Flattering Advice: Avoiding Disappointment in Advice-Giving

Amanda Chen[1] & David Hagmann[1]

[1] The Hong Kong University of Science and Technology

17 December, 2024

Author Note

Correspondence concerning this article should be addressed to Amanda Chen, Department of Management, The Hong Kong University of Science and Technology, Hong Kong. E-mail: zchengj@connect.ust.hk

Abstract

Good advice improves decision quality but often requires delivering unpleasant truths that may disappoint advisees. Across three pre-registered and incentivized experiments involving real adviser-advisee interactions (n = 3,900), we demonstrate that advisers prioritize avoiding disappointment, even at the expense of accuracy. In Study 1, advisers financially rewarded for accuracy still tailored recommendations to aspirational goals expressed by advisees, resulting in poorer advice. When incentivized to be liked, advisers provided even more flattering advice, and advisees rated these advisers as more likable, despite the advice being less honest and less accurate (Study 2). In Study 3, advisers incorporated advisees' subjective and uninformative performance expectations into recommendations, leading to systematically different advice for men and women. Because men expected to perform better than women, they received more aspirational advice, even when their gender was unknown to advisers. Advisers' efforts to avoid disappointment may thus contribute to systematic gender disparities in advice, with implications for downstream decision-making and equity.

Keywords: Gender difference, Advice, Interpresonal relationship

Word count: 6,104

Flattering Advice: Avoiding Disappointment in Advice-Giving

Introduction

Advice plays a critical role in both professional and personal decision-making, often serving as a crucial tool for navigating complex challenges and improving outcomes (Gino & Moore, 2007; Harvey & Fischer, 1997; Soll & Larrick, 2009). While prior research has extensively examined the *recipients* of advice (e.g., how they evaluate or act on recommendations) far less attention has been devoted to understanding the motivations and behaviors of the *givers* of advice (Yaniv, 2004; Yaniv & Choshen-Hillel, 2012; with a few exceptions: Blunden et al., 2019; Schaerer et al., 2018). In this paper, we shift the focus to advisers, proposing that their recommendations are shaped not only by accuracy considerations but also by interpersonal concerns. Specifically, we propose that advisers factor in advisees' subjective expectations, providing flattering advice that aligns with these expectations, even when doing so leads to objectively worse recommendations. Notably, this tendency persists in our experimental settings, where advisees' expectations are clearly uninformative and advisers are financially incentivized for accuracy.

Advice is often viewed as a mechanism to improve decision satisfaction or prediction accuracy, assuming both parties prioritize solving the problem at hand. However, prior research highlights that interpersonal dynamics can complicate these interactions (Brooks et al., 2015; Minson et al., 2011). Advice seekers may avoid asking for help due to fear of rejection (Bénabou et al., 2022), and people generally have been shown to avoid information that could lead to disappointment (Gill & Prowse, 2012; Golman et al., 2017). They are motivated to avoid unpleasant truths in an ethical dilemma (Shalvi et al., 2019) and dislike those who deliver unwelcome feedback even when they had no role in the outcome (John et al., 2019). These tendencies align with models of belief utility (Loewenstein & Molnar, 2018), which suggest that individuals experience hedonic costs when receiving bad news. A natural reference point for evaluating the valence of advice is the advisee's expectation in the

absence of advice (Kőszegi & Rabin, 2006). Consequently, advisers who are aware of expectations may strategically adapt their advice to avoid delivering bad news, even at the expense of accuracy.

This interpersonal dimension of advice-giving has significant implications for understanding systematic gender differences in advice quality. Men are more likely than women to receive aspirational advice and encouragement to pursue competitive or risky opportunities (Kanze et al., 2018). Conversely, women are often shielded from criticism (Jampol et al., 2023), which, while intended to be kind, may ultimately limit their professional growth. We propose a novel mechanism to explain these patterns: advisers calibrate their recommendations to match advisees' expectations. Because, in many tasks (including those in our experiments) men are more confident and have higher expectations about their performance than do women, they are more likely to receive flattering (but, in our setting, worse) advice.

We present evidence for this mechanism from three preregistered and incentivized experiments involving real interactions between advisers and advisees (N = 3,900). In Study 1, we experimentally manipulate advisee's expectations and show that advisers account for those expectations, even when they offer no relevant insight for the decision. In Study 2, we show that advisers put more weight on interpersonal concerns when they are incentivized to be liked. Advisees reward this flattery by rating advisers as more likeable and equally trustworthy compared to those offering more honest (and accurate) feedback. In Study 3, we measure expected performance on a math quiz and show that advisers take these expectations into account even when they have information about actual performance. Because men expect to score higher, they are more likely to be advised to compete against a stronger group for a higher potential payment than are women, even when advisers are unaware of the advisees' gender. Notably, this leads to worse advice for men.

Open Science Statement

We report all manipulations, measures, and data exclusion in this and the following study. The preregistration reports, screenshots of all experimental materials, and the analysis code to replicate all statistical analyses and figures are available on the Open Science Framework (https://osf.io/8r3d4/?view_only=5ad7bafcd16b4d4ba08bb28b0e2bd02d).

Study 1

We first examine whether advisers take into account the expectations of advisees when providing a recommendation. A group of participants ("Advisees") complete a quiz consisting of ten questions that draw on ego-relevant domains. They then express a non-binding preference to compete against a group of high performers or low performers on the same task, based on their past performance. We nudge participants towards picking either the low or the high performer group by selecting one of the two by default. A second group of participants ("Advisers") observe Advisees' score on the quiz and, in one treatment, also which group they had selected. We predicted that Advisees who were nudged toward the high performer group would be more likely to receive advice to compete against this high group when Advisers observed their non-binding choice.

Methods

In a preliminary stage, we recruited 50 participants from Prolific and gave them five minutes to complete a quiz consisting of ten items. The quiz included word puzzles, identifying emotions from photos, and selecting the best responses for hypothetical scenarios. These questions were adapted from surveys that measure problem-solving, emotional intelligence, and communication skills—which we pre-tested as being important in modern society and hence where participants might have a stake in doing well. Participants received a bonus of five cents for each correctly answered question. We then ranked them based on their score and label the 20 participants with the lowest (highest) scores as the Low (High) Performer Group.

