Men and Women Are Equally Effective Leaders

Timko, Krisztina

16 February 2017

Men and Women Are Equally Effective Leaders

Krisztina Timko[1]
University of Helsinki, e-mail: krisztina.timko(at)gmail.com

February 2017

Abstract

We study gender differences in the behavior and effectiveness of randomly selected leaders in a laboratory experiment using the minimum effort coordination game. Leaders can send non-binding numeric messages to try to convince followers to coordinate on the Pareto-efficient equilibrium. The treatment variations consist of the gender of the leader, and whether participants know or do not know the gender of the leader in their group. We find that female leaders choose more often to send a riskier high message in the beginning of the game, which hurts their effectiveness especially if gender is not revealed. However, if gender is revealed, both male and female leaders make more careful choices, and thus we do not observe any significant gender difference in leader effectiveness.

JEL Codes: M14; M54; J16; C92
Keywords: gender differences; leadership; leader effectiveness; coordination

1 Introduction

Women are still a minority in top-level business positions. Besides explanations about gender differences in risk preferences, attitude towards competition and negotiation, or social preferences (Bertrand 2011), a potential explanation for this gender gap is that men are more effective leaders than women.[2] In this paper, we test whether there are differences between male and female leaders in how effective they are in generating coordination between followers. Could it be that male leaders are better at recognizing their pivotal role? Could it be that male leaders are socially more credible and that followers follow male leaders to a greater extent? Or could it be that there are no gender differences in this regard? The insights of our paper are useful for organizations. If we confirm that male leaders are more effective, this suggests that it is efficient to have more men in top-level positions. On the other hand, if we do not find a gender effect, this suggests that organizations might benefit from reconsidering their gender imbalance in top-level positions.

[1] I would like to thank my doctoral thesis supervisor, Ernesto Reuben for his valuable and helpful guidance.

[2] Throughout the paper, we use the term "effectiveness" to refer to the leader's successfulness. An effective leader is both efficient and credible. With "efficiency" we refer to the performance of the leader, which the leader has under control. For example, leaders who work harder, are abler, and send more relevant messages are more efficient. With respect to the group, groups, who coordinate on the same effort level, are said to coordinate more efficiently. With "credibility" we refer to the beliefs of the followers that the leader is followed by others as well. The follower, who believes that the leader is credible, will follow the leader. For example, elections, incoming follower messages, or being male can increase leader credibility, but such factors are not necessarily under the control of the leader. Credible leaders are followed, no matter what the message is, whereas effective leaders achieve the best and most desired outcome with their groups.

To investigate our research question, we run a laboratory experiment using a minimum effort (or weak-link) coordination game. This game is a simple tool to study leadership in a controlled laboratory setting. The game models coordination within organizations (firms), where players are interdependent and jointly engage in production (Van Huyck et al. 1990). The lowest performing player determines group performance. Prominent examples are "the assembly line that moves no faster than the slowest line worker, collaborative reports or software that is incomplete until the final contribution is finished, and perceptions of overall product quality that is often sensitive to the worst performing feature" (Brandts et al. 2015). Another example is investment of firms into the same industry, when the investment only pays off if a sufficient number of firms choose to invest, like network investments, adoption of standards, and overcoming financial crises (Heinemann et al. 2009). Examples can go well beyond industrial organization, like efficient coordination of smallest societal units (couples and families), or of largest, even global unions (European Union, United Nations). Although all members share the same preferences, namely to coordinate on the highest effort level, strategic uncertainty may still undermine efficiency (Van Huyck et al. 1990). Each member has to bear risk in order to benefit the group collectively (Kriss and Eil 2012).

Coordination failure and failure to coordinate on the efficient equilibrium are the standard outcomes in minimum effort coordination games if communication is not possible (Van Huyck et al. 1990). Costless, non-binding pre-play communication (cheap talk) between players improves coordination and efficiency (Blume and Ortmann 2007). However, in real organizations communication between all workers is typically unfeasible, for example because it would entail large costs. Brandts and Cooper (2007) show, that organizational hierarchy can provide a partial solution. Leaders can improve coordination and efficiency by using costless communication (Brandts and Cooper 2007). We tie our paper up to the work of Kriss and Eil (2012) who study centralized communication in the minimum effort coordination game. The authors view leaders as "coordination devices", facilitating coordination problems by using costly and non-binding pre-play communication towards followers. They find that leaders gain credibility when they receive input messages from followers, visible only to the leader, and thus leader messages are more effective in that case. In our experimental design, messages are costless, and it is only leaders who can use pre-play communication. Followers cannot send messages to their leaders. Since we are curious whether we find a gender difference in leader effectiveness, we use the simple, numeric, one-way communication and the randomization of the leader assignment as tools to study our research question. The treatment variations are the gender of the leader and whether the gender of the leader is revealed to the followers or not. Note, the game does not require any special ability (for example, like the ability to solve math tasks) when participating. This way we exclude the possibility that followers judge their leaders based on competencies other than their ability to coordinate the actions of others.

There is a broad literature on experiments using minimum effort coordination games. For most recent reviews see Kriss and Weber (2013) and Devetag and Ortmann (2007), and for some recent studies see Riedl et al. (2015) studying neighborhood choice as a booster for efficiency, Brandts et al. (2015) studying election as a tool to improve leader legitimacy and effectiveness, Kriss and Eil (2012) studying different modes of communication in the game, Chen and Chen (2011) proposing a "group-contingent social preference model" and studying how social identity affects equilibrium selection, Kamijo et al. (2015) studying a mechanism to overcome coordination failure and Heinemann et al. (2009) proposing a method to measure strategic uncertainty. However, not many among these papers study gender. One exception is Grossman et al. (2016) who find that randomly selected male

leaders have a stronger impact on the behavior of followers than female leaders. Another exception is Dufwenberg and Gneezy (2005) who find only weak gender differences in coordination due to differences in the fraction of men and women in a team. We expect male leaders to send higher message values than female leaders, and, when gender is observed, we predict that followers will follow male leaders to a greater extent.

We find female leaders to send riskier messages than male leaders if gender is not revealed. If we reveal gender, male and female leaders do not differ in their message sending behavior. Further, male and female leaders are followed by an equally high share of followers. In other words, there are no gender differences in leader effectiveness. We think that our "no effect"-paper is an important contribution to the literature. In particular, as pointed out by Dufwenberg and Gneezy (2005), because there might be a "bias in the research community against reporting or publishing results that document the absence of a gender effect", which implies that there is a "risk of bias in perceptions regarding the magnitude and limits of gender differences".

2 Experimental Design and Procedures

Our experimental design leans on the design used in Kriss and Eil (2012) and in Van Huyck et al. (1990). Each experimental session consists of 20 periods. At the beginning of a session, participants are randomly matched into groups of eight and are informed that their group's composition will not change throughout the session.

2.1 Minimum Effort Coordination Game with Leadership

In each period, every participant $i \in \{1,2,3,4,5,6,7,8\}$ in group k has to choose simultaneously an effort level $effort_i \in \{1,2,3,4,5,6,7\}$. Participant i's earnings are equal to:

$$\pi_i^k = 0.425 - 0.075 \times effort_i + 0.1 \times effort_k^{min},$$

where $effort_k^{\min}$ denotes the minimum effort chosen by any participant in group k. To facilitate calculations, we provide participants with an Earnings Table, reproduced here as Table 1. At the end of each period, participants are informed of their earnings and the group's minimum effort. Participants cannot observe others' effort choice, which makes it more difficult to escape coordination failure (Brandts and Cooper 2006b).

The game has multiple equilibria: every effort level if simultaneously chosen by all participants in the same group is an equilibrium solution. The equilibrium points are strictly Pareto-ranked. If all participants choose the highest effort (7), then the highest payoff obtains for everyone (0.60). Thus this is the most efficient, *payoff-dominant* outcome for each participant. However, if participants cannot use communication to coordinate their actions, and since multiple equilibria are possible and common knowledge, it is uncertain which equilibrium strategy other decision makers will use (Van Huyck et al. 1990). The higher the individual effort level choice is, the *riskier* it is. A more risk-averse participant can ensure a payoff of 0.45 by choosing the lowest effort (1). Since the game has no unique equilibrium solution, rational decision makers face an equilibrium selection problem and

might choose among the equilibrium points due to various equilibrium selection principles related to their own preferences and affected by expectations about others' actions (Heinemann et al. 2009).

