Follow my Lead: Assertive Cheap Talk and the Gender Gap

Shanthi Manian
Washington State University

Ketki Sheth*
University of California, Merced

June 29, 2020

*We are grateful to Martin Abel, Jeff Butler, Erin Giffin, Rowena Gray, Jill McCluskey, Ted Miguel, Brigitte Roth Tran, Charlie Sprenger, Erin Troland, colleagues at UC Merced and Washington State University, and various seminar and conference participants for helpful comments and suggestions. We thank Ajay Barman and Jugal Marfatia for research assistance.

Abstract

A person's success often depends on whether others believe what they say. Growing evidence suggests that people are less likely to believe statements made by women rather than men. We consider whether assertive cheap talk, an important and widely used tool for increasing credibility, is a mechanism for this gender gap. If women face negative returns to assertive cheap talk, then they have less access to an effective tool for increasing their credibility. We provide evidence using a laboratory experiment and an online replication, both with real stakes, in an advice-following setting. We study whether assertive cheap talk affects advice following, whether subjects discriminate based on advisor gender, and whether there are differential returns to assertive cheap talk by gender. Subjects were randomly assigned to an unseen male or female team leader who were otherwise identical, and to different types of pre-scripted, increasingly assertive written communication from the leader. Assertive language significantly increased advice following, but we find no evidence for gender discrimination. We also find that assertive language had positive returns for both male and female leaders, despite subjects perceiving this language as more masculine. However, female subjects were still less likely to choose the self-promotional language. Thus, even in the absence of discrimination, this choice would reduce adherence to advice provided by women, generating a gender gap. Greater use of assertive language could be an effective strategy for women to increase their influence and credibility in the labor market.

1 Introduction

A person's success often depends on whether others believe what they say. A growing body of evidence supports the notion, widely asserted in popular discourse (Solnit, 2008), that people are less likely to believe statements made by women rather than by men. For example, Hoffmann and Tarzian (2001) review evidence that health providers are less likely to believe female patients' complaints; Mengel, Sauermann and Zölitz (2018) and Boring (2017) show that female instructors receive lower teaching evaluations; and Hengel (2019) shows that female economists face a higher bar in the academic publication process. We study whether this gap is due to gender discrimination. Then, we consider whether assertive cheap talk, an important and widely used tool for increasing one's credibility (Cooper and Kagel, 2016; Rudman and Glick, 2016; Charness, Rustichini and van de Ven, 2018), is a potential mechanism for this gender gap.

From formal presentations to casual discussions in team meetings, our society gives significant weight to an individual's own assertion of their value, even when that statement is costless to make and unverifiable (i.e., cheap talk) (Charness, Rustichini and van de Ven, 2018; Cooper and Kagel, 2016). Yet women receive conflicting messages about the use of assertive language. Many popular press articles advise women to stop using "weak" self-deprecating language, and instead use "stronger" assertive language, in order to get ahead (e.g., Libby (2016)). However, assertive and self-promoting behavior can be seen as masculine, and a large body of literature suggests that violating gender norms causes negative backlash (Rudman and Phelan, 2008; Phelan and Rudman, 2010; Williams and Tiedens, 2016).[1] That is, self-promotion through assertive cheap talk may actually reduce material payoffs for women. If so, then women may have less access to an effective tool for improving their credibility.

In this paper, we provide evidence about this cheap talk mechanism from a laboratory experiment at UC Merced and a replication administered through Amazon Mechanical Turk, an online platform. Both the laboratory experiment and the online replication offer real stakes in an advice-following setting. We study: 1) whether subjects are less likely to follow women's advice due to gender discrimination; 2) whether assertive cheap talk affects willingness to follow advice; and 3) whether

[1] Following Williams and Tiedens (2016), we refer to these negative returns as "backlash" throughout the paper. Williams and Tiedens (2016) use the term "backlash" to refer to any "negative outcome" in their meta-analysis of gender and backlash.

there are differential returns to assertive cheap talk by gender. We also explore whether women have different preferences for using assertive cheap talk, whether subjects expect gender discrimination by others in this advice-following setting, and whether such expectations can explain different preferences for assertive language.

In the lab experiment, university students were randomly assigned to a "team leader" who provided pre-scripted written advice on how best to play an incentivized game. The information presented to subjects about their team leaders was identical, except for their team leader's gender (male or female) and the assertiveness in the cheap talk that accompanied their team leader's advice. Specifically, though the substance of the advice was identical for all subjects, subjects were randomly assigned to different levels of assertiveness in the cheap talk used by their team leader: self-deprecating (*least assertive*), neutral (*moderately assertive*), or self-promoting (*most assertive*). We then replicated the experiment using Amazon Mechanical Turk. Results were remarkably similar across 1,011 subjects in the two experiments. Both the original experiment and the replication were preregistered prior to implementation.

We find no evidence for gender discrimination: on average, subjects were as likely to adhere to advice provided by a woman as that by a man. We can rule out a gender gap greater than 5.1 percentage points in adherence to advice. However, assertive cheap talk significantly increased adherence to advice, confirming that such cheap talk is an important tool for improving credibility and influence. The results suggest that assertive cheap talk is particularly effective when the quality of a statement is uncertain.[2]

We find no evidence for a negative response to women's use of assertive language. Among those assigned to the most assertive team leaders, team leader gender did not significantly affect adherence to advice. Subjects also rated assertive women and men similarly in a subjective evaluation, including on measures of both competence and likeability. Thus, assertive cheap talk increased adherence to advice for both male and female leaders.

Subjects were significantly more likely to characterize more assertive cheap talk as "more masculine." However, this gendered perception of more assertive cheap talk was malleable: the gender of

[2]These results are consistent with a large body of literature on advice taking, reviewed by Schotter (2003) and Bonaccio and Dalal (2006). Using similar laboratory environments, several studies have found that advisor confidence increases the likelihood that advice is followed. However, we are not aware of a study that focuses on differential responses to advisor confidence by gender.

the randomly assigned team leader affected perceptions of the language. Even though assertiveness was associated with masculinity, this norm was relaxed by exposure to violations.

Although more assertive language increased advice following for both male and female team leaders, female subjects were less likely to choose the most assertive cheap talk when asked which language type they would use if they were advising future subjects. Thus, even in the absence of discrimination, women's lower willingness to use assertive cheap talk would reduce adherence to advice provided by women, generating a gender gap. However, this reduced preference for assertive cheap talk cannot be explained by the language being less effective for women.

Although there was no gender discrimination in the experiment, subjects expected there to be. Even when incentivized to accurately report their beliefs, subjects overwhelmingly expected fewer of their peers to follow the advice of female leaders. However, there was little variation in the expected gender gap by language type; that is, subjects did not expect greater gender discrimination when the leader used assertive cheap talk.

Our paper contributes to a growing literature that suggests female expertise and advice are less heeded (e.g., Abel (2019), Ayalew, Manian and Sheth (2019), Bohren, Imas and Rosenberg (2019), Egan, Matvos and Seru (2017), Landsman (2018), Grossman et al. (2019), and Sarsons (2017)). We add to this literature by obtaining clean identification of the effect of gender discrimination using a real-stakes lab experiment. Research documenting gender gaps in natural settings identifies important differential responses to men versus women. Our paper builds on this literature by distinguishing whether the difference is driven by direct gender discrimination or by characteristics that differ by gender. This may explain why our results, which focus on isolating the effects of the former, differ from those that find a gender gap. For example, Grossman et al. (2019) find that subjects are less likely to adhere to women's advice when advisors are allowed to use their own words. Our results help identify whether such reduced adherence to advice is driven by discrimination against women *per se* or by average differences in how women provide advice (e.g., language choice). These results also suggest that providing more flexibility in language choice may result in a gender gap in adherence to leadership.

In addition, we contribute to a literature that aims to understand why women are less likely to assert themselves (Chakraborty and Serra, 2019; Exley and Kessler, 2019; Cooper and Kagel, 2016; Babcock et al., 2003; Moss-Racusin, Phelan and Rudman, 2010). We show that women exhibit a

preference against assertive language, even among those who do not expect backlash to assertive language, in a context where the payoff maximizing strategy is to use assertive language.

Finally, our paper adds to the literature on the implications of role congruity theory (Eagly and Karau, 2002). We use this theoretical foundation, in which women may face negative reactions for violating gender norms by behaving assertively (Rudman and Glick, 2016), to test implications for assertive language that is commonly used in the labor market.[3] Empirical evidence generally suggests that women may face backlash in terms of likeability when acting assertively, but are less likely to experience backlash on competency. This foundational work has given rise to a growing number of studies estimating the existence and magnitude of backlash on "downstream outcomes" such as hireability, salary, and other outcomes that require "complex judgments [that] likely combine both liking and competence evaluations, though perhaps not as equal inputs" (Williams and Tiedens, 2016). These downstream outcomes are of particular interest because they directly affect payoffs.

