On the Effectiveness of Elected Male and Female Leaders and Team Coordination

Ernesto Reuben

Krisztina Timko

On the Effectiveness of Elected Male and Female Leaders and Team Coordination

Ernesto Reuben

New York University Abu Dhabi, e-mail: ereuben@nyu.edu

Krisztina Timko

University of Helsinki, e-mail: krisztina.timko@gmail.com

ABSTRACT

We study the effect on coordination in a minimum-effort game of a leader's gender depending on whether the leader is democratically elected or is randomly-selected. Leaders use non-binding messages to try to convince followers to coordinate on the Pareto-efficient equilibrium. We find that teams with elected leaders coordinate on higher effort levels. Initially, the benefits of being elected are captured solely by male leaders. However, this gender difference disappears with repeated interaction because unsuccessful male leaders are reelected more often than unsuccessful female leaders.

JEL Codes: M14; M54; J16; C92

Keywords: gender differences; leadership; democracy effect; leader effectiveness; coordination;

Note: This is the authors' version of a work that was accepted for publication at the *Journal of Economic Science Association*. Changes resulting from the publishing process may not be reflected in this document. A final version is published at http://dx.doi.org/10.1016/10.1007/s40881-018-0056-3.

1. Introduction

One of the defining characteristics of effective leaders is their ability to coordinate the actions of team members through the successful communication of common goals (Bolton, Brunnermeier, and Veldkamp 2010). Historically, most leaders have been men, and to the present date, women are rare in top decision-making positions. Does this gender disparity persist because men are more effective leaders than women? In this paper, we run a laboratory experiment to investigate whether male and female leaders are equally effective in a setting where the only factor determining team success is coordination among team members. Moreover, we distinguish between teams where leaders are selected by team members and teams where they are appointed exogenously.

As many, we study team production with strong complementarities by using the minimum-effort game (Van Huyck, Battalio, and Beil 1990). In the game, team members simultaneously choose among different effort levels, but team output is determined by the lowest effort. Hence, team members face a tradeoff between exerting low effort, which produces little output, and exerting more effort but risking that their effort is wasted because other team members do not exert effort. In the laboratory, play in minimum-effort games often converges to the low-effort equilibrium (Devetag and Ortmann 2007).

A leader's role in the minimum-effort game is clear: convince other team members, the followers, that everyone will exert high effort. If followers believe others will be convinced by the leader, then it is in their best interest to exert high effort. Previous work has demonstrated that giving leaders the option to send messages can substantially increase team coordination on high effort, even when a team has failed to coordinate in the past (Brandts and Cooper 2007; Sahin, Eckel, and Komai 2015; Weber et al. 2001). We use the degree to which leaders can "turnaround" a team through non-binding pre-play messages as our measure of leader effectiveness.

We distinguish between two broad reasons why men might be more effective leaders than women. A first reason is that male leaders might behave differently than female leaders. In particular, for a given likelihood that followers will comply with the leader's message, if men are less risk averse (Croson and Gneezy 2009)[1] and more overconfident (Reuben et al. 2012) than

[1] Ertac and Gurdal (2012) demonstrate that female leaders make less risky decisions when making risky decisions for their team.

women, then male leaders might be more willing to request and exert high effort. A second reason is that followers might believe that others are more likely to follow messages sent by men. If this is the case, then male leaders would be more effective even if they send the same effort request as female leaders. Given that men have historically held more leadership roles, it is reasonable to expect such beliefs by participants in an experiment.[2]

Besides gender, a leader's credibility is affected by other important factors, such as the procedure used to select the leader. In recent work, Brandts, Cooper, and Weber (2015) show that democratically-elected leaders are more effective than randomly-selected leaders, presumably because elected leaders are more legitimate.[3] We extend this line of work by studying whether male and female leaders benefit equally from being elected.[4] Since elected leaders are typically male (e.g., members of parliament and heads of state), it is reasonable to hypothesize that the legitimacy of elected female leaders might differ from the legitimacy of elected male leaders.

To our knowledge, Grossman et al. (2017) and Heursen, Ranehill, and Weber (2018) are the only other studies of gender differences in leadership in a coordination game.[5] Using a design similar to ours, Grossman et al. (2017) find that randomly-selected male leaders have a stronger impact on the behavior of followers than female leaders. Moreover, for a given impact, male leaders receive better subjective evaluations and higher monetary bonuses from their followers. By contrast, Heursen, Ranehill, and Weber (2018) find no difference between the effectiveness of randomly-selected male and female leaders, irrespective of whether the leaders' gender was visible to the followers.