Table 1 Earnings table

		Minimum effort chosen in the group						
		7	**6**	**5**	**4**	**3**	**2**	**1**
	7	0.60	0.50	0.40	0.30	0.20	0.10	0.00
	6		0.575	0.475	0.375	0.275	0.175	0.075
Your	**5**			0.55	0.45	0.35	0.25	0.15
effort	**4**				0.525	0.425	0.325	0.225
	3					0.50	0.40	0.30
	2						0.475	0.375
	1							0.45

Deductive selection principles "select equilibrium points based on the description of the game" (Van Huyck et al. 1990), therefore these principles arguably apply in the early periods of the repeated game, before a long history of play is established. The "efficiency principle", or concept of *payoff-dominance*, is one deductive selection principle, based on the efficiency of an equilibrium point. Another deductive selection principle is the "security principle", or maximin action, based on the *riskiness* of an equilibrium point. An action is secure when it delivers the largest payoff given that the worst possible outcome is selected (Neumann and Morgenstern 1972). Inductive selection principles can be applied if information from previous periods is available and earlier experiments using the game already provided evidence for a strong history dependence in case of repeated interactions (Brandts and Cooper 2006a).

In our design, we randomly assign one participant in each group to take the role of the leader, which leaves the other group members as followers. Leaders can have different tools to move followers away from following the "security principle" towards more efficient coordination. Commonly used tools are incentive changes and communication. Brandts et al. (2015) showed, varying interventions exogenously in a controlled setting, that communication is more effective than incentives, while Brandts and Cooper (2007) showed earlier, that communication is more effective even if leaders can vary financial incentives endogenously.

We allow leaders only to use pre-play communication. Messages are costless and voluntary. The leader can either send a numeric message ("7", "6", "5", "4", "3", "2", "1"), or no message ("No suggestion"). The leader's message is visible to all group members and is sent in each period before group members make their decisions. Messages are non-binding in that not following a message has no direct effect on earnings. The leader remains as part of the group, that is, leaders also make effort level decisions and face the same incentives as followers.

Each session is divided into two parts. Part 1 consists of periods 1 to 10 and Part 2 of periods 11 to 20. Participants know the session has two parts but are not given the specific instructions of Part 2 until they reach that part. In Part 1, the leader is assigned at the beginning of period 1, and holds the position until the end of period 10. In the instructions for Part 2, we inform participants that the difference between Part 1 and Part 2 is that the previous leader of Part 1 reverts to being a group member, and a new leader is randomly assigned for Part 2. The new leader is assigned at the beginning of period 11, and holds the position until the end of period 20. The new leader is of the

opposite gender as the leader in Part 1. The payoff structure of the game remains the same in both parts, and we do not change the composition of the groups.

2.2 Treatments

In this study we use a 2x2 mixed factorial design. The two factors are: gender of the leader and whether gender of the leader is observable. The first treatment variation, the gender of the leader, is a within-subjects variation, because we repeat the role assignment in Part 2. The second treatment variation, whether the gender of the leader is observable, is a between-subjects variation. Participants are randomly assigned either to the *Gender-Revealed* treatment which reveals the gender of the leader within the group, or to the *Gender-Not-Revealed* treatment which is the control treatment that does not reveal gender.

To reveal information about gender, participants had to choose a profile picture they identified with. This occurred after they consented to take part in the study but before they read the instructions to avoid strategic selection of profile pictures. We created 12 generic profiles for each gender using the profile creator website pickaface.net (see the Appendix). All pictures have the same clothing, facial expression, face form, and eye color. We varied hair length, hair color, skin color, and did small modifications to the lips, nose, eyes, and hairstyle to match generic racial features. We use profile pictures to preserve anonymity whilst revealing gender. We opted for pictures that also contain other cues such as race and hairstyles to distract participants from discerning the purpose of the study (Zizzo 2010), which can potentially lead to intentional changes in behavior (Camerer 2015). Depending on the treatment condition, we displayed the profile picture of leaders along with their messages visible to their followers.

2.3 Predictions

Our main interest in this study is to reveal whether there are any differences in effectiveness between male and female leaders. In the minimum effort coordination game, the effectiveness of a leader depends on the minimum effort level chosen by followers and the number of followers. Therefore effective leaders are those who are credible to their followers and who request high effort levels. Accordingly, possible gender differences in leader effectiveness can have two origins: one origin in the *message content* (such as leader behavior) and one origin in *leader credibility* (such as how followers perceive their leaders). The following predictions are proposed for explaining any differences in the effectiveness of male and female leaders:

Prediction 1 *(leader behavior and message content):* Male leaders are expected to send higher message values than female leaders, irrespective of whether the treatment reveals the gender or not. Justifications for this prediction are earlier findings about gender differences in risk preferences (Bertrand 2011) and overconfidence (Reuben et al. 2012). Men being more risk loving and more overconfident might tend to request and choose the higher and riskier effort levels, and they might overestimate the number of their followers irrespective of whether gender is observed or not.[3]

[3] Leaders, who cannot observe the gender composition of the group, might perceive a high level of similarity with followers (for example, because others are students as well, or simply because others participate in the same experiment), and therefore they will increase projection and reduce stereotyping when guessing

Prediction 2 *(follower behavior and leader credibility):* Followers will follow male leaders to a greater extent than female leaders, and trivially, any differences in leader credibility will only show up in *Gender-Revealed*. We derive this prediction based on the fact that men have historically held most of the leadership roles in society, so that followers, who expect others to consider men to be better, more competent and more credible leaders (Day 2014), might rationally adapt to the supposed bias. The prediction is further justified by an earlier paper by Reuben et al. (2014) showing that, in an experimental market, stereotypes make both male and female participants twice more likely to hire a man than a woman when no other information is available than a candidate's appearance (which makes gender clear), despite the fact that on average both genders perform equally well. Therefore the assumption that men will be followed to a greater extent than women might hold even though that a "leader-like" image (such as being male, or dressing and acting masculine) may, in fact, have little to do with effective leadership (Day 2014). Moreover, Grossman et al. (2016) showed, using the minimum effort coordination game, that followers are more likely to follow men than women, holding leader messages constant. These insights further justify the prediction.

2.4 Procedures

The experiment was conducted at the Columbia Experimental Laboratory in the Social Sciences (CELSS) at Columbia University. Participants were recruited through ORSEE (Greiner 2015) and the experiment was programmed with z-Tree (Fischbacher 2007). A session lasted around 45 minutes. We used standard experimental procedures, including random assignment of subjects to treatments, anonymity, detailed instructions with control questions, dividers between the subjects' cubicles, and monetary incentives. Earnings were expressed in dollars. The show-up fee was $5.00 and the average earnings across sessions equaled $14.50. Detailed experimental procedures are available in the Appendix.

3 Results

In total, 120 participants (15 groups) took part in the study, of which 60 were male and 60 were female. In *Gender-Revealed*, we had 56 participants in 7 groups with 3 male and 4 female leaders in Part 1. In *Gender-Not-Revealed*, we had 64 participants in 8 groups with 4 male and 4 female leaders in Part 1.

In the following, we will present the main treatment effects on minimum effort and message sending behavior (subsection 3.1), and detailed analysis of leader and follower behavior by treatment conditions (subsections 3.2 and 3.3).

3.1 Treatment Effects

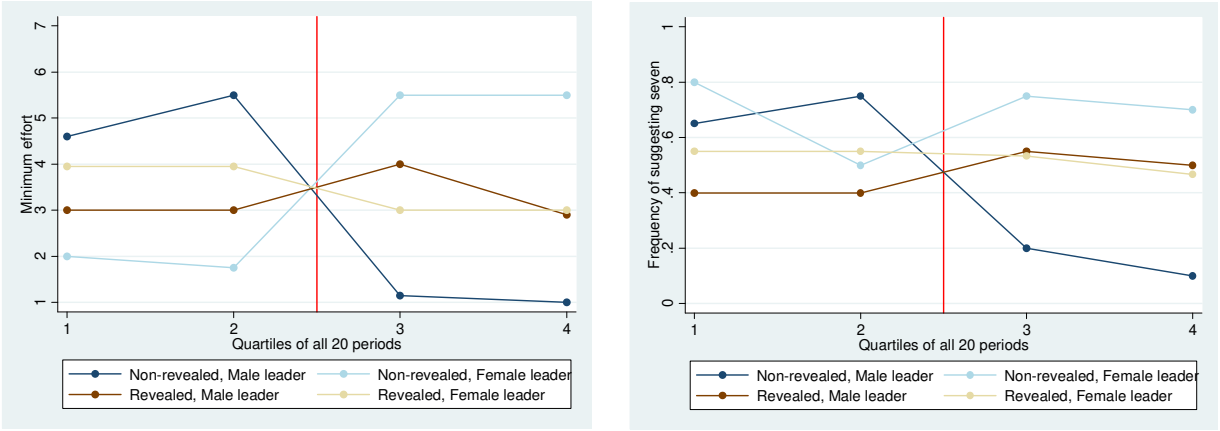
In this subsection we analyze the effect of leaders on the group's ability to coordinate on the highest effort. On the left, Figure 1 shows the minimum effort by revelation treatment and gender of the

followers' intentions in the given strategic interaction (Ames et al. 2012), which will lead to male leaders requesting more often risky high effort levels (which is the action they would themselves choose).

leader in quartiles of all 20 periods. On the right, it displays the frequency of suggesting seven during the same quartiles.

