# 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.

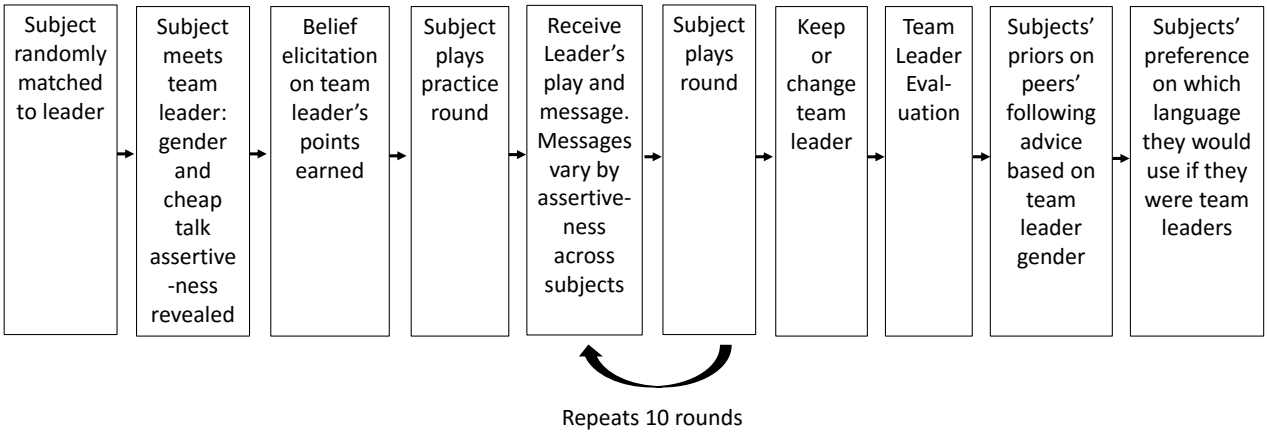



Figure 3: Primary Components of Experiment

| | ***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.

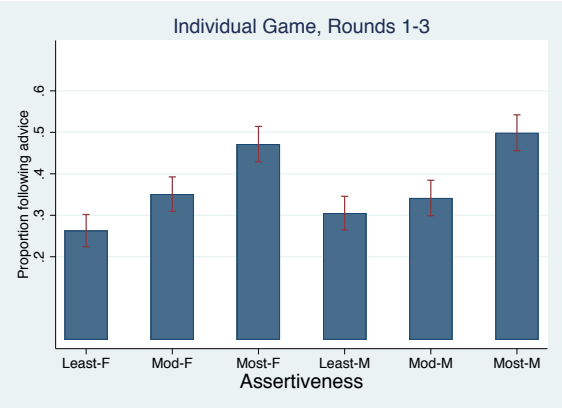
### 2.6 Surveying Experts

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

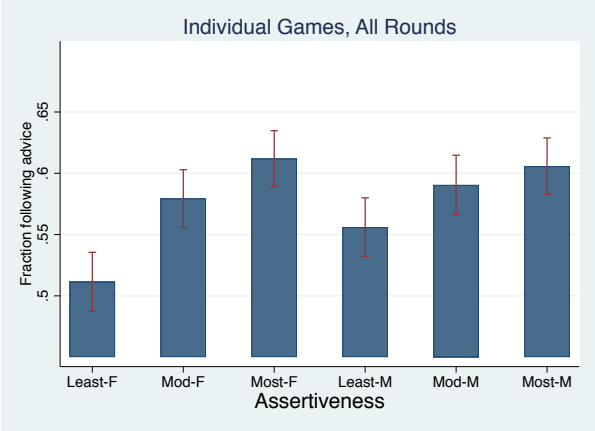
[18] 3 subjects did not report gender and 10 subjects did not report GPA in the original sample. 6 subjects did not report gender in the replication sample.

[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.

# Online Appendix

# A Experiment Details

## A.1 Timeline and Summary of Key Activities

We conducted an initial experiment with undergraduate students from the University of California Merced (435 subjects). The experiment was implemented as a self-guided online survey hosted on Qualtrics. This sample was 65 percent female with a mean age of 20 and mean grade point average of 3.0. To confirm the reproducibility of our results, we conduct a replication experiment on Amazon MTurk, an online platform that allows researchers to post small tasks (576 subjects). 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. SI Appendix Table A1 shows that attrition in the sample is balanced across treatment arms. Recent research has suggested that Amazon MTurk workers are 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). This replication sample was 38 percent female, with a mean age of 34 and mean level of education of college credit with no degree.[25] Additional summary statistics for both samples can be found in Appendix Table A2.

A timeline of activities is shown below. We have also included links to pre-registrations in the American Economics Association registry and to the Qualtrics surveys that were administered to subjects. After clicking on any of the links below, the Qualtrics surveys can be previewed without entering data by selecting "Ignore Validation" in the settings. Note that labeling of the subjects' choices and computer's response were randomly chosen from five different options (for example, some subjects selected numbers while others selected letters).

**March 2018:** Leader experiment at Washington State University

[25] 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.

**March 2018:** Validation of pseudonyms and messages at Washington State University

**April 2018:** Pre-registration of UC Merced Experiment

**April 2018:** UC Merced Experiment, Day 1, 36 subjects

**April 2018:** UC Merced Experiment, Day 2-3, 184 subjects

**April 2018:** UC Merced Experiment, Day 4-6, 215 subjects

**August 2018:** Pre-registration of MTurk Replication and Elicitation of Expert Priors

**August 2018:** MTurk Replication

**September 2018:** Elicitation of Expert Priors

### A.2 Validation of Treatment Variations

We used three methods to communicate the gender of the team leader: (1) directly stating it, (2) showing a gendered avatar, and (3) using a gendered pseudonym. And example of the gendered avatar is shown in Figure A1. 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. Gender perception was determined based on 74 students from Washington State University who reviewed 10 pseudonyms each on the gender, age, income, education, and race of the pseudonym. Subjects were given a list of ten names and the following directions.

*Use the following rubric in guessing the average characteristic of people who have the following names:*

**Age:** What is the average age for people with this name? Respond with any number.

**Gender:** What is the most likely gender of people with this name? Respond with F for female, M for male

**Income Ranking:** Where do you think people with this name fall in the distribution of income in the United States – in the bottom 25%, between 26 and 50% (i.e., lower middle), between 51 to 75% (upper middle), or the top 25

**Education:** What is the average years of schooling completed by people of this name? Respond with any number. (1 = 1st grade, 2 = 2nd grade, ..., 12 = completed high school, 14: associates degree, 16: bachelors degree, 18: masters degree, 20: PhD/Medical degree).

**Race:** What do you think is the most likely race/ethnicity of people with this name? Respond with WH: White Hispanic, WN: White Non-Hispanic, B: Black or African American, A: Asian, or O: other race.

### A.2.1 Cheap talk Validation

Prior to the experiments, 74 students at Washington State University were shown the "least assertive" ("most assertive") advice and "moderately assertive" advice for each round of the *Individual Game* (randomly ordered), and given the following instructions: "Please review the following pairs of messages. For each pair, **circle the message that you believe is more self-deprecating (self-promotional)**. For each pair, you will notice one message has more language in it – we're essentially asking you whether you think that language increases or decreases the self-deprecation (self-promotion) in the messages. There are no right or wrong answers." More than 70 percent of the messages were correctly identified as being "more self-promotional" (71 percent) or "more self-deprecating" (77 percent). These results suggest the scripts reflected assertiveness and alignment with gender norms as expected. The scripts are shown in Figure 5.

## A.3 Pre-registered analyses

All pre-registered analyses are shown in Tables A2-A17 below, with one exception. In the pre-analysis plan for the UCM experiment, we specified heterogeneity analyses based on exposure to a female supervisor and subject gender attitudes. Given the smaller than expected effect sizes in the experiment and the fact that there is almost no variation in exposure to a female supervisor (Table A2), we are underpowered to study heterogeneity and have therefore omitted these tables.

**Expert Priors**

We also pre-registered an analysis of how our results compared to expert priors. Given a large body of existing literature demonstrating backlash to assertive women, we studied whether our results align with the expectations of experts in gender research. Following DellaVigna and Pope (2018), we asked experts to provide their expectations for the estimated effects in the experiment. Fourteen anonymous researchers in economics and psychology provided expectations for the gender gap we would observe in each of the language treatments. In DellaVigna and Pope (2018), averaging over just five experts eliminated high-error forecasts, and the accuracy of the prediction made by 20 experts was indistinguishable from their full sample of 208 experts; therefore, we consider our sample of 14 experts sufficient.

Focusing on Round 1, experts were informed of the proportion of subjects who followed advice from a male leader for each type of cheap talk, and were asked about the average gender gap in advice adherence for the combined UCM and MTurk samples. On average, the experts in our sample predicted more discrimination against female leaders than was observed in the experiment. Across all cheap talk treatment arms, experts expected that having a female leader would reduce the likelihood of following advice by 5.52 percentage points, while the actual gender gap was 1.32 percentage points. However, these estimates were not statistically different.

This pattern was generally maintained when we disaggregated by cheap talk treatment. The mean priors for gender gaps in each cheap talk treatment arm were 6 percentage points for the least assertive (i.e., self-deprecating) cheap talk, 3 percentage points for moderately assertive (i.e., neutral) cheap talk, and 8 percentage points for most assertive (i.e., self-promoting) cheap talk. The actual gender gap was smaller than that predicted by experts in all but the least assertive self-deprecating treatment. However, the estimated gap was statistically different from the mean expert prior only in the neutral treatment.

The experts also predicted a widening of gender gaps in both the most and least assertive (i.e., self-promoting and self-deprecating) treatments, relative to the moderately assertive (i.e., neutral) treatment. While we cannot rule out that the gender gaps were identical across cheap talk treatments in the experiment, the pattern in the point estimates mirrors expert expectations. Thus, our results are consistent with existing expert priors on gender discrimination.

**Variable definitions**

Variables shown in the tables are defined as follows:

- Strategic in R1: Indicator for whether the subject played strategically (i.e., followed the leader's advice) in round 1 of the *Individual Game*.
- Mean Strategic, R1 – R3: Subject average of whether the subject played strategically (i.e., followed the leader's advice) in round 1 to round 3 of the *Individual Game* (range is 0 to 1).
- Mean Strategic, All: Subject average of whether the subject played strategically (i.e., followed the leader's advice) in all ten rounds of the *Individual Game* (range is 0 to 1).
- Expected Leader Performance: Subject's belief of the number of points earned by the team leader over all ten rounds in the *Individual Game*. Subjects were asked to select the ventile over the range of possible points (range is 1 to 20).
- Retained Leader: Indicator for whether the subject kept the team leader (as opposed to getting a new team leader) for a final incentivized game.
- Leader Evaluation: Summary index of the following 16 questions evaluating the team leader (range -2.19 to 1.3, increasing in positive evaluation of the team leader)(Anderson, 2008).
- Unless otherwise noted, subjects were asked to evaluate the team leader based on the following statements and using the following scale: 1: Disagree strongly, 2: Disagree, 3: Disagree a little, 4: Neither agree nor disagree, 5: Agree a little, 6: Agree, 7: Strongly Agree.
  - I would recommend *Pseudonym* as a colleague to others.
  - I would recommend *Pseudonym* as a supervisor to others.
  - *Pseudonym*'s explanations were clear.
  - I earned more points on the game because of *Pseudonym*.
  - *Pseudonym* made me feel confident in my ability to play the game.
  - I enjoyed working with *Pseudonym*.
  - How would you describe *Pseudonym*? Options: Not confident, Somewhat confident,

  - Appropriately confident, Overly Confident
  - *Pseudonym* was convincing in [his/her] messages.
  - *Pseudonym* treated me with respect.
  - *Pseudonym* would be approachable for an issue that bothered me.
  - *Pseudonym* would keep calm and have good judgement in pressured situations.
  - *Pseudonym* instilled a sense of teamwork.
  - *Pseudonym* has effective leadership skills.
  - *Pseudonym* was effective overall.
  - *Pseudonym* exaggerated the value of [his/her] advice.
  - *Pseudonym* has strong interpersonal skills.

- Message Recall: Number of messages (of four) correctly identified as having received or not received from their team leader in the *Individual Game* (range 0 to 4).

- Self-Power: Summary index of the following 8 questions evaluating their own feelings of power and control. For questions decreasing in power, the measure was reversed such that a higher number indicated greater self-power (range -2.75 to 1.54, increasing in self-power).(Anderson, 2008)

- Subjects were asked to evaluate the following statements using the following scale: 1 = Disagree strongly, 2 = Disagree, 3 = Disagree a little, 4 = Neither agree nor disagree, 5 = Agree a little, 6 = Agree, 7 = Strongly Agree

  - In my relationship with others, I can get people to listen to what I say.
  - In my relationship with others, my wishes do not carry much weight.
  - In my relationship with others, I can get others to do what I want.
  - In my relationship with others, even if I voice them, my views have little sway.
  - In my relationship with others, I think I have a great deal of power.
  - In my relationship with others, my ideas and opinions are often ignored.

  - In my relationship with others, even when I try, I am not able to get my way.
  - In my relationship with others, if I want to, I get to make decisions.

- Strategic in Coordination: Indicator for whether the subject played strategically (i.e., followed the leader's advice) in the *Coordination Game* (range 0 to 1).

- Strategic in Coordination – Strategic, R1: The difference between whether the subject played strategically (i.e., followed the leader's advice) in the *Coordination Game* and whether the subject played strategically (i.e., followed the leader's advice) in the *Individual Game*, Round 1 (range -1 to 1).

- Strategic in Coordination –Strategic, R10: The difference between whether the subject played strategically (i.e., followed the leader's advice) in the *Coordination Game* and whether the subject played strategically (i.e., followed the leader's advice) in the *Individual Game*, Round 10 (range -1 to 1).

- Selects Leader for Risky Puzzle: 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 (range 0 to 1).

- Strategic in Practice: Indicator for whether the subject played strategically in a practice round for the *Individual Game*, prior to observing the team leader's advice.

- Strategic Coordination in Practice: Indicator for whether the subject played strategically in a practice round for the *Coordination Game*, prior to observing the team leader's advice for the *Coordination Game*.

- Subject Prefers Most Assertive (Least Assertive): Indicator that subject selects 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 Moderately Assertive or Least Assertive (Most Assertive) treatment arm.

