varies with gender (Appendix Figure B.12). We find that male advisors are overall more likely than female advisors to indicate comparison with their advisee as a motive for their feedback choice (Fisher's exact tests: p-values = 0.003 and 0.051 for male and female advisees, respectively). However, for both male and female advisors, there is no difference with respect to the advisee's gender in their propensity to indicate comparison with their advisee as a motive (Fisher's exact tests: p-values = 1.000 and 0.368, respectively). Overall, we do not see a difference in the envy motive, based on the advisee's gender, that can explain our result.

[34]The IAT we use captures the rate at which advisors are able to assign names (i.e., Ben, Paul, Daniel, John, Jeffrey, Rebecca, Michelle, Emily, Julia, Anna) to either "male" or "female", as well as different adjectives (i.e., analytical, logical, objective, reasonable, scientific, perceptive, impulsive, empathetic, delicate, sensitive) to either "rational" or "emotional".

of their behavior.

Beliefs about the behavior of other advisors. Appendix Figure B.13 presents advisors' beliefs about the feedback decisions of other advisors. Specifically, they were asked to indicate whether they expect gender differences in feedback decisions when advisors observe a negative precise signal of "Rank 4" (panel a) or a positive precise signal of "Rank 1" (panel b), separately for non-instrumental and instrumental feedback. Interestingly, the figures reveal that advisors anticipate other advisors to be more likely to obscure both negative and positive feedback from women than from men (z-test: p-values < 0.001 for both negative and positive signals). However, there are no differences in the distribution of expectations between the non-instrumental and instrumental treatments (Kolmogorov-Smirnov test: p-values = 0.998 and 0.860, respectively, for both a negative and a positive signal). Finally, we find that Result 2 remains robust to the inclusion of advisors' beliefs about the feedback decisions of other advisors as controls (Appendix Table B.17). Hence we do not find that beliefs about descriptive norms provide any meaningful explanation of advisors' behaviour.

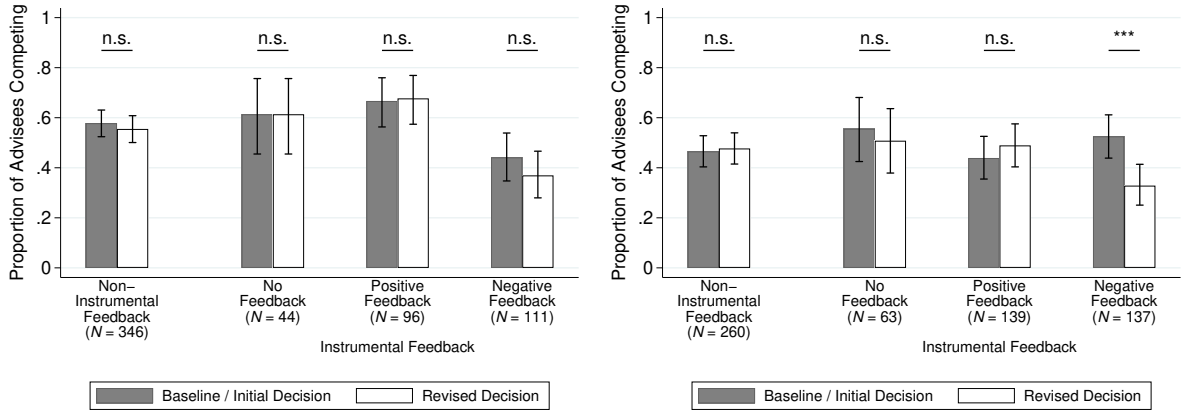
7 Advisees' Tournament-Entry Decisions

How does feedback influence advisees' willingness to compete? Recall that all advisees first make a tournament-entry decision prior to receiving any information about the advisor's task, thus providing us with their baseline competitive preferences. Then, advisees make a second (revised) tournament-entry decision, either before they receive the advisor's feedback (non-instrumental treatment), or after receiving the advisor's feedback (instrumental treatment). Hence, the advisee's revised decision is conditioned either on having no new information at all (in the non-instrumental treatment) or on the specific feedback provided by the advisor (in the instrumental treatment). Figure 5 presents advisees' initial and revised tournament-entry decisions, separately for male advisees (panel a) and female advisees (panel b).