Next, we recruited 201 participants who completed the same 10-item quiz, also earning five cents for each correct answer. We then asked them to express a non-binding preference for whether they would like to compete against the High or Low Performer Group. We informed them that they would be invited back at a later date when they could make a binding decision and they could earn a bonus based on whether their score on this quiz was equal to or higher than a randomly selected member from their comparison group. If they picked the High Performer Group and had an equal or higher score than a randomly selected member of that group, they would earn a bonus of 50 cents. If they picked the Low Performer Group, they could earn a bonus of 20 cents. If their score was lower in this comparison, they would not earn an additional bonus. We randomized which of the two groups (High or Low Performers) was selected by default, and participants were free to select the other group. Participants also made a guess as to their score, and the survey concluded with basic demographic questions.

We then recruited 1,000 participants for the role of advisers, and the focal part of our experiment. They were informed of the ability quiz that participants had completed, how the Low Performer Group and High Performer Groups were constructed, as well as the choice and incentives for the other participants. We then asked participants to give advice to ten participants on which group they should compete against. Advisers were randomly assigned to one of two treatments. In the "Performance" treatment, they observed the score of the other participant. In the "Performance + Expectation" treatment, they additionally observed the advisee's non-binding choice for which group to compete against. However, since the outcome was based solely on the past quiz score and advisers, but not advisees, know the score, this initial choice was not informative. Advisers could earn the same bonus as one of the participants they had given advice to and who returned to make a decision. The survey concluded with basic demographic questions.

Finally, we invited advisees back for the follow-up survey. Following our

preregistration, we kept the survey open for 7 days. In total, 176 participants returned. The short survey reminded them of the task they completed in the previous survey, informed them that other participants from Prolific had observed their real score and given them advice against which group to compete against, and finally reminded them of how many questions they guessed they had answered correctly. Importantly, they were not informed of their true score or the score of the groups they could compete against. Participants then observed the advice from a randomly selected adviser and made their decision.

Results

We begin by examining the performance of the Stage 1 participants. Because our default treatment takes place after participants completed the ability quiz, we do not expect a difference in performance across the Low and High Default treatments. Indeed, the two groups scored no different from one another (5.10 and 5.18 for the Low Default and High Default treatments, respectively, $t(199) = 0.31$, $p = .756$). The default manipulation affected their initial choices. In the "Low Default" treatment, only 22% of participants initially preferred competing against the High Performer Group, compared to 65% of participants who saw the High Performer Group highlighted ($\chi^2(1, \text{n} = 201) = 36.62$, $\text{p} < .001$).

We now examine whether advisers took the advisees' expectations into account. Recall that advisees who were defaulted to competing against the High Performer Group were indeed more likely to choose this group, which, from the advisers' perspective, could signal high confidence in their abilities. Our theory predicts that showing expectations should lead to more flattering advice— specifically, a recommendation to compete against the High Performer Group— for advisees with a high default choice compared to those with a low default choice. Column 1 of Table 1 reports a linear probability model with advice to compete against the high group as the outcome measure, incorporating the default choice advisees saw, whether expectations were shown to the adviser, and the interaction of the two. Because each adviser made ten recommendations, we cluster standard errors at the adviser

Table 1

Advice to compete against the High Performer Group in Study 1. Displaying the non-binding choice of those for whom the Low Performer Group was selected by default makes it more likely that they are advised to compete against the High Performer Group (Column 1). Column 2 restricts the analysis to advisers who observed advisees' initial non-binding choice and controls for advisees' score on the quiz. Standard errors are clustered at the adviser level.

	(1)	(2)
High Default	0.011	0.025*
	(0.013)	(0.011)
Expectation Shown	−0.041*	
	(0.016)	
Expectation x High Default	0.037+	
	(0.019)	
Performance		0.168***
		(0.004)
Constant	0.389***	−0.503***
	(0.012)	(0.020)
N	10 000	5000

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

level. As predicted, we find a marginally significant interaction effect: advisees with a default choice to compete against the High Performer Group (nudged to have high expectations, though not necessarily holding them) were more likely to receive advice to compete against stronger competitors when their expectations were shown compared to when they were hidden. We reported a linear probability model on advising to compete against the High Performer group for advisers in the "Expectations" treatment, controlling for the true performance of the advisee in Column 2 of Table 1. We observe a significant main effect of the default treatment, where a high default increased the likelihood of receiving flattering advice. It is worth noting that, since only 65 % of advisees in the high default group actually stuck with their default choice, this represents a conservative test of the effect of high expectations on the likelihood of receiving flattering advice.