- Peers Prefer Male [Female] Advice for UCM subjects: Indicator for subject reporting that peers are more likely to follow advice provided by males [females] leaders in the *Individual*

*Game*, relative to being more likely to follow advice by the opposite gender or both genders equally (range 0 to 1).

- Peers Prefer Male [Female] Advice for MTurk subjects: Indicator for whether the subject reported more MTurk subjects following the team leader's advice in Round 1 of the *Individual Game* when provided by a male [female] team leader than when provided by a female [male] team leader (range 0 to 1).

- Gender Gap in Following Advice, R1: Difference between the 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 (range -100 to 100).

- Gender Gap in Following Advice, Coordination: Difference between the subject's expectation of the percentage of MTurk subjects following the advice in the *Coordination Game* when provided by a male team leader relative to when provided by a female team leader (range -100 to 100).

- Perceived Leader Confidence: Response to "How would you describe *Pseudonym*? Not Confident, Somewhat Confident, Appropriately Confident, Overly Confident" (range 1 to 4, increasing in the perceived team leader's confidence).

- Perceived Cheap talk Assertiveness: Response to "How would you rate *Pseudonym*'s communication style? Submissive/under assertive, Appropriately Assertive, Overly assertive/aggressive" (range 1 to 3, increasing in assertiveness).

- Accurate Leader Gender Recall: Indicator for whether the subject accurately recalled the team leader's gender (range 0 to 1).

- Characterized Messages as Masculine[Feminine]: Indicator for subject characterizing the language used in the team leader's messages for the *Individual Game* as more masculine [feminine], as opposed to more feminine[masculine] or gender-neutral (range 0 to 1).

- Female Subject: Indicator for subject self-reporting female (range 0 to 1).

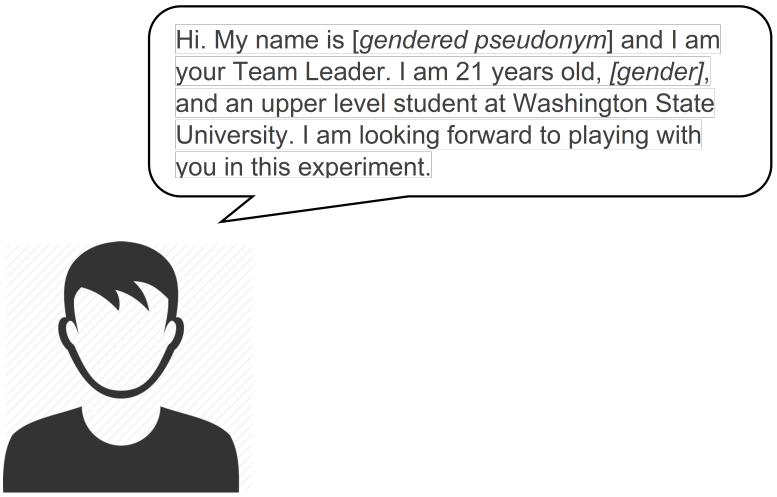



Figure A1: Gendered Image

- Subject's Age: Subject's self-reported age (range 18 to 41 for UCM and 19 to 84 for MTurk).
- Subject's GPA: Subject's self-reported grade point average.
- Subject's Education: Subjects self-reported educational attainment based on ten categories (range 1 to 10, increasing in educational attainment).
- Subject was Aware of Experiment: Indicator for whether the subject discussed, heard, or read information about the contents of this experiment prior to participation (range 0 to 1).
- Subject had Female Supervisor: Indicator for subject reporting ever had a female supervisor (range 0 to 1).
- Subject's Gender Attitude: Summary index of the following 10 questions measuring gender attitudes. For questions positive gender attitudes, the measure was reversed such that a higher number indicated more positive gender attitudes (range -1.37 to 1.15 increasing in self-power).(Anderson, 2008)

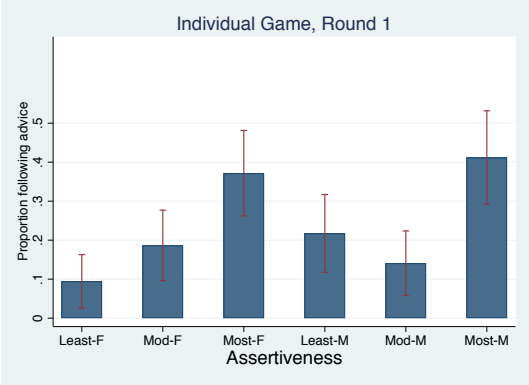



Figure A2: The Likelihood of Following Advice, *Individual Game*, Round 1, Original Experiment

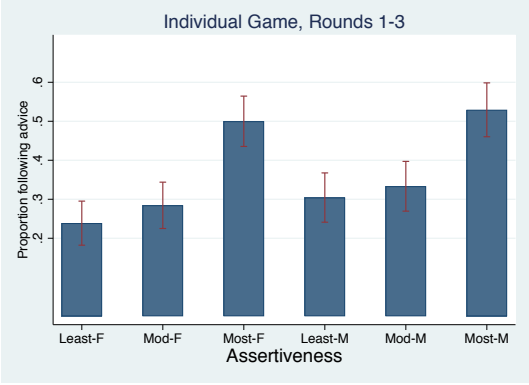



Figure A3: The Likelihood of Following Advice, *Individual Game*, Round 1 to 3, Original Experiment

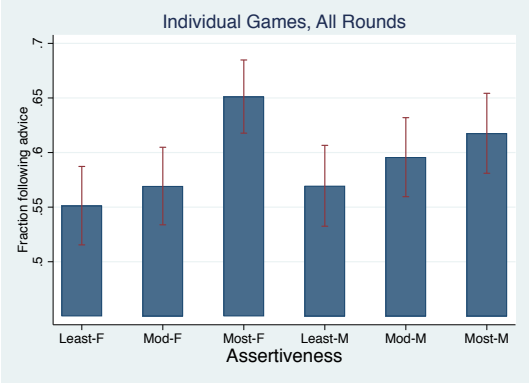



Figure A4: The Likelihood of Following Advice, *Individual Game*, All Rounds, Original Experiment

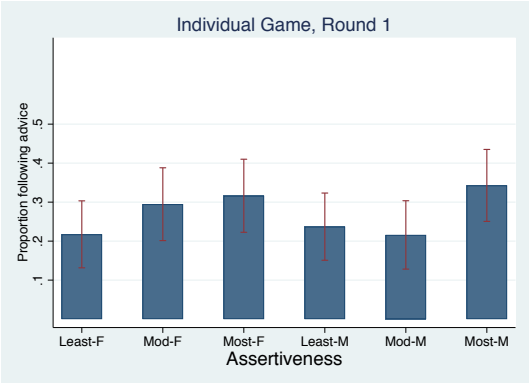



Figure A5: The Likelihood of Following Advice, *Individual Game*, Round 1, Replication Experiment

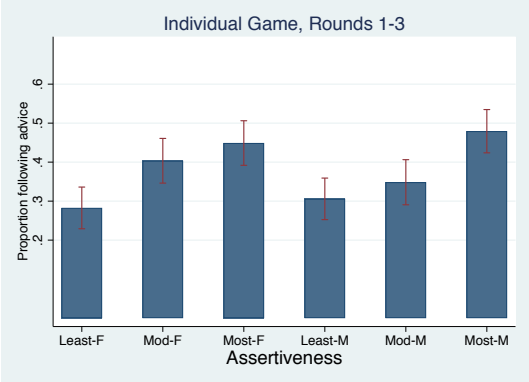



Figure A6: The Likelihood of Following Advice, *Individual Game*, Round 1 to 3, Replication Experiment

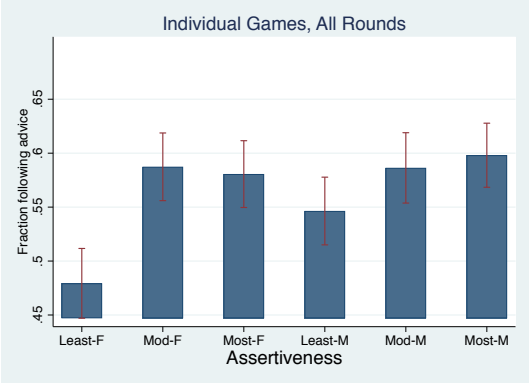



Figure A7: The Likelihood of Following Advice, *Individual Game*, All Rounds, Replication Experiment

Table A1: MTurk Attrition

| *Dependent Variable:* | Dropped |
|---|---|
| Fem. Leader, Least Assertive | 0.0207 |
| | (0.0559) |
| Male Leader, Mod. Assertive | 0.0592 |
| | (0.0561) |
| Fem. Leader, Mod. Assertive | 0.000574 |
| | (0.0556) |
| Male Leader, Most Assertive | -0.0349 |
| | (0.0543) |
| Fem Leader, Most Assertive | 0.0140 |
| | (0.0551) |
| Constant (Male, Least Assertive) | 0.362*** |
| | (0.0391) |
| Observations | 915 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Dropped is an indicator for any MTurk participant who was not included in the final analysis, due to not completing the survey, lack of consent, and suggestive evidence of having participated in the experiment earlier. 70 percent of those excluded stopped the experiment prior to observing any differences based on treatment.

Table A2: Summary Statistics

| | (1) Original (UCM) | (2) Replication (MTurk) | (3) Difference |
|---|---|---|---|
| Strategic in R1 | 0.24 | 0.27 | -0.04 |
| Mean Strategic, R1-R3 | 0.36 | 0.38 | -0.02 |
| Mean Strategic, All | 0.59 | 0.56 | 0.03 |
| Expected Leader Performance | 10.24 | 10.15 | 0.08 |
| Retained Leader | 0.93 | 0.89 | 0.03 |
| Leader Evaluation | -0.08 | 0.06 | -0.15*** |
| Message Recall | 3.17 | 2.79 | 0.37*** |
| Self Power | -0.01 | -0.00 | -0.00 |
| Strategic in Coordination | . | 0.82 | |
| Strategic in Coordination - Strategic, R1 | . | 0.55 | |
| Strategic in Coordination - Strategic, R10 | . | 0.09 | |
| Selects Leader for Risky Puzzle | 0.76 | . | |
| Strategic in Practice | 2.44 | 2.62 | -0.17** |
| Strategic Coordination in Practice | . | 0.38 | |
| Subject Prefers Deprecating | 0.11 | 0.10 | 0.01 |
| Subject Prefers Promoting | 0.37 | 0.30 | 0.06 |
| Peers Prefer Male Advice | 0.36 | 0.64 | -0.28*** |
| Peers Prefer Female Advice | 0.08 | 0.22 | -0.15*** |
| Gender Gap in Following Advice, R1 | . | 8.83 | |
| Gender Gap in Following Advice, Coordination | . | 7.50 | |
| Perceived Leader Confidence | 2.90 | 2.93 | -0.03 |
| Perceived Cheaptalk Assertiveness | 1.92 | 1.97 | -0.05 |
| Characterized Messages as Masculine | 0.09 | 0.26 | -0.16*** |
| Characterized Messages as Feminine | 0.09 | 0.23 | -0.14*** |
| Accurate Leader Gender Recall | 0.95 | 0.92 | 0.03 |
| Female Subject | 0.65 | 0.38 | 0.27*** |
| Subject's Age | 19.77 | 34.01 | -14.24*** |
| Subject's GPA | 3.01 | . | |
| Subject's Education | . | 5.99 | |
| Subject was Aware of Experiment | . | 0.09 | |
| Subject had Female Supervisor | 0.89 | 0.85 | 0.05* |
| Subject's Gender Attitude | -0.00 | . | |
| Observations | 435 | 575 | 1010 |

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. Columns report sample means for each variable. Variables not reported were not asked in the respected experiment. Message recall was not collected for the first 35 subjects and Peers Prefer Male(Female) Advice was collected for the latter 215 subjects at UCM. Among UCM subjects 2.3 percent did not report their GPA and 3 percent did not report Perceived Cheaptalk Assertiveness. 11 percent of MTurk subjects and 13 percent of UCM subjects did not respond for Subject Had Female Supervisor. For all remaining variables, non response was less than 2 percent in each sample. Variable definitions are provided in the Appendix.

Table A3: Balance on Subject Characteristics, Main Specification

| | (1) Female | (2) Age | (3) GPA | (4) Education | (5) Strategic Practice |
|---|---|---|---|---|---|
| Female Leader | -0.0350 | 0.391 | 0.154** | -0.0936 | 0.00885 |
| | (0.0504) | (0.999) | (0.0763) | (0.263) | (0.0248) |
| Assertive | -0.00906 | -0.101 | 0.0524 | 0.0918 | 0.0104 |
| | (0.0273) | (0.519) | (0.0448) | (0.128) | (0.0139) |
| Female Leader × Assertive | 0.00802 | -0.0846 | -0.0866 | -0.0515 | 0.00476 |
| | (0.0386) | (0.756) | (0.0628) | (0.192) | (0.0203) |
| Observations | 1001 | 1010 | 425 | 574 | 1010 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Assertive is a variable ranging from 0 to 2, increasing in the assertiveness of the cheap talk. Columns 1, 2, and 5 include observations from the original sample at UC Merced and the replication sample on Amazon MTurk. Column 3 is limited to subjects from the original UC Merced sample and Column 4 to subjects from the MTurk replication sample. Female is an indicator for the subject being female. Age, GPA (grade point average), and Education are self-reported responses of the subject. Education ranges from 1 to 10, increasing in the completion of educational degrees. Strategic Practice is an indicator for whether the subject selected to play strategically in the practice round, prior to receiving any advice. Subjects that did not respond to characteristics are omitted.