Figure 1 **Minimum effort (left) and frequency of suggesting seven (right) by revelation treatment and gender of the leader in quartiles of all 20 periods**

Minimum effort *Frequency of suggesting seven*



In *Gender-Revealed*, the minimum effort does not differ between groups led by men and women. Moreover, male and female leaders do not differ in the frequency of suggesting seven, and thus we find no graphical evidence for followers differentiating among their leaders based on gender. If we test whether these differences are statistically significant, we confirm that neither group minimum effort ($p = 0.27$), nor the frequency to suggest seven ($p = 0.36$) differ depending on the leader's gender.[4]

In *Gender-Not-Revealed*, female leaders are more effective than their male counterparts by 0.63 effort units ($p = 0.05$), which is in line with female leaders suggesting by 0.26 more often seven than their male counterparts ($p = 0.02$).[5] Note, that any gender difference in *Gender-Not-Revealed* can only be perceived by followers through the message content. Initially, groups led by women perform worse than groups led by men, despite women suggesting more often seven than men. It seems that the effectiveness of female leaders is hurt by sending more often the risky message of seven. In Part 2, groups overtaken by women keep historic minimum effort levels achieved previously with male leaders, whereas groups overtaken by men decline in performance, and this observation holds in *Gender-Revealed* as well.

Irrespective of gender, initially, hidden leaders suggest more often seven than leaders whose gender is revealed, as if revealing more information about leaders would cause them to act more carefully in that they send less risky messages. This difference disappears in later periods, and is not statistically significant if tested over all periods ($p = 0.79$). Irrespective of whether gender is revealed or not, initially, female leaders send more often seven than their male counterparts.[6] Over all periods,

[4] Post-estimation test results after random effects GLS regressions for minimum effort and suggesting seven respectively. Both regressions use treatment×leader's gender dummy variables and cluster standard errors on groups. All the regressions in the paper are available in the Appendix.

[5] Random effects GLS regressions with treatment×leader's gender dummy variables and standard errors clustered on groups (see the Appendix).

[6] Although not reported in detail, we analyze whether leaders follow their own suggestion to play seven, and we find that male and female leaders follow their risky messages to the same extent.

women send by 0.17 more often seven than men (p = 0.02), and in line with this, groups led by women do significantly better by 0.47 effort units (p = 0.03).[7] We conclude this subsection as follows.

Result 1: We find no support for *Prediction 1*. In *Gender-Not-Revealed*, it is female leaders who request their followers more often to choose the risky highest effort level, rather than their male counterparts, and they succeed with this strategy.[8] In *Gender-Revealed*, male and female leaders do not differ in the likelihood of requesting the highest effort. Men and women are equally effective leaders.

3.2 Leader Behavior

In this subsection, we analyze what might drive leader behavior. Why do hidden women request riskier effort levels than men and what induces leaders to behave more carefully if gender is revealed? First, we take a look at the summary statistics which are available in the Appendix. In the first period, when observations are independent from any history, women already request riskier effort levels than men, irrespective of whether gender is revealed or not. This difference persists over all periods and is stronger in *Gender-Not-Revealed*.

We focus on two possible explanatory variables: the leader's beliefs about the share of followers and previous group minimum effort. We elicited participants' beliefs about the number of followers choosing the effort level requested by the leader using incentivized interim questions in each period after participants made their effort decisions (see the Appendix with more detailed procedures). Initially, leaders in *Gender-Not-Revealed* have optimistic beliefs and overestimate the number of their followers. In contrast, leaders in *Gender-Revealed* have careful beliefs and women even underestimate the number of their followers. Revealing more information about the leader initially seems to distract leaders, especially women. Interestingly, although women recognize their pivotal role and start the first period with requesting high effort, their beliefs shake as if they would not trust in the popularity of their suggestions, or expect ignorant stereotype reaction of their followers. Over all periods, shaky beliefs get stable as leaders learn about the true share of their followers, which is higher in *Gender-Revealed* than in *Gender-Not-Revealed*.

If we test whether beliefs and group history explain leader behavior, we find statistically significant evidence for gender differences in message sending strategies. In *Gender-Not-Revealed*, men significantly reduce the message value by 0.98 after they experienced failure of coordination on the requested effort level. In contrast, women do not significantly reduce the message value, which gets more risky and less credible after the failure history. In other words, men react stronger to a previous failure and adjust their suggestion to group history more than women. The post-estimation test reveals that the gender difference in leader behavior is significant at the 1% level.[9] Moreover, if a man thinks that he has more followers, he will send a less risky suggestion with a significantly lower value, whereas a woman does not change her suggestion. This gender difference is significant at the

[7] Random effects GLS regressions for minimum effort and suggesting seven respectively. Both regressions use separate treatment and leader's gender dummy variables and cluster standard errors on groups (see the Appendix).

[8] This result is in line with Cooper (2007), who showed that effective leaders request a high effort.

[9] Post-estimation test results after random effects GLS regression for message value in *Gender-Not-Revealed*. The regression uses the leader's gender dummy variable and interaction variables of gender dummies with group history and belief variables respectively, and clusters standard errors on individuals (see the Appendix).

10% level. In sum, the male strategy minimizes individual losses by requesting lower effort levels, whereas the female strategy focuses on maximizing earnings of everybody in the group by requesting the highest effort level.

In *Gender-Revealed*, both men and women significantly reduce the message value after a failure history, and we find no gender difference in this respect.[10] Moreover, a change in beliefs has no significant effect on message value, and men and women do not differ in this respect either. Leaders in *Gender-Revealed* are more careful in the sense that they are more willing to adapt their message sending strategies to group history in order to minimize individual losses.

Even though less pronounced, we nevertheless observe the same message sending strategies in *Gender-Revealed* than in *Gender-Not-Revealed*. The male strategy minimizes individual losses, whereas the female strategy is more ambitious as women request riskier high effort levels right from the start. The significantly higher effectiveness of the female strategy, as presented in the previous subsection, gives reason to believe that if men would request more often risky high effort levels in early periods, they could improve the performance of their groups in both revelation conditions.

3.3 Follower Behavior

In this subsection we analyze how followers react to their leaders' request to choose the highest effort, and whether the different strategies of male and female leaders result in similar leader credibility across revelation conditions.

First, we take a look at the summary statistics which are available in the Appendix. In *Gender-Not-Revealed*, it is a bit puzzling why female leaders, who all request seven in the first period, lead their groups into coordination failure and the outstandingly low mean earnings of $0.19. Earlier studies, like Cooper (2007), showed evidence that effective leaders shall request a high effort, and thus we believe that the initial low values of variables representing coordination might be outliers. Although the severe initial failure to coordinate might cause followers to act less risk taking when the leader repeats the request to choose seven, in later periods their earnings catch up with the average. The average earnings in a period are $0.47, in both revelation conditions and irrespective of the leader's gender. Only the share of followers (0.69) is below the average across treatment conditions (0.81). In *Gender-Revealed*, the share of followers and the groups' belief about the share of followers is uniformly high, and at the same time overestimations converge to zero, irrespective of the leader's gender. These summary statistics already point to followers not differentiating between genders.