Before examining the impact of feedback through our two treatments, we briefly summarize baseline preferences. Overall, 53% of advisees initially choose to enter the tournament, which is substantially higher than the expected 25%, but comparable to the 54% found in Niederle and Vesterlund (2007). Splitting the sample by gender, we find that 57% of male advisees enter the tournament, compared to 48% of female advisees, a significant gap of 9 percentage points (Fisher's exact test: p-value = 0.003).[35] Given the substantial baseline over-entry by both genders, we can note that there may be limited scope for positive feedback to alter decisions.

We now focus on the case where feedback is non-instrumental. By design, advisors' feedback in the non-instrumental treatment should not influence advisees' tournament-entry decisions. That is, since advisees do not receive any new information between their two tournament-

[35] While substantially less than the gap found in Niederle and Vesterlund (2007), this result is similar to the average gap of 13 percentage points, as found in a recent meta-analysis by Markowsky and Beblo (2022).



(a) Male Advisees

(b) Female Advisees

Note: Error bars represent 95% confidence intervals. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, n.s. $p > 0.10$.

Figure 5: Proportion of advisees choosing to enter tournament

entry decisions, there should be no difference in the rate of revisions across feedback conditions within the non-instrumental treatment.[36] Indeed, we do not observe statistically significant differences in the proportion of advisees revising their competition choices across the different types of feedback received from advisors (Fisher's exact test: p-value = 0.595).

When feedback is instrumental *and* advisees receive no feedback (rank between 1 and 4) from their advisors, we observe that there are no statistically significant differences between advisees' initial and revised decisions (Wilcoxon signed-rank tests: p-values = 1.000 and 0.508 for male and female advisees, respectively). In other words, when advisees receive no feedback from their advisors even when advisors may have the opportunity to do so, they do not revise their tournament-entry decisions.[37]

We next consider advisees' decisions based on the valence of the feedback received. We consider the case where they receive feedback that is positive (pooling both precise and vague), or feedback that is negative. Figure 5 reveals that when advisees receive negative feedback, female advisees revise their decisions by opting out of the tournament (Wilcoxon signed-rank test: p-value < 0.001), but male advisees do not revise their tournament-entry decisions (Wilcoxon signed-rank test: p-value = 0.115). On the other hand, neither male nor female advisees respond to positive feedback (Wilcoxon signed-rank tests: p-values = 1.000 and 0.311 for male and female advisees, respectively). As previously noted, this lack of response may be related to aggregate over-entry.[38]

[36] It may be possible that forcing advisees to make another decision may affect their willingness to compete (e.g., Fu and Zhong, 2025). Nonetheless, we would not expect any difference in the propensity to revise their decisions in either direction.

[37] Non-revision ends up being optimal in the study, given our earlier findings that advisors obscure both positive and negative signals in the instrumental treatment. This is in contrast to our theoretical framework, which predicts that no feedback should be interpreted as a negative signal given the (predicted) behavior of advisors.

[38] An alternative classification is to pool advisees who receive precise feedback that they are ranked 2, 3, or