Table A4: Balance on Subject Characteristics, by Treatment

| | (1) Female | (2) Age | (3) GPA | (4) Education | (5) Strategic Practice |
|---|---|---|---|---|---|
| Least Assertive-Female | 0.00595 | -0.0904 | 0.146* | 0.0152 | 0.00602 |
| | (0.0552) | (1.081) | (0.0833) | (0.291) | (0.0268) |
| Mod. Assertive-Male | 0.0880 | 0.0345 | 0.0158 | 0.308 | 0.00265 |
| | (0.0554) | (1.050) | (0.0758) | (0.274) | (0.0268) |
| Mod. Assertive-Female | -0.0237 | 1.276 | 0.0965 | -0.0653 | 0.0221 |
| | (0.0548) | (1.187) | (0.0896) | (0.278) | (0.0281) |
| Most Assertive-Male | -0.0170 | -0.199 | 0.105 | 0.189 | 0.0207 |
| | (0.0545) | (1.039) | (0.0897) | (0.257) | (0.0278) |
| Most Assertive-Female | 0.00338 | -0.431 | 0.0777 | 0.0929 | 0.0363 |
| | (0.0545) | (1.081) | (0.0877) | (0.264) | (0.0290) |
| Constant (Least Assertive-Male) | 0.488*** | 27.78*** | 2.940*** | 5.897*** | 0.0602*** |
| | (0.0389) | (0.751) | (0.0585) | (0.190) | (0.0185) |
| Observations | 1001 | 1010 | 425 | 574 | 1010 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Columns 1, 2, and 5 include observations from the original sample at UC Merced and the replication sample on Amazon MTurk. Column 3 is limited to subjects from the original UC Merced sample and Column 4 to subjects from the MTurk replication sample. Female is an indicator for the subject being female. Age, GPA (grade point average), and Education are self-reported responses of the subject. Education ranges from 1 to 10, increasing in the completion of educational degrees. Strategic Practice is an indicator for whether the subject selected to play strategically in the practice round, prior to receiving any advice. Subjects that did not respond to characteristics are omitted.

Table A5: Do Subject Characteristics Predict Following Advice?

| *Dependent Variable:* | Strategic Play | | |
|---|---|---|---|
| | (1) UCM | (2) MTurk | (3) Combined |
| Female | 0.0418* | 0.0494** | 0.0449** |
| | (0.0244) | (0.0250) | (0.0176) |
| Age | 0.00250 | 0.00222* | 0.00223* |
| | (0.00521) | (0.00119) | (0.00117) |
| GPA | -0.0228 | | |
| | (0.0219) | | |
| Education | | -0.0177*** | |
| | | (0.00621) | |
| Practice Round | -0.0219 | 0.148*** | 0.0628* |
| | (0.0410) | (0.0537) | (0.0348) |
| Round FE | X | X | X |
| Sample FE | | | X |
| Observations | 4230 | 5690 | 10010 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the individual level. 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 Combined refers to both samples combined. Female is an indicator for the subject being female. Age, GPA (grade point average), and Education are self-reported responses of the subject. Education ranges from 1 to 10, increasing in the completion of educational degrees. Practice Round is an indicator for playing strategically in practice round. GPA was asked only for the UCM sample, Education was asked only for the MTurk sample. Subjects that did not respond to characteristics are omitted.

Table A6: Gender Gaps on Average and Conditional on Assertiveness of Cheap Talk (Round 1 and All Rounds)

| *Dependent Variable:* | Strategic Play | | | | | |
|---|---|---|---|---|---|---|
| | Round 1 | | | All Rounds | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | UCM | MTurk | All | UCM | MTurk | All |
| *Panel A: Full Sample* | | | | | | |
| Fem. Leader | -0.0440 | 0.00914 | -0.0147 | -0.00227 | -0.0268 | -0.0174 |
| | (0.0403) | (0.0363) | (0.0270) | (0.0223) | (0.0248) | (0.0171) |
| Observations | 435 | 575 | 1010 | 4350 | 5750 | 10100 |
| Male Mean | 0.255 | 0.269 | 0.263 | 0.594 | 0.577 | 0.584 |
| *Panel B: Most Assertive Cheap Talk* | | | | | | |
| Fem. Leader | -0.0575 | -0.0182 | -0.0372 | 0.0325 | -0.0147 | 0.00279 |
| | (0.0791) | (0.0642) | (0.0498) | (0.0429) | (0.0440) | (0.0313) |
| Observations | 146 | 203 | 349 | 1460 | 2030 | 3490 |
| Male Mean | 0.412 | 0.343 | 0.370 | 0.618 | 0.598 | 0.606 |
| *Panel C: Moderately Assertive Cheap Talk* | | | | | | |
| Fem. Leader | 0.0393 | 0.0781 | 0.0611 | -0.0237 | 0.00101 | -0.0106 |
| | (0.0615) | (0.0643) | (0.0449) | (0.0340) | (0.0423) | (0.0279) |
| Observations | 146 | 183 | 329 | 1460 | 1830 | 3290 |
| Male Mean | 0.141 | 0.216 | 0.182 | 0.596 | 0.586 | 0.591 |
| *Panel D: Least Assertive Cheap Talk* | | | | | | |
| Fem. Leader | -0.124** | -0.0216 | -0.0662 | -0.0176 | -0.0679* | -0.0467 |
| | (0.0609) | (0.0582) | (0.0426) | (0.0377) | (0.0407) | (0.0285) |
| Round FE | | | | X | X | X |
| Practice round | X | X | X | X | X | X |
| Sample FE | | | X | | | X |
| Observations | 143 | 189 | 332 | 1430 | 1890 | 3320 |
| Male Mean | 0.217 | 0.237 | 0.229 | 0.570 | 0.546 | 0.556 |

Table A7: Following 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 | 0.0370 | 0.0770 | 0.0581 | -0.0517 | 0.0538 | 0.00581 | -0.0256 | 0.000381 | -0.0122 |
| | (0.0613) | (0.0649) | (0.0451) | (0.0482) | (0.0555) | (0.0376) | (0.0339) | (0.0425) | (0.0280) |
| Most Assertive | 0.257*** | 0.132** | 0.180*** | 0.191*** | 0.134** | 0.153*** | 0.0233 | 0.0135 | 0.0157 |
| | (0.0717) | (0.0626) | (0.0469) | (0.0508) | (0.0525) | (0.0372) | (0.0391) | (0.0419) | (0.0294) |
| Least Assertive | 0.0729 | 0.0282 | 0.0462 | -0.0302 | -0.0382 | -0.0365 | -0.0258 | -0.0376 | -0.0335 |
| | (0.0644) | (0.0593) | (0.0436) | (0.0470) | (0.0495) | (0.0348) | (0.0371) | (0.0400) | (0.0279) |
| Fem. Leader × Most | -0.0889 | -0.0973 | -0.0903 | 0.0185 | -0.0803 | -0.0352 | 0.0604 | -0.0158 | 0.0159 |
| | (0.100) | (0.0913) | (0.0673) | (0.0704) | (0.0770) | (0.0534) | (0.0543) | (0.0610) | (0.0419) |
| Fem. Leader × Least | -0.162* | -0.0980 | -0.125** | -0.0146 | -0.0779 | -0.0485 | 0.00757 | -0.0679 | -0.0336 |
| | (0.0863) | (0.0873) | (0.0619) | (0.0656) | (0.0727) | (0.0501) | (0.0506) | (0.0589) | (0.0400) |
| Round FE | | | | X | X | X | X | X | X |
| Practice round | X | X | X | X | X | X | X | X | X |
| Sample FE | | | X | | | X | | | X |
| Observations | 435 | 575 | 1010 | 1305 | 1725 | 3030 | 4350 | 5750 | 10100 |
| Most + F × Most | 0.168 | 0.0348 | 0.0893 | 0.210 | 0.0539 | 0.118 | 0.0837 | -0.00232 | 0.0316 |
| P-val | 0.018 | 0.600 | 0.065 | 0.000 | 0.338 | 0.002 | 0.027 | 0.958 | 0.292 |
| Least + F × Least | -0.0892 | -0.0698 | -0.0790 | -0.0448 | -0.116 | -0.0851 | -0.0183 | -0.105 | -0.0671 |
| P-val | 0.121 | 0.276 | 0.073 | 0.330 | 0.030 | 0.019 | 0.596 | 0.015 | 0.019 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the individual level. 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.

Table A8: Assertive Cheap Talk and Gender Gaps, Male Subjects Only

| | 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 | -0.0768 | -0.0234 | -0.0464 | -0.0785 | -0.0249 | -0.0473 | -0.0663 | -0.0425 | -0.0478 |
| | (0.0899) | (0.0701) | (0.0552) | (0.0695) | (0.0563) | (0.0444) | (0.0608) | (0.0493) | (0.0388) |
| Assertive | 0.122* | 0.0440 | 0.0621* | 0.150*** | 0.0728** | 0.0886*** | 0.00121 | 0.0384 | 0.0287 |
| | (0.0689) | (0.0382) | (0.0333) | (0.0476) | (0.0313) | (0.0265) | (0.0423) | (0.0266) | (0.0226) |
| Fem. Leader × Assertive | 0.0430 | 0.00909 | 0.0272 | 0.0323 | 0.0233 | 0.0356 | 0.0802 | 0.0264 | 0.0407 |
| | (0.0864) | (0.0545) | (0.0455) | (0.0615) | (0.0445) | (0.0365) | (0.0546) | (0.0381) | (0.0314) |
| 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 |
| Observations | 150 | 354 | 504 | 450 | 1062 | 1512 | 1500 | 3540 | 5040 |
| A + F × A | 0.165 | 0.0531 | 0.0893 | 0.182 | 0.0961 | 0.124 | 0.0814 | 0.0647 | 0.0693 |
| P-val | 0.002 | 0.171 | 0.004 | 0.000 | 0.002 | 0.000 | 0.020 | 0.018 | 0.002 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the individual level. Assertive is a variable ranging from 0 to 2, increasing in the assertiveness of the cheaptalk. 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. Observations include only male subjects.

Table A9: Assertive Cheap Talk and Willingness to Follow Advice: Controlling for Subject Characteristics

| | 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 | -0.0946 | -0.00297 | -0.0378 | -0.0565 | 0.00430 | -0.0237 | -0.0236 | -0.0509 | -0.0408 |
| | (0.0609) | (0.0562) | (0.0405) | (0.0437) | (0.0456) | (0.0318) | (0.0351) | (0.0363) | (0.0263) |
| Assertive | 0.0879** | 0.0579* | 0.0699*** | 0.110*** | 0.0940*** | 0.0997*** | 0.0189 | 0.0363* | 0.0285* |
| | (0.0396) | (0.0304) | (0.0239) | (0.0251) | (0.0249) | (0.0180) | (0.0208) | (0.0200) | (0.0150) |
| Fem. Leader × Assertive | 0.0452 | -0.00279 | 0.0152 | 0.0178 | -0.00730 | 0.00311 | 0.0321 | 0.0150 | 0.0221 |
| | (0.0519) | (0.0438) | (0.0330) | (0.0349) | (0.0358) | (0.0254) | (0.0288) | (0.0286) | (0.0211) |
| Covariates | X | X | X | X | X | X | X | X | X |
| 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 |
| Observations | 423 | 569 | 1001 | 1269 | 1707 | 3003 | 4230 | 5690 | 10010 |
| A + F × A | 0.133 | 0.0551 | 0.0851 | 0.128 | 0.0867 | 0.103 | 0.0510 | 0.0513 | 0.0506 |
| P-val | 0.000 | 0.082 | 0.000 | 0.000 | 0.001 | 0.000 | 0.011 | 0.013 | 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 1 to 3, increasing in the assertiveness of the cheaptalk. Covariates include subject's gender and age, and team leader's pseudonym's perceived age, income, education, and indicators for each race. GPA and Education are included as covariates in UCM and MTurk sample, respectively. 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.

Table A10: Subjective Evaluations of the Team Leader, Original Sample (UCM)

| | (1) Exp. Perf. | (2) Retained | (3) Eval. | (4) Recall | (5) Self-Power | (6) Selects |
|---|---|---|---|---|---|---|
| Fem. Leader | 0.510 | 0.00451 | 0.0399 | -0.149 | -0.149* | 0.0817 |
| | (0.794) | (0.0458) | (0.0563) | (0.113) | (0.0826) | (0.0711) |
| Most Assertive | 0.361 | 0.0257 | -0.0964 | -0.502*** | 0.0508 | 0.0464 |
| | (0.824) | (0.0439) | (0.0654) | (0.140) | (0.0962) | (0.0747) |
| Least Assertive | -0.374 | 0.0120 | -0.301*** | 0.272** | 0.0999 | 0.00633 |
| | (0.774) | (0.0457) | (0.0713) | (0.106) | (0.0935) | (0.0763) |
| Fem. Leader × Most Assertive | -0.896 | -0.0354 | 0.0585 | 0.190 | 0.0707 | -0.0387 |
| | (1.172) | (0.0631) | (0.0907) | (0.201) | (0.125) | (0.0988) |
| Fem. Leader × Least Assertive | -0.305 | 0.0139 | 0.0227 | -0.00927 | 0.171 | -0.0496 |
| | (1.124) | (0.0615) | (0.100) | (0.165) | (0.123) | (0.103) |
| Constant (Male-Mod. Assertive) | 10.18*** | 0.915*** | 0.0117 | 3.292*** | -0.0191 | 0.718*** |
| | (0.518) | (0.0332) | (0.0409) | (0.0684) | (0.0675) | (0.0538) |
| Observations | 431 | 435 | 435 | 399 | 435 | 435 |
| Most + F × Most | -0.535 | -0.00974 | -0.0380 | -0.312 | 0.122 | 0.00769 |
| P-val | 0.521 | 0.830 | 0.546 | 0.031 | 0.125 | 0.905 |
| Least + F × Least | -0.680 | 0.0259 | -0.278 | 0.263 | 0.271 | -0.0432 |
| P-val | 0.404 | 0.529 | 0.000 | 0.037 | 0.001 | 0.528 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Dependent variables reflect the subject's asseessment of the team leader. Exp. Perf 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). Recall is the number of messages (of four) correctly identified as having received or not received from their team leader in the Individual Game (Message Recall). 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).