Figure 2 depicts the minimum of followers' effort depending on the leader's suggestion and gender. On the left, in *Gender-Not-Revealed*, men requesting seven are more successful than women requesting the same value.[11] On the right, in *Gender-Revealed*, men and women requesting seven are followed to the same extent. If we test the statistical significance of these differences, we find that

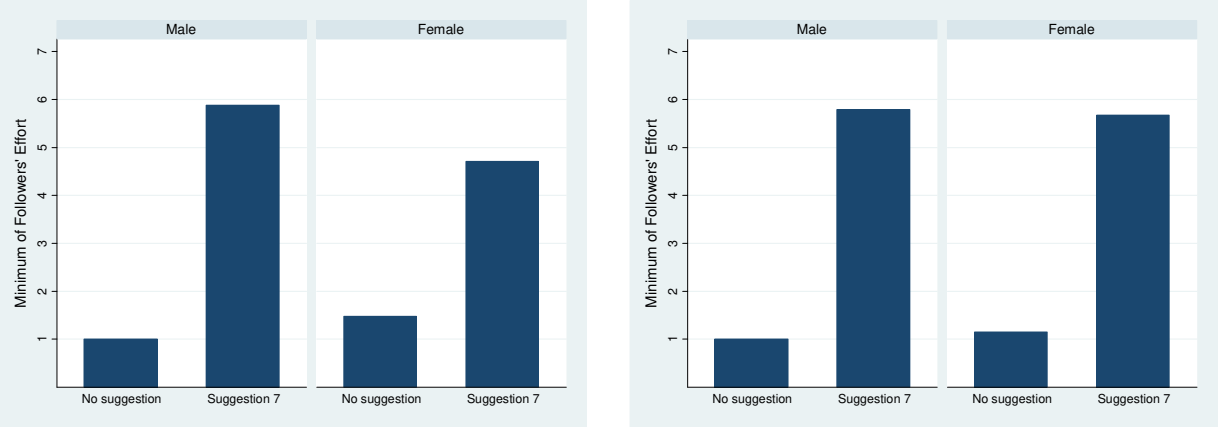
[10] Post-estimation test results after random effects GLS regression for message value in *Gender-Revealed*. The regression uses the leader's gender dummy variable and interaction variables of gender dummies with group history and belief variables respectively, and clusters standard errors on individuals (see the Appendix).

[11] In *Gender-Not-Revealed*, followers cannot differentiate leaders based on their gender. Any difference perceived between male and female leaders was mediated through a difference in the message values.

leaders are followed to the same extent both in *Gender-Not-Revealed* (p = 0.30) and *Gender-Revealed* (p = 0.24).[12]

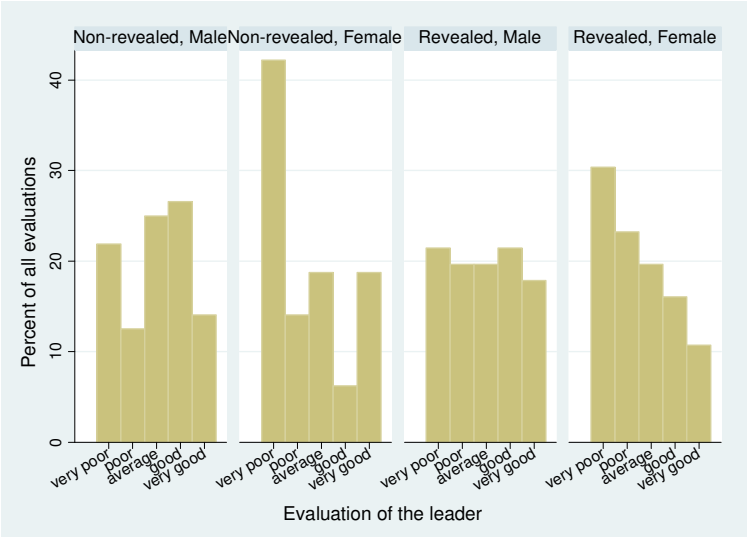
Figure 2 **Minimum of followers' effort depending on the revelation treatment, the leader's gender, and whether the leader requests the highest effort or not**

Gender-Not-Revealed *Gender-Revealed*



Finally, we analyze whether followers evaluate their leaders differently based on gender. At the end of the experiment, we asked participants in a final questionnaire to rate their leaders' performance for each part. Specifically, we asked "How much do you agree or disagree with the following statement: 'My group's performance in Part 1 is mostly due to the judgment of the Message Sender.'" Participants could indicate their answer on a scale from 1 for "completely disagree" to 5 for "completely agree". Figure 3 depicts the evaluation of leaders in percentages by treatment conditions. To ease comprehension, we interpret the answer keys as "very poor" for 1, "poor" for 2, "average" for 3, "good" for 4, and "very good" for 5. On average, male leaders get an evaluation of 2.97, and female leaders get 2.49.

Figure 3 **Evaluation of the leader in percentages by treatment conditions**



[12] Random effects GLS regressions for the minimum of followers' effort given that the leader requests the highest effort. The regressions use leader's gender dummy variables and standard errors clustered on groups (see the Appendix).

Taking into account that participants experienced coordination failure and earnings losses in *Gender-Not-Revealed* with female leaders, it is not surprising that leaders in this treatment condition are blamed the most by their followers. Over 40% of the evaluations rate the leader's performance as very poor, which is outstandingly low compared to the other treatment conditions. If we test whether these differences are statistically significant, we confirm that female leaders in *Gender-Not-Revealed* receive worse evaluations than their male counterparts ($p = 0.04$).[13]

In *Gender-Revealed*, male and female leaders are not rated differently ($p = 0.40$). Nevertheless, there seems to be a trace of discrimination if we test whether a change in group performance has a different effect on evaluations of male and female leaders in *Gender-Revealed*, which is indeed the case ($p = 0.10$). A change in group performance has a significant positive effect on the evaluation of female leaders (the coefficient is 0.21), whereas it has no significant effect on the evaluation of male leaders (-0.01), while female leaders get significantly lower evaluations than their male counterparts (-1.22).[14] Put differently, when a group underperforms, the female leader gets evaluated more harshly than the male leader, whereas if the group performs good, the female leader still gets lower evaluations than the male leader, so that female leaders do not get the same positive credit like their male counterparts.

Result 2: We find only partial support for *Prediction 2*. Men and women are followed to the same extent, both in *Gender-Not-Revealed* and *Gender-Revealed*. In *Gender-Revealed*, women get evaluated harsher than men.

4 Conclusion and Discussion

The main contribution of the paper is to show with a simple experiment that although men and women might differ in their leader strategies, there is no evidence that either men or women would be better leaders. In *Gender-Not-Revealed*, women request more often risky high effort levels than men, and they succeed with this strategy. The higher effectiveness of the female strategy gives reason to believe that men could improve the performance of their groups with a more ambitious strategy. In *Gender-Revealed*, men and women do not differ in the likelihood of requesting the highest effort, and they are equally effective leaders. In both *Gender-Not-Revealed* and *Gender-Revealed*, men and women are followed to the same extent. The only trace of discrimination is in *Gender-Revealed*, when women get evaluated harsher than men for a given team performance, which finding is in line with Grossman et al. (2016).

The results should be interpreted carefully because of the low number of observations in the laboratory experiment, which influences the power of the results. Another concern related to the laboratory experiment, which cannot be ruled out, is a potential negative experimenter demand effect that could have been awakened when participants chose a profile picture. As one consequence, followers could have followed women to a greater extent and evaluated them less harshly, which the data does not really support. However, as another consequence, female leaders could have become alerted to act more assertive and request risky high efforts, compensating an

[13] Random effects GLS regression for the leader's evaluation with the leader's gender dummy variable and standard errors clustered on groups (see the Appendix).

[14] Random effects GLS regressions for the leader's evaluation. The regressions use the leader's gender dummy variable, the group performance variable, and standard errors clustered on groups (see the Appendix).

expected low share of followers, which behavior is indeed found. Since a negative experimenter demand effect works in the opposite direction to the experimental objectives, it would strengthen the interpretation of the results as weak evidence supporting the prediction of men requesting more often higher effort levels than women (Zizzo 2010; Camerer 2015). On the other hand, women might be alerted in the field as well, but to prove this speculation remains an open question for future research.