Our study complements the work of Grossman et al. (2017) and Heursen, Ranehill, and Weber (2018) in important ways. First, we consider gender differences between elected as well as randomly-selected leaders. Second, participants in our experiment can opt out from becoming a

---

[2] For evidence that widely-held gender stereotypes affect behavior in experiments, even when they do not hold in the laboratory see Reuben, Sapienza, and Zingales (2014) and Bohnet, van Geen, and Bazerman (2016).

[3] In the context of cooperation, Levy et al. (2011) show that messages from democratically-elected leaders have a bigger impact on contributions to a public good game than those of randomly-selected leaders. More generally, there is a growing body of work showing that institutions are more effective if they are democratically-chosen (Dal Bó 2014).

[4] Even though followers do not select their leaders in many organizations, they are often consulted by those that do.

[5] Dufwenberg and Gneezy (2005) study differences in coordination (without leaders) depending on the fraction of men and women in a team. Other papers study gender differences in leading by example in public good games. For example, Arbak and Villeval (2013) find that men and women are equally likely to lead as long as the leaders' gender is visible.

leader. A potential disadvantage of this design choice is that men and women who self-select to be leaders might not be representative. On the other hand, outside the laboratory leaders are rarely appointed without their consent. Therefore, studying a population of people who self-select to be leaders might be more appropriate. Moreover, if a leader's legitimacy depends on their willingness to lead it is important for followers to know that nobody is forced to be a leader. Third, while in Grossman et al. (2017) and Heursen, Ranehill, and Weber (2018) teams had the same leader throughout the experiment, in this paper, teams select a leader every three periods. This is an interesting extension because, when combined with elections, it gives teams the opportunity to dispense of leaders who do not perform well. Studying gender differences in reelection is valuable precisely because of the finding in Grossman et al. (2017) that male leaders are evaluated more leniently than female leaders. This bias could cause lower reelection rates for successful female leaders, which exacerbates gender differences, but also increase the reelection rate of unsuccessful male leaders, which diminishes them. A final noticeable difference between our studies is that leaders make an effort decision in our experiment, which makes our setting more applicable to teams where leaders directly impact the team's output. This means that, in our experiment, there is one additional channel through which gender differences in leadership can emerge. Namely, the team members' belief that the leader will follow his or her own recommendation.

We find that democratically-elected leaders are more effective than randomly-selected leaders, but the benefit of being elected initially accrues only to male leaders. Over time, this gender difference disappears because unsuccessful male leaders are reelected at higher rates than unsuccessful female leaders.

2. The experiment

Each experimental session consists of 26 periods. At the beginning of a session, participants are randomly matched into teams of five and are informed that their team's composition will not change throughout the session. In each period, every participant $i \in \{1,2,3,4,5\}$ in a team simultaneously chooses an effort level $e_i \in \{0,10,20,30,40\}$. Participant i's earnings are equal to $\pi_i = 200 - 5e_i + 6\min(e_1, e_2, e_3, e_4, e_5)$, where $\min(e_1, e_2, e_3, e_4, e_5)$ is the smallest effort chosen in the team. At the end of each period, participants are informed of their earnings and the team's minimum effort. Each session is divided into two parts. Part 1 consists of periods 1 to 8 and Part

2 of periods 9 to 26. 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, participants play the minimum effort game without a leader. Given previous evidence, we expect teams will end up coordinating on the lowest effort level by the end of Part 1. Failing to coordinate on high effort levels makes the introduction of a leader more meaningful.

In Part 2, we introduce leaders. Specifically, every three periods, one participant in each team is selected to be the team's leader, which leaves the other team members as followers. The leader writes one message that is shown to all followers. The message is displayed before the first period of the three-period leadership term. Leaders can write anything they wish, including a blank message. Messages are non-binding in that not following a message has no direct effect on earnings. Leaders make effort decisions and face the same incentives as followers.