Table 3: OLS regression results of advisees' tournament entry decisions

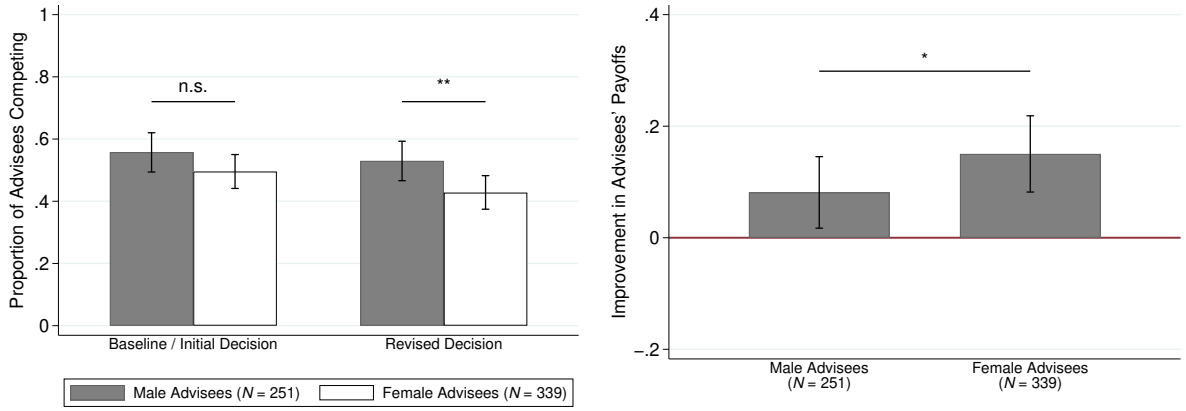
	(1)	(2)	(3)
Dependent Variable: Advisee revises decision toward tournament entry			
Positive instrumental feedback	0.043	0.034	0.037
	(0.033)	(0.048)	(0.049)
Negative instrumental feedback	−0.132***	−0.049	−0.055
	(0.030)	(0.042)	(0.044)
No instrumental feedback	−0.019	0.023	0.026
	(0.036)	(0.048)	(0.048)
Female advisee	−0.005	0.035	0.039
	(0.022)	(0.024)	(0.024)
Female advisee × Positive instrumental feedback		0.005	0.007
		(0.067)	(0.067)
Female advisee × Negative instrumental feedback		−0.160***	−0.156***
		(0.061)	(0.060)
Female advisee × No instrumental feedback		−0.082	−0.084
		(0.070)	(0.070)
Constant	−0.006	−0.023	0.144
	(0.015)	(0.015)	(0.108)
Controls	N	N	Y
Observations	1,196	1,196	1,196
R^2	0.025	0.033	0.038

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Robust standard errors in parentheses.
Other controls include the advisee's age, education level, posterior beliefs about their performance, and absolute score in the task.

Table 3 presents estimates from OLS regressions of changes in advisees' tournament-entry decisions. The dependent variable takes the value of 0 if the advisee does not change their decision, +1 if the advisee initially chooses piece rate but later revises this to tournament entry, and -1 if the advisee initially chooses tournament entry but later revises this to piece rate. The baseline comparison group is advisees in the non-instrumental treatment. Individual dummy variables control for whether advisees in the instrumental treatment receive positive feedback, negative feedback, or no feedback.

The estimates in the table confirm our conclusions from Figure 5. Column (1) reveals that advisees on average do not change their tournament-entry decisions when they receive

4 as having received "negative" feedback, since these advisees should be opting out of tournament given the information they receive. Results with this alternative classification are reported in Appendix Figure B.14 and Appendix Table B.18. Here, we find that there is a significant response by both male and female advisees to positive feedback in that advisees are more likely to opt into the tournament given positive feedback (Wilcoxon signed-rank tests: p-values = 0.002 and < 0.001 for male and female advisees, respectively). However, male and female advisees do not differ in their response to positive feedback (p-values = 0.501 and 0.481 in columns 2 and 3 of Appendix Table B.18, respectively). Additionally, while male advisees respond to negative feedback by being more likely to opt out of tournament (Wilcoxon signed-rank test: p-value = 0.002), female advisees are still more likely to respond to negative feedback than male advisees (p-values = 0.015 and 0.020 in columns 2 and 3 of Appendix Table B.18, respectively).



(a) Gender Competition Gap

(b) Change in Advisees' Earnings

Note: Error bars represent 95% confidence intervals. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, n.s. $p > 0.10$.

Figure 6: Advisees' decisions and earnings in the Instrumental treatment

no feedback (p-value = 0.595) or positive feedback (p-value = 0.198), but they opt out of the tournament after receiving negative feedback (p-value < 0.001). Columns (2) and (3) include interactions between the advisees' gender and the valence of feedback, which allow us to determine whether male and female advisees differ in their response to each type of feedback. The coefficient estimates affirm that female advisees are more responsive to negative feedback than male advisees by revising their decision away from tournament entry after feedback (p-values = 0.008 and 0.010, respectively). However, male and female advisees do not differ in their response to no feedback (p-values = 0.241 and 0.230, respectively) or to positive feedback (p-values = 0.937 and 0.915, respectively).

We summarize our findings as follows.