We address the open question of how commonly used self-promotional assertive language affects the downstream outcome of adherence to advice in a real-stakes environment. Many decisions, if not most, are made by individuals after they have consulted others. The ubiquity of advice in real-world decision-making makes the response to advice one of the most critical downstream outcomes to study. Advice is also an excellent example of the interplay between competence and likeability judgments. The decision to follow advice is frequently subjective, depending both on how much we view the advice-giver as competent and on how much we like the advice-giver. For example, Casciaro and Lobo (2008) show that likeability is more important than competence when deciding whom to ask for advice in an organization, though both aspects matter. We find no evidence that commonly used assertive language triggers backlash toward women in terms of likeability or the downstream outcome of following advice.

Similar to other studies on gender stereotypes and assertive behavior, we confirm that increased assertiveness in cheap talk is associated with masculinity (Williams and Tiedens, 2016). However, we do not find evidence that subjects are less likely to heed the advice provided by women who violate the norm by using such language. This is true for both incentivized advice following and non-incentivized subjective evaluation questions. This result is consistent with a number of studies that

[3]See Rudman and Phelan (2008) and Williams and Tiedens (2016) for reviews of gender role congruity theory and backlash.

also do not find evidence of backlash for dominant or assertive women, highlighting the importance of heterogeneous effects at the intersection of assertiveness and gender. In a recent meta-analysis of gender backlash, Williams and Tiedens (2016) conclude that implicit forms of assertiveness, in which the subject is not aware of the dominance behavior, may not result in backlash. It may be that self-promotional assertive language that is commonly used in the labor market is not explicit enough to trigger backlash, especially in contexts in which subjects face real consequences. Thus, our results yield support for the theory that moderately counter-stereotypical behaviors that do not trigger conscious awareness of gender norm violations will not result in backlash.

The rest of the paper proceeds as follows. In Section 2, we describe the experimental design. We then present the results of the experiment in Section 3. Section 4 discusses implications of the results and concludes.

2 Study Design

The entire experiment is conducted on a computer, where subjects complete an online self-guided "survey" hosted on Qualtrics.

We first describe the core game played in the experiment. This is followed by a description of the elements of the subject's experience, including the pairing with team leaders and the chronological components of the game that were identical for all subjects. We then describe the randomized experimental variations, and the two samples from which participants were recruited for the original and replication experiments.

2.1 The *Individual Game*

The primary task in the experiment is a signaling game adapted from Cooper and Kagel (2005). We refer to this game as the "*Individual Game*." Subjects played the Individual Game 10 times (i.e., 10 rounds).

The *Individual Game* is a two-player game. To explain the payoffs, we refer to the players as Player 1 (P1) and Player 2 (P2). The game proceeds as follows: Nature selects a type for P1, Type A or Type B, each with equal probability. P1 observes his type, then chooses a number between 1 and 4. P2 sees P1's selected number, but not P1's type, and plays *Left* or *Right*. P1's payoff

Player 1

Type A				Type B			
A's choice	Left	Right		B's choice	Left	Right	*Expected Payout (not shown)*
1	168	444		1	276	568	*299*
2	150	426		2	330	606	*393*
3	132	426		3	352	628	*465*
4	-188	-38		4	316	592	*573*

Player 2

Player 2's choice	Type A	Type B
Left	500	200
Right	250	250

Figure 1: *Individual Game*, Subject Payoff (expected payoffs not shown to subjects)

is based on his type, his chosen number, and P2's response. P2's payoff is based on her chosen response and P1's type. The payoff structure of the game is shown in Figure 1. This information, excluding the expected payout, was shown to all subjects.

This game is characterized by asymmetric information: P1 knows his type, but P2 does not, even though P1's type determines both players' payoffs. This provides an incentive for strategic play in the following way. First, notice that it is always preferable for P2 to select *Left* when playing against P1-Type A, and to select *Right* when playing against P1-Type B. Similarly, P1 is always better off if P2 selects *Right*, regardless of his type. This means that P1-Type B would like P2 to know that he is Type B, so that P2 will play *Right*.

There are two other key features to notice for P1's payoff structure. First, P1-Type B generally prefers higher numbers than P1-Type A. And second, selecting the number 4 is a dominated strategy for P1-Type A: he is always better off choosing another number, regardless of how P2 responds. However, this is not true for P1-Type B. These features allow P1-Type B to perfectly distinguish himself from P1-Type A by choosing 4. In other words, P1-Type B signals his type to P2 by selecting 4, because this is a dominated strategy for P1-Type A. Once P2 knows that P1 is Type

B, she will play *Right*. So, selecting 4 provides P1-Type B the highest expected payoff.

In our experiment, all subjects played as P1-Type B. P2 was played by a computer programmed to play as university students had in the original Cooper and Kagel (2005) experiment. For example, in the original experiment by Cooper and Kagel (2005), university students playing as P2 responded *Right* 41% of the time when P1 played 3. The computer in our experiment was therefore programmed to select *Right* with 41% probability when P1 plays 3. Subjects were informed that "Player 2 is played by a computer that mimics real life players" and that "the computer has been programmed to mimic how real life university students have played this game as Player 2."[4] They were also informed, accurately, that the computer only considers their choice in the current round when making its decision and does not keep a record of their previous choices from earlier rounds.

2.2 Team Leaders and Advice

The *Individual Game* is useful for studying responsiveness to advice because it poses a difficult question that has an unambiguous correct answer, but the answer itself is not obvious. Thus, there is a clear and important role for advice.

In the experiment, subjects were paired with a team leader whose role was "to provide advice to [the subject] during the game." Before each round of the *Individual Game*, the team leader showed the subject how he or she (the team leader) had played in that round, and sent a message to the subject. Subjects were informed, accurately, that the team leader's compensation was based in part on how well the team leader's team members played the *Individual Game*, where the subject was one team member.[5] Subjects were also told that the team leader was selected from among students who participated in the experiment at Washington State University. The leader selection process is described in detail in Section 2.2.1 below.

As mentioned, all subjects played the game on a computer. Therefore, the subjects did not physically see the team leader. The team leader sent messages to the subject through the computer and was represented on the computer screen by an avatar face (see Appendix Figure A1). When the subjects were introduced to the team leader, they were informed that a previous version of the

[4]Similarly, the computer is programmed to play *Right* with 8% probability when P1 played 1, 23% probability when P1 played 2, and 93% probability when P1 played 4.

[5]The team leaders were paid a bonus based on the average number of points earned by their team members, at a conversion rate of 100 points to $1 USD. We did not provide this level of detail to the subjects.

experiment had been played at Washington State University and that their leaders were selected from among those students. The introduction provided a pseudonym for the team leader and some basic demographic characteristics of the team leader: gender, age, and year at Washington State University. Subjects were told that the name was a pseudonym to preserve anonymity but that all other information was true and accurate. Free response questions asked about the leader toward the end of the experiment suggest that subjects believed the team leaders were real and believed the information provided about the team leaders. The large majority of subjects described their leader using the correct pronoun, and many described their leader as "trustworthy" or "helpful". Others said they felt a sense of loyalty toward their leader.

All interactions between subjects and team leaders were identical across all subjects (except the experimental variations described in subsequent sections). Each team leader played 4 in every round and advised the subject to also play 4 in every round.[6] Within each treatment arm, the wording of this advice was also identical. This was good advice: recall that 4 was the selection with the highest expected payoff. Our purpose in providing the same advice across all treatments was to ensure consistency across all treatment arms. We chose to provide good advice, as opposed to bad advice, because we were particularly interested in how subjects respond when advisors are actually useful, as opposed to when advisors make mistakes in their advice (see Landsman (2018) and Sarsons (2017) for examples suggesting that women are punished more for mistakes).

After each round, the team leaders sent a message explaining in progressively greater detail why 4 is the optimal choice. Though the team leader's advice was to select the correct answer, the advice appeared subjective to the subject. Subjects would have maximized their points by following the team leader's advice *on average*, but on any given round, the advised play may not have been successful. The stochastic response of the computer made it possible that a subject could earn more points by disregarding the team leader's advice. Therefore, the quality of the advice was not immediately obvious and may not have become clear even after multiple rounds. Similar to Cooper and Kagel (2005), we find that even when advised to play the optimal strategy, a significant proportion of subjects (25 percent in our sample) did not select this strategy even in their final round of the game. Thus, just like in "real world" labor market settings, subjects had to determine the quality of the advice they were provided and could not unequivocally confirm the quality of

[6] As described below, we selected two well-performing leaders.

that advice.