2.1. Leader selection

Leaders are selected with a two-step process. In the first step, each participant decides whether to be a candidate for the leader position. In the second step, a leader is selected among the available candidates. The candidates selected as the leader receives 50 additional points. Participants who are not candidates play a lottery that pays 50 additional points with a 0.5 probability.

We randomly assigned teams to one of two leader-selection procedures. In the *Random* treatment, one of the candidates is randomly assigned to be the leader, each with equal probability. In the *Election* treatment, team members elect the leader by ranking each candidate. Specifically, if there are $C > 1$ candidates, each team member i assigns a unique rank $r_i^c \in \{1, \dots, C\}$ to each candidate c, where the 1^{st} rank indicates the most preferred candidate and the C^{th} rank the least preferred. The candidate with the lowest average rank wins the election.[6] In case of a tie, the winner is chosen randomly among the tied candidates. In both *Random* and *Election*, if there are no candidates then the team has no leader for the next three periods. Moreover, if there is only one candidate then that participant automatically becomes the leader.

[6] Brandts, Cooper, and Weber (2015) use majority voting. We decided to use the Borda count because it generally elects leaders who have broad support within the team rather than the favorite of a simple majority.

2.2. Gender information

To convey information about gender, we had participants chose 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 profile pictures for each gender using the profile creator website pickaface.net (see Figure A1 in the online Appendix). A specific participant saw only the 12 pictures consistent with their gender to avoid participants from strategically selecting a picture of the opposite gender (Charness, Cobo-Reyes, and Sanches 2018). All pictures had 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).

We displayed the profile picture of all team members during the leader-selection process and of the leader in all subsequent screens. Importantly, we provide this information in both *Random* and *Election*. Finally, individual team members were also identifiable by a unique numeric ID, which avoids confusion in case two team members choose the same profile picture.

2.3. Procedures

The experiment was conducted at the Vernon Smith Center of Experimental Economics (VSCEE) at Francisco Marroquín University. Participants were recruited through ORSEE (Greiner 2015) and the experiment was programed with z-Tree (Fischbacher 2007). Sessions lasted around one hour. 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 points and were converted to Guatemalan quetzals at a rate of 10 quetzals per 500 points. Average earnings equaled GTQ 101.83 ($14.76). Detailed experimental procedures are available in the online Appendix.

3. Results

In total, 75 participants (15 teams) took part in *Random*, of which 36 were male and 39 were female. In *Election*, we had 120 participants (24 teams), of which 63 were male and 57 were female. Throughout the paper, we run regressions to test whether treatment differences are statistically significant. We always cluster standard errors on teams. All the reported repressions are available in the online Appendix.

As expected, without a leader, none of the teams managed to coordinate at high effort levels. By the last period of Part 1, the minimum effort in all 39 teams was zero. Moreover, if we compare either effort or earnings in Part 1 depending on the treatment and gender of the leader at the beginning of Part 2, we do not find any statistically significant differences ($p > 0.330$ for effort and $p > 0.273$ for earnings).[7] In other words, since all teams are in the same situation at the beginning of Part 2, any subsequent differences in behavior can be attributed to differences that occurred after leaders are introduced. Nonetheless, to be certain that our results are not affected by behavior in Part 1, we control for the teams' mean minimum effort in Part 1 in all subsequent regressions.

We first analyze the initial effects of leadership, namely periods 9 to 11, in subsection 3.1. Subsequently, in subsection 3.2, we analyze the same effects in periods 12 to 26. That is, after there have been opportunities for leaders to change. Given that we have clear directional hypotheses concerning the effect of elections and the leader's gender, from here on, we report p-values of one-tailed tests. Finally, since the profile pictures also conveyed the leaders' race, if we include the leaders' gender as a regressor, then we also control for their race.

3.1. Initial effects of leadership (periods 9 to 11)

We start our analysis by looking at the decision to become a candidate. Is there a gender difference in the fraction of men and women who nominate themselves? In *Random*, 78% of men become candidates compared to 67% of women, while in *Election*, it is 81% of men and 75% of women. These gender differences are not statistically significant ($p = 0.101$ in *Random* and $p =$

[7] Ordered probit regression for effort and OLS regression for earnings, both using treatment×leader's gender dummy variables (see Table A2).