Result 3. *On average, advisees respond to feedback by adjusting their tournament-entry decisions accordingly in response to feedback. However, female advisees respond more to negative feedback than male advisees by opting out of tournament entry.*

What are the implications of Result 3? We first investigate the impact of advisors' feedback on the gender competition gap. By design, this refers only to the instrumental feedback treatment. Panel (a) of Figure 6 presents the gender competition gap before and after advisees receive advisors' feedback. The figure reveals that, prior to receiving feedback, female advisees are slightly less likely to enter into a tournament than male advisees, although this difference is not statistically significant (Fisher's exact test: p-value = 0.156). Importantly, the gender competition gap widens and becomes statistically significant after advisees receive their advisors's feedback (p-value = 0.016).

How does this translate to gender differences in advisees' earnings? Panel (b) of Figure 6 presents the difference in advisees' implied earnings between their revised and initial

tournament-entry decisions given instrumental feedback. A positive difference implies that the advisee's revised decision leads to higher earnings than their initial decision, while a negative difference implies that advisees' earnings have decreased as a result of revising their tournament-entry decision.

The figure reveals that, on average, there is an improvement in advisees' earnings arising from advisees revising their tournament-entry decisions after advisors' feedback (Wilcoxon signed-rank test: p-values = 0.001 and < 0.001 for male and female advisees, respectively). However, this positive improvement is greater for female advisees than for male advisees (Wilcoxon rank-sum test: p-value = 0.094). Hence, despite the widening gender competition gap, female advisees enjoy a greater improvement in their earnings compared to male advisees given their advisor's feedback.

Figure 6 highlights two key takeaways arising from gender differences in advisors' provision of instrumental feedback as well as advisees' response to negative feedback. First, the gender competition gap increases when advisors provide instrumental feedback. Second, while advisees' payoffs improve in general as a result of instrumental feedback, the increase in payoffs is greater for female advisees than it is for male advisees.

8 Discussion

Feedback provision is a fundamental process in many workplace and education settings. Yet, anecdotal and survey evidence suggests that performance information is often withheld in these contexts. Our findings confirm that about 25% of advisors partially or completely obscure performance information, even when feedback is of immediate instrumental use for decision making by the advisee. Contrary to our theoretical model, but consistent with survey evidence by Solomon (2016) and Zenger and Folkman (2017), positive feedback is obscured nearly as much as negative feedback. Likewise, although our framework predicts that advisors' altruism toward advisees will shape feedback decisions, the motives reported by advisors in our follow-up experiment suggest otherwise. Most advisors' decisions to send versus obscure positive feedback are closely associated with a preference for transparency and a sense of duty to their advisees. This result also emerges in the analysis of free-form text responses in our main experiment. Moreover, given that the experiment is conducted in an anonymous, one-shot online setting, this may represent a lower bound on the extent to which a sense of duty can motivate advisors' behavior. Finally, for a sizable minority, their decision to obscure positive feedback is shaped by their perceptions about the usefulness of feedback, or by self- and social-image concerns, such as envy or concern about how the advisee will perceive them.

When feedback is not immediately instrumental for decision making, we find that (vague) negative signals of performance are significantly more likely to be obscured than positive signals of performance. We also find suggestive evidence that the obscuring of negative feedback depends on the advisee's gender. While advisors obscure vague negative feedback from male

advisees 17.5% of the time, they obscure it nearly twice as much (34.2%) from female advisees. Exploring our data in greater detail, we find that within the set of advisors matched to women, the tendency to obscure performance information is driven by those who believe that the female advisee they are matched with has above-median confidence. In our follow-up experiment, advisors' concern for advisees' ego utility influences the feedback that they give to female advisees, but not for male advisees. We conjecture that this suggests a rationale of ego-protection, whereby some advisors are "shielding" their female advisees from news that would lead to a substantial negative shock to their (ego) beliefs about their own ability. In line with this, analysis of advisors' stated motives show substantially more ego-related concerns by advisors for female advisees than for male advisees.