2.2.1 Selection of the team leaders

Two team leaders were selected from 15 real individuals who participated in an experiment at Washington State University focused on the *Individual Game* (i.e., the same signaling game described in Section 2.1). We call these subjects the leader pool. In the experiment at WSU, leader pool subjects learned prior to playing that they could be selected to be a "team leader" for future subjects. We told them that playing well was one criterion for being selected, and that if selected, they would receive a bonus based on how future subjects (their team members) performed in the game.[7] We then provided detailed information on the best way to play and allowed subjects in the leader pool to play twenty rounds; ten rounds were considered practice and the decisions in the remaining ten rounds were shown to the subjects in the *Individual Game*.[8]

After playing, the potential team leaders agreed to share their decisions in the game and to send pre-scripted messages to their team members (i.e., the subjects). Those in the leader pool were shown all potential messages that would be used in the experiment and asked whether they would be willing to send the messages to their potential future team.[9] The purpose of having team leaders send pre-scripted messages was to ensure that the messages provided to subjects were identical. The team leaders consented to sending any and all of the pre-scripted messages shown to them.[10] The experimenter decided which message was sent in which round.

We selected two well-performing leader pool subjects, one male and one female, to serve as the team leaders. These two subjects made identical choices in the *Individual Game* and had identical outcomes.[11] They were also the same age and year at the university. Our use of real individuals

[7] The bonus was the average compensation that subjects earned in playing the game, ranging from USD 2.76 to USD 6.28. All those in the leader pool were also compensated in the same way the subjects were compensated in the primary experiment.

[8] After this training, 7 out of 15 leader pool subjects played the best option.

[9] The exact wording in the experiment was as follows: "Please see below a series of potential messages that could be sent to your team if you're selected as Team Leader. Part of future rounds in this study is to understand nuances in language, so many of the messages are similar with slight changes in the style of language used. Would you be willing to send these messages to your potential future team?" Out of the 15 leader pool subjects, 12 consented to send the messages.

[10] Due to a programming error, three messages were unintentionally omitted from the list shown to leaders. These messages, shown in row 1 of Figure 5 below, were very similar to those that the leaders agreed to.

[11] The potential team leaders played 20 rounds of the *Individual Game*; only the first 10 rounds were shown to their team members (i.e., the subjects). The selected leaders both played 4 in all 20 rounds. In general, across the 20 rounds, subjects in the leader pool played 4 in 75.3% of rounds. Six out of 15 subjects played 4 in all 20 rounds.

who had identical choices and outcomes allowed us to ensure that no deception was involved in the experiment, while also obtaining clean identification of gender discrimination. The team leaders played the games as described to the subjects, received bonuses based on subjects' performance, and all demographic characteristics described to the subjects were true. We did not describe the team leader experiment to the subjects in detail, so subjects were not explicitly told that the team leaders received training prior to playing or that messages were pre-scripted. We highlight these details here to improve clarity of the study to the reader.

2.3 Subject Experience

The study was implemented as a self-guided online survey in Qualtrics. That is, the experimenters did not interact with or provide verbal directions to the subjects. All interactions between subjects and team leaders happened via Qualtrics. Subjects completed the entire study on a computer screen.[12] After providing informed consent, subjects first learned the rules of the *Individual Game*. Before viewing an introduction to their team leader or receiving any advice, subjects played a practice round of the *Individual Game*. Subjects then did the following in chronological order: viewed an introduction to the team leader, received an introductory message from their team leader including their pseudonym, provided an incentivized expectation of their team leader's performance in the *Individual Game*, and then played 10 rounds of the *Individual Game*. One round of the game was randomly selected for payment.[13]

Prior to each round, subjects observed how their team leader had played in the round and their team leader's points earned for that round. In most rounds, they then viewed a message from their team leader. This message was shown in a speech bubble alongside a gendered avatar (see Appendix Figure A1, which shows the male avatar as an example). On the same screen where the new message was shown, all previous messages from the team leader were shown below. Subjects then moved on to making their own selection of which number to play. When making this selection, they saw a table with the following information from all previous rounds: their choice, their points earned,

[12] As described below, UC Merced students participated in the study in a lab, and Amazon MTurk members participated in the study from their own locations. However, both types of participants completed the study online, and the experimenters did not interact with either type of subject.

[13] As described below, the study includes subjects from both UC Merced and Amazon MTurk. The exchange rate of points to USD differed across the two subject pools. In both cases, one round was randomly selected for subject compensation. At UC Merced, subjects were compensated at an exchange rate of 100 points = USD 2. On MTurk, they were compensated based on the formula points/200 - USD 1.

Round 6

Player 1's Points (You)

Type A			Type B		
	If Player 2 selects:			If Player 2 selects:	
	Left	Right		Left	Right
If Type A selects: 1	168	444	If Type B selects: 1	276	568
2	150	426	2	330	606
3	132	408	3	352	628
4	-188	-38	4	316	592

Conversion of Points to Payment: Points/2 - $1.00

(e.g., 568 points = $5.68/2 - $1.00 = $1.84)

Player 2's Points

	If Player 1 is Type A:	If Player 1 is Type B:
If Player 2 selects Left:	500	200
If Player 2 selects Right:	250	250

Round	Your Choice	Your Points	Ashley's Choice	Ashley's Points
1	3	628	4	592
2	3	352	4	592
3	3	352	4	592
4	3	352	4	592
5	4	592	4	592
6			**4**	**592**
7				
8				
9				
10				
Total		**2276**		**2960**

What is your selection?

1

2

3

4

Prefer not to respond (0 points)

Figure 2: Screenshot of *Individual Game*
Note: Options were randomly re-labeled.

their team leader's choice, and their team leader's points (see Figure 2).

After playing the *Individual Game*, subjects had the opportunity to either keep their team leader or be matched to a new team leader before playing a subsequent incentivized game.[14] The new leader was described as having "performance on this final game [that] is similar to [your current leader]." Subjects who chose to change their team leader were matched to a different male leader. The only purpose of this subsequent game was to incentivize the choice of whether to keep or change the leader.[15]

Finally, subjects completed a questionnaire on several topics, in the following order. First, they provided their evaluation of their team leader. Then, we elicited the subject's expectations of how *other* subjects in the experiment followed their team leader's advice as a function of the team leader's gender. This second question was asked as a hypothetical question to a subset of the sample in the original experiment and as an incentivized question, in which subjects earned a bonus based on their accuracy, to all subjects in the replication experiment. These predictions were incentivized using a quadratic scoring rule, with a bonus amount ranging from 0 to USD .25. Third, we elicited the subject's own preference for the type of cheap talk they would use if they were hypothetically selected to play the role of a team leader to future subjects (the cheap talk types are described in detail below). No subjects actually played as team leaders—we only asked what they would do *if* they were hypothetically a team leader in the same experiment. The primary components of the experiment are graphically depicted in Figure 3.

2.4 Treatment Variations

The description of the experiment to this point is identical for all subjects. We now discuss the elements of the experiment that varied for subjects; i.e., the experimental variation introduced by the researchers. We implemented a cross-cutting randomization of team leader gender (male or female) and type of cheap talk used by the team leader, where the type of cheap talk was increasing in

[14] At UC Merced only, subjects also had the opportunity to allow their leader to complete a "risky puzzle." They earned a bonus if their leader completed it correctly and points were deducted if the leader did not complete it correctly. Subjects were not told what the puzzle would be, and the subjects themselves did not complete the puzzle. This was intended as an incentivized measure of beliefs about leader ability as a potential mechanism for explaining discrimination. Because there was no gender difference in this measure at UC Merced, we did not include the question in the MTurk replication so that we could pursue alternative explanations better aligned with the initial results. The puzzle was a difficult math problem. The selected leaders both answered it incorrectly.

[15] The game was one question from the Raven's matrices cognitive test. Subjects who answered correctly earned $1 at UC Merced and $.10 on MTurk.