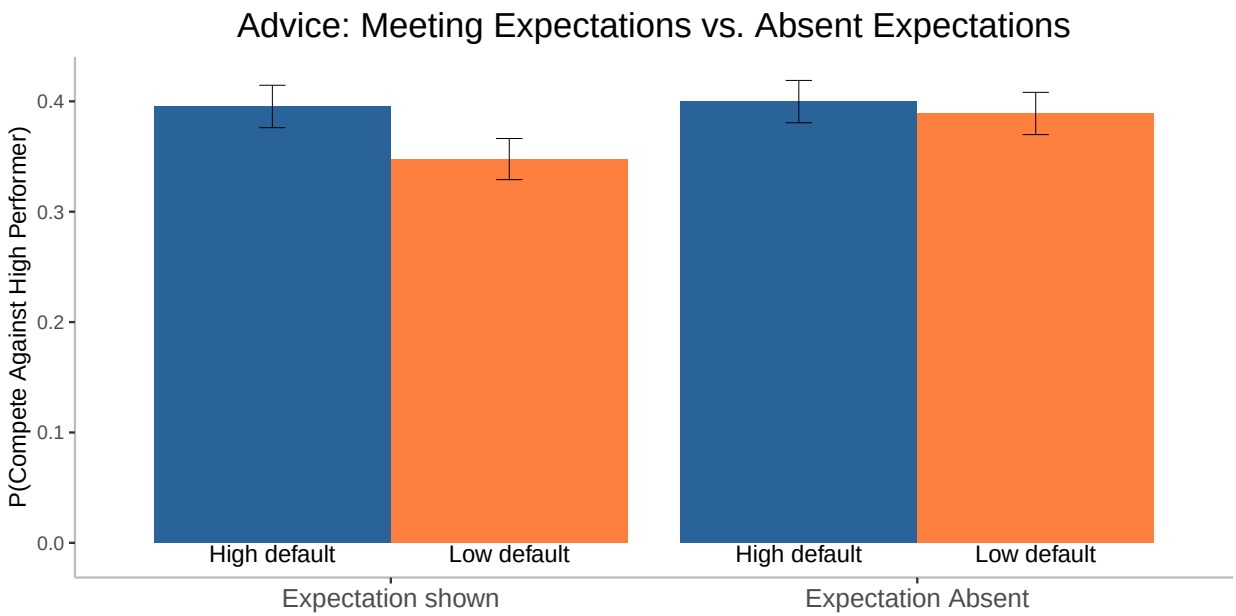

Figure 1

Advisees who were nudged to express high expectations were more likely to receive advice to compete against the High Performer Group.

Discussion

When the quality of a decision does not depend on someone's expectations, the adviser's awareness of the advisee's expectations should not influence the advice given. However, in line with our argument that advisers take into account the belief utility of the advisee and prefer to flatter rather than disappoint, we find that displaying the advisee's expectations does make a difference. Specifically, advisees who expressed a preference for stronger competitors were more likely to receive 'flattering' advice, encouraging them to pursue that option. One possible explanation for this finding is that, rather than trying to flatter, advisers may simply be telling advisees what they want to hear. In the next study, we aim to investigate the mechanism behind this discrepancy in advice-giving when expectations are revealed, exploring the question: Why do people give flattering advice?

Study 2

So far, we have shown that displaying expectations influences the advice given. Specifically, advisers were more likely to recommend competing against a group of high

performers to advisees with higher performance expectations compared to those without. We now directly test whether this effect is driven by interpersonal concerns. In particular, we explore whether more flattering advice stems from a desire to be liked, and whether advisees, in turn, like advisers more when they receive flattering advice.

We conducted a three-stage experiment in which we experimentally manipulated the incentives for advisers. We began by inviting a group of advisees to upload photos of themselves ("selfies") and informed them that they would be rated on their attractiveness. We grouped them with nine other participants of the same gender and recruited participants of the opposing gender to rank them from most to least attractive and to provide advice. Specifically, we asked them to advise the participant they ranked as the 7th most attractive (4th least attractive) on what rank they should bet they were ranked by a larger group of raters. Advisers were randomly assigned to two treatments, receiving a bonus payment either if the advisee guessed their rank accurately or if the advisee evaluated the adviser as likeable as measured by an unincentivized scale response. We hypothesize that advisers who want to be liked by advisees will recommend that they bet on a lower rank, i.e. that they are more attractive.

Methods

We recruited 300 participants from Prolific and, after asking demographic questions, invited them to upload photos of themselves (selfies) to be rated by other participants on attractiveness. 100 men and 107 women agreed to do so and uploaded pictures that adhered to our instructions (e.g., did not include other people). We selected the first 100 photos from women to arrive at a gender-balanced sample ($M_{\text{Age}} = 39.37$ years). Participants were informed that their selfies would be randomly grouped with those of nine other participants of their gender and ranked in terms of attractiveness by a group of new Prolific participants of the opposite gender. Lastly, they guessed their rank (unincentivized).

Next, we recruited a new, gender-balanced sample for the role of advisers. 472

participants from Prolific ($M_{\text{Age}} = 41.03$ years; 49.79% Female) started by providing demographic information, then were matched to a group of the opposite gender. They then ranked participants from most to least attractive using a drop-down menu next to each picture. Because of a limitation with the survey software, participants were able to select the same rank for multiple participants, rather than rank them uniquely. We remove 115 participants who did not follow instructions and provide a unique ranking.

After submitting their ratings, they saw the photo of the participant they had ranked as the 7th most attractive (i.e., the 4th least attractive). They were reminded of the rank they had given to that person and informed that this participant would be invited back and could earn a $1 bonus if they guessed their rank correctly. The rank was determined by the aggregate ratings of all participants who had ranked this group. Their task was to give advice to this participant about the rank they should bet on based on having observed all ten selfies and their own assessment. We randomly assigned participants to one of two incentivization schemes. In the "Accuracy" treatment, they received a bonus identical to the advisee: $1 if they guessed their rank correctly. In the "Likeability" treatment, we informed them that they would be rated by the advisee on a 5-point Likert scale for how likeable they thought they were. Each point on the scale would translate to a bonus of 20 cents. They then selected a rank that they would recommend the advisee to bet on.

Finally, we invited participants from Stage 1 and were ranked as the 7th attractive by at least one adviser (so that they received advice) back for the follow-up survey. Following our preregistration, we kept the survey open for 7 days. In total, 146 participants (77 men, 69 women) returned. They were reminded of the selfie they uploaded in Stage 1 and informed that a group of 10 selfies, including theirs, had been rated by other participants from Prolific. They then saw the advice from a randomly selected adviser and made their estimate with a $1 incentive for guessing accurately. They then saw the rank they had been advised to bet on one more time and were asked to rate the adviser's likability, warmth,

friendliness, good-naturedness, trustworthiness, and sincerity on 5-point Likert scales (adapted from Fiske et al. (2007)).