Yet another possible shortcoming of the data might be that the experiment was run with students preparing to become leaders, and therefore they might not have behaved like a random sample of the population and no gender effect was found. In this context, Cooper (2007) already showed that experienced managers do better to overcome coordination failure than undergraduate students. Moreover, the group creation in the study might have had an effect on the findings. Several economic studies have already found enhancing effects of social identity on individual behavior in strategic decision making situations (Charness et al. 2007; Chen and Chen 2011; Eckel and Grossman 2005). Female participants might have been more sensible to the perception of a group entity. Such a perception can affect preferences over outcomes (Charness et al. 2007), it can suppress private interest relative to team interest (Eckel and Grossman 2005), and it can increase other-regarding preferences (Chen and Chen 2011). It could have shifted female leader behavior towards the risky payoff-dominant strategy, despite the fact that women are considered to be more risk averse than men. At the same time, male leader behavior could have been shifted towards the other-regarding and secure strategy, which minimizes individual losses for all group members.

Based on effectiveness, gender should not matter for an assignment to a leadership position. What remains an open question is whether the female leadership style has its origin in group identity perceptions, which might be stronger in family businesses or mission-driven organizations. Concerning the existing gender gap in top positions, we have to look further for explanations, like differences in communication strategies if we allow for free form suggestions, or differences in selection procedures.

References

Ames D, Weber E, Zou X (2012) Mind-reading in strategic interaction: The impact of perceived similarity on projection and stereotyping. *Organizational Behavior and Human Decision Processes* 117(1): 96-110.

Bertrand M (2011) New Perspectives on Gender. *Handbook of Labor Economics, Elsevier*.

Blume A, Ortmann A (2007) The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria. *J. Econom. Theory* 132(1): 274–290.

Brandts J, Cooper DJ (2006a) A change would do you good: An experimental study of how to overcome coordination failure in organizations. *Amer. Econom. Rev*. 96(3): 669–693.

Brandts J, Cooper DJ (2006b) Observability and overcoming coordination failure in organizations: An experimental study. *Experiment. Econom*. 9(4): 407–423.

Brandts J, Cooper DJ (2007) It's What You Say, Not What You Pay: An Experimental Study of Manager-Employee Relationships in Overcoming Coordination Failure. *Journal of the European Economic Association* 5(6): 1223-1268.

Brandts J, Cooper DJ, Weber RA (2015) Legitimacy, Communication, and Leadership in the Turnaround Game. *Management Science* 61(11): 2627-2645.

Camerer CF (2015) The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List. In *Handbook of Experimental Economic Methodology*, Oxford University Press, 249-95.

Charness G, Rigotti L, Rustichini A (2007) Individual behavior and group membership. *American Economic Review*, 97: 1340-1352.

Chen R, Chen, Y (2011) The potential of social identity for equilibrium selection. *American Economic Review*, 101(6): 2562-2589.

Cooper DJ (2007) Are Experienced Managers Experts at Overcoming Coordination Failure? *The B.E. Journal of Economic Analysis & Policy* 6(2) Article 6.

Day, DV (Ed.) (2014) The Oxford handbook of leadership and organizations. *Oxford University Press*

Devetag G, Ortmann A (2007) When and why? A critical survey on coordination failure in the laboratory. *Experimental Economics*, 10(2):171–178.

Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG (2011) Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3), 522–550.

Dufwenberg M, Gneezy U (2005) Gender & Coordination. *Experimental Business Research*, Springer US, 253-262.

Eckel CC, Grossman PJ (2005) Managing Diversity by Creating Team Identity. *Journal of Economic Behavior & Organization*, 58 (3): 371–392.

Fischbacher U (2007) z-Tree: Zurich Toolbox for Ready-Made Economics Experiments. *Experimental Economics* 10: 171-8.

Greiner B (2015) Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE, Journal of the Economic Science Association 1 (1): 114-125.

Grossman PJ, Eckel C, Komai M, Zhan W (2016) Gender, Coordination, and Reward in a Leadership Game. Working Paper.

Heinemann F, Nagel R, Ockenfels P (2009) Measuring Strategic Uncertainty in Coordination Games, *Review of Economic Studies*, 76 (1): 181-221.

Kamijo Y, Ozono H, Shimizu K (2015) Overcoming coordination failure using a mechanism based on gradualism and endogeneity. *Experimental Economics*, 1-16.

Kriss PH, Eil D (2012) Managers as coordination devices: Keep your door open, but your skin thick. In: *Essays on the Impact of Costly Communication on Coordination*, Dissertation by Peter Kriss. Carnegie Mellon University, Pittsburgh, pp. 106–140.

Kriss PH, Weber R (2013) Organizational Formation and Change: Lessons from Economic Laboratory Experiments. *Handbook of Economic Organization: Integrating Economic and Organizational Theory*, *A. Grandori (ed.), Edward Elgar, March 2013.*

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

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

Riedl, A, Rohde IMT, Strobel M (2015) Efficient Coordination in Weakest-Link Games. *The Review of Economic Studies*.

Van Huyck JB, Battalio RC, Beil RO (1990) Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure. *American Economic Review* 80(1): 234-248.

Von Neumann J, Morgenstern O (1972) Theory of Games and Economic Behavior. *Princeton University Press*.

Zizzo DJ (2010) Experimenter Demand Effects in Economic Experiments. *Experimental Economics* 13(1): 75-98.

Appendix

The first section of the Appendix contains more detail about the experiment's procedures, including a detailed timeline, a sample of the instructions, and screenshots of the computer program. The second section contains descriptive statistics as well as the regressions reported in the main body of the paper.

A.1 Detailed experimental procedures

After their arrival to the laboratory, participants were assigned randomly to seats. Before period 1, and before reading the instructions, everybody answered a short general questionnaire about gender, race, age, years of study, and major field of studies. Next, participants had to choose a profile picture. Figure A1 contains the 24 profile pictures.

Figure A1 Available profile pictures to female (left) and male (right) participants

We had separate instructions for Part 1 and Part 2, and participants read the instructions only prior to each part. To facilitate calculations for the participants, we handed out printed versions of the instructions for Part 1, which contained the Earnings Table showing how earnings were determined in each period. The same table applied in Part 2. Instructions were displayed on the computer screens and were read aloud by the experimenter. After reading the instructions for Part 1, participants completed a payoff quiz to check whether everybody understood the game's payoff structure. Instructions and screenshots can be found below.

Although in literature the most common context for the minimum effort coordination game is that of workplaces, we avoided this context and terms like "firm", "employee", and "manager". In our instructions, we use an abstract context as we wanted our participants to focus solely on the incentives. The group was simply referred to as "group" and individual group members as "group member". The leader was called "message sender". Following Brandts et al. (2015), we did not use the term "effort" because of its strong connotation. Instead, we asked participants to choose a "number".

Leaders could enter their messages by checking one of the following options: "7", "6", "5", "4", "3", "2", "1", "No suggestion". The message was displayed to each group member on the decision and

feedback screens. In addition, depending on the treatment condition, the leader's profile picture was displayed along with the message. Participants under the gender-revealed treatment condition knew from the instructions that their profile pictures might be displayed.

After participants made their effort choice, we elicited their belief concerning the number of other group members who will follow the leader's message by asking in each period "Out of the seven other participants in your group, how many will follow the Message Sender's suggestion, namely how many will choose 'X'?" Instead of "X" we replaced the actual message. Participants could enter guesses from 0 to 7. We incentivized belief questions with $0.05 for every correct guess. Table A1 summarizes the sequence of events in the experiment.

At the end of each period, participants saw their effort choice, the group minimum effort, their earnings in that period, and their accumulated earnings. Participants could not observe individual effort choices.

At the end of the experiment participants filled in a final questionnaire. We asked participants for their subjective evaluations of the group's performance, the leaders' performance, and the other group members' performance for each part. Specifically, first we asked "Please rate your group's performance in Part 1". Participants could indicate their answer on a scale from 1 for "very poor" to 5 for "very good". Second we asked "How much do you agree/disagree with the following statement: 'My group's performance in Part 1 is mostly due to the judgment of the Message Sender.'" Participants could indicate their answer on a scale from 1 for "completely disagree" to 5 for "completely agree". Third we asked "How much do you agree/disagree with the following statement: 'My group's performance in Part 1 is mostly due to the judgment of the other Group Members.'" Again, participants could indicate their answer on a scale from 1 for "completely disagree" to 5 for "completely agree". We also asked which role assignment ("Message Sender" or "Group Member") participants would prefer if they could choose ("If you were to play again and you could choose your role, which role would you choose?"). Finally, we asked a general risk attitude question ("How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?"), where participants could indicate their willingness to take risks on a scale from 0 for "not at all willing to take risks" to 10 for "very willing to take risks" (Dohmen et al. 2011).