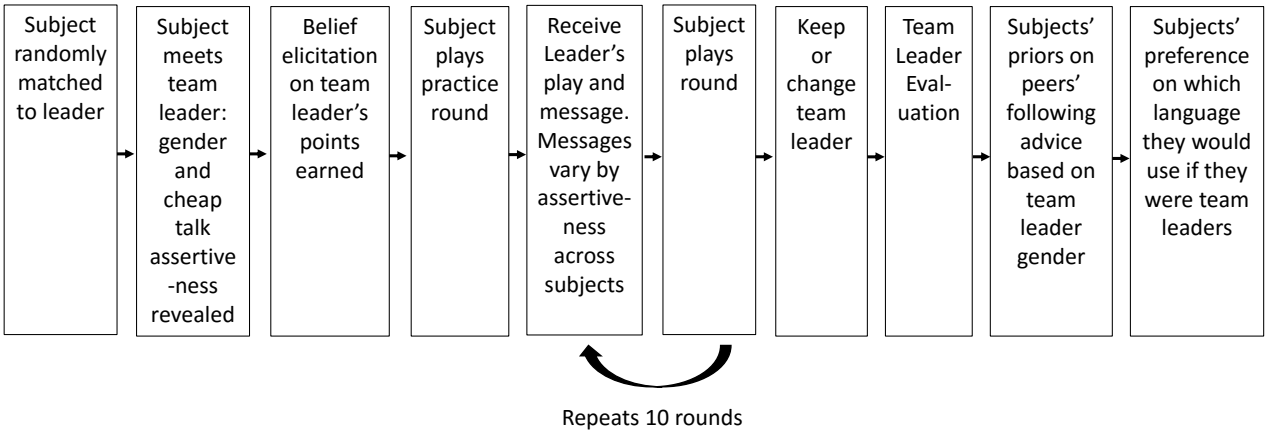


Figure 3: Primary Components of Experiment

	Cheap Talk Type		
	Least assertive	**Moderately assertive**	**Most assertive**
Male team leader	Male team leader using the least assertive cheap talk (i.e., self-deprecating)	Male team leader using moderately assertive cheap talk (i.e., neutral)	Male team leader using the most assertive cheap talk (i.e., self-promoting)
Female team leader	Female team leader using the least assertive cheap talk (i.e., self-deprecating)	Female team leader using moderately assertive cheap talk (i.e., neutral)	Female team leader using the most assertive cheap talk (i.e., self-promoting)

Figure 4: Cross-randomized Experimental Treatment Variation

assertiveness (self-deprecating, neutral, or self-promoting). See Figure 4 for a graphical depiction of the different treatment arms. This cross-cutting design allows us to test whether subject willingness to follow advice changed based on the assertiveness in the cheap talk accompanying the advice and based on the gender of the advice-giver. It also allows us to test whether the assertiveness in the cheap talk changed the gender gap in willingness to follow advice. All other information about the team leader and all interactions with the team leader were identical across subjects, including the substantive component of the advice provided.

2.4.1 Gender

Leader gender is made salient in three ways: (1) directly informing the subject of the team leader's gender during the team leader's online introduction; (2) using a gendered pseudonym and corresponding pronouns in all descriptions of the team leader and all interactions with the team leader;

and (3) showing an image of a gendered face (i.e., avatar) when relaying advice from the team leader. Appendix Figure A1 shows an example of how this information was presented for the male team leader.

Team leaders were randomly assigned one out of 26 possible gendered pseudonyms. These pseudonyms were selected from the most popular 52 names for each gender in the 1990s according to Social Security data. We further narrowed the list by selecting 26 names for which the gender was unambiguous and the perception of age, income, education, and race were balanced across gender perception. We determined how the names were perceived by conducting a validation exercise with 74 undergraduate students from Washington State University. These students each reviewed 10 pseudonyms and were asked to guess the gender, age, income, and education for a typical person with that name. The instructions provided to the students for this validation exercise are shown in Appendix Section A.2.

The team leader's gendered pseudonym was mentioned on over 15 pages of the online survey. At the end of the experiment, over 90 percent of subjects correctly recalled the gender of their team leader.

2.4.2 Cheap Talk

Randomized assertiveness in cheap talk was used in the team leader's online introduction and in advice messages provided throughout the games. As previously described, the team leader provided the same substantive advice in all treatments: the team leader always advised the subject to play 4. However, the treatment arms differed in their level of assertiveness. In the most assertive treatment (self-promoting), the team leader directly asserted that his or her advice was highly valuable. In the moderately assertive treatment (neutral), the team leader provided the advice without additional statements. In the least assertive treatment (self-deprecating), the team leader was humble and communicated some uncertainty about the added value of the advice. The messages used for each treatment arm are shown in Figure 5. The phrases within the introduction and advice in Figure 5 that differed across treatments are bolded for clarity. The subjects did not see any bolding or italics.

The scripts were written to mirror language that is commonly used in professional settings. The most assertive language (i.e., self-promoting) was drawn from example cover letters provided

	Cheap Talk Type		
	Least Assertive	**Moderately Assertive**	**Most Assertive**
Introduction Message to Subjects:	Hi. My name is _____ and I am your Team Leader. I am 21 years old, *[gender*], and an upper level student at Washington State University. **I am not sure how good a leader I will be**, but I am looking forward to playing with you in this experiment.	Hi. My name is _____ and I am your Team Leader. I am 21 years old, *[gender*], and an upper level student at Washington State University. I am looking forward to playing with you in this experiment.	Hi. My name is _____ and I am your Team Leader. I am 21 years old, *[gender*], and an upper level student at Washington State University. **If you listen to my advice, I can assure you that my skills and experiences will help you perform well in this game. I am looking forward to playing with you in this experiment.**
Round 1 Advice:	I played 4. **I am not sure whether it is a good choice.**	I played 4.	I played 4. **I am convinced that it is the best choice.**
Round 2 Advice:	I played 4, **but I'm not an expert at this game.**	I played 4.	I played 4. **Follow my lead and choose 4 too.**
Round 3 Advice:	**I might be missing something in choosing to play 4.**	My decision was to play 4.	**I have figured out how to make good decisions in this experiment. Trust me, 4 is the best play.**
Round 4 Advice:	**You probably have better problem-solving skills than I do, but here is what I am thinking:** When you play 4, Player 2 guesses you are Type B.	When you play 4, Player 2 guesses you are Type B.	**I have extremely strong problem-solving skills and my advice can be an asset to you.** When you play 4, Player 2 guesses you are Type B.
Round 5 Advice:	**I am not sure, but it might help** to make Player 2 think you are Type B. **Am I making sense?**	You need to make Player 2 think you are Type B.	**My strengths include exceptional analytical thinking.** You need to make Player 2 think you are Type B.
Round 6 Advice:	**I don't know, but playing 4 seems like it's working.**	When you play 4, Player 2 is more likely to think that you are Type B.	**I have earned the trust of my peers in the past, and I strongly encourage you to play 4.**
Round 7 Advice:	**Sorry, but I was just thinking that if you play 4, Player 2 might be more likely to play Out, so you might earn more. I'm not sure if that makes sense.**	If you play 4, Player 2 is more likely to play Out, so you earn more.	**I pride myself on my ability to deliver, and I want you to succeed. I figured out that if you play 4, Player 2 is more likely to play Out, so you earn more. That is why I recommend you play 4.**
Round 8 Advice:	**I don't know if this is helpful, but my thought is that maybe** you can make the computer think you are Type B by playing 4.	You can make the computer think you are Type B by playing 4.	**I gravitate naturally to instruction, and I am keen to help you. The smart move is to play 4. See,** you can make the computer think you are Type B by playing 4.

Figure 5: Advice Messages by Cheap Talk Treatment

Note: No text was bolded for subjects. Advice in Round 5 and 8 are edited to be more concise. No advice messages were provided in Rounds 9 and 10.

by university career centers. The least assertive language (i.e., self-deprecating) was based on "common language to avoid" on career advice internet sites. We confirmed through survey questions that the experimental manipulation worked as intended. Subjects in the most assertive treatment (i.e., self-promoting) were significantly more likely to rate their leader's cheap talk style as "overly assertive/aggressive" (20 percent versus 6 percent) and characterized the advice messages as being "more masculine" (36 percent versus 10 percent). Those in the least assertive treatment arm (i.e., self-deprecating) were significantly more likely to rate the cheap talk style as "submissive/under assertive" (44 percent versus 2 percent) and characterized the messages as "more feminine" (30 versus 10 percent).[16]

We also validated the cheap talk treatments with our sample of 74 undergraduate students at Washington State University. The students were asked to compare the moderately assertive advice with either the least assertive (self-deprecating) or most assertive (self-promoting) advice in the experiment. They were asked to circle the message they believed to be more self-deprecating (i.e., least assertive) in the former comparison, or more self-promotional (i.e., most assertive) in the latter comparison. More than 70 percent of the messages were correctly identified as being "more self-promotional" (71 percent) or "more self-deprecating" (77 percent). The instructions provided to the students for this validation exercise are shown in Appendix Section A.2.

2.5 Sample

We report the results of an initial experiment and a replication of that experiment. We first conducted a laboratory experiment with 435 undergraduate students at the University of California, Merced (UCM). To study the reproducibility of our results, we replicated this experiment with 576 Amazon Mechanical Turk workers. Mechanical Turk (MTurk) is an online platform that allows researchers to post small tasks.[17] Recent research has suggested that Amazon MTurk workers are

[16]For statements on gendered language, subjects could select from the following choices to characterize the language used by their team leader: 1) more masculine (i.e., more representative of the type of language used by males), 2)gender neutral (i.e., equally representative of the type of language used by males and females), 3) more feminine (i.e., more representative of the type of language used by females), or 4) prefer not to answer. For statements on assertiveness, subjects could select from the following choices to rate their team leader's cheap talk style: 1) Submissive/under assertive, 2) Appropriately assertive, 3) Overly assertive/aggressive, and 4) prefer not to respond.