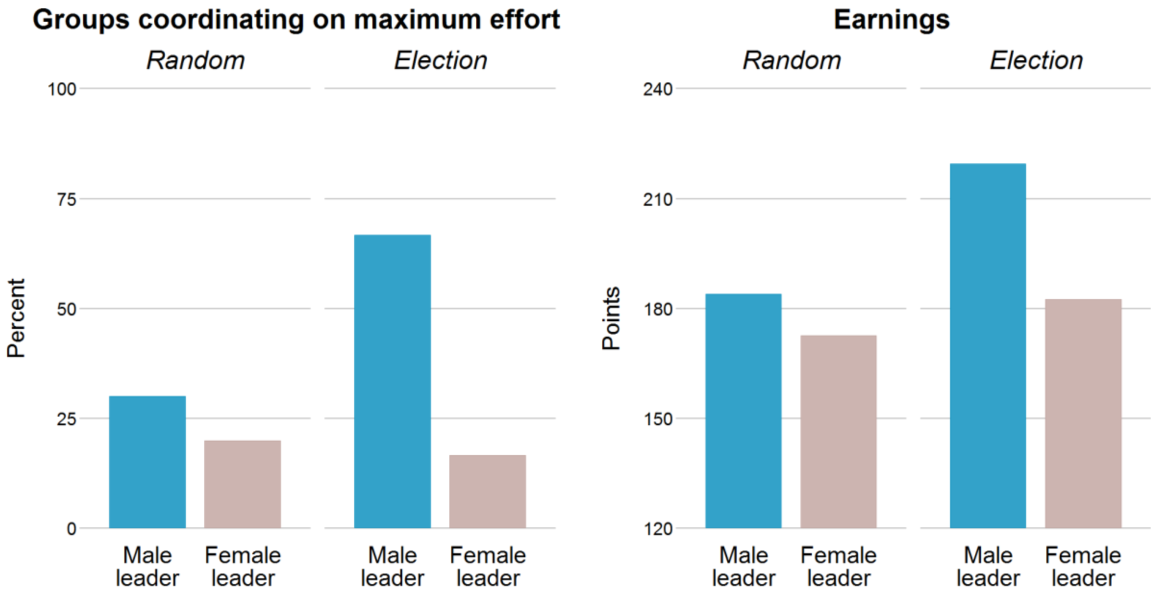

Figure 1. Fraction of teams that coordinate on the highest effort (left) and mean earnings per period (right) in periods 9 to 11 depending on the treatment and the leader's gender.

0.261 in *Election*),[8] suggesting that men and women are similarly willing to become leaders.[9] In *Election*, we can also ask the question, are male candidates more likely to be elected as leaders? We find that the answer is no, 24% of male and 28% of female candidates became leaders resulting in 12 teams lead by a female leader and 12 teams by a male leader.[10] Importantly, there were at least two nominations per team such that all leaders in *Election* were selected through voting and all leaders in *Random* were selected with the random mechanism.

We now turn to the effect of leaders on the teams' ability to coordinate. On the left, Figure 1 depicts the fraction of times teams coordinate on the highest effort in periods 9 to 11. On the right, it displays mean earnings during the same periods. We concentrate on these two variables as they represent the best test for a leader (coordination on the highest effort) and the teams' welfare (earnings). Additional summary statistics are available in Table A3 in the online Appendix.

---

[8] Probit regressions of the participants' decision to become a candidate on their own gender (see Table A10). In the online Appendix, we show that participants initially nominate themselves too often but converge over time to nomination rates consistent with a simple theoretical framework with risk-neutral players (see Tables A10, A11, and Figure A2).

[9] Unlike papers that study entry into tournaments (Niederle and Vesterlund 2011), we do not find that women shy away from becoming candidates. The reason might be that participants think that nominating themselves benefits others as gender differences in tournament entry have been shown to diminish when individuals compete for their team (Healy and Pate 2011).

[10] In *Random*, 10 teams were randomly assigned a male leader and 5 teams a female leader.

Table 1. Fraction of leaders who ask for the highest effort and subsequent effort choices in periods 9 to 11 depending on the treatment and the leader's gender.