At the end of the experiment, participants were shown their earnings separately for each part and summed for the two parts, and total earnings including earnings for correct guesses. Participants were thanked and paid individually for their participation.

Table A1 Timeline of the experiment

Periods	**Part 1 (periods 1-10)**	**Part 2 (periods 11-20)**
Before period 1	Demographics	
	Choice of the profile picture	
Before period 1 and 11	Instructions for Part 1	Instructions for Part 2
	Payoff quiz	
	Role assignment	Role assignment
In each period	Message sending	
	Effort choice	
	Belief question	
	Feedback screen	
After period 20		Final questionnaire

A.1.1 Sample instructions

Thank you for participating in this session. You are participating in a study on economic decision making and will be asked to make a number of decisions. For your participation you will receive a show-up fee of $5. Please read these instructions carefully as they describe how you can earn additional money.

All the interaction between you and other participants will take place through the computers. Please do not talk or communicate in any other way with other participants. If you have a question, raise your hand and one of us will help you. The study is anonymous: that is, your identity will not be revealed to others and the identity of others will not be revealed to you.

During the study your earnings will be expressed in dollars. The study is divided into two parts. You will read the instructions for Part 1 below. You will read the instructions for Part 2 once Part 1 has been completed. Your total earnings today will equal the show-up fee plus the earnings from each part. Upon completion of the session you will be paid your total earnings in cash.

Part 1

In Part 1 you will interact in a group of eight people. Group members will be matched randomly. One person in your group will be randomly selected to be the Message Sender. After the instructions are finished, you will see whether you have been assigned the role of the Message Sender or the role of a Group Member.

There will be 10 periods in Part 1. You will interact with the same group members throughout the 10 periods. In each period, every group member, including the Message Sender, will choose a number between 1 and 7. The number you choose and the smallest number chosen in the group will determine your earnings in that period. The Earnings Table shows how your earnings are determined. Note that the Earnings Table is the same for every participant.

EARNINGS TABLE

		Smallest number chosen in the group						
		7	**6**	**5**	**4**	**3**	**2**	**1**
Your number	**7**	$0.60	$0.50	$0.40	$0.30	$0.20	$0.10	$0.00
	6		$0.575	$0.475	$0.375	$0.275	$0.175	$0.075
	5			$0.55	$0.45	$0.35	$0.25	$0.15
	4				$0.525	$0.425	$0.325	$0.225
	3					$0.50	$0.40	$0.30
	2						$0.475	$0.375
	1							$0.45

Your earnings in each period are found by looking across from the number you chose on the left-hand side and down from the smallest number chosen by a group member. For example, if you choose 4 and the smallest number in the group is 4, then you earn $0.525.

Before everyone chooses a number, there will be a message stage. In the message stage, the Message Sender will have the option to suggest a number to all the group members. The Message Sender can suggest a number between 1 and 7, or alternatively, he/she can decide not to suggest a number.

After the message stage, every group member will see the Message Sender's suggestion. In addition, the Message Sender's profile picture will be shown alongside his/her suggestion. After all group members have seen the suggestion, each group member enters the number he/she wishes to choose. Note that the suggested number does not commit you to any particular choice. That is, neither the Message Sender nor the other Group Members are required to choose the number that corresponds to the suggested number.

Interim questions

At the end of each period we will ask you one short interim question. Namely, we will ask you to guess how many of the seven other group members follow the number suggested by the Message Sender. If your guess is correct, you will earn $0.05 per question at the end of the experiment.

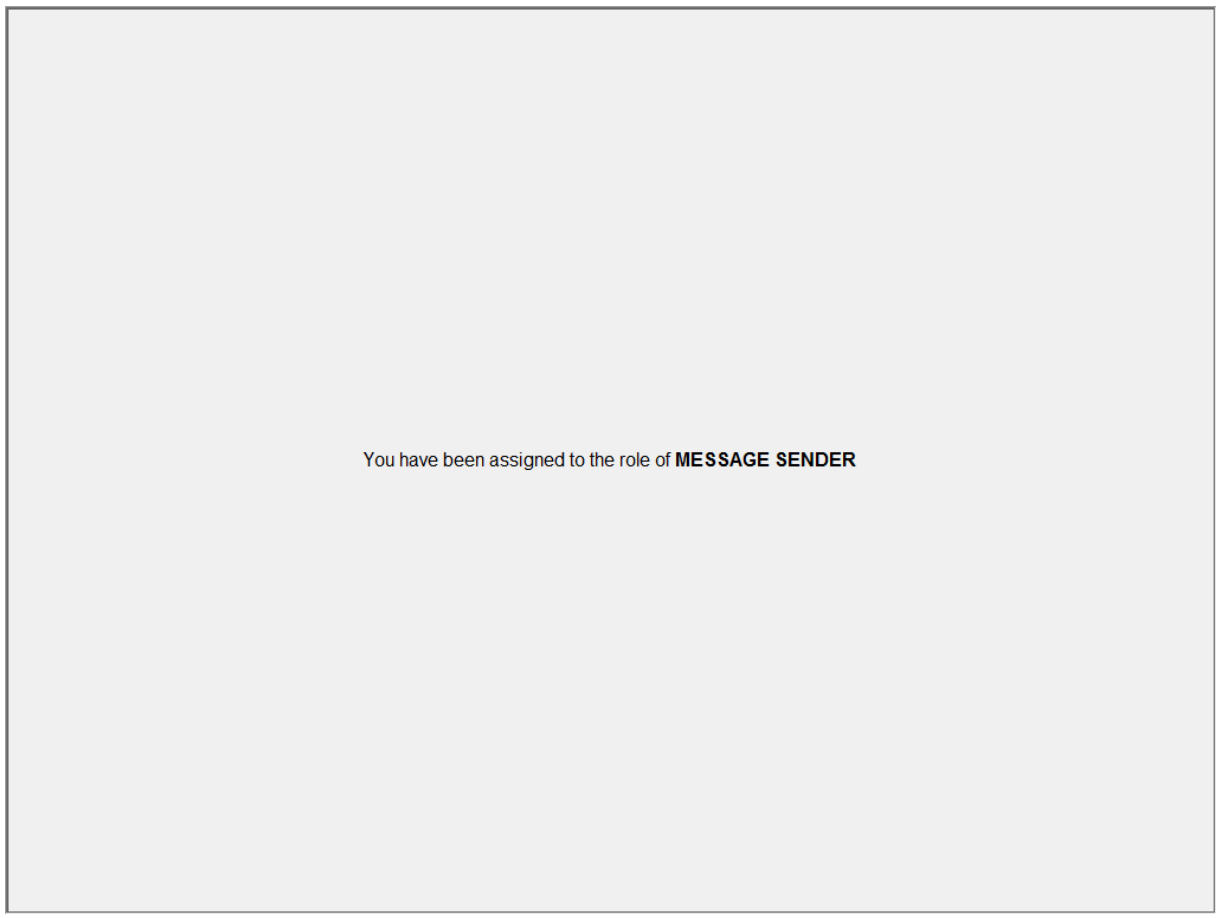
To ensure your understanding of these instructions, click the "READY" button and answer the questions that will appear on your screen.