With recent evidence suggesting that women are no less eager to receive performance feedback than men (Coffman and Klinowski, 2025; Erkal et al., 2024), our findings imply that the differential treatment of female advisees is unwarranted.[39] While we are unable to fully identify intent, the patterns suggesting gender biases could be more pronounced among male advisors are in line with a form of paternalistic discrimination (Viki et al., 2003; Buchmann et al., 2023) or benevolent sexism (Glick and Fiske, 2001; Erkal et al., 2023). Overall, the differential treatment along gender lines raises important questions about its broader implications for female participation in competitive environments (such as leadership or career progression) or in educational attainment. The greater likelihood of hiding negative news from female advisees signifies an even larger welfare penalty, given we observed stronger responsiveness to feedback from female advisees.

Additionally, we find that men and women respond differently to negative feedback, with men having a tendency to discount such feedback and choosing to compete despite it being against their best interests to do so. This is consistent with findings in the literature which suggests that women compete too little and men compete too much (e.g., Niederle and Vesterlund, 2007). While most of the literature on gender and competition has focused on the suboptimal tournament-entry decisions by women and policies to mitigate such behavior (e.g. Niederle and Vesterlund, 2011), our results highlight an additional source of inefficiency arising from men's behavior. In particular, the findings suggest that men may be more resistant to behavioral adjustment following negative feedback. An important implication is that feedback provision and workplace training programs may need to account for gender differences in responses to negative feedback in order to mitigate suboptimal decision-making.

Taking stock of our findings, it is important to explore methods that can address biases that may exist in the feedback provision process. Although feedback is an instrumental tool for development in organizations and education, our findings reveal that a substantial proportion of

[39] Similarly, Castagnetti and Schmacker (2022) also do not find gender differences in demand for information about ability. An exception is Sharma and Castagnetti (2023), who find that women opt for less informative feedback in an explicitly male-stereotyped task.

feedback is obscured from recipients. This withholding appears to be driven mainly by advisors who place less weight on transparency and duty, allowing other motives, e.g., image concerns, to come into play. This suggests that, in addition to designing feedback provision processes that promote objective and transparent communication, institutions should cultivate strong norms about transparency and duty to their workforce. Framing precise feedback as a hallmark of effective mentorship may further align transparency with advisors' self-image and improve the overall quality of feedback provision.

In addition to the above recommendations, implementing standardized feedback mechanisms, such as structured feedback forms or rating scales, may further offer a way to reduce the role of personal biases and preferences in the feedback process. Additionally, the idea of gender-neutral feedback platforms, where the gender of the advisee is not disclosed, presents a potential area for investigation. Finally, the setting that we consider is one where advisors' financial incentives are not directly tied to their advisee's performance, which mirrors advisor-advisee relationships such as mentorship that we see in many organizations. Investigating advisors' behavior in a setting where their financial incentives are tied to the advisee's performance or behavior is left for future work.

References

Abeler, J., D. Nosenzo, C. Raymond (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

Alan, S., S. Ertac, E. Kubilay, G. Loranth (2020). Understanding gender differences in leadership. *The Economic Journal*, 130(626):263–289.

Alempaki, D., V. Burdea, D. Read (2025). Deceptive communication: Direct lies vs. ignorance, selective truth and silence. *Journal of Political Economy: Microeconomics*, online early.

Banerjee, R., N. D. Gupta, M.-C. Villeval (2020). Feedback spillovers across tasks, self-confidence and competitiveness. *Games and Economic Behavior*, 123:127–170.

Benistant, J., F. Galeotti, M.-C. Villeval (2022). Competition, information, and the erosion of morals. *Journal of Economic Behavior & Organization*, 204:148–163.

Berlin, N., M.-P. Dargnies (2016). Gender differences in reactions to feedback and willingness to compete. *Journal of Economic Behavior & Organization*, 130:320–336.

Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, 86(5):1549–1575.

Brandts, J., V. Groenert, C. Rott (2015). The impact of advice on women's and men's selection into competition. *Management Science*, 61(5):1018–1035.

Brandts, J., C. Rott (2021). Advice from women and men and selection into competition. *Journal of Economic Psychology*, 82:102333.

Buchmann, N., C. Meyer, C. D. Sullivan (2023). Paternalistic discrimination. *Working Paper*.

Castagnetti, A., R. Schmacker (2022). Protecting the ego: Motivated information selection and updating. *European Economic Review*, 142:104007.

Chaudhuri, A., S. Graziano, P. Maitra (2006). Social learning and norms in a public goods experiment with inter-generational advice. *The Review of Economic Studies*, 73(2):357–380.