Results

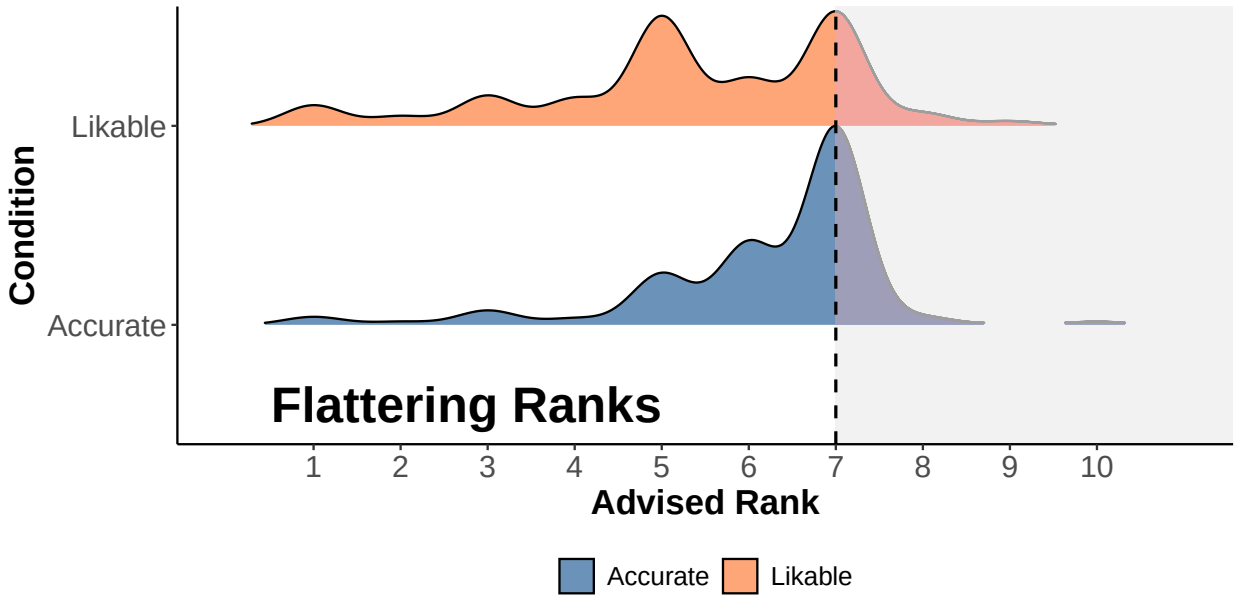

Figure 2

Advisers were asked to suggest what rank to bet on for someone they ranked as the 7th most attractive out of ten people. When incentivized for accuracy, most of them advised betting on 7, which was based on their evaluation of the recipient's attractiveness. However, when incentivized for likability, many participants gave flattering advice by suggesting betting on higher ranks that imply a higher evaluation of attractiveness.

We begin by examining the prior beliefs of advisees who uploaded their selfies. On average, men guessed they ranked 5.76 in their group of 10 and women guessed that they ranked 6.37. Participants overall underestimate their attractiveness relative to the benchmark average of 5.5. Moreover, women do so more than men $t(205) = 2.15$, $p = .033$. Notably, people's self-perceptions correlated strongly with the aggregate ratings of the advisers (0.43, $t(193) = 6.70$, $p < .001$). However, there was substantial heterogeneity in perceptions of attractiveness. Of the 200 participants, 149 were ranked as 7th most attractive by at least one adviser. On average, men in this subset estimated they were ranked 5.94th and women estimated they were ranked 6.59th ($t(143) = 1.94$, $p = .054$).

Next, we turn our attention to the advisers (Figure 2). In the Accuracy condition, those uploading selfies were advised to bet on rank 6.19. Notably, this is significantly more attractive than the 7th rank those advisers had themselves guessed just on the prior screen ($t(234) = -8.82$, $p < .001$). This suggests that even when incentivized for accuracy, participants offered flattering advice.[1] Importantly, and as predicted, we find that advisers in the Likeable treatment recommend betting on a lower rank, communicating that they think the participant in the selfie is more attractive (5.38, $t(470) = 5.44$, $p < .001$). This suggests that advisers inferred that advising someone that they were more attractive would make the adviser appear more likeable and therefore provided advice that communicated a more favorable impression of the participants' attractiveness. The distribution shown in Figure 2 shows that participants do not simply tell participants that they are the most attractive person in the group. They may infer that flattering advise needs to be somewhat realistic to be believable. We return to this in the general discussion.

Table 2

When individuals receive advice that implies a high level of attractiveness (lower rank), they tend to perceive the advice giver as more likable (Column 1), warm (Column 2). Column 3 shows that advisors are rated as more trustworthy when they advice lower ranks, but this relationship is only directional.

	(1)	(2)	(3)
Advised Rank	−0.109*	−0.133**	−0.054
	(0.046)	(0.044)	(0.042)
Constant	3.777***	3.903***	3.397***
	(0.279)	(0.266)	(0.257)
N	146	146	146

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Finally, we examine whether flattering advice indeed leads to more positive evaluations of advisers, or whether flattering advice is dismissed as insencere. Following our preregistration, we average the ratings on likeability, warmth, friendliness, and

[1] This shading could be due to concerns of avoiding disappointment. However, it could also be that advisers are uncertain about the rankings they had given and make a recommendation that combines their own belief with a uniform prior.

good-naturedness to create a scale of likeability ($\alpha = 0.92$); and we create a scale of trustworthiness by averaging the ratings of trustworthiness and sincerity ($\alpha = 0.87$).