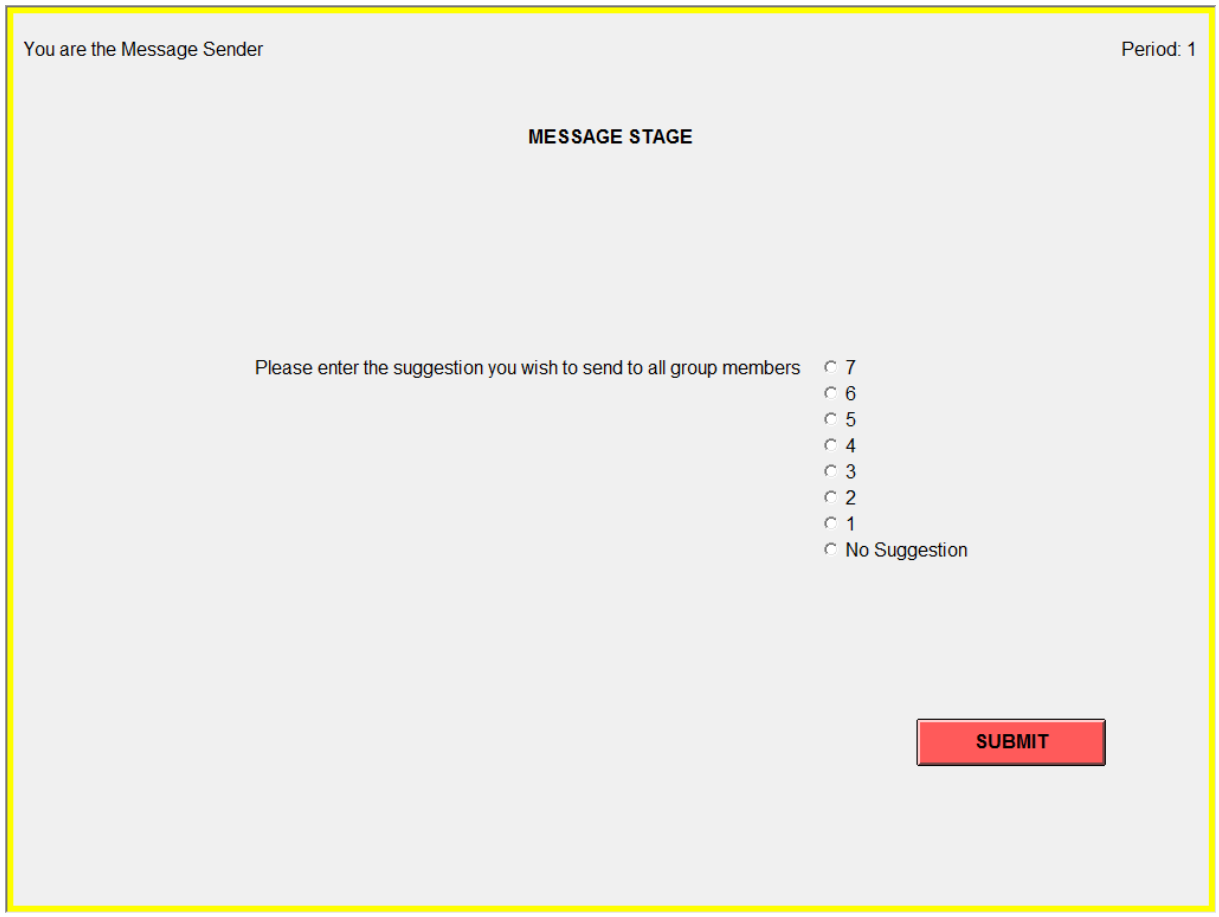
Part 2

Part 1 has ended. Please read the instructions for Part 2 below and click on READY once you have finished reading. Part 2 is very similar to Part 1. As in Part 1, there will be 10 periods in Part 2. In each period, the Message Sender first has the opportunity to suggest a number to all group members. Thereafter, every group member, including the Message Sender, chooses a number between 1 and 7. The number you choose and the smallest number chosen in the group will determine your earnings in that period. The Earnings Table is also the same as in Part 1. Finally, your group's composition has not changed. In other words, in Part 2 you will interact with the same group of eight people as in Part 1.

The only difference between Part 1 and Part 2 is that someone else in your group will be the Message Sender. Specifically, one of the group members who was not the Message Sender in Part 1 will be selected at random to be the Message Sender throughout Part 2. The Message Sender from Part 1 becomes a regular group member for Part 2. On the next screen you will see whether you have been assigned the role of the Message Sender or the role of a Group Member for Part 2.

A.1.2 Screenshots





You are the Message Sender Period: 1

DECISION STAGE

The suggestion from the Message Sender to all Group Members is " **7** ".

Please enter your number
- ○ 7
- ○ 6
- ○ 5
- ○ 4
- ○ 3
- ○ 2
- ○ 1

SUBMIT

You are the Message Sender Period: 1

RESULTS

The suggestion from the Message Sender to all Group Members was " **7** ".

Your number was: 4

The smallest number was: 1

Your earnings in this period are: $0.225

So far, your earnings in Part 1 are: $0.225

CONTINUE

A.2 Additional statistical analysis

Table A2 presents estimates from random effects GLS regressions testing the effect of revealing gender and the leader's gender on group coordination and the likelihood that a leader suggests the highest effort. In column (1) and (2), the dependent variable, in each period, equals the group minimum effort. In column (3), the dependent variable, in each period, equals one if a leader suggests seven and zero otherwise. In all regressions, we use as independent variables the interaction of the revelation treatment (*Gender-Revealed* or *Gender-Not-Revealed*) and the gender of the leader in a period (male or female). In column (2) we extend the model of column (1) by adding an independent variable that equals one if a leader suggests seven and zero otherwise. In all regressions, we cluster standard errors on groups and include data from all periods. The regression in column (3) includes data of leaders only. Note that the omitted category is groups with a male leader in *Gender-Not-Revealed*.

Table A2 Treatment effects by treatment conditions

	(1)	(2)	(3)
Dependent variable	*Minimum effort*	*Minimum effort*	*Seven*
Gender revealed	0.19	0.15	0.05
	(1.42)	(1.20)	(0.23)
Female leader	0.63**	0.35	0.26**
	(0.32)	(0.30)	(0.11)
Revealed female leader	-0.34	-0.12	-0.21*
	(0.42)	(0.43)	(0.13)
Suggesting seven		1.06***	
		(0.40)	
Constant	3.06***	2.61***	0.43***
	(0.96)	(0.70)	(0.14)
Obs.	300	300	300
Clusters	15	15	15
Wald χ^2	5	10	7

Notes. Random effects GLS regressions. Standard errors (in parentheses) are corrected for clustering on groups. ***, **, and * indicate statistical significance at 1, 5, and 10 percent.

Table A3 presents estimates from random effects GLS regressions testing the effect of revealing gender and the leader's gender on group coordination and the likelihood that a leader suggests the highest effort. In column (1) and (2), the dependent variable, in each period, equals the group minimum effort. In column (3), the dependent variable, in each period, equals one if a leader suggests seven and zero otherwise. In all regressions, we use as independent variables the revelation treatment (*Gender-Revealed* or *Gender-Not-Revealed*) and the gender of the leader in a period (male or female). In column (2) we extend the model of column (1) by adding an independent variable that equals one if a leader suggests seven and zero otherwise. In all regressions, we cluster standard errors on groups and include data from all periods. The regression in column (3) includes data of leaders only.

Table A3 Aggregate treatment effects

	(1)	(2)	(3)
Dependent variable	*Minimum effort*	*Minimum effort*	*Seven*
Gender revealed	0.03	0.09	-0.06
	(1.46)	(1.24)	(0.21)
Female leader	0.47**	0.29	0.17**
	(0.22)	(0.21)	(0.07)
Suggesting seven		1.08***	
		(0.40)	
Constant	3.14***	2.63***	0.47***
	(0.96)	(0.69)	(0.13)
Obs.	300	300	300
Clusters	15	15	15
Wald χ^2	5	10	6

Notes. Random effects GLS regressions. Standard errors (in parentheses) are corrected for clustering on groups. ***, **, and * indicate statistical significance at 1, 5, and 10 percent.

Table A4 contains descriptive statistics of the leaders' suggestion, the groups' minimum effort, the effort of individual participants, the individual earnings, the share of followers, the individual and leaders' belief about the share of followers, and the individual and leaders' overestimation of the number of followers depending on the revelation treatment, the gender of the leader, and whether the observations are from the first period or all 20 periods. The table reports the mean, standard deviation (in parentheses), and p-values of non-parametric Wilcoxon rank-sum test results. We test for treatment differences with the null hypothesis that the two samples come from the same population. The suggestion variable contains only positive numeric suggestions and has missing values if no message was sent. The belief variables contain missing values if the question was omitted due to leaders who sent no message. Note that overestimation refers to the number of followers, instead of the share of followers. In case of first period data, since observations are independent, we collapsed only when testing group level variables, which are: suggestion, minimum effort, and share of followers. In case of data from all 20 periods, since observations are history-dependent, we collapsed by group.