[17]We recorded an additional 339 MTurk workers who began the survey, but did not complete it. However, the significant majority (72 percent) stopped prior to the revelation of the team leader. The remaining 94 are balanced across treatment arms making it unlikely that treatment assignment resulted in differential demand to complete the experiment. Appendix Table A1 shows that attrition in the sample is balanced across treatment arms.

more representative of the US population than American university students and that the resulting data is no less reliable than when obtained in a traditional lab (Horton, Rand and Zeckhauser, 2011; Paolacci, Chandler and Ipeirotis, 2010; Buhrmester, Kwang and Gosling, 2011). Both studies were pre-registered in the American Economic Association social science registry. Links to the pre-registrations in the American Economic Association registry and to the experiment (i.e., an online survey hosted on Qualtrics) can be found in Appendix Section A.1.

The original UC Merced sample was 65 percent female with a mean age of 20 and mean grade point average of 3.0. The replication sample was 38 percent female, with a mean age of 34 and mean level of education of college credit with no degree.[18] Summary statistics for the two samples are shown in Appendix Table A2.

We confirm the validity of our randomization by conducting a series of balance tests on our experimental sample, using the data we collected about our subjects.[19] Appendix Table A3 shows balance using our main specification, where assertiveness increases from 0 to 2 to reflect the assertive ordering of the cheap talk treatment arms. In addition, Appendix Table A4 tests for statistically significant differences in subject characteristics across the six treatment groups shown in Figure 4.[20] While generally well-balanced, we do observe a significant difference in GPA for those assigned to female leaders at UC Merced. To explore whether this could affect our results, we estimate whether subject characteristics predict following advice and playing strategically, our primary outcome of interest (Appendix Table A5). We do not find evidence that GPA predicts adherence to advice, suggesting that the imbalance we observe is not driving our results. In addition, we observe no statistical difference across any treatment arm on how subjects' played in the practice round, prior to receiving any advice.

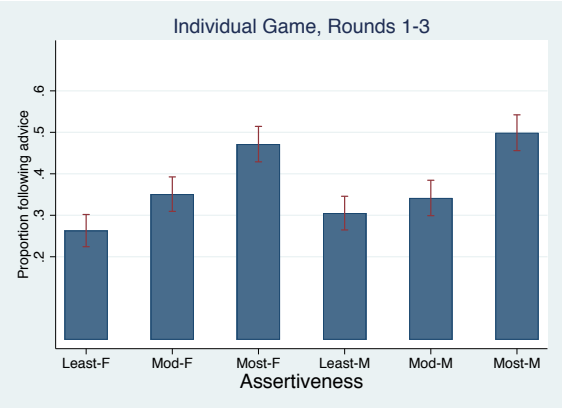
2.6 Surveying Experts

We surveyed 14 experts in gender-related research in economics and psychology on their expectations of how subjects would follow advice in the game based on the type of cheap talk and the gender of the leader. Experts were shown a summary of the experimental design and had the ability to observe

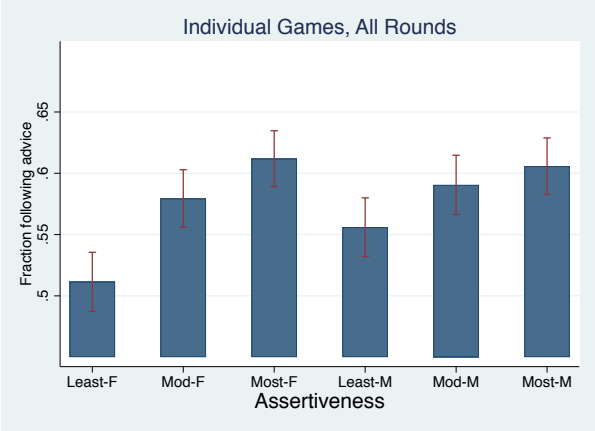
[18] 3 subjects did not report gender and 10 subjects did not report GPA in the original sample. 6 subjects did not report gender in the replication sample.

[19] We collected relatively limited data on the subjects in order to reduce the length of the overall experiment.

[20] In addition to the comparisons shown, there are 15 potential pairwise comparisons between treatment groups. We do not show these for brevity, but the results are consistent with the results shown.



(a) Rounds 1 to 3



(b) All Rounds

Figure 6: Likelihood of Following Advice, *Individual Game*

exactly what subjects experienced when stating their expectations. This process was pre-registered in the American Economic Association social science registry (see Appendix Section A.1 for links to the pre-registration). All analysis and results from this exercise are shown in Appendix Section A.3.

3 Results

In the following sections, we present combined results for the original and replication experiments, unless otherwise noted. Results for each separate experimental sample, and all regressions specified in our pre-analysis plan, are shown in the Appendix.

3.1 Gender Does Not Matter, but Cheap Talk Does

We find no evidence that team leader gender affected the likelihood of subjects following the advice given, the primary outcome of the experiment (Figure 6, Table 1). However, more assertive cheap talk increased willingness to follow advice for both men and women. Figure 6 shows the percentage of subjects that followed the team leader's advice for each treatment arm in rounds 1 to 3 (Figure 6a) and in all rounds (Figure 6b) of the *Individual Game*. Table 1 shows the regression estimates of the effect of gender and assertiveness.

Aggregating across all cheap talk types and rounds, we can rule out discrimination against female leaders of more than 5.1 percentage points (i.e., 5.1 percentage points is the lower bound of

Table 1: Assertive Cheap Talk and Gender Gaps in Willingness to Follow Advice

	Round 1			Round 1 to 3			All Rounds		
	(1) UCM	(2) MTurk	(3) All	(4) UCM	(5) MTurk	(6) All	(7) UCM	(8) MTurk	(9) All
Fem. Leader	-0.0827	0.0103	-0.0320	-0.0671	0.00177	-0.0293	-0.0295	-0.0528	-0.0426
	(0.0567)	(0.0543)	(0.0394)	(0.0417)	(0.0442)	(0.0310)	(0.0340)	(0.0376)	(0.0261)
Assertive	0.0913**	0.0529*	0.0674***	0.110***	0.0867***	0.0953***	0.0245	0.0254	0.0245
	(0.0384)	(0.0300)	(0.0238)	(0.0246)	(0.0251)	(0.0180)	(0.0209)	(0.0209)	(0.0151)
Fem. Leader × Assertive	0.0371	-0.000788	0.0167	0.0173	-0.00207	0.00647	0.0267	0.0257	0.0247
	(0.0502)	(0.0434)	(0.0328)	(0.0341)	(0.0356)	(0.0253)	(0.0284)	(0.0300)	(0.0211)
Round FE	X	X	X	X	X	X	X	X	X
Practice round	X	X	X	X	X	X	X	X	X
Sample FE			X			X			X
Obs (Individual*Round)	435	575	1010	1305	1725	3030	4350	5750	10100
A + F × A	0.128	0.0521	0.0842	0.128	0.0847	0.102	0.0512	0.0510	0.0492
P-val	0.000	0.096	0.000	0.000	0.001	0.000	0.008	0.018	0.001

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the individual level. Assertive is a variable ranging from 0 to 2, increasing in the assertiveness of the cheap talk. Practice Round is an indicator for playing strategically in practice round. Round FE are fixed effects for each of the ten rounds played. Sample fixed effects are fixed effects reflecting whether the subject participated in the original experiment at UC Merced or the replication experiment on Amazon MTurk. UCM refers to the original sample at UC Merced, MTurk refers to the replication sample on Amazon MTurk, and All refers to both samples combined.

the 95 percent confidence interval, as shown in Appendix Table A6). There were also no statistically significant gender differences within each cheap talk treatment. For both male and female team leaders, greater assertiveness increased adherence to advice. Appendix Table A7 shows the corresponding regression estimates of statistical differences across leader gender and each level of assertiveness.

The gendered pattern of effects of assertiveness in cheap talk is inconsistent with backlash to gender norm violations in terms of following advice. We define backlash as occurring when a subject is less likely to follow advice from a team leader who violates gender norms. On average, the subjects viewed the more assertive cheap talk types as more masculine. As the assertiveness of cheap talk increased, subjects characterized their team leader's language as being more "masculine" and less "feminine" than subjects assigned to less assertive treatment arms (Column 1 and 2 in Table 2). Thus, female leaders were increasingly violating gender norms, as perceived by the subjects, in more assertive cheap talk treatment arms.

Because assertive cheap talk was viewed as more masculine, if backlash were present, adherence to advice would have decreased as female leaders became more assertive. However, subjects partnered with a moderately assertive or most assertive female leader were *more* likely to follow her advice. This is true for both the increase from least assertive to moderately assertive, and from moderately assertive to most assertive, as shown in Figure 6.

Interestingly, the random assignment to a female leader reduced the likelihood that subjects characterized the language they observed as masculine, and increased the likelihood that they characterized it as feminine (Columns 3 and 4 in Table 2). This suggests that gendered perceptions of assertive language may be quite malleable.