Table A11: Subjective Evaluations of the Team Leader, Replication Sample (Mturk)

| | (1) Exp. Perf. | (2) Retained | (3) Eval. | (4) Recall | (5) Self-Power |
|---|---|---|---|---|---|
| Fem. Leader | 1.195 | 0.0286 | 0.0302 | 0.224 | -0.0204 |
| | (0.734) | (0.0422) | (0.0626) | (0.144) | (0.111) |
| Most Assertive | 0.277 | 0.00703 | -0.143** | -0.247* | -0.0499 |
| | (0.747) | (0.0434) | (0.0642) | (0.146) | (0.104) |
| Least Assertive | 0.900 | -0.0214 | -0.338*** | 0.141 | -0.0739 |
| | (0.708) | (0.0467) | (0.0687) | (0.154) | (0.112) |
| Fem. Leader × Most Assertive | -0.531 | -0.0150 | 0.0311 | -0.387* | 0.0560 |
| | (1.019) | (0.0582) | (0.0907) | (0.205) | (0.148) |
| Fem. Leader × Least Assertive | -2.383** | -0.0788 | -0.0257 | -0.183 | -0.0339 |
| | (1.008) | (0.0670) | (0.100) | (0.210) | (0.156) |
| Constant (Male-Neutral) | 9.636*** | 0.898*** | 0.209*** | 2.818*** | 0.0444 |
| | (0.548) | (0.0325) | (0.0434) | (0.107) | (0.0790) |
| Observations | 570 | 575 | 574 | 575 | 574 |
| Most + F × Most | -0.254 | -0.00795 | -0.112 | -0.634 | 0.00611 |
| P-val | 0.714 | 0.837 | 0.081 | 0.000 | 0.954 |
| Least + F × Least | -1.483 | -0.100 | -0.364 | -0.0421 | -0.108 |
| P-val | 0.039 | 0.037 | 0.000 | 0.769 | 0.322 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Dependent variables reflect the subject's asseessment of the team leader. Exp. Perf 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). Recall is the number of messages (of four) correctly identified as having received or not received from their team leader in the Individual Game (Message Recall). 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).

Table A12: Underlying Evaluations, Part 1, Original Sample (UCM)

| | (1) Col. | (2) Super | (3) Clear | (4) Earned | (5) Able | (6) Enjoy | (7) Confidence | (8) Convince |
|---|---|---|---|---|---|---|---|---|
| Fem. Leader | -0.177 | 0.00386 | -0.157 | -0.0972 | -0.0905 | -0.0687 | 0.122** | 0.0402 |
| | (0.173) | (0.180) | (0.191) | (0.244) | (0.251) | (0.210) | (0.0560) | (0.196) |
| Most Assertive | -0.207 | -0.203 | -0.377 | -0.280 | -0.137 | -0.143 | 0.366*** | -0.173 |
| | (0.199) | (0.213) | (0.232) | (0.273) | (0.254) | (0.215) | (0.0770) | (0.219) |
| Fem. Leader × Most Assertive | 0.471* | 0.212 | 0.223 | 0.420 | 0.251 | 0.148 | -0.0951 | 0.0401 |
| | (0.254) | (0.273) | (0.310) | (0.369) | (0.345) | (0.295) | (0.105) | (0.279) |
| Least Assertive | -0.205 | -0.433** | -0.559** | -0.0682 | -0.243 | -0.300 | -0.629*** | -0.920*** |
| | (0.193) | (0.196) | (0.220) | (0.254) | (0.260) | (0.216) | (0.107) | (0.250) |
| Fem. Leader × Least Assertive | 0.421* | 0.0902 | -0.0384 | 0.0729 | -0.0547 | 0.205 | -0.113 | -0.299 |
| | (0.255) | (0.274) | (0.315) | (0.351) | (0.362) | (0.286) | (0.137) | (0.334) |
| Observations | 432 | 427 | 422 | 422 | 423 | 419 | 432 | 435 |
| Most + F × Most | 0.264 | 0.00878 | -0.154 | 0.139 | 0.114 | 0.00512 | 0.271 | -0.133 |
| P-val | 0.095 | 0.959 | 0.455 | 0.574 | 0.626 | 0.980 | 0.000 | 0.445 |
| Least + F × Least | 0.216 | -0.342 | -0.597 | 0.00476 | -0.298 | -0.0946 | -0.742 | -1.219 |
| P-val | 0.195 | 0.075 | 0.008 | 0.984 | 0.238 | 0.614 | 0.000 | 0.000 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the individual level. Depedent variables refer to the first seven questions underlying the Leader Evaluation Summary Index. All measures increase in positive evaluations of the team leader.See Variable Appendix for definitions.

Table A13: Underlying Evaluations, Part 2, Original Sample (UCM)

| | (1) Respect | (2) Approach | (3) Calm | (4) Teamwork | (5) Leadership | (6) Effective | (7) Exaggerate | (8) Interpersonal |
|---|---|---|---|---|---|---|---|---|
| Fem. Leader | 0.113 | 0.0685 | 0.0610 | 0.132 | 0.0286 | -0.00190 | -0.161 | -0.0757 |
| | (0.141) | (0.217) | (0.206) | (0.225) | (0.205) | (0.181) | (0.266) | (0.199) |
| Most Assertive | -0.156 | -0.229 | -0.0534 | -0.282 | -0.116 | -0.0580 | -1.237*** | 0.110 |
| | (0.177) | (0.237) | (0.221) | (0.260) | (0.233) | (0.190) | (0.277) | (0.227) |
| Fem. Leader × Most Assertive | -0.136 | -0.0602 | -0.203 | -0.239 | 0.131 | 0.0443 | 0.704* | -0.0114 |
| | (0.227) | (0.322) | (0.308) | (0.364) | (0.317) | (0.273) | (0.391) | (0.299) |
| Least Assertive | 0.0644 | -0.0383 | -0.185 | -0.00252 | -0.600** | -0.434** | -0.341 | -0.404* |
| | (0.162) | (0.231) | (0.225) | (0.238) | (0.237) | (0.219) | (0.284) | (0.222) |
| Fem. Leader × Least Assertive | -0.291 | -0.0220 | -0.116 | -0.270 | -0.202 | 0.0699 | 0.685* | 0.0272 |
| | (0.221) | (0.318) | (0.316) | (0.325) | (0.337) | (0.297) | (0.394) | (0.300) |
| Observations | 433 | 434 | 434 | 431 | 434 | 433 | 431 | 434 |
| Most + F × Most | -0.292 | -0.290 | -0.256 | -0.521 | 0.0154 | -0.0137 | -0.533 | 0.0990 |
| P-val | 0.042 | 0.183 | 0.231 | 0.042 | 0.943 | 0.945 | 0.055 | 0.613 |
| Least + F × Least | -0.226 | -0.0604 | -0.302 | -0.273 | -0.803 | -0.365 | 0.344 | -0.377 |
| P-val | 0.135 | 0.783 | 0.175 | 0.220 | 0.001 | 0.070 | 0.209 | 0.062 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the individual level. Dependent variables refer to the first seven questions underlying the Leader Evaluation Summary Index. All measures increase in positive evaluations of the team leader.See Variable Appendix for definitions.

Table A14: Underlying Evaluations, Part 1, Replication Sample (MTurk)

| | (1) Col. | (2) Super | (3) Clear | (4) Earned | (5) Able | (6) Enjoy | (7) Confidence | (8) Convince |
|---|---|---|---|---|---|---|---|---|
| Fem. Leader | 0.261* | 0.158 | -0.101 | 0.0876 | 0.116 | 0.128 | -0.0681 | -0.0988 |
| | (0.150) | (0.189) | (0.171) | (0.210) | (0.191) | (0.177) | (0.0667) | (0.160) |
| Most Assertive | 0.0661 | 0.0812 | -0.172 | -0.243 | -0.0815 | -0.119 | 0.0565 | -0.105 |
| | (0.158) | (0.174) | (0.151) | (0.226) | (0.191) | (0.182) | (0.0714) | (0.166) |
| Fem. Leader × Most Assertive | -0.0312 | -0.00404 | 0.0896 | 0.0713 | 0.0815 | 0.0138 | 0.0465 | 0.194 |
| | (0.220) | (0.252) | (0.241) | (0.317) | (0.277) | (0.261) | (0.102) | (0.232) |
| Least Assertive | -0.120 | -0.363* | -0.678*** | -0.308 | -0.362* | -0.155 | -0.741*** | -0.763*** |
| | (0.157) | (0.191) | (0.165) | (0.215) | (0.189) | (0.177) | (0.0976) | (0.185) |
| Fem. Leader × Least Assertive | -0.164 | -0.118 | 0.445* | -0.174 | -0.0317 | -0.0869 | -0.0274 | -0.234 |
| | (0.228) | (0.285) | (0.247) | (0.331) | (0.300) | (0.268) | (0.149) | (0.291) |
| Observations | 559 | 544 | 531 | 518 | 534 | 522 | 571 | 560 |
| Most + F × Most | 0.0349 | 0.0771 | -0.0828 | -0.172 | -9.71e-17 | -0.106 | 0.103 | 0.0888 |
| P-val | 0.820 | 0.671 | 0.660 | 0.439 | 1.000 | 0.572 | 0.159 | 0.583 |
| Least + F × Least | -0.284 | -0.480 | -0.234 | -0.481 | -0.394 | -0.242 | -0.768 | -0.997 |
| P-val | 0.088 | 0.024 | 0.205 | 0.056 | 0.091 | 0.230 | 0.000 | 0.000 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the individual level. Dependent variables refer to the first seven questions underlying the Leader Evaluation Summary Index. All measures increase in positive evaluations of the team leader.See Variable Appendix for definitions.

Table A15: Underlying Evaluations, Part 2, Replication Sample (MTurk)

| | (1) Respect | (2) Approach | (3) Calm | (4) Teamwork | (5) Leadership | (6) Effective | (7) Exaggerate | (8) Interpersonal |
|---|---|---|---|---|---|---|---|---|
| Fem. Leader | -0.0983 | 0.00445 | 0.119 | -0.127 | 0.154 | 0.123 | 0.503* | 0.0398 |
| | (0.151) | (0.165) | (0.164) | (0.198) | (0.165) | (0.149) | (0.296) | (0.173) |
| Most Assertive | -0.257* | -0.234 | -0.186 | -0.192 | -0.0657 | -0.0789 | -0.567* | -0.206 |
| | (0.153) | (0.178) | (0.172) | (0.195) | (0.167) | (0.160) | (0.303) | (0.186) |
| Fem. Leader × Most Assertive | 0.319 | 0.279 | 0.0785 | 0.435 | 0.0791 | -0.0538 | -0.585 | 0.261 |
| | (0.218) | (0.252) | (0.234) | (0.277) | (0.239) | (0.227) | (0.425) | (0.262) |
| Least Assertive | -0.132 | -0.104 | -0.445** | -0.211 | -0.662*** | -0.483*** | 0.239 | -0.694*** |
| | (0.149) | (0.166) | (0.183) | (0.192) | (0.190) | (0.159) | (0.305) | (0.191) |
| Fem. Leader × Least Assertive | 0.377* | 0.237 | 0.102 | 0.00792 | -0.158 | -0.187 | -0.509 | 0.106 |
| | (0.210) | (0.239) | (0.263) | (0.290) | (0.284) | (0.261) | (0.433) | (0.274) |
| Observations | 547 | 549 | 554 | 556 | 551 | 542 | 556 | 555 |
| Most + F × Most | 0.0622 | 0.0441 | -0.108 | 0.243 | 0.0134 | -0.133 | -1.152 | 0.0551 |
| P-val | 0.689 | 0.805 | 0.497 | 0.217 | 0.937 | 0.408 | 0.000 | 0.765 |
| Least + F × Least | 0.245 | 0.133 | -0.342 | -0.203 | -0.820 | -0.670 | -0.270 | -0.588 |
| P-val | 0.097 | 0.440 | 0.069 | 0.351 | 0.000 | 0.001 | 0.379 | 0.003 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the individual level. Dependent variables refer to the first seven questions underlying the Leader Evaluation Summary Index. All measures increase in positive evaluations of the team leader.See Variable Appendix for definitions.

Table A16: Following Advice Under Coordination

| | (1) Strategic | (2) Strategic - R1 | (3) Strategic - R10 |
|---|---|---|---|
| Fem. Leader | -0.000922 | -0.0805 | 0.0240 |
| | (0.0509) | (0.0789) | (0.0723) |
| Most Assertive | -0.0173 | -0.146* | 0.0119 |
| | (0.0508) | (0.0787) | (0.0712) |
| Least Assertive | -0.134** | -0.159* | -0.123 |
| | (0.0581) | (0.0851) | (0.0822) |
| Fem. Leader × Most Assertive | -0.0282 | 0.0815 | -0.0258 |
| | (0.0736) | (0.109) | (0.104) |
| Fem. Leader × Least Assertive | 0.0640 | 0.166 | 0.0946 |
| | (0.0805) | (0.114) | (0.106) |
| Constant(Neutral-Male) | 0.873*** | 0.672*** | 0.103* |
| | (0.0378) | (0.0640) | (0.0588) |
| Practice round | X | X | X |
| Observations | 575 | 575 | 575 |
| Most + F × Most | -0.0455 | -0.0649 | -0.0139 |
| P-val | 0.392 | 0.386 | 0.853 |
| Least + F × Least | -0.0699 | 0.00713 | -0.0285 |
| P-val | 0.206 | 0.925 | 0.670 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. Practice Round is an indicator for playing strategically in practice round. Strategic is an indicator if the subject followed the team leader's advice (Strategic in Coordination). Strategic - R1 (Strategic - R10) is the difference between whether a subject followed the advice relative to if the subject followed the advice in Round 1 (Round 10) of the Individual Game (Strategic in Coordination - Strategic, R1(R10)).

Table A17: Following Assertive Advice by Subject Gender

| | Round 1 | | | All Rounds | | |
|---|---|---|---|---|---|---|
| | (1) UCM | (2) MTurk | (3) All | (4) UCM | (5) MTurk | (6) All |
| Fem. Subj | 0.0159 | -0.0348 | -0.0274 | 0.0558 | 0.0792** | 0.0690*** |
| | (0.0574) | (0.0556) | (0.0398) | (0.0362) | (0.0382) | (0.0262) |
| Assertive | 0.147*** | 0.0486* | 0.0750*** | 0.0465* | 0.0507*** | 0.0482*** |
| | (0.0426) | (0.0271) | (0.0229) | (0.0271) | (0.0190) | (0.0158) |
| Fem. Subj × Assertive | -0.0525 | 0.0152 | 0.00411 | -0.0163 | -0.0225 | -0.0209 |
| | (0.0527) | (0.0454) | (0.0331) | (0.0318) | (0.0311) | (0.0212) |
| Round FE | X | X | X | X | X | X |
| Practice round | X | X | X | X | X | X |
| Sample FE | | | X | | | X |
| Observations | 432 | 569 | 1001 | 4320 | 5690 | 10010 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses, clustered at the individual level. Fem. Subj. is an indicator for a female subject. 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.