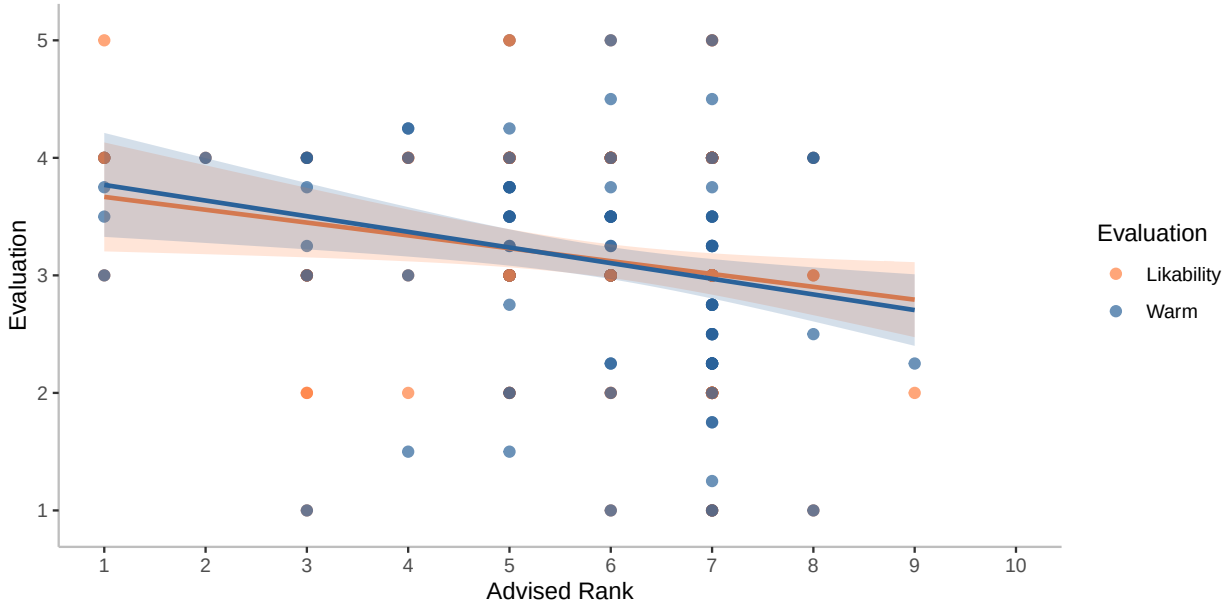

Figure 3

Participants tend to rate advisors who suggest betting on a higher rank (implying greater attractiveness) as more likable and warm.

As seen in Figure 3, we found that advisees who suggested that the advisee was more attractive (lower rank) were indeed rated as more likeable and warm (b = -0.109, p < .05; b = -0.133, p < .001, respectively; See Columns 1 and 2 of Table 2). When using their self-expectations as a reference point, advisors who suggest a rank lower than that point are perceived as less likable and warm (see Figure 4). Interestingly, these benefits are not at the cost of being hypocritical; advisers who recommend a more favorable rank are viewed as no less trustworthy (Column 3 of Table 2).

We were not powered to do a comparison across the two experimental groups and did not preregister such a difference. Indeed, we find no difference in likeability and warmth across the two treatments ($t(144) = 1.09$, $p = .275$, and $t(144) = 0.58$, $p = .562$, respectively). We also assess the quality of advice by measuring the discrepancy between the advised rank and the final rank. Our analysis revealed no significant difference between the two

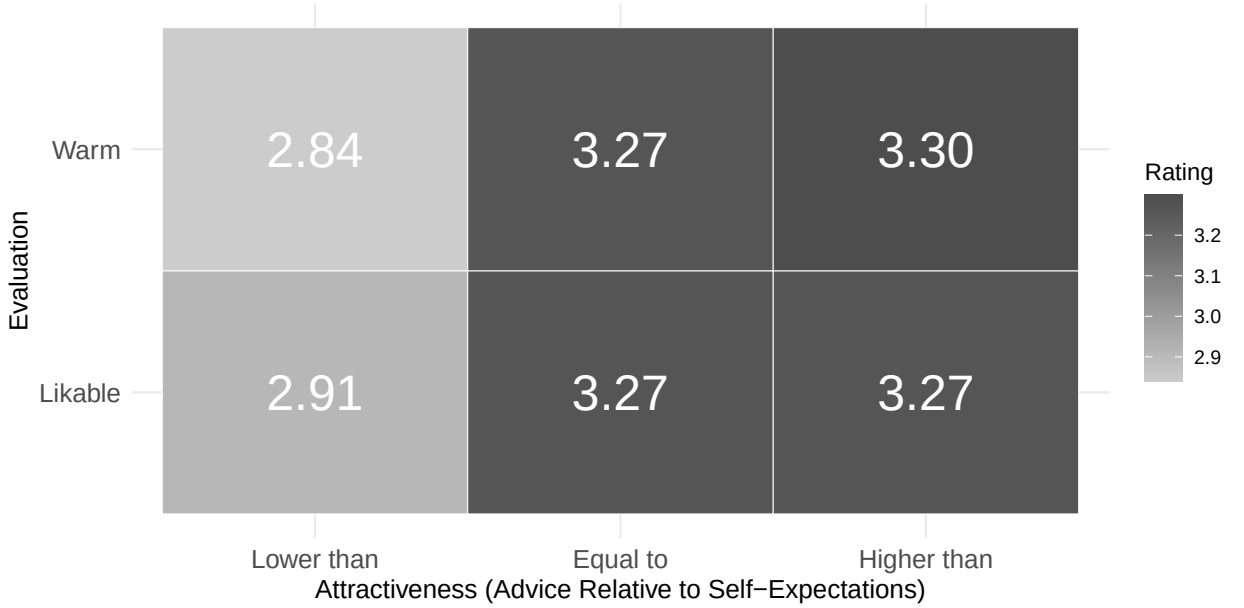

Figure 4

After participants received advice that suggested betting on a rank that was lower than, equal to, or higher than their self-expectations, they evaluated advisors based on their likability and warmth. Participants rated advisors who gave flattering advice and delivered a high evaluation of their attractiveness as more likable and warm than those whose advice implied a lower evaluation of their attractiveness.

treatments in terms of how helpful the advice was for advisees in making accurate guesses, $t(144) = -0.29$, $p = .776$.

Discussion

In another context where advice communicates ego-relevant information (here, people's attractiveness), we find that the advice people give is contingent on their incentives. Specifically, when they get rewarded for being more likeable, they recommend that the advisee bet on a more favorable rank than when they are incentivized for accuracy. Importantly, advisees do not discount flattering advice and instead evaluate people who advise them to bet on a more attractive rank as warmer and more likeable. These gains to interpersonal perceptions do not come at the cost of trustworthiness, even when the advice is substantially inflated (e.g., a recommendation that the advisee is the most attractive in the group).