In contrast to the predictions generated by a theory of backlash, increasing assertiveness improved adherence to advice for *both* male and female leaders despite more assertive language being seen as more masculine (Table 1).[21] This effect was largest in early rounds of the game. This is notable because subjects had less objective evidence of the quality of the team leader's advice in earlier

[21]This specification is slightly different from the main estimating equation in our pre-analysis plan. The pre-specified equation estimated the marginal effect of the least assertive (i.e., self-deprecating) and most assertive (i.e., self-promoting) cheap talk separately. This reflected an expectation of nonlinearity in the effects of assertive cheap talk, based on role congruity theory. Since we do not find evidence for nonlinearity, for clarity in presentation, we show the average marginal effect of increased assertiveness in cheap talk, where assertiveness increases from 0 to 2 to reflect the assertive ordering of the cheap talk treatment arms. The pre-specified estimation is shown in Appendix Table A7.

Table 2: Language Characterization by Subjects

	(1) Masculine	(2) Feminine	(3) Masculine	(4) Feminine
Assertive	0.140***	-0.0977***	0.140***	-0.0980***
	(0.0146)	(0.0149)	(0.0142)	(0.0144)
Fem. Leader			-0.158***	0.164***
			(0.0225)	(0.0219)
Sample FE	X	X	X	X
Observations	1010	1010	1010	1010
Mean	0.187	0.167	0.187	0.167

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Dependent variables reflect the subject's perception of the team leader's advice: Masculine(Feminine) is an indicator for Characterized Messages as Masculine (Feminine). Assertive is a variable ranging from 0 to 2, increasing in the assertiveness of the cheap talk. Sample fixed effects are fixed effects reflecting whether the subject participated in the original experiment at UC Merced or the replication experiment on Amazon MTurk.

rounds. In the first round, subjects had observed only one round of play by the team leader, and each incremental increase in leader assertiveness raised the likelihood of the advice being followed by 7 percentage points (Column 3). The estimated effect of assertive cheap talk is 2 percentage points higher for female leaders, though this difference is not statistically significant. The effect of assertiveness diminishes as we include subsequent rounds of play, where subjects saw that the team leader consistently earned high payoffs through strategic play (Columns 4-9). Yet, even when combining all rounds, the effect of assertive cheap talk remains positive and significant for female leaders (Column 9). These results are robust to controlling for subject characteristics (Appendix Table A9).

We additionally confirm that our results are not masking backlash against assertive female team leaders that is driven by only male subjects. In Appendix Table A8, we limit analysis to only male subjects. Similar to the main results, among male subjects, the effect of assertiveness is not statistically different across male versus female leaders. Indeed, the point estimates suggest that more assertive cheap talk may be more effective for female leaders.

Though we did not find evidence for backlash in terms of adherence to advice, previous literature suggests that violating gender norms could have negative consequences along other dimensions. In

other words, perhaps there was no backlash in willingness to follow advice, but there was backlash on other important aspects of leadership, such as likeability. Table 3 estimates the effects of the type of cheap talk and leader gender on several additional outcomes. In Columns 1 and 2, we study beliefs about the capacity of leaders. We elicited the expected performance of the team leader in the *Individual Game* (Column 1), and subjects received a bonus based on the accuracy of their guess (see Figure 3). Second, we offered subjects the opportunity to either keep or change the team leader providing advice in a subsequent game (Column 2), which was also incentivized. We also study a summary index of subjective assessments of the leaders (Column 3) and feelings of self-power (Column 4). Results for each underlying measure in the summary index can be found in Appendix Tables A12-A15.

We do not find robust differences in perceived ability by team leader gender, or by assertive cheap talk (Column 1 of Table 3). Though not statistically significant, the point estimates are inconsistent with backlash predicted by role congruity theory. Increasing assertiveness has a much larger and positive effect for female leaders, despite the greater likelihood of assertiveness in the language being characterized as masculine. Similarly, the least assertive (i.e., self-deprecating) female leaders were expected to perform worse than their male counterparts using the same language, despite such self-deprecating language being characterized as more feminine. Indeed, when subjects had the opportunity to request a different team leader for a final task, they were equally likely to choose to keep working with male and female leaders, regardless of language type (Column 2 of Table 3). We generally see the same pattern for the subjective evaluations and self-power.

Much of the existing literature on gender role congruity has focused specifically on backlash in terms of likeability. The subjective evaluations included measures of both competence and likeability. Thus, if assertive language increased perceived competence and decreased likeability for female leaders, this could be hidden in the composite measures in Table 3. In Table 4, we focus specifically on measures of likeability in the subjective evaluations. Again, we find no evidence of backlash in terms of likeability towards more assertive female leaders.

In Table 3, assertiveness improves the overall subjective evaluations for both men and women. Table 4 shows which components of this measure drive the results. We see that assertiveness increases the perception of stronger interpersonal skills and raises the likelihood the leader will be recommended to others. However, these average effects of assertiveness do mask some variation in

Table 3: Subjective Evaluations of the Team Leader

	(1) Exp. Perf.	(2) Retained	(3) Eval.	(4) Self-Power
Fem. Leader	-0.225	-0.00871	0.0269	-0.0381
	(0.478)	(0.0309)	(0.0508)	(0.0674)
Assertive	-0.0340	0.0112	0.0981***	-0.00364
	(0.266)	(0.0157)	(0.0266)	(0.0363)
Fem. Leader × Assertive	0.399	0.00626	0.0235	0.00240
	(0.377)	(0.0227)	(0.0376)	(0.0506)
Sample FE	X	X	X	X
Observations	1001	1010	1009	1009
A + F × A	0.365	0.0174	0.122	-0.00124
P-val	0.172	0.289	0.000	0.972

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Assertive is a variable ranging from 0 to 2, increasing in the assertiveness of the cheap talk. Dependent variables reflect the subject's assessment of the team leader. Exp. Perf is subject's belief of the ventile of points earned by the team leader over all ten rounds in the Individual Game (Expected Leader Performance). Retained is an indicator for whether the subject kept the team leader (as opposed to getting a new team leader) for a final incentivized game (Retained Leader). Eval. is the summary index of 16 questions evaluating the team leader, increasing in positive evaluation (Leader Evaluation). Self-Power: Summary index of 8 questions evaluating the subject's own feelings of power and control. Selects is an indicator for whether the subject selected the team leader to play an incentive logic game in which the subject gained earnings if the team leader performed well and lost earnings if the team leader performed poorly (Selects Leader for Risky Puzzle). Missing observations reflect subject preferring not to respond to the questions used to construct the dependent variable. Sample fixed effects are fixed effects reflecting whether the subject participated in the original experiment at UC Merced or the replication experiment on Amazon MTurk.

the response to assertiveness. Appendix Tables A10-A15 show that both the least assertive and the most assertive leaders receive lower overall evaluations than the moderately assertive (i.e., neutral) leaders. The underlying measures suggest that different dimensions of likeability can be reduced both for not being assertive enough and for being too assertive. However, we continue to find no robust differences by leader gender for any level of assertiveness.

3.2 Are these results surprising? Subject Priors

The lack of gender discrimination in the experiment does not align with the expectations of subjects themselves (Table 5). In our original experiment (UCM), we asked a subset of subjects whether they expected their peers' willingness to follow advice to depend on the team leader's gender.[22] As shown in Table 5, Panel A, a significant proportion of the sample believed that their peers would be more likely to follow the advice provided by a male team leader. In contrast, only a handful of subjects expected that adherence would be higher for female team leaders.

Similarly, in the MTurk replication experiment, we asked subjects to estimate the proportion of other subjects they believed would follow the same advice they observed for Round 1 of the *Individual Game*, when the advice was provided by a male team leader versus by a female team leader. We provided subjects with a bonus based on the accuracy of their estimates. Regardless of their cheap talk treatment status, subjects believed that female team leaders' advice was less likely to be followed (Table 5, Panel B). Column 1 of the last row of the table shows that when combining all types of cheap talk, we can reject that the gender gap in the experiment was as large as expected by subjects. This expected gender gap did not increase as the leaders used more assertive language.

Given this discrepancy between beliefs and outcomes, we implemented a *Coordination Game* in the replication experiment, in which subjects were better off following the leader's advice only if they believed their partner would as well. Thus, if subjects expected others to discriminate against female team leaders, then subjects with female team leaders would be better off also ignoring the advice. Appendix Table A16 presents the results of the Coordination Game; in summary, we again do not find gender gaps in willingness to follow the leader, suggesting that subjects' expectation of others' discrimination did not translate into increased discrimination against female leaders. However,

[22] We added this question after we started the original experiment, and as a result, we only have responses from subjects that participated in the later days of the experiment at UC Merced.