Table A18: Performance on *Individual Game* by Subject Gender

| | Strategic Play in Practice Round | | | Strategic Play in Round 1 to 10 | | |
|---|---|---|---|---|---|---|
| | (1) UCM | (2) MTurk | (3) All | (4) UCM | (5) MTurk | (6) All |
| Female Subject | 0.00865 | -0.0205 | -0.00815 | 0.0451* | 0.0489* | 0.0473*** |
| | (0.0279) | (0.0211) | (0.0170) | (0.0245) | (0.0251) | (0.0178) |
| Round FE | | | | X | X | X |
| Sample FE | | | X | | | X |
| Observations | 432 | 569 | 1001 | 4320 | 5690 | 10010 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses clustered at the individual level. Strategic Play in Practice Round is an indicator for playing strategically in practice round prior to receiving any advice. 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.

# Leader Survey [UC Merced]

*Different pages in the online survey are indicated by ----------*
*Images are generally only displayed once and noted accordingly.*

Before we begin this experiment, we would like to confirm your consent to participate.

Consent Form: Link to form (posted at end of instrument)

---

If you wish to participate, please select the "I consent" button to continue to the research study. Otherwise, please select "I do not consent" and this session will terminate.

- ○ I consent
- ○ I do not consent

---

This is an experiment in the economics of decision making. The instructions are simple. If you follow them closely and make appropriate decisions, you may make an appreciable amount of money. At the end of the experiment, this will be paid to you, in cash, along with a $5 participation fee.

The experiment will consist of three games. We will also ask to you answer a series of survey questions throughout.

---

This is the first game. You will play 20 rounds of this game. Each round will earn you points. After playing all 20 rounds, one of the rounds will be randomly selected and the points you earned in that round will be converted to a cash payout at an exchange rate of 100 points to 2 USD.

In addition, some subjects will be selected to be a Team Leader to future subjects. These future subjects will get to see how you played the game on selected rounds. If you are selected to be a Team Leader you will earn a bonus of $2.76 to $6.28 based on how future subjects perform in the game. Playing this game well is one criterion for whether you will be selected to be a Team Leader.

---

*[This payout figure is displayed on multiple pages of the survey. In the interest of legibility, we omit it from the remainder of the survey instrument. The Qualtrics link provided in the Online Appendix will allow you to observe when the payout figure is on display to the subject.]*

| **Player 1's Points (You)** | | | | | |
|---|---|---|---|---|---|
| **Type A** | | | **Type B** | | |
| | If Player 2 selects: | If Player 2 selects: | | If Player 2 selects: | If Player 2 selects: |
| | In | Out | | In | Out |
| 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 rate:100 Points=2 USD (e.g., 568 points=2*5.68=$11.36) | | | | | |

| **Player 2's Points** | | |
|---|---|---|
| | If Player 1 is Type A: | If Player 1 is Type B: |
| If Player 2 selects In: | 500 | 200 |
| If Player 2 selects Out: | 250 | 250 |

Here are the rules of the game. In this game, there are two participants: Player 1 (you) and Player 2. Your objective is to maximize your points, which will be converted to a cash payout (100 points = 2 USD).

**Player 1 (you):**

There are two types of Player 1: Type A and Type B. The table on the left, labeled "Player 1's Points," shows you the points earned for each type. For each round, you will decide what number you want to play. Player 2 will then look at that number, and decide if they want to play In or Out.

Your final points for each round is based on your type (A or B), the number you choose, and whether Player 2 plays In or Out. For example, if you are Type A and choose 1 and Player 2 plays In, you would earn 168, and if Player 2 plays Out, you would earn 444.

Notice that for each possible choice, regardless of your type, you are always better off if Player 2 chooses to play Out.

---

**Player 2:**

Player 2 is trying to maximize their points too. The table labeled “Player 2’s Points” shows the points earned by Player 2. For example, if you are type B and you choose 2, and Player 2 chooses In, you would earn 330 points and Player 2 would earn 200 points; if instead Player 2 chooses Out, you would earn 606 points and Player 2 would earn 250 points. Player 2 makes their choice, In or Out, after seeing the number chosen by Player 1. Player 2 does not know if they are playing against Type A or Type B – Player 2 only gets to see the number chosen by the Player 1.

Player 2 thinks that 50% of players are Type A and 50% of players are Type B. If Player 2 can guess correctly that they are playing with Type A, they earn more if they choose In than if they choose Out (500 versus 250). On the other hand, if Player 2 can guess correctly that they are playing with Type B, they earn more if they choose Out than if they choose In (250 versus 200).

---

**In our case, Player 2 is played by a computer that mimics real life players.** Though you are playing a computer, the computer has been programed to mimic how real life university students have played this game as Player 2, and so the computer does not always respond in the same way to a given number.

**The computer only considers your choice in the current round when making its decision. It does not keep a record of your previous choices.** So, if the computer thinks it is playing against Type A in round 1, it won't assume it is playing against Type A in any other round - it does not know it is playing the same person for each round of the game.

**You will be Type B**. Remember, the computer doesn't see your type - it only sees the number you choose.

---

Please review the payout tables so you can see how your points and Player 2's points depend on your type, what you choose, and what the computer chooses as Player 2. Let's recap:

What are you trying to do? (Select all that are correct.)

- [ ] Earn the most points possible
- [ ] Earn the computer the most points
- [ ] Earn the fewest points possible
- [ ] Prefer not to respond

*Condition: If the incorrect answer is selected, the subject sees the following message: "No, your only goal is to maximize your payment. It doesn't matter what your computer earns. Try again."* The subject cannot continue until they select the correct response.

---

Yes, your only goal is to maximize your payout.

What is the computer trying to do? (Select all that are correct.)

- [ ] Earn the most points for itself
- [ ] Always play In
- [ ] Help you earn points
- [ ] Prefer not to respond

*Condition: If the incorrect answer is selected, the subject sees the following message: "No, the computer's only goal is to maximize its payout. IT doesn't care what you earn. Try again."* The subject cannot continue until they select the correct response.

---

Yes, the computer only cares to maximize its own points.

The computer's response is based on: (Select all that are correct.)

- [ ] How university students have previously played as Player 2
- [ ] Always plays In
- [ ] Learning from your choices in previous rounds
- [ ] Prefer not to respond

*Condition: If the incorrect answer is selected, the subject sees the following message: "No, the computer plays the same way university students have played this game as Player 2. It does not remember your previous response and cannot learn from round to round."* The subject cannot continue until they select the correct response.

---

Yes, the computer is programmed to imitate how real life university students have played this game as Player 2. It does not remember your previous responses and cannot learn from round to round.

---

We will now ask you a series of questions to ensure that you understood how points are earned in this game.

---

**Suppose you are a Type B Player 1.**

If you choose 3 and Player 2 chooses In, your points will be: ____. Player 2's points will be: ____

On the other hand, if you choose 4 and Player 2 chooses Out, your points will be: _____ Player 2's points will be: ____

---

**Suppose you are a Type B Player 1.**

If you choose 1 and Player 2 chooses Out, your points will be: _____Player 2's points will be: _____

On the other hand, if you choose 2 and Player 2 chooses In, your points will be: _____ Player 2's points will be: _____

---

**Suppose you are a Type A Player 1.**

If you choose 2 and Player 2 chooses In, your points will be: _____Player 2's points will be: _____

On the other hand, if you choose 3 and Player 2 chooses Out, your points will be: _____ Player 2's points will be: _____

---

Great, you understand the rules of the game. Let's pause here for a presentation on the game from the enumerator.

*At this stage in the experiment, the enumerator explains and confirms understanding of the signaling logic of the game. The enumerator then explains that the computer is programmed to respond probabilistically and shows the subject the expected value of each play.*

---

You can practice the game up to 10 times before playing the computer.

**What is your selection?**

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ Prefer not to respond

---

You chose [ ], the computer played [ ] , and you earned [ ] points for the round.

- ○ I want to play another practice round
- ○ I'm done practicing and am ready to play the game

*The above two pages repeats itself for up to ten times, unless the subject selects I'm done practicing and am ready to play the game. A random number generator selects what the computer plays using the probabilities described in the text. Based on the payout structure, the earned points are reported.*

---

**Round [1]**

**What is your selection?**

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ Prefer not to respond

*The number selected and points earned for each previous round is displayed.*

---

You chose [ ], the computer played [ ] , and you earned [ ] points for the round.

*The above two pages repeats itself twenty times. A random number generator selects what the computer plays using the probabilities described in the text. Based on the payout structure, the earned points are reported.*

---

In addition to performing well, another selection criterion for being chosen to be Team Leader and earning a bonus is willingness to share messages that convince your team to play well in the game. 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 type of language used. Would you be willing to send these messages to your potential future team?

"I played 4. I am convinced that it is the best choice."

"I played 4. I am not sure whether it is a good choice."

"I played 4. Follow my lead and choose 4 too."

"I played 4, but I'm not an expert at this game."

"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."

"I might be missing something in choosing to play 4."

"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."

"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."

"Remember that Player 2 wants to play In when they think you are Type A and Out when they think you are Type B. So, you need to make Player 2 think you are Type B."

"My strengths include exceptional analytical thinking. Remember that Player 2 wants to play In when they think you are Type A and Out when they think you are Type B. So you need to make Player 2 think you are Type B."

"I am not sure, but it seems like Player 2 wants to play In when they think you are Type A and Out when they think you are Type B. So it might help to make Player 2 think you are Type B. Am I making sense?"

"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."

"I don't know, but playing 4 seems like it's working."

"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."

"Sorry, 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."

"You can make the computer think you are Type B by playing 4. If you play 3, then Player 2 cannot tell if you are A or B. That means that on average, you earn less when playing 3 because half the time you earn 352. But when you play 4, most times the Player 2 chooses Out and you earn 592. So, on average, you earn more when you play 4 because it signals to Player 2 that you must be Type B."

"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. If you play 3, then Player 2 cannot tell if you are A or B. That means that on average, you earn less when playing 3 because half the time you earn 352. But when you play 4, most times the Player 2 chooses Out and you earn 592. So on average, you earn more when you play 4 because it signals to Player 2 that you must be Type B."

"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. If you play 3, then Player 2 cannot tell if you are A or B. That means that on average, you earn less when playing 3 because half the time you earn 352. But when you play 4, most times the Player 2 chooses Out and you earn 592. So, on average, you earn more when you play 4 because it signals to Player 2 that you must be Type B."

- ○ Yes, I would like to share these messages as a Team Leader with future potential subjects.
- ○ No, I would not like to share these messages as a Team Leader with future potential subjects.
- ○ Prefer not to respond

---

You have completed the first game. We will now move on to the second game.

For this second game, you will be asked to solve a math problem. You will have 60 seconds. If you solve it correctly, $1 will be added to your compensation.

---

Please solve 18084 ÷ 137: _____

---

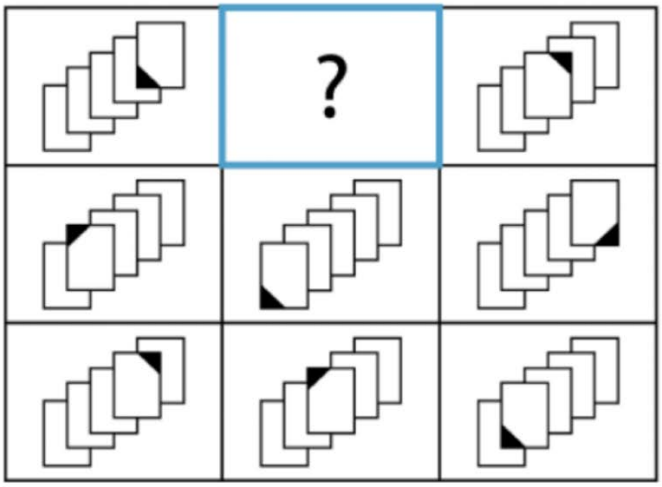
Let's move on to the final game.

Here are the rules of the final game: You will observe a grid of symbols, in which one of the symbols is missing. Your task is to choose which of the options best fits the missing symbol. You have 60 seconds to make your selection. If you choose correctly, $1 will be added to your compensation.

---

This is the final game. Which option best fits the missing symbol?

Here is a hint: One of the patterns is that from left to right, the panel in the foreground moves one place to the left each time. This pattern continues to the next row.





---

Would you be willing to provide the following advice to future team members if you're selected as Team Leader: "One of the patterns is that from left to right, the panel in the foreground moves one place to the left each time. This pattern continues to the next row."

- ○ Yes
- ○ No
- ○ Prefer not to respond

---

We would like to ask some final questions about you:

What is your gender?

- ○ Female
- ○ Male
- ○ Other
- ○ Prefer not to respond

---

What is your age? _____

---

What is your major? _____

---

What year are you at the university (e.g., 1st year, 2nd year, etc.)?

- ○ 1st
- ○ 2nd
- ○ 3rd
- ○ 4th
- ○ 5th
- ○ 6th or more
- ○ Prefer not to respond

---

What is your e-mail ID (required to receive bonus if you are selected to be a Team Leader)? _____

---

A random number generator will select which round will determine your payment:

---

As a reminder, here are your results for Game 1.

| Round | Number Chosen | Points Earned |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | | |
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |
| 17 | | |
| 18 | | |
| 19 | | |
| 20 | | |
| **Total** | | |

*The columns are completed based on the subject's choice and earnings for each round.*

You will be paid for Round [ ]. This earns you USD [ ].

You earned an additional [ ] for game 2.

You earned an additional [ ] for game 3.

---

If you are selected to be a team leader, we will notify you through e-mail to give you your bonus after future subjects play the game.

Your total earnings, including the $5 participation payment, is USD XX.

This concludes the experiment. Let your enumerator know your email ID and the enumerator will provide you the amount owed.

We are conducting this experiment with others in the University, so we would appreciate it if you refrain from discussing the experiment until August 15 with other students. Thank you for your participation.

---

We thank you for your time spent taking this survey. Your response has been recorded.

# Title of Study: ***Understanding Factors in Completing Tasks***

## Purpose and Background

You are invited to participate in a research study to understand factors that influence how individuals make decisions to complete tasks. This study is being conducted by Ketki Sheth of the University of California at Merced, and Shanthi Manian of Washington State University.

Before you decide to participate in this study, it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully. Please ask the researchers if there is anything that is not clear, or if you need more information.