Study 3

In Study 3, we replicate the main effect in a different ego-relevant domain: intelligence. The experimental design is conceptually identical to Study 1, but this time we utilize a more direct measure of expectation by asking participants to estimate their scores.

This three-stage experiment begins with measuring actual performance. Participants are then anchored to low or high performance expectations and asked to guess how many questions they answered correctly. As in previous studies, we recruit advisers who are presented with either just the advisee's true performance or both the true performance and the number of questions the advisee believes they answered correctly. Advisers are then asked to provide advice on choosing a competitor. Finally, we present the advice to the advisees and ask them to make their final choice. Our key hypotheses are: (1) advisees anchored to high expectations will be more likely to receive advice to compete against the high-performing group, and (2) men will be more likely than women to be advised to compete against the high-performing group when expectations are disclosed.

Methods

We begin by first recruiting a sample of 50 participants from Prolific to complete a 10-question multiple choice mathematics quiz. The questions were taken from a paper-version of the ASVAB standardized exam, such that answers were not available online. Participants had five minutes to answer the quiz and were paid 10 cents for each correctly answered question. Like before, we define top 20 scores as "High Performers" and the bottom 20 scorers as the "Low Performers." On average, participants answered 4.42 questions correctly, High Performers scored between 5 and 10, and Low Performers scored between 0 and 3. In order to anchor the expectations of participants in our main experiment, we simulated 1,000 pairings of groups of five participants, with the 5th percentile of groups scoring an average of 2.60 and the 95 th percentile scoring an average of 6.40. We report these averages to participants in the Low Expectations and High Expectations treatment, respectively.

We then recruited 1,002 participants for the role of advisees in Stage 1 of our main experiment. To arrive at a gender-balanced sample, we dropped the last two male participants to complete the survey, ending up with a sample of 500 men and 500 women ($M_{\text{Age}} = 42.16$). Participants completed the same 10-item mathematics quiz as the earlier participants and were informed that their performance would affect their bonus earnings in a follow-up stage to be conducted a few days later. After completing the quiz, we informed them of the average score of a group of five participants from the preliminary survey. We randomly assigned them to learn about the 5th percentile of groups, which scored 2.60. ("Low Expectations" treatment) or the 95th percentile of groups, which scored 6.40 ("High Expectations" treatment). Participants then made a guess (unincentivized) about how many questions they think they answered correctly. The survey concluded with basic demographic questions.

We then recruited 1,000 participants for Stage 2, placing them in the role of advisers. We began by informing them of the mathematics quiz that participants in the preliminary study and Stage 1 had completed, and informed them of the average score of all participants in the preliminary study. Advisers had to recommend whether an advisee should compete against the Low Performers or the High Performers (we used these terms in the survey). We anticipated that being told to compete against High Performers is more flattering and hence being told to compete against the Low Performers would be disappointing if one had expected to do well. Advisees could earn a bonus if their score was equal to or higher than that of a randomly selected member from their chosen group. Competing against the High Performers was more challenging, resulting in an average bonus of 50 cents, compared to 30 cents when competing against other groups.

Like before, advisers were randomly assigned to one of two treatments. In the "Baseline" treatment, they only observed the score of the advisee on the mathematics quiz. In the "Expectation" treatment, they observed the score as well as the advisee's guess for

how many questions they answered correctly. We want to emphasize that sice the outcome is determined only by the past score on the quiz, the advisee's guess is immaterial to which group they should compete against. Moreover, in neither treatment did they receive any demographic information about their advisee. Participants gave recommendations to 10 advisees, which unbeknownst to them were five men and five women matched to have identical performance on the test.[2] They were informed that if their advice was shown to a participant who returned for the follow-up survey, they would receive the identical bonus as that participant. The survey then concluded with basic demographic questions.

Finally, we invited participants from Stage 1 back for the follow-up survey. Following our preregistration, we kept the survey open for 7 days. In total, 951 participants (481 men, 470 women) returned. The brief survey reminded them of the task they completed in Stage 1, informed them that other participants from Prolific had observed their real score and given them advice against which group to compete against, and finally were reminded them of how many questions they guessed they had answered correctly. Importantly, they were not informed of their true score or the score of the groups they could compete against. Participants then observed the advice from a randomly selected adviser and made their decision.

Results

We begin by examining the performance of the Stage 1 participants. Because our treatment takes place after participants completed the mathematics quiz, we would not expect a difference in performance across the Low and High Expectations treatments. Indeed, the two groups scored no different from one another (4.79 and 4.55 for the Low Expectations and High Expectations treatments, respectively, $t(998) = -1.64$, $p = .102$). The expectations treatment did, however, affect how well they thought they performed. Participants in the "Low Expectations" treatment guessed a score of 4.79 vs. 4.42 in the

[2] We made this decision to account for the possibility of gender differences in performance.

High Expectations treatment ($t(998) = 3.78$, $p < .001$), showing that the manipulation was successful, albeit small. Contrary to our expectations, we did find a gender difference in performance: men scored 5.02 on average, while women scored 4.33 ($t(998) = -4.73$, $p < .001$). Consistent with this difference, men thought they answered more questions correctly than did women (4.75 vs 3.56, $t(998) = -8.60$, $p < .001$). This difference, however, does not affect the interpretation of our findings, which will rely on an interaction of gender with an experimental treatment for Stage 2 participants.

We define a measure of "overconfidence" as (Performance - Estimate). We observe that women underestimate their performance by 0.76 points, while men do so by only 0.27 points $\Delta M = -0.49$, 95% CI $[-0.73, -0.25]$, $t(997.87) = -4.03$, $p < .001$. Although neither gender is overconfident, men on average are more confident in their performance than are women, as we had expected.