Chaudhuri, A., A. Schotter, B. Sopher (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., M. Schonger, C. Wickens (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Chen, S., H. Schildberg-Hörisch (2019). Looking at the bright side: The motivational value of confidence. *European Economic Review*, 120:103302.

Coffman, K., D. Klinowski (2025). Gender and preferences for performance feedback. *Management Science*, 71(4):3497–3516.

Correll, S., C. Simard (2016). Vague feedback is holding women back. *Harvard Business Review*, 94(1):2–5.

Croson, R., U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.

Danz, D., L. Vesterlund, A. J. Wilson (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review*, 112(9):2851–2883.

Deversi, M., A. Ispano, P. Schwardmann (2021). Spin doctors: An experiment on vague disclosure. *European Economic Review*, 139:103872.

Ding, T., A. Schotter (2019). Learning and mechanism design: An experimental test of school matching mechanisms with intergenerational advice. *The Economic Journal*, 129(623):2779–2804.

Drobner, C., S. J. Goerg (2024). Motivated belief updating and rationalization of information. *Management Science*, 70(7):4583–4592.

Drouvelis, M., P. Paiardini (2021). Feedback quality and performance in organisations. *The Leadership Quarterly*:101534.

Dugar, S., A. Mitra, Q. Shahriar (2019). Deception: The role of uncertain consequences. *European Economic Review*, 114:1–18.

Eriksson, T., A. Poulsen, M.-C. Villeval (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6):679–688.

Erkal, N., L. Gangadharan, B. H. Koh (2020). Replication: Belief elicitation with quadratic and binarized scoring rules. *Journal of Economic Psychology*, 81:102315.

Erkal, N., L. Gangadharan, B. H. Koh (2023). Do women receive less blame than men? attribution of outcomes in a prosocial setting. *Journal of Economic Behavior & Organization*, 210:441–452.

Erkal, N., B. H. Koh, L. Xu (2024). Gender and demand for social comparison information. *Working Paper*.

Ertac, S., M. Gümren, L. Koçkesen (2019). Strategic feedback in teams: Theory and experimental evidence. *Journal of Economic Behavior & Organization*, 162:1–23.

Ertac, S., L. Koçkesen, D. Ozdemir (2016). The role of verifiability and privacy in the strategic provision of performance feedback: Theory and experimental evidence. *Games and Economic Behavior*, 100:24–45.

Exley, C. L., K. Nielsen (2024). The gender gap in confidence: Expected but not accounted for. *American Economic Review*, 114(3):851–885.

Fu, J., S. Zhong (2025). Visceral influences and gender difference in competitiveness. *Journal of Economic Psychology*, 107:102788.

Gallen, Y., M. Wasserman (2021). Informed choices: Gender gaps in career advice. *Working Paper*.

Gibbs, M. J. (1991). An economic approach to process in pay and performance appraisals. *manuscript, University of Chicago, Graduate School of Business*.

Gill, D., Z. Kissová, J. Lee, V. Prowse (2019). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science*, 65(2):494–507.

Gill, D., V. Prowse (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, 5(2):351–376.

Glick, P., S. T. Fiske (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, 56(2):109.

Gneezy, U., C. Gravert, S. Saccardo, F. Tausch (2017). A must lie situation–avoiding giving negative feedback. *Games and Economic Behavior*, 102:445–454.

Gneezy, U., J. Potters (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.

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

Greve, H. R., et al. (2003). *Organizational learning from performance feedback: A behavioral perspective on innovation and change*. Cambridge University Press.

Gürtler, O., C. Harbring (2010). Feedback in tournaments under commitment problems: Experimental evidence. *Journal of Economics & Management Strategy*, 19(3):771–810.

Hannan, R. L., R. Krishnan, A. H. Newman (2008). The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *Accounting Review*, 83(4):893–913.

Hattie, J., H. Timperley (2007). The power of feedback. *Review of Educational Research*, 77(1):81–112.

Ho, T.-H., C. Yeung (2014). Giving feedback to clients. *Management Science*, 60(8):1926–1944.

Hossain, T., R. Okui (2013). The binarized scoring rule. *Review of Economic Studies*, 80(3):984–1001.