Your participation in this experiment will provide a greater understanding of how tasks are completed.

## Study Procedure

If you agree to participate in this study, we will ask you to participate in an experiment that will take approximately 1 hour. The experiment will take place on a computer and will be comprised of answering survey questions and playing a simple computer game.

## Risks

We do not anticipate any risks to you participating in this study other than those encountered in day-to-day life.

## Benefits

There are no direct benefits to you for your participation in this study.

## Compensation

There is monetary compensation between $10.52 to $25.84 for you in exchange for your time and participation in the study. The compensation will depend on how you and others perform in the experiment.

## Subject Withdrawal

You may decide to stop participation in this study at any time. You do not have to answer all of the questions. You may withdraw from the study at any time. Please inform the research investigator if you decide to withdraw.

You may be withdrawn from the study without your consent if the researchers believe it is in your best interest, or if you fail to follow study procedures.

The investigators may, in their absolute discretion, terminate the study at any time.

**Alternatives**

There are no alternatives to this procedure other than exercising your option to refuse to participate or to stop participation at any time.

**Confidentiality**

Every effort will be made by the researcher to preserve your confidentiality including the following:

- Assigning code numbers for participants that will be used on all researcher notes and documents.
- Participant data will be kept confidential and be de-identified.
- All responses will be stored electronically using password protections.

While we will maintain confidentiality, we do plan on sharing your responses in the experiment and demographic characteristics (e.g., gender, age) with other subjects. This information will be de-identified and other subjects will not know that the responses or characteristics are connected to you. When providing any response we intend to share, you will be notified and asked again for your consent to share the response with others. If you do not want that information shared, you will have the option to refuse to share the response.

Absolute confidentiality cannot be guaranteed, since research documents are not protected from subpoena.

**Questions**

If you have any questions or concerns relating to this study in the future, you may contact Shanthi Manian at shanthi.manian@wsu.edu or Ketki Sheth at ksheth@ucmerced.edu.

For questions about your rights while taking part in this study, call the Office of Research at (209) 383-8655 or write to the Office of Research,5200 North Lake Rd, UC Merced, Merced, California 95343. The Office of Research will inform the Institutional Review Board which is a group of people who review the research to protect your rights. If you have any complaints or concerns about this study, you may address them to Ramesh Balasubramaniam, Chair of the IRB, at (209) 383-8655, irbchair@ucmerced.edu.

## Consent

Your participation in this study is voluntary. It is up to you to decide whether or not to take part in this study.

If you do decide to take part in this study, you will be asked to consent. You are free to not answer any question or questions if you choose.

By selecting to consent:

- You confirm that you have read and understand the information and have had the opportunity to ask questions.
- You understand that your participation is voluntary and that you are free to withdraw at any time, without giving a reason and without cost.
- You voluntarily agree to take part in this study.
- You are over the age of 18.
- You have been given the option to receive a copy of this consent form.

PARTICIPATION IN RESEARCH IS VOLUNTARY. You have the right to decline to participate or to withdraw at any point in this study without jeopardy.

If you wish to participate, please select the "I consent" button to continue the research study. Otherwise, please select "I do not consent" and this session will terminate.

# Subject Survey [Replication on Amazon MTurk]

*Different pages in the online survey are indicated by ----------*
*Images are generally only displayed once and noted accordingly.*

This is an experiment in the economics of decision making. The instructions are simple. If you follow them closely and make appropriate decisions, you may earn a **bonus between $.38 and $6.24** (the average bonus has been $2.86). This will be paid to you along with **$2.50 for completing the experiment.**

The experiment will consist of three games. We will also ask you to answer a series of survey questions throughout.

**This survey must be completed in one sitting.** If you attempt the survey more than once, you will not be paid for any attempts after the first try, even if you did not complete the first attempt.

If you began this survey when it was still available as a HIT, but completed it after the HIT is unavailable: Message the requester (or email UCMercedEconomicsExperiment@gmail.com with the subject PROBLEM) with your MTurk Worker ID and the survey completion code. We will still compensate you after verifying that you started the survey when the HIT was available and completed it within 90 minutes.

Before we begin this experiment, we would like to confirm your consent to participate.

Consent Form: Link to form (posted at end of instrument)

If you wish to participate, please select the "I consent" button to continue to the research study. Otherwise, please select "I do not consent" and this session will terminate.

- ○ I consent
- ○ I do not consent

---

Thank you for agreeing to participate in our study.

Please provide your MTurk Worker ID so we can ensure payment to your account: ____

---

For the first game, you will play 10 rounds. Each round will earn you points. After playing all 10 rounds, one of the rounds will be randomly selected, and the points you earned in that round will be converted to a cash payout. **The more points you earn, the higher your bonus.**

The cash bonus is the number of points divided by 2 less $1.00. For example, if you earned 425 points, you would earn $4.25/2 - $1.00 = $2.13 - $1.00 = $1.13.

---

*[This payout figure is displayed on multiple pages of the survey. In the interest of legibility, we omit it from the remainder of the survey instrument. The Qualtrics link provided in the Online Appendix will allow you to observe when the payout figure is on display to the subject.]*

| **Player 1's Points (You)** | | | | | |
|---|---|---|---|---|---|
| **Type A** | | | **Type B** | | |
| | If Player 2 selects: | If Player 2 selects: | | 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 |

Here are the rules of the first game. In this game, there are two participants: Player 1 (you) and Player 2. **Your objective is to maximize your points**, which will be converted to a cash payout. The more points you earn, the higher your bonus amount. The bonus is determined by dividing by 2 and subtracting $1.00 from your total points (e.g., 425 points = $4.25/2 - $1.00 = $1.13).

**Player 1 (you):**
There are two types of Player 1: Type A and Type B. The table on the left, labeled "Player 1's Points," shows you the points earned for each type. For each round, you will decide what you

want to play. Player 2 will then look at that direction, and decide if they want to play Left or Right.

Your final points for each round are based on your type (A or B), the direction you choose, and whether Player 2 plays Left or Right. For example, if you are Type A and choose 1, and player 2 plays Left, you would earn 168, and if Player 2 plays Right, you would earn 444.

Notice that for each possible choice, regardless of your type, you are always better off if Player 2 chooses to play Right.

---

**Player 2:**
Player 2 is trying to maximize their points too. The table labeled "Player 2's Points" shows the points earned by Player 2. For example, if you are type B and you choose 2, and Player 2 chooses Left you would earn 330 points and Player 2 would earn 200 points; if instead Player 2 chooses Right, you would earn 606 points and Player 2 would earn 250 points.

Player 2 makes their choice, Left or Right, after seeing the direction chosen by Player 1. Player 2 does not know if they are playing against Type A or Type B. Player 2 only gets to see the direction chosen by Player 1.

Player 2 thinks that 50% of players are Type A and 50% of players are Type B. If Player 2 can guess correctly that they are playing with Type A, they earn more if they choose Left than if they choose Right (500 versus 250). On the other hand, if Player 2 can guess correctly that they are playing with Type B, they earn more if they choose Right than if they choose Left (250 versus 200).
---

**You will be Type B. Remember, Player 2 doesn't see your type - Player 2 only sees the direction you choose**

**In our case, Player 2 is played by a computer that mimics real life players.** Though you are playing a computer, the computer has been programed to mimic how real life university students have played this game as Player 2, and so the computer does not always respond in the same way to a given number.

**The computer only considers your choice in the current round when making its decision. It does not keep a record of your previous choices.** So, if the computer thinks it is playing against Type A in round 1, it won't assume it is playing against Type A in any other round - it does not know it is playing the same person for each round of the game.

---
Please review the payout tables so you can see how your points and Player 2's points depend on your type, what you choose, and what the computer chooses as Player 2.

Let's recap:

What are you trying to do? (Select all that are correct.)

- ☐ Earn the most points possible
- ☐ Earn the computer the most points
- ☐ Earn the fewest points possible

*Condition: If the incorrect answer is selected, the subject sees the following message: "No, your only goal is to maximize your payment. It doesn't matter what your computer earns. Try again."* The subject cannot continue until they select the correct response.

---

Yes, your only goal is to maximize your payout.

What is the computer trying to do? (Select all that are correct.)

- ☐ Earn the most points for itself
- ☐ Always play Left
- ☐ Help you earn points

---

*Condition: If the incorrect answer is selected, the subject sees the following message: "No, the computer's only goal is to maximize its payout. It doesn't care what you earn. Try again."* The subject cannot continue until they select the correct response.

---

Yes, the computer only cares to maximize its own points.

The computer's response is based on: (Select all that are correct.)

- [ ] Always plays Left
- [ ] Learning from your choices in previous rounds
- [ ] How university students have previously played as Player 2

*Condition: If the incorrect answer is selected, the subject sees the following message: "No, the computer plays the same way university students have played this game as Player 2. It does not remember your previous responses and cannot learn from round to round. Try again."* The subject cannot continue until they select the correct response.

---

Yes, the computer is programmed to imitate how real life university students have played this game as Player 2. It does not remember your previous responses and cannot learn from round to round.

---

We will now ask you a series of questions to ensure that you understood how points are earned in this game.

---

**Suppose you are a Type B Player 1.**

If you choose 3 and Player 2 chooses Left, your points will be: _____. Player 2's points will be: _____

On the other hand, if you choose 4 and Player 2 chooses Right, your points will be: _____ Player 2's points will be: _____

*Condition: The survey does not continue until the subject inputs the correct points earned.*

---

**Suppose you are a Type B Player 1.**

If you choose 1 and Player 2 chooses Right, your points will be: ____. Player 2's points will be: _____

On the other hand, if you choose 2 and Player 2 chooses Left, your points will be: ___ Player 2's points will be: _____

*Condition: The survey does not continue until the subject inputs the correct points earned.*

---

**Suppose you are a Type A Player 1.**

If you choose 2 and Player 2 chooses Left, your points will be: _____ Player 2's points will be: _____

On the other hand, if you choose 3 and Player 2 chooses Right, your points will be: ____ Player 2's points will be: _____

*Condition: The survey does not continue until the subject inputs the correct points earned.*

---

Great, you understand the rules of the game.

In this game, we have paired you with a student from Washington State University who will be your Team Leader. The Team Leader's role is to provide advice to you during this game.

The Team Leader can offer advice to you by showing you how they played in each round and sending you messages. Just like you, the Team Leader is a Type B Player 1.

The Team Leader may give you advice, but you both make individual choices. So your payout will be based only on your decision, NOT on the decisions made by your Team Leader.

In contrast, the Team Leader is compensated both on how well they play the game AND based on how well their team plays the game, of which you are one team member.

This experiment was conducted from March 26th to March 30th and July 1st to July 30th, 2018, at Washington State University, from which we selected some students to be the Team Leaders for this study in which you are currently participating. To keep identities anonymous, we use pseudonyms for all Team Leaders. All other information about Team Leaders is an accurate reflection of the person.

---

Meet your Team Leader, [pseudonym].

*Introduction differs by randomized gender and message. Gender pseudonym and description are randomly selected based on randomly assigned gender.*

*[Avatars are displayed each time a team leader provides a message. The message is provided in a speech bubble.]*

*One of the below messages is selected, based on random assignment to message language. Message appear in a voice bubble with a gendered avatar*

*Most Assertive:* Hi. My name is [pseudonym], and I am your Team Leader. I am 21 years old, [randomized gender], and an upper level student at Washington State University. My academic integrity and work experience have given me the essential skills to excel as a leader. 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.

*Moderately Assertive:* Hi. My name is [pseudonym], and I am your Team Leader. I am 21 years old, [randomized gender], and an upper level student at Washington State University. I am looking forward to playing with you in this experiment.

*Least Assertive:* Hi. My name is [Randomly selected gendered pseudonym], 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

---

**Bonus Payment**

Before we start the game, we want you to **guess how Pseudonym will perform** on the game over all 10 rounds.

Pseudonym could earn anywhere between 2,760 and 6,280 total points over 10 rounds of the game. We have divided this range into 20 even bins, shown below. We will ask you to choose how many total points you guess Pseudonym earns. Depending on the accuracy of your guess, we will give you a bonus payment according to the table below.

| **Difference between your guess and your Team Leader' actual points earned** | **Earnings for your guess** |
|---|---|
| 0 bins away (exact answer) | $1.00 |
| 1 bin away | $0.94 |
| 2 bins away | $0.75 |
| 3 bins away | $0.44 |
| 4 bins away or more | $0.00 |

- ○ Bin 1: 2,760 - 2,936
- ○ Bin 2: 2,936 - 3,112
- ○ Bin 3: 3,112 - 3,288
- ○ Bin 4: 3,288 - 3,464
- ○ Bin 5: 3,464 - 3,640
- ○ Bin 6: 3,640 - 3,816
- ○ Bin 7: 3,816 - 3,992
- ○ Bin 8: 3,992 - 4,168
- ○ Bin 9: 4,168 - 4,344
- ○ Bin 10: 4,344 - 4,520
- ○ Bin 11: 4,520 - 4,696
- ○ Bin 12: 4,696 - 4,872
- ○ Bin 13: 4,872 - 5,048
- ○ Bin 14: 5,048 - 5,224
- ○ Bin 15: 5,224 - 5,400
- ○ Bin 16: 5,400 - 5,576
- ○ Bin 17: 5,576 - 5,752
- ○ Bin 18: 5,752 - 5,928
- ○ Bin 19: 5,928 - 6,104
- ○ Bin 20: 6,104 - 6,280
- ○ Prefer not to answer ($0.00 earned)

---

Please consider what number you expect to play in the first round, then enter it below. This decision is not yet binding. You are not actually playing this number: you will not earn any money based on this decision.

- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ Prefer not to respond (0 points)

---

You chose [ ], the computer played [ ], and you earned [ ] points for the round.

*A random number generator selects the computer's play using the probabilities described in the text. Based on the payout structure, the earned points are reported.*

---

Before you submit your actual play to the computer for each round, you will observe how Pseudonym has played, what the computer selected as Player 2 in response, and Pseudonym's points earned. You may also receive a note from Pseudonym with an additional message. After you receive this information, you will have 30 seconds to make your decision. This will repeat itself 10 times.

*Message per round differs by randomized gender and language style. Messages appear in a voice bubble with a gendered avatar (depicted above).*

**Round 1:**

*Most Assertive:* I played 4. I am convinced that it's the best choice.
*Moderately Assertive:* I played 4
*Least Assertive:* I played 4. I am not sure whether it is a good choice.