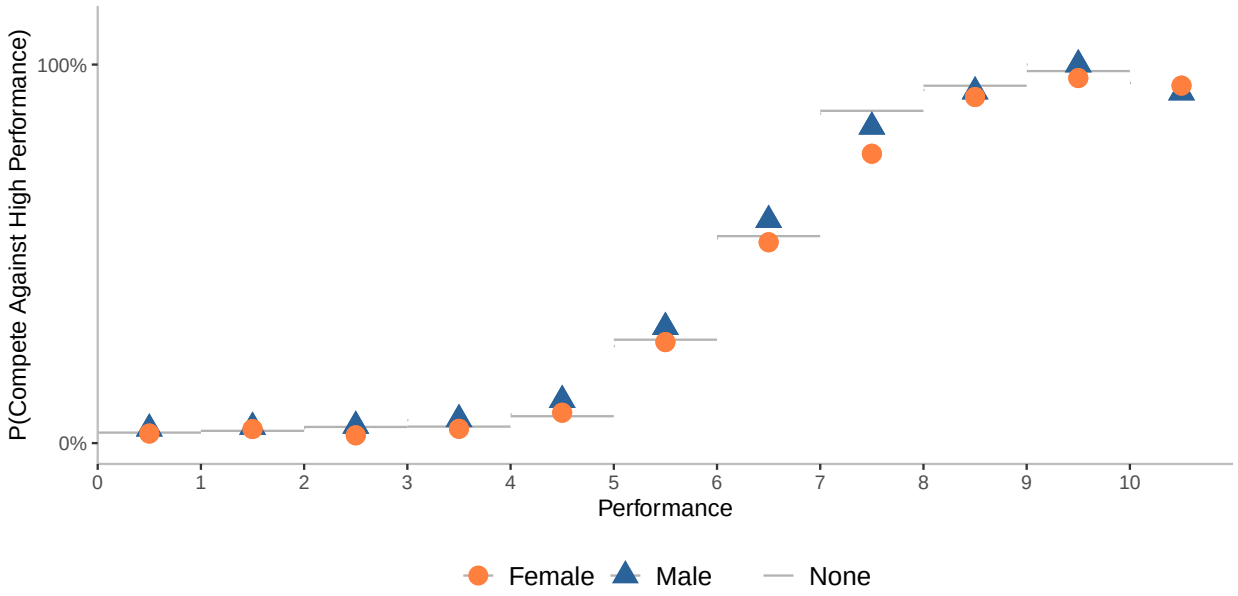

Figure 5

Advisers observed real performance and the participants' estimated performance, but not their gender. As a result of their higher expectations, men (blue) were advised to compete against the high performance group more often than women (red). The grey line shows advice absent expectations, which did not differ by gender.

Next, we examine whether advisers took the advisee's expectations into account.

Table 3

Column 1 displays the advice given to compete against a High Performance group based on whether the advisees were primed with low or high expectations and the expectation level shown to the advisor. Column 2 displays the advice given to compete against a High Performance group based on the gender of the targets and the expectation level shown to the advisor. Column 3 displays the expected bonus of the advice received based on the gender of the targets and the expectation level shown to the advisor. All standard errors clustered at the level of the advisor.

	(1)	(2)	(3)
High Expectation	0.012		
	(0.011)		
Performance	0.128***		
	(0.003)		
Expectation Shown		−0.034*	0.000
		(0.013)	(0.003)
Advisee Male		0.002	0.000
		(0.005)	(0.001)
Expectation x Male		0.032***	−0.005**
		(0.009)	(0.002)
Constant	−0.263***	0.339***	0.258***
	(0.015)	(0.010)	(0.002)
N	5000	10 000	10 000

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Column 1 of Table 3 reports a linear probability model on advising to compete against the High Performer group for advisers in the "Expectations" treatment, controlling for the true performance of the advisee. Because each adviser made ten recommendations, we cluster standard errors at the adviser level. However, contrary to our expectations, we do not see a significant effect of the expectations treatment. This may be because the induced difference in expectations was too small.

However, recall that men were more confident in their performance than were women. Our theory thus predicts that showing expectations should lead to more flattering advice (i.e., a recommendation to compete against the High Performer group) for men than for women. We report a linear probability model with advice to compete against the high group as the outcome measure, and the advisee's gender, whether expectations were shown to the

adviser, and the interaction of the two in Column 2 of Table 3. As predicted, we find a significant interaction effect: men are more likely to be advised to compete against the High Performers when expectations are shown than when they are hidden ($p < 0.001$). We show this result graphically in Figure 5).

To examine the quality of advice, we computed the expected bonus earnings for someone who followed the recommendations. For example, if an adviser suggested competing against the High Performer group, we matched the advisee against all 20 members of that group and determined how often their score matched or exceeded that of the member. We then multiplied this number by the respective bonus earnings (50 cents and 30 cents for the High and Low Performer group, respectively). To see if including expectations leads to worse advise for men, we report a linear probability model with the experimental treatment of the adviser, the gender of the advisee, and their interaction in Column 3 of Table 3. Displaying expectations led men to be advised to compete against the High Performer group more often, and this advise turned out to be bad: men receive worse advise than do women when expectations are displayed, but not in their absence.[3]

To determine whether flattering advice is truly costly, however, we need to examine the outcome of the advisees. In particular, they could ignore flattering advice, recognizing it as such and thus failing to adhere to it. Align with our prediction, participants who were primed with high expectations earned less when their expectations were conveyed to the advisers, although this result is only directional (see Column 1 of Table 4). Similarly, as shown in Column 2 of Table 4, we also found that male participants received less reward when their expectations were presented, even when their performance was the same. These findings suggest that flattering advice is not without consequences. Furthermore, our

[3] This analysis was not preregistered, and we note here that the reduction in expected earnings is small. However, it is interesting that expectations have a negative effect for men who underestimate their performance on average. One possibility is that advisers suggest the High Performer group more often than is optimal. We return to the possibility that advice is overall biased to be flattering in Study 2.

Table 4

Column 1 displays the actual bonus based on the performance and whether the participant is primed with high expectations. Column 2 displays the actual bonus based on the performance and gender of the participant. Columns 3 and 4 display the chance of adopting the advice based on the gender of the participants, whether the advisee sees expectations, and whether the advice is to compete against high performers. The former considers only the main effect, while the latter also includes the interactive effect.