Huang, L., Z. Murad (2020). Feedback spillover effect on competitiveness across unrelated tasks. *Behavioral Research in Accounting*, 32(1):69–85.

Jampol, L., V. Zayas (2021). Gendered white lies: Women are given inflated performance feedback compared with men. *Personality and Social Psychology Bulletin*, 47(1):57–69.

Jin, G. Z., M. Luca, D. Martin (2021). Is no news (perceived as) bad news? an experimental investigation of information disclosure. *American Economic Journal: Microeconomics*, 13(2):141–73.

Kessel, D., J. Mollerstrom, R. van Veldhuizen (2021). Can simple advice eliminate the gender gap in willingness to compete? *European Economic Review*:103777.

London, M. (2014). *The power of feedback: Giving, seeking, and using feedback for performance improvement*. Routledge.

Lovász, A., E. Cukrowska-Torzewska, M. Rigó, Ágnes Szabó-Morvai (2022). Gender differences in the effect of subjective feedback in an online game. *Journal of Behavioral and Experimental Economics*, 98:101854.

Markowsky, E., M. Beblo (2022). When do we observe a gender gap in competition entry? a meta-analysis of the experimental literature. *Journal of Economic Behavior & Organization*, 198:139–163.

Niederle, M., L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Niederle, M., L. Vesterlund (2011). Gender and competition. *Annual Review of Economics*, 3(1):601–630.

Prendergast, C., R. Topel (1993). Discretion and bias in performance evaluation. *European Economic Review*, 37(2-3):355–365.

Prue, D. M., J. A. Fairbank (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management*, 3(1):1–16.

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

Schotter, A. (2003). Decision making with naive advice. *American Economic Review*, 93(2):196–201.

Schotter, A., B. Sopher (2007). Advice and behavior in intergenerational ultimatum games: An experimental approach. *Games and Economic Behavior*, 58(2):365–393.

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*:317–344.

Serra-Garcia, M., E. Van Damme, J. Potters (2011). Hiding an inconvenient truth: Lies and vagueness. *Games and Economic Behavior*, 73(1):244–261.

Sharma, K., A. Castagnetti (2023). Demand for information by gender: An experimental study. *Journal of Economic Behavior & Organization*, 207:172–202.

Silva-Goncalves, J., R. van Veldhuizen (2020). Subjective judgment and gender bias in advice: Evidence from the laboratory. *Working Paper*.

Solomon, L. (2016). Two-thirds of managers are uncomfortable communicating with employees. *Harvard Business Review*:2–5.

Sun, K.-K., S. Papadokonstantaki (2023). Lying aversion and vague communication: An experimental study. *European Economic Review*, 160:104611.

van Veldhuizen, R. (2022). Gender differences in tournament choices: Risk preferences, overconfidence, or competitiveness? *Journal of the European Economic Association*, 20(4):1595–1618.

Viki, G. T., D. Abrams, P. Hutchison (2003). The "true" romantic: Benevolent sexism and paternalistic chivalry. *Sex Roles*, 49:533–537.

Villeval, M.-C. (2020). Performance feedback and peer effects. In K. F. Zimmermann (ed.), *Handbook of Labor, Human Resources and Population Economics*. Cham: Springer International Publishing, pp. 1–38.

Villeval, M.-C. (2023). Cognitive and motivational effects. In T. Eriksson (ed.), *Elgar Encyclopedia of Labour Studies*. Edward Elgar Publishing, pp. 160–163.

Villeval, M.-C. (2025). Honesty and lie aversion. In S.-H. Chuah, R. Hoffmann, A. Neelim (eds.), *Elgar Encyclopedia of Behavioural and Experimental Economics*, chapter 70. Edward Elgar Publishing, pp. 237–238.

Weeks, J. W., R. G. Heimberg, T. L. Rodebaugh (2008). The fear of positive evaluation scale: Assessing a proposed cognitive component of social anxiety. *Journal of Anxiety Disorders*, 22(1):44–55.

Wozniak, D., W. T. Harbaugh, U. Mayr (2014). The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics*, 32(1):161–198.

Zenger, J., J. Folkman (2017). Why do so many managers avoid giving praise. *Harvard Business Review*, 2.