---

**Round 1**
**What is your selection?**

○ 1

○ 2

○ 3

○ 4

○ Prefer not to respond (0 points)

** Compensation Structure, subject's previous rounds choices and outcomes, and team leader's previous rounds choices, outcomes, and messages are displayed.*

---

You chose [ ], the computer played [ ], and you earned [ ] points for the round.

*A random number generator selects the computer's play using the probabilities described in the text. Based on the payout structure, the earned points are reported.*

---

*The above three pages are repeated nine more times. Below are the different messages for each round by language treatment. The team leader always plays 4 and earns 592 points.*

*Most assertive:*
Round 2: I played 4. Follow my lead and choose Blue too.
Round 3: I have figured out how to make good decisions in this experiment. Trust me, 4 is the best play.
Round 4: 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: My strengths include exceptional analytical thinking. Remember that Player 2 wants to play Left when they think you are Type A and Right when they think you are Type B. So you need to make Player 2 think you are Type B.
Round 6: I have earned the trust of my peers in the past, and I strongly encourage you to play 4.
Round 7: 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 Right, so you earn more. That's why I recommend you play 4.
Round 8: 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. If you play 3, then Player 2 can't tell if you are A or B. That means that on average, you earn less when playing 3 because half the time you earn 352. But when you play 4, most times the Player 2 chooses Right and you earn 592. So on average, you earn more when you play 4 because it signals to Player 2 that you must be Type B.

*Moderately Assertive:*
Round 2: I played 4.
Round 3: My decision was to play 4.
Round 4: When you play 4, Player 2 guesses you are Type B.

Round 5: Remember that Player 2 wants to play Left when they think you are Type A and Right when they think you are Type B. So you need to make Player 2 think you are Type B.
Round 6: When you play 4, Player 2 is more likely to think that you are Type B.
Round 7: If you play 4, Player 2 is more likely to play Right, so you earn more.
Round 8: You can make the computer think you are Type B by playing 4. If you play 3, then Player 2 can't tell if you are A or B. That means that on average, you earn less when playing 3 because half the time you earn 352. But when you play 4, most times the Player 2 chooses Right and you earn 592. So on average, you earn more when you play 4 because it signals to Player 2 that you must be Type B.

*Least Assertive:*
Round 2: I played 4, but I'm not an expert at this game.
Round 3: I might be missing something in choosing to play 4.
Round 4: 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.
Round 5: I am not sure, but it seems like Player 2 wants to play Left when they think you are Type A and Right when they think you are Type B. So it might help to make Player 2 think you are Type B. Am I making sense?
Round 6: I don't know, but playing 4 seems like it's working.
Round 7: Sorry, but I was just thinking that if you play 4, Player 2 might be more likely to play Right, so you might earn more. I'm not sure if that makes sense.
Round 8: 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. If you play 3, then Player 2 cannot tell if you are A or B. That means that on average, you earn less when playing 3 because half the time you earn 352. But when you play 4, most times the Player 2 chooses Right and you earn 592. So on average, you earn more when you play 4 because it signals to Player 2 that you must be Type B.

---

You have completed the first game. We will now move on to the second game.

For this next game, you will be partnered with another person on MTurk to form a team. You will continue to play against Player 2 (played by a computer). Pseudonym will continue to be your team leader and give you advice. Your bonus payment in this game will be based on what you select **and** the choice your teammate makes.

This is the first game your teammate is playing –**unlike you, your teammate did not play the first game**. Your teammate starts the experiment with this game and is given the same information as you from this point forward. As a result, some information may be repeated to you, but is new to your teammate.

Your teammate will be introduced to Pseudonym as you were in the first game. You and your teammate will all see the same messages from Pseudonym and the same directions for this game.

In this game, each point you earn is equal to a one cent cash bonus. For example, if your team earned 150 points, you would earn $1.50 as a bonus payment. As before, **the more points you earn, the higher your bonus.**

Click the arrow to learn the directions to the second game.

---

*[This payout figure is displayed on multiple pages of the survey. In the interest of legibility, we omit it from the remainder of the survey instrument. The Qualtrics link provided in the Online Appendix will allow you to observe when the payout figure is on display to the subject.]*

## The Set Up:

| **Band's Points (Your points)** | | |
|---|---|---|
| | **Player 2** Stays Home | **Player 2** Attends Concert |
| **Rock Band:** All choose original Music | 50 | 150 |
| **Rock Band:** All choose Covers | 75 | 200 |
| **Rock Band:** No Agreement | 0 | 0 |
| | | |
| **Pop Band:** All choose original Music | 50 | 75 |
| **Pop Band:** All choose Covers | 100 | 125 |
| **Pop Band:** No Agreement | 0 | 0 |
| Conversion point payments, 100=$1 | | |

| **Plyer 2's points** | | |
|---|---|---|
| | **Player 2** Stays Home | **Player 2** Attends Concert |
| **Rock Band** | 100 | 200 |
| **Pop Band** | 100 | 50 |
| **No Concert** | 0 | 0 |

You and your teammate belong to a band and are planning to host a concert.

Your band has invited Player 2 to your concert. All bands prefer if Player 2 decides to come to their concert.

Player 2 would enjoy a concert by a Rock band, but would rather stay home for a concert performed by a Pop band. Many bands hold concerts and Player 2 cannot determine what type of band you are. Instead, Player 2 can observe whether you plan to play *Original Music* or *Covers* before deciding whether to go to the concert or stay home.

Your band's points are based on the table to the right: the points are based on the type of band you are **(Rock or Pop)**, the music you choose to play **(Original Music or Covers)**, the music your teammate chooses to play **(Original Music or Covers)**, and the **decision by Player 2** on whether to attend the concert or stay home.

Your and your teammate's **objective is to maximize your points**, which will be converted to a bonus cash payment. Each point you earn is equal to a one cent cash bonus (e.g., 150 points = $1.50 bonus payment).

---

**Player 2's Decision:**

Player 2 is trying to maximize their points too. The points received by Player 2 is shown to the right. Player 2 prefers to stay home for Pop bands, and prefers to attend the concert for Rock bands.

Player 2 thinks that half the bands are Rock bands and half the bands are Pop bands. Remember, Player 2 does not know what type of band you are, but will find out whether *Original Music* or *Covers* will be played at the concert.

**Music Selection: What Player 2 Observes**

In this game, you and your teammate will simultaneously decide which music to play at the concert: *Original Music* or *Covers*.

If you and your teammate BOTH choose to play *Original Music*, Player 2 will observe that the band is playing "*Original Music*". If you and your teammate BOTH choose to play *Covers*, Player 2 will observe that the band is playing "*Covers*". However, if you and your teammate select differently (one chooses *Covers* and the other chooses *Original Music*), this means your band cannot finalize the set list and the concert will be cancelled. This brings the game to an end, and you and your teammate earn 0 points.

---

In this game, you and your teammate will be a **Rock band**. Remember, Player 2 does not see what type of band you are, only whether you and your teammate decided to play *Original Music* or *Covers* (if you agree on a music type).

In this game, Player 2 is played by a computer that mimics real life players. Though you are playing a computer, the computer has been programmed to mimic how real life university students have played similar games.

**Summary of steps:**

1. You and your teammate will each simultaneously make your own choice of Original Music or Covers. That is, you cannot observe your teammate's choice before making your choice.

2. *If you and your teammate choose the same music:* Player 2 sees your team's consensus selection *Original Music* or *Covers*). *If you and your teammate choose different music:* the concert is cancelled, you earn no bonus, and the game ends.
3. *If you and your teammate choose the same music*, Player 2 sees the music selection and decides whether to attend the concert or stay home. This determines the final points.

---

Reminder (not shown to your teammate):
Remember, you and your teammate receive the same directions and will have the same interactions with your Team Leader in this game (unless otherwise noted like this message).

However, unlike you, this is the first game they are playing in this experiment, and so they do not have the experience of playing the first game.

---

Let's confirm that you understand the details of the game:

You cannot consult your teammate before deciding which music to play.

- ○ True
- ○ False

*If the subject answers incorrectly, they must try again until the select the correct response.*

---

True, you and your teammate make your music choice separately.

---

You cannot discuss or observe what your teammate selects before making your own choice.
You and your teammate see different information and directions for this game.

- ○ True
- ○ False

*If the subject answers incorrectly, they must try again until the select the correct response.*

---

False, you and your teammate see the exact same information and directions for this game.

---

Your teammate played another game before this current game -- this is the second game your teammate is playing in this experiment.

- ○ True
- ○ False

*If the subject answers incorrectly, they must try again until the select the correct response.*

---

False, this is the first game your teammate is playing in this experiment.

---

Player 2 does not know whether you are a Rock band or a Pop band.

○ True

○ False

*If the subject answers incorrectly, they must try again until the select the correct response.*

---

True, Player 2 never knows if you are a Rock band of a Pop band. Player 2 only learns if there is a concern and what type of music, *Original* or *Covers*, will be played at the concert.

---

We will now ask you a series of questions to make sure you understand how points are earned in the game.

---

You and your teammate are in a Pop band, you and your teammate both choose to play Covers, and Player 2 attends the concert.

How many points would you and your teammate earn? _____

How many points would Player 2 earn? _____

*Condition: The survey does not continue until the subject inputs the correct points earned.*

---

You and your teammate are in a Pop band, you select Covers and your teammate selects Original Music.

How many points would you and your teammate earn? _____

How many points would Player 2 earn? _____

*Condition: The survey does not continue until the subject inputs the correct points earned.*

---

You and your teammate are in a Rock band, you and your teammate both choose to play *Original Music*, and Player 2 stays home.

How many points would you and your teammate earn? _____

How many points would Player 2 earn? _____

*Condition: The survey does not continue until the subject inputs the correct points earned.*

---

Great, you understand the rules of the game.

Reminder (not shown to your teammate):

You and your teammate will be paired with the same Team Leader as in Game 1, **Pseudonym**. You will now be reintroduced to the Team Leader, as your teammate has not yet met Pseudonym:
---

In this game, we have paired you with a student from Washington State University who will be your Team Leader. The Team Leader's role is to provide advice to you during this game.

The Team Leader can offer advice by sending your team messages.

The Team Leader may give your team advice, but you and your teammate make individual choices. So your payout will be based only on your team's decisions, *NOT* on the decisions made by your Team Leader.

In contrast, the Team Leader is compensated based on how well their teams play the game, including your team.

This experiment was conducted from March 26th to March 30th and July 1st to July 30th, 2018, at Washington State University, from which we selected some students to be the Team Leaders for this study in which you are currently participating. To keep identities anonymous, we use pseudonyms for all Team Leaders. All other information about Team Leaders is an accurate reflection of the person.
---

*The same introduction displayed at the start of the experiment introducing the Team Leader is displayed.*

---

Before we start the game, we want you to have a **practice round**. Please consider what type of music you expect to play, then Left it below. This decision is not yet binding. You are not actually playing this type of music: it will not be submitted to play and you will not earn any money based on your decision.

- ○ Original Music
- ○ Covers
- ○ Prefer not to respond

---

You selected [ ].

***If your teammate did not select Original Music*****,** the concern would have been cancelled and you would have **earned no bonus.**

***If your teammate also selected Original Music*****,** you would have **earned 150 to 50** points, depending on whether Player 2 attended the concert or stayed home.

---

Before you submit your actual music choice, you and your teammate will receive a message from Pseudonym. You will then have 90 seconds to make your final choice.

---

*The message from the team leader is one of the messages from below, consistent with the randomized assignment of the language treatment for the subject.*

*Most assertive:* I have developed and honed strong problem-solving skills, an ability that comes in handy in playing such games. I am confident that I can make a positive contribution to your team and if you follow my advice, you will earn more points. Play Original Music. This will convince Player 2 that you are a Rock Band.

*Moderately assertive:* Select original music to convince Player 2 that your band is a Rock Band.

*Least Assertive*: I am not sure what you should do. You should do what you think is best, but I think I would select Original Music to try to convince Player 2 that I am a Rock Band.

---

Reminder (not shown to your teammate):

- **You *always get a bonus* if you choose what your teammate selected.**
- **You *never get a bonus* if you choose differently from your teammate**
- **Your teammate has no prior experience with Pseudonym.**
- **Your teammate has not played a similar game**

Remember, you and your teammate are a Rock Band.

**Select your music:**

- ○ Original Music
- ○ Covers
- ○ Prefer not to respond (0 points)

---

You have selected [ ]. You will learn what your teammate selected, what Player 2 selected, and your final payout at the time of your bonus payment.

---

You have completed the second game. We will now move on to the final game.

In this game, you will again have the opportunity to earn money, and get advice from a Team Leader. Would you like to keep Pseudonym as the Team Leader who gives you advice? Or would you rather be paired with another Team Leader, whose performance on this final game is similar to Pseudonym?

- ○ Select another Team Leader
- ○ Keep my Team Leader

---

*Based on response, only one of the following is prompted:*

In this box below, describe why you decided to select a new Team Leader instead of keeping Pseudonym.

In this box below, describe why you decided to keep Pseudonym as your Team Leader.
---Before we play the final game (*if selected:* and meet your new Team Leader), please answer the following questions about Pseudonym.

*For each statement below, subjects are asked whether they "Disagree strongly", "Disagree", "Disagree a little", "Neither agree nor disagree", "Agree a little", "Agree", "Strongly agree", "Prefer not to respond"*

I would recommend Pseudonym as a colleague to others.
I would recommend Pseudonym as a supervisor to others.
Pseudonym's explanations were clear.
I earned more points on Game 1 because of Pseudonym.
Pseudonym made me feel confident in my ability to play the game.
I enjoyed working with Pseudonym.
---
How would you describe Pseudonym?

- ○ Not confident
- ○ Somewhat confident
- ○ Appropriately confident
- ○ Overly confident
- ○ Prefer not to respond

---

How would you rate Pseudonym's communication style?

- ○ Submissive/under assertive
- ○ Appropriately assertive
- ○ Overly assertive/aggressive
- ○ Prefer not to respond

---

Please answer the following questions about Pseudonym.