	Random		*Election*	
	Male leader	Female leader	Male leader	Female leader
% of leaders asking for the highest effort	70	60	67	42
% of teams coordinating on the highest effort				
Leader does not ask for the highest effort	0	0	0	0
Leader asks for the highest effort	43	33	100	40
% of followers choosing the highest effort				
Leader does not ask for the highest effort	0	0	6	2
Leader asks for the highest effort	63	50	100	67

Even though all leaders produce some coordination on the highest effort, male leaders in *Election* do considerably better. The same can be said about average earnings. If we test whether these differences are statistically significant, we confirm that teams in *Election* do better with male leaders than with female leaders ($p = 0.002$ for coordination, $p = 0.007$ for earnings) while teams in *Random* do not differ depending on the leader's gender ($p = 0.313$ for coordination, $p = 0.308$ for earnings). Moreover, while male leaders do better when they are elected ($p = 0.028$ for coordination, $p = 0.017$ for earnings), female leaders do not differ depending on the leader-selection procedure ($p = 0.476$ for coordination, $p = 0.309$ for earnings).[11] In other words, we replicate previous findings that leadership improves coordination on high effort and more so with elected leaders (Brandts, Cooper, and Weber 2015). However, we find that the benefit of being elected accrues only to male leaders.

Next, we take a closer look at the leaders' messages and followers' behavior. We concentrate on one type of message: explicitly asking followers to choose the highest effort. The benefit of looking at this type of message is that it is easily coded and it has been shown by Brandts, Cooper, and Weber (2015) to be crucial for team coordination.[12]

The first row of Table 1 shows that fraction of leaders who ask for the highest effort ranges from 70% of male leaders in *Random* to 42% of female leaders in *Election*. Even though this is a

[11] Probit regression for coordination on the highest effort and OLS regression for earnings. Both regressions use treatment×leader's gender dummy variables (see Table A4). These results are robust to using team averages as the units of observation or running the regressions solely with data from period 9 (see Table A4-R).

[12] In Timko (2017) we present a detailed analysis of the messages' content that demonstrates that explicitly asking for the highest effort is the only type of message that varies between treatments and predicts team coordination.

noticeable difference in the leaders' behavior, it is not statistically significant.[13] The subsequent rows of Table 1 show the fraction of teams that coordinate on the highest effort and the fraction of followers who choose the highest effort depending on the leader's message. Clearly, if a leader does not ask for the highest effort then followers do not choose high effort irrespective of the treatment or the leader's gender. By contrast, if a leader asks for the highest effort, all followers of elected male leaders follow the leader's request whereas a considerably smaller fraction does so if the leader is female or randomly appointed.[14]

This result poses the intriguing question, why are men and women equally likely to be elected as leaders but are not followed equally when they ask for the highest effort? A difficulty of answering this question is that there is no variance in the following rate of elected male leaders who ask for the highest effort. Therefore, we are unable to test directly whether the determinants of following this type of message varies across genders. Fortunately, in addition to their choices, we also asked participants to self-report the number of followers they expect will comply with the leader's suggestion.[15] After a message asking for the highest effort, 81% of followers of male-led groups expect the team will coordinate on the highest effort compared to 60% of followers in female-led groups ($p = 0.044$). To understand this difference, we classify followers as having preference for a male or a female leader by looking at the gender of their highest-ranked candidate (excluding themselves). Followers who prefer female leaders have similar beliefs irrespective of the elected leader's gender: with a male leader, 76% expect coordination on the highest effort, while with a female leader 70% do ($p = 0.377$). By contrast, followers who prefer male leaders have more trouble believing others will follow female leaders: 84% think the team

[13] Male vs. female leaders in *Election* ($p = 0.104$) and in *Random* ($p = 0.353$). *Election* vs. *Random* among male ($p = 0.450$) and female ($p = 0.751$) leaders. Probit regression with treatment×leader's gender dummy variables (see Table A4).

[14] Male vs. female leaders in *Election* ($p = 0.004$ for coordination, $p = 0.041$ for effort) and in *Random* ($p = 0.481$, for coordination, $p = 0.364$ for effort). *Election* vs. *Random* among male ($p < 0.001$, for coordination, $p = 0.001$ for effort) and female ($p = 0.418$, for coordination, $p = 0.266$ for effort) leaders. Since followers react only when the leader asks for the highest effort, we use probit regressions with sample selection based on the type of message sent and treatment×leader's gender dummy variables (see Table A5).

[15] Belief elicitation was not incentivized. Participants simply self-reported their beliefs in the first period of each leadership term. See the online Appendix for details.