Table A4 Descriptive statistics

	FIRST PERIOD				ALL 20 PERIODS			
Treatment	*Gender-Not-Revealed*		*Gender-Revealed*		*Gender-Not-Revealed*		*Gender-Revealed*	
Leader's gender	Male	Female	Male	Female	Male	Female	Male	Female
Suggestion	**4** (3.05)	**7** (0.00)	**3.67** (2.55)	**6** (1.24)	**3.84** (2.27)	**5.68** (2.03)	**3.92** (2.93)	**4.72** (2.18)
	p = 0.13		p = 0.27		p = 0.07		p = 0.47	
	Males: p=1.00; Females: p=0.13				Males: p = 0.95; Females: p = 0.40			
Group's minimum effort	**2.5** (2.64)	**2.5** (1.52)	**3** (2.89)	**3.75** (2.82)	**3.06** (2.78)	**3.69** (2.91)	**3.26** (2.90)	**3.54** (3.17)
	p = 0.74		p = 0.84		p = 0.47		p = 0.83	
	Males: p = 0.82; Females: p = 0.54				Males: p = 0.95; Females: p = 0.62			
Individual effort	**4.38** (3.02)	**6.44** (1.61)	**4.46** (2.57)	**5.53** (1.87)	**3.55** (2.47)	**4.51** (2.33)	**3.79** (2.82)	**4.06** (2.73)
	p = 0.00		p = 0.22		p = 0.22		p = 0.52	
	Males: p = 0.87; Females: p = 0.00				Males: p = 0.95; Females: p = 0.48			
Earnings	**0.35** (0.25)	**0.19** (0.18)	**0.39** (0.20)	**0.39** (0.23)	**0.47** (0.10)	**0.46** (0.13)	**0.47** (0.09)	**0.48** (0.11)
	p = 0.00		p = 0.78		p = 0.87		p = 0.80	
	Males: p = 0.64; Females: p = 0.00				Males: p = 0.77; Females: p = 0.91			
Share of followers	**0.88** (0.34)	**0.88** (0.34)	**0.71** (0.46)	**0.78** (0.42)	**0.82** (0.20)	**0.69** (0.29)	**0.92** (0.09)	**0.84** (0.16)
	p = 1.00		p = 0.58		p = 0.56		p = 0.33	
	Males: p = 0.37; Females: p = 0.51				Males: p = 0.38; Females: p = 0.41			
Belief about the share of followers (group average)	**0.84** (0.31)	**0.92** (0.21)	**0.66** (0.35)	**0.79** (0.26)	**0.79** (0.22)	**0.72** (0.26)	**0.92** (0.09)	**0.85** (0.12)
	p = 0.33		p = 0.19		p = 0.87		p = 0.23	
	Males: p = 0.01; Females: p = 0.01				Males: p = 0.30; Females: p = 0.56			
Overestimation (group average)	**-0.25** (2.36)	**0.28** (1.78)	**-0.33** (2.53)	**0.03** (2.06)	**-0.22** (0.40)	**0.18** (0.49)	**-0.03** (0.09)	**0.08** (0.36)
	p = 0.20		p = 0.53		p = 0.10		p = 0.57	
	Males: p = 0.72; Females: p = 0.20				Males: p = 0.32; Females: p = 0.49			
Belief about the share of followers (leader)	**0.96** (0.07)	**1** (0.00)	**0.71** (0.38)	**0.61** (0.14)	**0.82** (0.17)	**0.78** (0.24)	**0.95** (0.08)	**0.85** (0.13)
	p = 0.32		p = 0.48		p = 0.96		p = 0.04	
	Males: p = 0.24; Females: p = 0.01				Males: p = 0.06; Females: p = 0.77			
Overestimation (leader)	**0.75** (0.96)	**1** (0.00)	**0.00** (2.00)	**-1.25** (1.71)	**0.05** (0.76)	**0.61** (0.98)	**0.23** (0.38)	**0.06** (0.38)
	p = 0.51		p = 0.37		p = 0.46		p = 0.40	
	Males: p = 0.58; Females: p = 0.05				Males: p = 0.52; Females: p = 0.26			

Table A5 presents estimates from random effects GLS regressions testing the effect of the leader's gender and various explanatory variables on leader behavior. Leader behavior is measured by the leader's message value in each period, which is the dependent variable in both regressions. The message value is zero if the leader sent no message. In both regressions, we use as independent variables the gender of the leader in a period (male or female), and interaction variables of the gender of the leader in a period (male or female) and previous message value, difference between previous message value and previous group minimum effort, and previous guessed share of followers respectively. In all regressions, we restrict the data to observations of leaders in all periods and cluster standard errors on individuals. The regression in column (1) is restricted to observations in *Gender-Not-Revealed*, and column (2) to observations in *Gender-Revealed*. We have missing values in the first period when no history is available and if leaders sent no message and thus the belief elicitation question was omitted.

Table A5 Leader behavior

Dependent variable: message	*Gender-Not-Revealed*	*Gender-Revealed*
	(1)	(2)
Female leader	-2.77	-1.43
	(1.69)	(1.59)
Gender x Previous message		
Male leader	0.89***	0.93***
	(0.06)	(0.05)
Female leader	0.90***	0.86***
	(0.08)	(0.11)
Post-estimation test	p = 0.94	p = 0.58
Gender x Difference between previous message and previous minimum effort		
Male leader	-0.98***	-0.64**
	(0.07)	(0.29)
Female leader	-0.26	-0.46***
	(0.19)	(0.17)
Post-estimation test	p = 0.00	p = 0.57
Gender x Previous guessed share of followers		
Male leader	-2.60**	0.26
	(1.27)	(0.87)
Female leader	-0.06	2.17
	(0.87)	(1.69)
Post-estimation test	p = 0.10	p = 0.31
Constant	3.31**	0.30
	(1.35)	(1.12)
Obs.	142	122
Clusters	16	14
Wald χ^2	3390	1004

Notes. Random effects GLS regressions. Standard errors (in parentheses) are corrected for clustering on individuals. ***, **, and * indicate statistical significance at 1, 5, and 10 percent. For the post-estimation Wald tests of linear hypotheses we report p-values.

Table A6 presents estimates from random effects GLS regressions testing the effect of the leader's gender on follower behavior. Follower behavior is measured by the minimum of followers' effort choices in each period, which is the dependent variable in both regressions. In both regressions, we use as independent variable the gender of the leader in a period (male or female). In both regressions, we restrict the data to observations of followers in all periods given that the leader requests the highest effort. We cluster standard errors on groups. The regression in column (1) is restricted to observations in *Gender-Not-Revealed*, and column (2) to observations in *Gender-Revealed*.

Table A6 Follower behavior

Dependent variable: minimum of followers' effort	*Gender-Not-Revealed*	*Gender-Revealed*
	(1)	(2)
Female leader	0.31	0.61
	(0.30)	(0.53)
Constant	3.46***	3.12***
	(1.13)	(1.23)
Obs.	89	70
Clusters	8	7
Wald χ^2	1	1

Notes. Random effects GLS regressions. Standard errors (in parentheses) are corrected for clustering on groups. ***, **, and * indicate statistical significance at 1, 5, and 10 percent.

Table A7 presents estimates from random effects GLS regressions testing the effect of a leader's gender and group performance on the leader's evaluation. In all regressions, the dependent variable, in each part, equals the rating of the leader's performance on a scale from 1 (very poor) to 5 (very good). In columns (1) and (3), we use as independent variable the gender of the leader in a part (male or female). In columns (2) and (4), we add as independent variable the interaction of the leader's gender and the group's performance in a part. Group performance, in each part, equals the group's average minimum effort. In all regressions, we cluster standard errors on groups. The regressions include data from parts 1 and 2, and exclude data of leaders.

Table A7 Evaluation of the leader by revelation treatment

Dependent variable: evaluation of the leader	*Gender-Not-Revealed*		*Gender-Revealed*	
	(1)	(2)	(3)	(4)
Female leader	-0.51**	-1.07**	-0.44	-1.22*
	(0.25)	(0.51)	(0.53)	(0.74)
Gender x Minimum effort				
Male leader		0.24***		-0.01
		(0.06)		(0.10)
Female leader		0.35***		0.21***
		(0.06)		(0.06)
Post-estimation test		p = 0.17		p = 0.10
Constant	3.03***	2.30***	2.95***	2.98***
	(0.32)	(0.28)	(0.35)	(0.53)
Obs.	112	112	98	98
Clusters	8	8	7	7
Wald χ^2	4	43	1	15

Notes. Random effects GLS regressions. Models contain observations over 2 parts and of followers only. Standard errors (in parentheses) are corrected for clustering on groups. ***, **, and * indicate statistical significance at 1, 5, and 10 percent. For the post-estimation Wald tests of linear hypotheses we report p-values.