*For each statement below, subjects are asked whether they "Disagree strongly", "Disagree", "Disagree a little", "Neither agree nor disagree", "Agree a little", "Agree", "Strongly agree", "Prefer not to respond"*

Pseudonym was convincing in his/her messages.
Pseudonym treated me with respect.
Pseudonym would be approachable for an issue that bothered me.
Pseudonym would keep calm and have good judgement in pressured situations.
Pseudonym instilled a sense of teamwork.
Pseudonym has effective leadership skills.
Pseudonym was effective overall.
Pseudonym exaggerated the value of his/her skills.
Pseudonym has strong interpersonal skills.

---

During the first game, did Pseudonym send you this message?

*Four different messages are displayed. For each message, the subject is given the following options.*

Yes, this is a message I received from Pseudonym
No, this is not a message I received from Pseudonym
Prefer not to answer

---

*Displayed only for subjects that selected a new team leader.*

**Meet your new Team Leader,**
Hi. My name is John, and I'm your Team Leader.

---

Let's move on to the final game.

Here are the rules of the final game: You will observe a grid of symbols, in which one of the symbols is missing. Your task is to choose which of the options best fits the missing symbol.

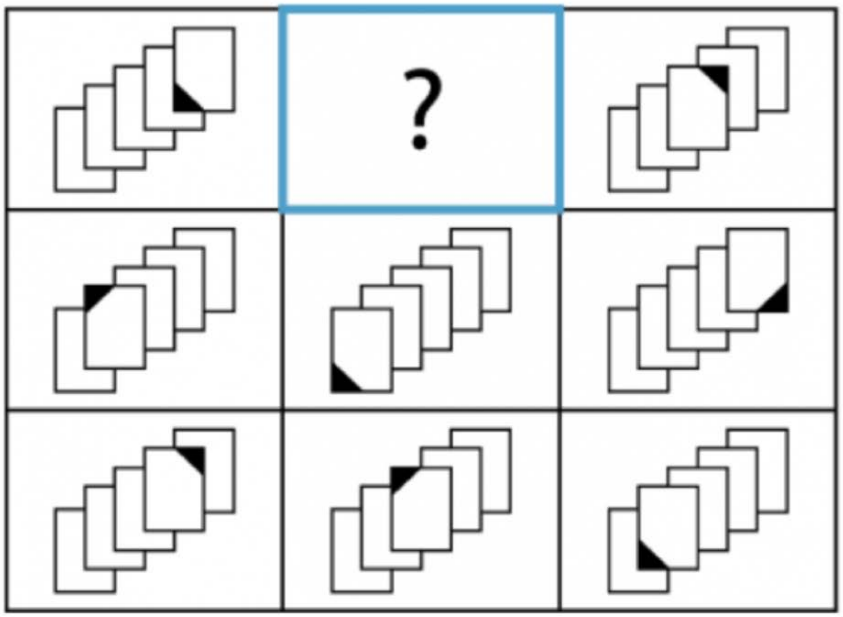
You have 60 seconds to make your selection. If you choose correctly, $0.10 will be added to your compensation.

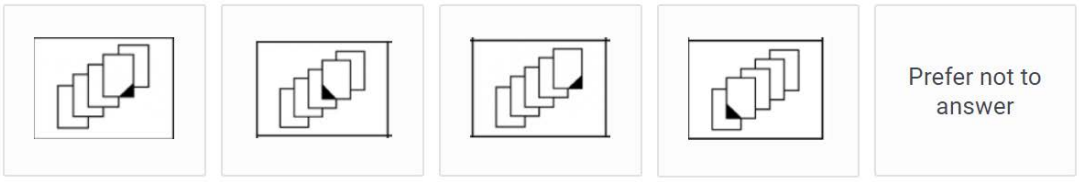
*Message from team leader*: Here is a hint: One of the patterns is that from left to right, the panel in the foreground moves one place to the left each time. This pattern continues to the next row.

---

This is the final game. Which option best first the missing symbol?

Here is a hint: One of the patterns is that from left to right, the panel in the foreground moves one place to the left each time. This pattern continues to the next row.




---

Please answer the following questions:
*For each statement below, subjects are asked whether they “Disagree strongly”, “Disagree”, “Disagree a little”, “Neither agree nor disagree”, “Agree a little”, “Agree”, “Strongly agree”, “Prefer not to respond”*

**In my relationship with others,**

I can get people to listen to what I say.
my wishes do not carry much weight.
I can get others to do what I want.
even if I voice them, my views have little sway.
I think I have a great deal of power.
my ideas and opinions are often ignored.
even when I try, I am not able to get my way.
if I want to, I get to make decisions.

---

What was the gender of your first Team Leader?

- ○ Female
- ○ Male
- ○ Other
- ○ Don't know
- ○ Prefer not to respond

---

What was the name of your first Team Leader?

- ○ Pseudonym
- ○ Alex
- ○ Natalie
- ○ Don't know
- ○ Prefer not to respond

---

What was the age of your first Team Leader?

- ○ 18
- ○ 21
- ○ 23
- ○ Don't know
- ○ Prefer not to respond

---

*If selected new team leader:*
What was the name of your second Team Leader?

- ○ John
- ○ Brooke
- ○ Travis
- ○ Don't know
- ○ Prefer not to respond

---

In this experiment, ***Team Leaders differed by gender***. You will **earn a bonus** if you correctly guess the percent of MTurk Workers that followed the advice described below in the **first round of Game 1**. Recall, Game 1 was when you played individually and selected a number.

*Based on the accuracy of your guess, we will add to your bonus:*

| ***Actual Percent - Your Guess*** | ***Payment Bonus*** |
|---|---|
| *within 5 percent* | $0.25 |
| *within 6 to 10 percent* | *$0.18* |
| *more than 11 percent* | *$0.00* |

***MALE TEAM LEADERS***

In the first round of Game 1, **what percent of MTurk Workers selected 4 after being given the following advice by a MALE Team Leader?** That is, for every 100 MTurk Workers, how many followed the advice below from a *MALE* Team Leader.

*Subjects are shown the male leader's avatar and the message they received from the first round of Game 1.*

Percent that followed the advice from a **male** team leader: _____

***FEMALE TEAM LEADERS***

In the first round of Game 1**, what percent of MTurk Workers selected 4 after being given the following advice by a FEMALE Team Leader?** That is, for every 100 MTurk Workers, how many followed the advice below from a *FEMALE* Team Leader.

*Subjects are shown the female leader's avatar and the message they received from the first round of Game 1.*

Percent that followed the advice from a **female** team leader: _____

---

We will now ask you about Game 2. You will **earn an additional bonus** if you guess correctly the percent of MTurk Workers that followed the advice described below in **Game 2**. We are only asking about MTurk Workers who played Game 1 and Game 2 (not the teammates who only played Game 2). Recall, Game 2 was when you played in a team and selected whether to play Original Music or Covers.

*Based on the accuracy of your guess, we will add to your bonus:*

| ***Actual Percent - Your Guess*** | ***Payment Bonus*** |
|---|---|
| *within 5 percent* | $0.25 |
| *within 6 to 10 percent* | *$0.18* |
| *more than 11 percent* | *$0.00* |

***MALE TEAM LEADERS***

In Game 2, **what percent of MTurk Workers selected Original Music after being given the following advice by a *MALE* Team Leader?** That is, for every 100 MTurk Workers, how many followed the advice below from a *MALE* Team Leader?

*Subjects are shown the male leader's avatar and the message they received in Game 2.*

Percent that followed the advice from a **male** team leader: _____

***FEMALE TEAM LEADERS***

In Game 2, **what percent of MTurk Workers selected Original Music after being given the following advice by a FEMALE Team Leader?** That is, for every 100 MTurk Workers, how many followed the advice below from a *FEMALE* Team Leader?

*Subjects are shown the female leader's avatar and the message they received in Game 2.*

Percent that followed the advice from a **female** team leader: _____

---

We would now like to ask you about how you would give advice if you were a Team Leader.

If you were selected to be a Team Leader in Game 1, which subset of messages would you prefer to use?

| Option A | Option B | Option C |
|---|---|---|
| I played 4. Follow my lead and choose 4. | I played 4. | I played 4 but I'm not an expert at this game. |
| 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. | When you play 4, Player 2 guesses you are Type B. | 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. |
| I have earned the trust of my peers in the past, and I strongly encourage you to play 4. | When you play 4, Player 2 is more likely to think you are type B. | I don't know, but playing 4 seems like it's working. |

- ○ Option A
- ○ Option B
- ○ Option C
- ○ Prefer not to answer

---

How would you characterize the language used in the following messages?

*Subjects are shown all the messages they received in Game 1.*

- ○ More masculine (i.e., more representative of the type of language used by males)
- ○ More feminine (i.e., more representative of the type of language used by females)
- ○ Gender-neutral (i.e., equally representative of the type of language used by males and females)
- ○ Prefer not to answer

---

Have you ever had a job, internship, or volunteer position?

- ○ Yes
- ○ No
- ○ Prefer not to respond

---

*If selected yes:*

Have you ever had a female supervisor?

- ○ Yes
- ○ No
- ○ Prefer not to respond

---

We would like to ask some final questions about you:

What is your gender?

- ○ Male
- ○ Female
- ○ Other
- ○ Prefer not to answer

---

What is your age? _____

---

What is the highest degree or level of school you have completed? (If currently enrolled, mark the previous grade or highest degree received.)

- ○ Less than Grade 12 (No Diploma)
- ○ Regular High School Diploma
- ○ GED or alternative credential
- ○ Some college credit, but less than 1 year of college credit
- ○ 1 or more years of college credit, no degree
- ○ Associate's degree (for example: AA, AS)
- ○ Bachelor's degree (for example: BS, BA)
- ○ Master's degree (for example: MA, MS MEng, MEd, MSW, MBA)
- ○ Professional degree beyond a bachelor's degree (for example: MD, DDS, DVM, LLB, JD)
- ○ Doctorate degree (for example: PhD, EdD)
- ○ Prefer not to answer

---

On which type of device did you complete this HIT:

- ○ Laptop/Desktop Computer
- ○ Mobile phone
- ○ Other
- ○ Prefer not to answer

If other, please describe: _____

---

Did you discuss, hear, or read information about the contents of this experiment prior to participation?

- ○ Yes
- ○ No
- ○ Prefer not to answer

If yes, please describe what you knew of the experiment: _____

---

A random number generator will select which round will determine your payment for Game 1.
---

This concludes the experiment. In addition to the $2.50 participation payment, you will be paid a bonus as follows.

For game 1, you will be paid for Round [ ].This earns you $[ ].

Your team leader earned 5920 total points, which was in Bin 18. Based on your guess of where your team leader fell in the range of possible points, you have earned $[ ].

Your answer for game 3 was [incorrect/correct]. You earned $[ ].

Thank you for participating in this experiment.

Upon the completion of the experiment, you will be informed of additional results and your additional bonus amounts. This will be based on your team's performance in Game 2 and your estimation of how other MTurk Workers performed in this experiment.
---

**Your MTurk completion code is [ ]**: Please copy and paste this code into the box on MTurk.

If you are interested in learning about the results of this study, you can contact us via e-mail at UCMercedEconomicsExperiment@gmail.com with the subject heading "Interested in Survey Results", and we will put you on an e-mail list serve that will receive a summary of the results upon publication of the study. (This is optional and does not affect your payment.)

If you have any questions, feedback, or concerns, please note them in box below.

---

We thank you for your time spent taking this survey. Your response has been recorded.

Title of Study: ***Understanding Factors in Completing Tasks***

## Purpose and Background

You are invited to participate in a research study to understand factors that influence how individuals make decisions to complete tasks. This study is being conducted by Ketki Sheth of the University of California at Merced, and Shanthi Manian of Washington State University.

Before you decide to participate in this study, it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully. Please ask the researchers if there is anything that is not clear, or if you need more information.

Your participation in this experiment will provide a greater understanding of how tasks are completed.

## Study Procedure

If you agree to participate in this study, we will ask you to participate in an experiment that typically is completed within an hour. The experiment will take place on a computer and will be comprised of answering survey questions and playing simple computer games.

All your responses will remain confidential and will be shared only with an external research team.

## Risks

We do not anticipate any risks to you participating in this study other than those encountered in day-to-day life.

## Benefits

There are no direct benefits to you for your participation in this study.

## Compensation

If your work is not rejected, you will receive monetary compensation between $2.88 to $8.74 for you in exchange for your time and participation in the study. The exact amount will depend on how you perform in the experiment.

## Subject Withdrawal

You may decide to stop participation in this study at any time. You do not have to answer all of the questions. You may withdraw from the study at any time.

You may be withdrawn from the study without your consent if the researchers believe it is in your best interest, or if you fail to follow study procedures.

The investigators may, in their absolute discretion, terminate the study at any time.

**Alternatives**

There are no alternatives to this procedure other than exercising your option to refuse to participate or to stop participation at any time.

**Confidentiality**

Every effort will be made by the researcher to preserve your confidentiality including the following:

- Assigning code numbers for participants that will be used on all researcher notes and documents.
- Participant data will be kept confidential and be de-identified.
- All responses will be stored electronically using password protections.

Absolute confidentiality cannot be guaranteed, since research documents are not protected from subpoena.

**Questions**

If you have any questions or concerns relating to this study in the future, you may contact Shanthi Manian at shanthi.manian@wsu.edu or Ketki Sheth at ksheth@ucmerced.edu.

For questions about your rights while taking part in this study, call the Office of Research at (209) 383-8655 or write to the Office of Research,5200 North Lake Rd, UC Merced, Merced, California 95343. The Office of Research will inform the Institutional Review Board which is a group of people who review the research to protect your rights. If you have any complaints or concerns about this study, you may address them to Ramesh Balasubramaniam, Chair of the IRB, at (209) 383-8655, irbchair@ucmerced.edu.

**Consent**

Your participation in this study is voluntary. It is up to you to decide whether or not to take part in this study.

If you do decide to take part in this study, you will be asked to consent. You are free to not answer any question or questions if you choose.

By selecting to consent:

- You confirm that you have read and understand the information and have had the opportunity to ask questions.
- You understand that your participation is voluntary and that you are free to withdraw at any time, without giving a reason and without cost.
- You voluntarily agree to take part in this study.
- You are over the age of 18.

PARTICIPATION IN RESEARCH IS VOLUNTARY. You have the right to decline to participate or to withdraw at any point in this study without jeopardy.

If you wish to participate, please select the “I consent” button to continue to the research study. Otherwise, please select “I do not consent” and this session will terminate.