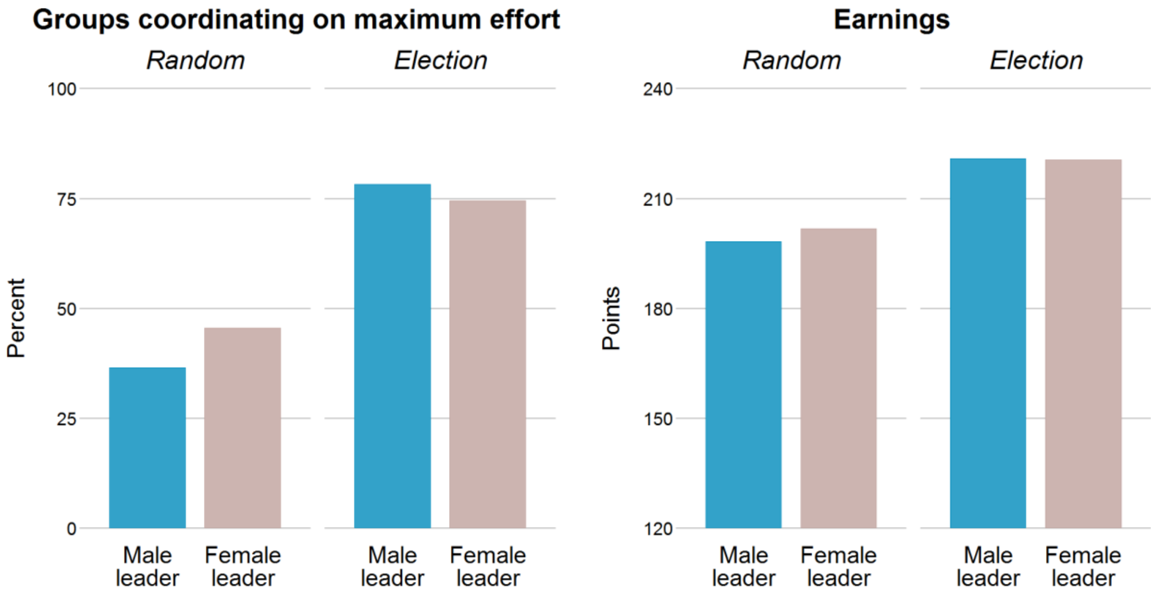

Figure 2. Fraction of teams that coordinate on the highest effort (left) and mean earnings per period (right) in periods 12 to 26 depending on the treatment and the leader's gender.

will coordinate on the highest effort if a male leader suggests it but only 49% hold that belief if a female leader suggests it ($p = 0.038$).[16] We summarize these findings as our first result.

Result 1: *Teams that elect a male leader initially do better than teams that elect a female leader and teams with randomly-chosen leaders of either gender. This is mostly due to the behavior of followers, who are more likely to follow the requests for high effort of elected male leaders.*

3.2. Effects of leadership after reselection (periods 12 to 26)

Next, we evaluate what happens after leader reselection has taken place. Figure 2 and Table 2 present the same statistics as Figure 1 and Table 1 but for periods 12 to 26.

With repetition, the advantage of elected male leaders over randomly-selected leaders still exists. By contrast, the performance of elected female leaders improves to the point that it equals that of their male counterparts (see Figure 2).[17] Table 2 reveals why this is the case. Both male

[16] Probit regressions with sample selection based on the type of message sent and leader's gender×preference for the leader's gender dummy variables (see Table A6).

[17] *Election* vs. *Random* among male (coordination: $p = 0.001$; earnings: $p = 0.017$) and female (coordination: $p = 0.036$; earnings: $p = 0.063$) leaders. Male vs. female leaders in *Election* (coordination: $p = 0.769$; earnings: $p = 0.718$) and in *Random* (coordination $p = 0.615$; earnings: $p = 0.510$ in *Random*). Same regressions as footnote 11 (see Table A4).

Table 2. Fraction of leaders who ask for the highest effort and subsequent effort choices in periods 12 to 26 depending on the treatment and the leader's gender.