	(1)	(2)	(3)	(4)
Performance	0.038***	0.039***		
	(0.003)	(0.003)		
High Expectation	−0.020			
	(0.013)			
Advisee Male		−0.002	−0.045+	−0.144***
		(0.013)	(0.026)	(0.041)
Expectation x Male				0.113*
				(0.053)
Expectation Shown			−0.002	−0.057
			(0.026)	(0.040)
Advice: High Performer			−0.138***	−0.207***
			(0.028)	(0.051)
Advice: High Performer x Male				0.131*
				(0.056)
Expectation x Advice: High Performer				−0.009
				(0.056)
Constant	0.087***	0.076***	0.857***	0.906***
	(0.016)	(0.015)	(0.024)	(0.030)
N	483	483	951	951

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

research indicates that because males are more likely to follow such advice (see Column 4 of Table 4, their tendency towards competition intensifies the cost of such advice.

Discussion

We found supportive evidence for our argument in a different domain. Advisors tend to consider advisees' self-expectations and match their advice accordingly, aiming to avoid disappointment. Consistent with the confidence gap narrative, men tend to underestimate their mathematics test scores less than women. When these expectations are displayed to advisors, they are more likely to advise men to compete against high performers. Notably,

this turns out to be poor advice: men whose advisors were aware of their expectations received worse advice. Our findings suggest that men with the same scores as their female counterparts ended up earning less, although this result is only directional. This discrepancy may be due to men receiving more favorable advice, being more likely to follow it, and ultimately facing worse outcomes.

General Discussion

Advice has the potential to shape people's career and personal outcomes. Honest feedback, however, may be painful to learn if it falls short of one's expectations. As prior work notes, this may motivate people to avoid information and avoid seeking help (Bénabou et al., 2022; Golman et al., 2017; Jaroszewicz et al., 2021). We present evidence from two experiments that advisers are also cognizant of this cost. As a result, they present flattering advice that avoids disappointing the recipient, and correctly anticipate that this boosts how advisees perceive them. However, this flattering advice comes at a cost to advisees, who would do worse if they followed it blindly, as we show in Study 1.

Moreover, a desire to avoid disappointment also means that advisers have to take into account the expectations of the advisee. We document that men are more optimistic about their performance than women are. As a result, they receive more flattering advice and are more likely to be told to aim higher. Notably, in the context of our experiment, this turns out to be bad advice ex post.

Our findings have implications for organizational practice, where mentoring and advice-giving may take into account an employees' expectation. We document this as a novel source of gender bias. Organizations could reduce this bias by calibrating employees' expectations to reduce overconfidence.

Participants in our experiment were paired anonymously. Even so, we document this desire to avoid disappointment. We anticipate that advice would be more flattering in

face-to-face communication and without anonymity. Moreover, existing relationships might make it even more difficult for advisers to be honest.

In our experiments, participants only received advice once and did not evaluate the adviser after observing the outcome of their decision. For example, advice that leads to bad outcomes may undermine the interpersonal benefits of flattery. Alternatively, people may still like the flattering advice and not fault the adviser for the bad outcome. Moreover, our setting involved only a single piece of advice on one task. Future research could examine whether people return to those who gave them flattering advice, or if they prefer someone who gave them the honest (but unpleasant) truth.

Advice has long been studied from the perspective of the receiver. Similarly, research on belief utility has examined how recipients respond to the valence of the information they receive. Here, we show that advisers, too, take into account the psychological impact of the information they convey. They may have even greater motivations to avoid conveying bad news, because they incur the interpersonal costs of delivering unfavorable information without reaping the benefits from helping someone make a better choice.

References

Bénabou, R., Jaroszewicz, A., & Loewenstein, G. (2022). *It Hurts to Ask*. National Bureau of Economic Research.

Blunden, H., Logg, J. M., Brooks, A. W., John, L. K., & Gino, F. (2019). Seeker beware: The interpersonal costs of ignoring advice. *Organizational Behavior and Human Decision Processes*, *150*, 83–100.

Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science*, *61*(6), 1421–1435.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83.

Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, *102*(1), 469–503.

Gino, F., & Moore, D. A. (2007). Effects of task difficulty on use of advice. *Journal of Behavioral Decision Making*, *20*(1), 21–35.

Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, *55*(1), 96–135. https://doi.org/10.1257/jel.20151245

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*(2), 117–133.

Jampol, L., Rattan, A., & Wolf, E. B. (2023). A bias toward kindness goals in performance feedback to women (vs. men). *Personality and Social Psychology Bulletin*, *49*(10), 1423–1438.

Jaroszewicz, A., Loewenstein, G., & Bénabou, R. (2021). *The pain of asking and being asked for informal help*.

John, L. K., Jeong, M., Gino, F., & Huang, L. (2019). The self-presentational consequences of upholding one's stance in spite of the evidence. *Organizational Behavior and Human Decision Processes*, *154*, 1–14.

Kanze, D., Huang, L., Conley, M. A., & Higgins, E. T. (2018). We ask men to win and women not to lose: Closing the gender gap in startup funding. *Academy of Management Journal*, *61*(2), 586–614.

Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, *121*(4), 1133–1165.

Loewenstein, G., & Molnar, A. (2018). The renaissance of belief-based utility in economics. *Nature Human Behaviour*, *2*(3), 166–167.

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to tango: Effects of collaboration and disagreement on dyadic judgment. *Personality and Social Psychology Bulletin*, *37*(10), 1325–1338.

Schaerer, M., Tost, L. P., Huang, L., Gino, F., & Larrick, R. (2018). Advice giving: A subtle pathway to power. *Personality and Social Psychology Bulletin*, *44*(5), 746–761.

Shalvi, S., Soraperra, I., Weele, J. J. van der, & Villeval, M. C. (2019). *Shooting the messenger? Supply and demand in markets for willful ignorance*.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 780.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93*(1), 1–13.

Yaniv, I., & Choshen-Hillel, S. (2012). When guessing what another person would say is better than giving your own opinion: Using perspective-taking to improve advice-taking. *Journal of Experimental Social Psychology*, *48*(5), 1022–1028.