Table 4: Likeability of the Team Leader

	(1) Enjoyed	(2) Respect	(3) Approachable	(4) Interpersonal	(5) Rec. Colleague	(6) Rec. Supervisor
Fem. Leader	0.0645	0.0426	0.114	0.0326	0.118	0.0605
	(0.128)	(0.102)	(0.128)	(0.138)	(0.114)	(0.134)
Assertive	0.0417	-0.0833	-0.0794	0.247***	0.0538	0.178**
	(0.0706)	(0.0602)	(0.0763)	(0.0782)	(0.0648)	(0.0707)
Fem. Leader × Assertive	0.0161	0.0203	0.00288	0.0347	0.0446	0.0509
	(0.0995)	(0.0813)	(0.103)	(0.107)	(0.0872)	(0.0987)
Sample FE	X	X	X	X	X	X
Observations	941	980	983	989	991	971
A + F × A	0.0579	-0.0630	-0.0765	0.282	0.0983	0.229
P-val	0.409	0.249	0.265	0.000	0.092	0.001

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the individual level. Assertive is a variable ranging from 0 to 2, increasing in the assertiveness of the cheap talk. Dependent variables are the level of agreement with the following statements: *I enjoyed working with –* ; *– treated me with respect*; *– would be approachable for an issue that bothered me*; *– has strong interpersonal skills*; *I would recommend – as a colleague to others*; *I would recommend – as a supervisor to others.* All measures increase in positive evaluations of the team leader. Missing observations reflect subject preferring not to respond to the questions used to construct the dependent variable.

Table 5: Subject Priors

		Assertiveness		
	(1)	(2)	(3)	(4)
	Full sample	Most	Mod.	Least
Panel A: Unincentivized Priors, Individual Game				
Pct: Female preferred	7.51	8.33	8.70	5.56
Pct: Male Preferred	36.15	40.28	31.88	36.11
Pct: Equal	56.34	51.39	59.42	58.33
Obs (Individuals)	213	72	69	72
Panel B: Incentivized Priors, Individual Game R1				
Fem. Leader	-0.0865***	-0.0710***	-0.0768***	-0.113***
	(0.00836)	(0.0134)	(0.0133)	(0.0164)
Male Mean	0.692***	0.695***	0.708***	0.673***
	(0.00907)	(0.0154)	(0.0166)	(0.0152)
Obs (Individuals*Prior per Gender)	1141	400	365	376
Estimated Effect	0.00914	-0.0182	0.0781	-0.0216
P value: est. effect = prior	0.00855	0.412	0.0170	0.115

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Panel A reports the percent of subjects reporting they expected peers were more likely to follow advice in the *Individual Game* when provided by female team leaders, male team leaders, or both genders equally. Panel B estimates the difference between subject's expectation of the percentage of MTurk subjects following the advice in Round 1 of the *Individual Game* when provided by a male team leader relative to when provided by a female team leader and uses robust standard errors. Panel A is calculated on the subset of subjects in the original sample to whom the question was asked; Panel B is estimated on the replication sample. The observations reflect two priors per subject: beliefs about the expected adherence to male team leaders and to female team leaders.

subject priors in both the *Individual Game* and *Coordination Game* show that they expected others to follow advice at much higher rates than they themselves did: while only 27 percent of subjects followed the leader's advice in Round 1 of the *Individual Game*, subjects expected over 60 percent of other subjects would follow the advice. Given a binary choice to follow the leader's advice or not, the expectation that most others would follow advice likely dominated discriminatory behaviors that could be observed in more continuous decisions.

3.3 Cheap Talk Preferences

The results indicate that the most assertive cheap talk was the optimal strategy for leaders to increase adherence to their advice in this experiment. We asked a subset of subjects in the original experiment, and all subjects in the replication, what type of cheap talk they would prefer to use if they were hypothetically selected to be team leaders.[23] Subjects were shown the advice messages from each of the three cheap talk treatment arms, and they were asked to select which set of messages they would prefer to use if they were a team leader. This was shown to subjects only after they finished playing the *Individual Game* and had responded to their own team leader's advice.

Despite the high potential return to assertive cheap talk, Table 6 shows that less than half of subjects chose the most assertive, self-promotional messages. The aversion to assertive language was stronger among female subjects. Only 32.7 percent of female subjects chose the most assertive messages, compared to 42.9 percent of male subjects. This difference was statistically significant at the 1 percent level. Even fewer subjects selected the least assertive language, and we see no consistent pattern for differences in this preference by subject gender.

Importantly, there were no significant differences in how male and female subjects responded to assertive cheap talk (Appendix Table A17). Thus, female subjects were less likely to choose assertive cheap talk, even though their response to assertiveness was similar to that of male subjects. Because assertive cheap talk increases credibility so effectively, these differences in cheap talk preferences would generate gender gaps in willingness to follow advice.

[23] We added this question after starting the original experiment at UC Merced, and as a result, we only have responses from subjects that participated in the later days of the experiment at UC Merced.

Table 6: Subject Preference for Assertive Cheap Talk

	(1) Assertive Preference	(2) Prefers Most Assertive	(3) Prefers Least Assertive
Female Subject	-0.126***	-0.102***	0.0239
	(0.0452)	(0.0336)	(0.0234)
Constant (Male Mean)	1.330***	0.429***	0.0981***
	(0.0529)	(0.0392)	(0.0263)
Sample FE	X	X	X
Observations	778	778	778

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Assertive Preference is a variable ranging from 0 to 2, increasing in the assertiveness of the cheap talk. Prefers Most Assertive (Least Assertive) is an indicator for selecting messages from the Most Assertive (Least Assertive) treatment arm if he/she were a team leader in the Individual Game, instead of messages from the More Assertive or Least Assertive (Most Assertive) treatment arm. Female Subject is an indicator for the subject being female and the Male Mean is the mean of male subjects. Sample fixed effects are fixed effects reflecting whether the subject participated in the original experiment at UC Merced or the replication experiment on Amazon MTurk. Observations include a subset of subjects from the original experiment at UC Merced and all subjects in the replication experiment on Amazon MTurk.

4 Discussion and Conclusion

A large body of literature shows that women are less likely to engage in assertive self-promotion (Exley and Kessler, 2019; Cooper and Kagel, 2016; Babcock et al., 2003; Moss-Racusin, Phelan and Rudman, 2010). One hypothesis is that this choice reflects expectations that assertive self-promotion will be ineffective for women (Amanatullah and Morris, 2010; Moss-Racusin, Phelan and Rudman, 2010). However, assertiveness may be a useful tool in signaling the level of certainty and confidence in one's statements. In this paper, we present a real-stakes setting where assertive cheap talk is highly effective and there is no evidence for backlash to assertive cheap talk by women, either in terms of likeability or willingness to follow advice.

Although subjects did not discriminate against women, they expected others would. This discrepancy cannot be explained by the nature of our game because we elicited beliefs on how subjects expected peers to respond in this specific experiment. It also cannot be explained by men performing differently than women in the game itself. If there were gender differences in performance, it might be the case that subjects rationally expected female leaders to perform worse, and thus

expected other subjects to follow their advice less often. We find no evidence that women were less skilled at the game. Female and male subjects selected the strategic play at equal rates in the practice round, prior to receiving advice. Moreover, in later rounds of the *Individual Game*, female subjects' performance was better than that of male subjects (see Appendix Table A18). Of course, this difference may be due either to differences in ability or differences in willingness to follow advice. In either case, the lack of differences by subject gender on performance in early rounds of the *Individual Game*, along with *better* performance by female subjects in later rounds of the game, suggests that accurate beliefs on ability differences by gender are an unlikely explanation for subjects' beliefs that female team leaders' advice is less likely to be followed.

Other research has also found a mismatch between expectations of gender discrimination and gender discrimination in reality. Alston (2019) finds that both male and female MTurk workers expect gender discrimination in a hiring context in which such discrimination does not occur. She finds that male job applicants were willing to pay to reveal their gender to a hiring manager. In contrast, female job applicants were willing to pay to *hide* their gender from a hiring manager. However, when gender was known, male and female job applicants were hired at the same rate.

One plausible explanation for these discrepancies between subject expectations and subject behavior is that subjects have difficulty determining in which contexts gender discrimination will occur. There is significant evidence that gender discrimination exists in numerous contexts.[24] When discrimination is occurring in the broader environment, this may generate uncertainty about whether discrimination is occurring in a specific context. Indeed, this uncertainty is commonly described as a tax that women and minorities have to cope with. McIntosh (1988) describes one of the effects of white privilege as the following belief held by white individuals: "If I have low credibility as a leader I can be sure that my race is not the problem," and she highlights that male privilege operates the same way. This suggests that wider discrimination in society could generate expectations of discriminatory behavior by others even in contexts where there is no discrimination.