	Random		*Election*	
	Male leader	Female leader	Male leader	Female leader
% of leaders asking for the highest effort	69	70	93	87
% of teams coordinating on the highest effort				
Leader does not ask for the highest effort	0	0	0	0
Leader asks for the highest effort	53	65	85	86
% of followers choosing the highest effort				
Leader does not ask for the highest effort	0	0	0	0
Leader asks for the highest effort	62	74	90	90

and female leaders almost always request the highest effort in *Election* but less so in *Random*.[18] Moreover, the differences in the reaction of followers to high-effort requests disappeared.

To conclude our analysis, we take a closer look at behavior in elections to understand why female leaders catch up with male leaders. Specifically, we look at whether the leaders' gender impacts the fraction of followers who challenge the leader by becoming candidates and the likelihood that the leader is reelected. Table 3 presents both statistics depending on whether the team coordinated on the highest effort during the leader's three-period term. Not surprisingly, leaders of successful teams are not challenged frequently and are likely to be reelected. For these teams, we do not observe significant gender differences ($p > 0.340$). By contrast, while unsuccessful male and female leaders are challenged at similar rates ($p = 0.278$), unsuccessful male leaders have a significantly higher chance of being reelected ($p = 0.004$).[19] These findings are stated as our second result.

Result 2: *With repetition, although elected leaders do better than randomly-appointed leaders, male and female leaders perform equally well. An important contributing factor is that unsuccessful female leaders are reelected less often than unsuccessful male leaders, which reduces the gender difference in performance of the remaining leaders.*

---

[18] *Election* vs. *Random* among male ($p = 0.011$) and female ($p = 0.115$) leaders. Same regressions as footnote 13 (see Table A4).

[19] Probit regressions with dummy variables identifying, for the previous three periods, the gender and race of the leader and whether the team was successful or unsuccessful (see Table A7). Since it is possible that unobserved team characteristics explain the interaction between the leader's gender, team success, and the probability of reelection, we also ran regressions including team fixed effects. We obtain very similar results.

Table 3. Likelihood of followers challenging and of the leader being reelected in periods 12 to 26 depending on the leader's gender and team coordination on the highest effort.

	Leader was	
	Male	Female
Team did not coordinate on high effort in the last three periods		
Fraction of followers who become candidates	49	56
Fraction of leaders who are reelected	58	14
Team coordinated on high effort in the last three periods		
Fraction of followers who become candidates	28	26
Fraction of leaders who are reelected	88	78

4. Conclusions

We study gender differences in the effectiveness of non-binding messages sent by team leaders in a minimum-effort game depending on whether a leader is elected or randomly selected.

Our findings are partly consistent with Grossman et al. (2017) and Heursen, Ranehill, and Weber (2018). Like Heursen, Ranehill, and Weber (2018), we do not find a gender difference in the effectiveness of randomly-appointed leaders. Instead, we find a gender difference in the effectiveness of elected leaders, which Grossman et al. (2017) find for randomly-appointed leaders. Unfortunately, there are numerous design differences between these studies to pinpoint the precise reason for these disparities. However, these findings do call for caution as they suggest that gender differences in leadership effectiveness are sensitive to the precise context in which they are studied.

Like Grossman et al. (2017), we find compelling evidence that male and female leaders are evaluated differently for a given team performance. In our study, this bias results in weaker selection pressure among male leaders. Interestingly, lenience towards unsuccessful male leaders has two consequences. On one hand, it makes it easier for men to continue being leaders, which means that the fraction of male leaders increases over time. The gender difference in the probability that a participant is elected to be the team's leader is small in the first election (19% for men vs. 21% for women, $p = 0.786$) but increases in subsequent elections (25% for men vs. 13% for women, $p = 0.041$).[20] On the other hand, lenience towards males contributes to the disappearance of the initial gender difference in effectiveness. This result is a useful reminder

[20] Probit regressions of being a leader on gender (see Table A8). A similar pattern is observed in *Random* but the gender differences are not statistically significant.

that observing gender parity in an organization might in fact be due to gender disparities in evaluations. It also suggests that policies such as gender quotas might not only increase female representation but also the competence of male leaders (Baltrunaite et al. 2014).

Lastly, we should point out that some of our results might be sensitive to the way we implemented the election. In particular, Kanthak and Woon (2015) find that women are less likely to nominate themselves as candidates unless nominating oneself does not involve monetary risk and candidates do not need to publicly advocate for themselves. It is therefore important for future research to determine how male and female leaders perform when elections bias the gender composition of the candidates.