Although subjects expected gender discrimination, they did not expect such discrimination to be exacerbated by assertive cheap talk. Yet, women expressed a disproportionate preference for less assertive language. Thus, we present a setting where women's preference for less assertive language

[24]See Bertrand and Duflo (2017) and Neumark (2018) for reviews of the literature in economics, and Eagly (2013) for a review of the psychology literature on gender discrimination in the workplace.

cannot be explained by backlash avoidance. In future research, a better understanding of the factors driving women's preference for less assertive language will be important for understanding gender gaps in adherence to leadership.

In addition, our results suggest that the language leaders choose may be an important factor in explaining gender gaps. Identifying the role of the choice itself is also an important direction for future research. Individuals' beliefs about how much flexibility a leader has in the language they use may affect their responsiveness to leadership by gender. For example, if subjects believe that leaders have little choice over the language used, women's use of more assertive language may not be viewed as a gender norm violation. In contrast, when individuals believe that the leaders have significant choice in the language they are using, then there may be more concern about backlash towards women for violating gender norms.

Our results are consistent with and contribute to the gender backlash literature's differentiation between implicit and explicit assertiveness (Williams and Tiedens, 2016). Even though subjects themselves identified the language in our experiment as being more masculine or feminine, they did not exhibit backlash towards women or men who violated those norms in terms of either willingness to follow advice or subjective evaluations of likeability, where backlash is commonly found. This suggests that the level of assertiveness required to enter consciousness and trigger backlash may be quite high. In addition, the perception of gender norms with respect to assertive language was itself quite malleable: whether subjects perceived the language as more masculine or more feminine was influenced by the randomized leader gender assignment. It is possible that a more explicit form of assertiveness or dominance behavior could trigger backlash in terms of willingness to follow advice.

However, our results show that assertive language that is commonly used in the labor market does not appear to trigger backlash, particularly in terms of the critical downstream effect of following advice. Assertive language has strong positive returns in terms of advice following for both men and women, though it can reduce likability. Increasing the use of assertive language may be an effective strategy for women to increase their influence and credibility in many labor market settings.

References

Abel, Martin. 2019. "Do Workers Discriminate against Female Bosses?" https://www.dropbox.com/s/nnq7i69em0g4edm/Abel_2019_Gender%20Discrimination%20Managers.pdf.

Alston, Mackenzie. 2019. "The (Perceived) Cost of Being Female: An Experimental Investigation of Strategic Responses to Discrimination." 42.

Amanatullah, Emily T., and Michael W. Morris. 2010. "Negotiating Gender Roles: Gender Differences in Assertive Negotiating Are Mediated by Women's Fear of Backlash and Attenuated When Negotiating on Behalf of Others." *Journal of Personality and Social Psychology*, 98(2): 256–267.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–1495.

Ayalew, Shibiru, Shanthi Manian, and Ketki Sheth. 2019. "Discrimination from Below: Experimental Evidence from Ethiopia." https://drive.google.com/file/d/1_39vu-dnNkyr6D22jLcBMRgD8MTSl9Np/view.

Babcock, Linda, Sara Laschever, Michele Gelfand, and Deborah Small. 2003. "Nice Girls Don't Ask." *Harvard Business Review*.

Bertrand, Marianne, and Esther Duflo. 2017. "Field experiments on discrimination." In *Handbook of economic field experiments*. Vol. 1, 309–393. Elsevier.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg. 2019. "The Dynamics of Discrimination: Theory and Evidence." *American Economic Review*, 109(10): 3395–3436.

Bonaccio, Silvia, and Reeshad S. Dalal. 2006. "Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences." *Organizational Behavior and Human Decision Processes*, 101(2): 127 – 151.

Boring, Anne. 2017. "Gender biases in student evaluations of teaching." *Journal of Public Economics*, 145: 27–41.

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High Quality, Data?" *Perspectives in Psychological Science*, 6(1): 3–5.

Casciaro, Tiziana, and Miguel Sousa Lobo. 2008. "When Competence Is Irrelevant: The Role of Interpersonal Affect in Task-Related Ties." *Administrative Science Quarterly*, 53(4): 655–684.

Chakraborty, Priyanka, and Danila Serra. 2019. "Gender differences in top leadership roles: Does worker backlash matter?" http://people.tamu.edu/ dserra/ChakrabortySerraLeadershipFeb2019.pdf.

Charness, Gary, Aldo Rustichini, and Jeroen van de Ven. 2018. "Self-confidence and strategic behavior." *Experimental Economics*, 21(1): 72–98.

Cooper, David J., and John H. Kagel. 2005. "Are two heads better than one? Team versus individual play in signaling games." *American Economic Review*, 95(3): 477–509.

Cooper, David J., and John Kagel. 2016. "A failure to communicate: an experimental investigation of the effects of advice on strategic play." *European Economic Review*, 82: 24–45.

DellaVigna, Stefano, and Devin Pope. 2018. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy*, 126(6): 2410–2456.

Eagly, Alice H. 2013. "Women as Leaders: Leadership Style Versus Leaders' Values and Attitudes." In *Gender and work: Challenging conventional wisdom*. Harvard Business School Press. https://www.hbs.edu/faculty/conferences/2013-w50-research-symposium/Documents/eagly.pdf.

Eagly, Alice H., and Steven J. Karau. 2002. "Role congruity theory of prejudice toward female leaders." *Psychological Review*, 109(3): 573–598.

Egan, Mark L, Gregor Matvos, and Amit Seru. 2017. "When Harry Fired Sally: The Double Standard in Punishing Misconduct." *NBER Working Paper Series*, 23242.

Exley, Christine L, and Judd B Kessler. 2019. "The Gender Gap in Self-Promotion." National Bureau of Economic Research Working Paper 26345.

Grossman, Philip J., Catherine Eckel, Mana Komai, and Wei Zhan. 2019. "It pays to be a man: Rewards for leaders in a coordination game." *Journal of Economic Behavior and Organization*, 161.

Hengel, Erin. 2019. "Publishing while female: Are women held to higher standards? Evidence from peer review." http://www.erinhengel.com/research/publishing_female.pdf.

Hoffmann, Diane E., and Anita J. Tarzian. 2001. "The Girl Who Cried Pain: A Bias against Women in the Treatment of Pain." *The Journal of Law, Medicine & Ethics*, 28(4_suppl): 13–27.

Horton, John, David G. Rand, and Richard J. Zeckhauser. 2011. "The online laboratory: conducting experiments in a real labor market." *Experimental Economics*, 14(3): 399–425.

Landsman, Rachel. 2018. "Gender Differences in Executive Departure." https://drive.google.com/file/d/0B4Gus2bxOyznZXM3TVlfTzZ2TWM/view.

Libby, Heather. 2016. "3 things women say that weaken the power of their words." https://www.upworthy.com/3-things-women-say-that-weaken-the-power-of-their-words.

McIntosh, Peggy. 1988. "White privilege: Unpacking the invisible knapsack."

Mengel, Friederike, Jan Sauermann, and Ulf Zölitz. 2018. "Gender Bias in Teaching Evaluations." *Journal of European Economic Association*, 17(2): 535–566.

Moss-Racusin, Corinne A., Julie E. Phelan, and Laurie A. Rudman. 2010. "When men break the gender rules: Status incongruity and backlash against modest men." *Psychology of Men & Masculinity*, 11(2): 140–151.

Neumark, David. 2018. "Experimental Research on Labor Market Discrimination." *Journal of Economic Literature*, 56(3): 799–866.

Paolacci, Gabriele, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. "Running experiments on Amazon Mechanical Turk." *Judgment and Decision Making*, 5(5): 411–419.

Phelan, Julie E., and Laurie A. Rudman. 2010. "Prejudice Toward Female Leaders: Backlash Effects and Women's Impression Management Dilemma." *Social and Personality Psychology Compass*, 4(10): 807–820.

Rudman, L.A., and P. Glick. 2016. "Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers." *Journal of Personality and Social Psychology*, 82: 1004–1010.

Rudman, Laurie A., and Julie E. Phelan. 2008. "Backlash effects for disconfirming gender stereotypes in organizations." *Research in Organizational Behavior*, 28: 61–79.

Sarsons, Heather. 2017. "Interpreting Signals in the Labor Market: Evidence from Medical Referrals." https://drive.google.com/file/d/1bDV1Tqhl6SX2CtM6Sf1c95PF5eloDJtr/view.

Schotter, Andrew. 2003. "Decision Making with Naive Advice." *American Economic Review*, 93(2): 196–201.

Solnit, Rebecca. 2008. "Men Explain Things to Me: Facts Didn't Get in Their Way." http://www.tomdispatch.com/blog/174918/tomgram%3A_rebecca_solnit%2C_the_archipelago_of_arroganc

Williams, Melissa J., and Larissa Z. Tiedens. 2016. "The subtle suspension of backlash: A meta-analysis of penalties for women's implicit and explicit dominance behavior." *Psychological Bulletin*, 142(2): 165–197.