References

Arbak, Emrah, and Marie-Claire Villeval. 2013. "Voluntary Leadership: Motivation and Influence." *Social Choice and Welfare* 40(3): 635–62.

Baltrunaite, Audinga, Piera Bello, Alessandra Casarico, and Paola Profeta. 2014. "Gender Quotas and the Quality of Politicians." *Journal of Public Economics* 118: 62–74.

Bohnet, Iris, Alexandra van Geen, and Max Bazerman. 2016. "When Performance Trumps Gender Bias: Joint vs. Separate Evaluation." *Management Science* 62(5): 1225–34.

Bolton, Patrick, Markus K Brunnermeier, and Laura Veldkamp. 2010. "Economists' Perspectives on Leadership." In *Handbook of Leadership Theory and Practice*, eds. Nitin Nohria and Rakesh Khurana. Boston, MA: Harvard Business School Press, 239–64.

Brandts, Jordi, and David J Cooper. 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): 1223–68.

Brandts, Jordi, David J Cooper, and Roberto A Weber. 2015. "Legitimacy, Communication, and Leadership in the Turnaround Game." *Management Science* 61(11): 2627–45.

Camerer, Colin F. 2015. "The Promise and Success of Lab–Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List." *Handbook of Experimental Economic Methodology*: 249–95.

Charness, Gary, Ramón Cobo-Reyes, and Ángela Sanches. 2018. "Anticipated Discrimination, Choices, and Performance: Experimental Evidence." School of Business Administration Working paper 02-02/2018. American University of Sharjah.

Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 47(2): 448–74.

Dal Bó, Pedro. 2014. "Experimental Evidence on the Workings of Democratic Institutions." In *Institutions, Property Rights, and Economic Growth: The Legacy of Douglass North*, eds. Sebastian Galiani and Itai Sened. New York, NY: Cambridge University Press, 266–88.

Devetag, Giovanna, and Andreas Ortmann. 2007. "When and Why? A Critical Survey on Coordination Failure in the Laboratory." *Experimental Economics* 10(3): 331–44.

Dufwenberg, Martin, and Uri Gneezy. 2005. "Gender & Coordination." In *Experimental Business Research*, eds. Rami Zwick and Amnon Rapoport. Boston, MA: Springer US, 253–62.

Ertac, Seda, and Mehmet Y Gurdal. 2012. "Deciding to Decide: Gender, Leadership and Risk-Taking in Groups." *Journal of Economic Behavior & Organization* 83(1): 24–30.

Fischbacher, Urs. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10(2): 171–78.

Greiner, Ben. 2015. "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association* 1(1): 114–25.

Grossman, Philip J, Catherine C Eckel, Mana Komai, and Wei Zhan. 2017. "It Pays to Be a Man: Rewards for Leaders in a Coordination Game." Working Paper 01-17. Monash University.

Healy, Andrew, and Jennifer Pate. 2011. "Can Teams Help to Close the Gender Competition Gap?" *The Economic Journal* 121(555): 1192–1204.

Heursen, Lea, Eva Ranehill, and Roberto A Weber. 2018. "Are Women Less Effective Leaders than Men? Evidence from Experiments Using Coordination Games." Unpublished manuscript.

Van Huyck, John B, Raymond C Battalio, and Richard O Beil. 1990. "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure." *American Economic Review* 80(1): 234–48.

Kanthak, Kristin, and Jonathan Woon. 2015. "Women Don't Run? Election Aversion and Candidate Entry." *American Journal of Political Science* 59(3): 595–612.

Levy, David M et al. 2011. "Leadership, Cheap Talk and Really Cheap Talk." *Journal of Economic Behavior & Organization* 77(1): 40–52.

Niederle, Muriel, and Lise Vesterlund. 2011. "Gender and Competition." *Annual Review of Economics* 3(1): 601–30.

Reuben, Ernesto, Pedro Rey-Biel, Paola Sapienza, and Luigi Zingales. 2012. "The Emergence of Male Leadership in Competitive Environments." *Journal of Economic Behavior & Organization* 83(1): 111–17.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. "How Stereotypes Impair Women's Careers in Science." *Proceedings of the National Academy of Sciences* 111(12): 4403–8.

Sahin, Selhan Garip, Catherine C Eckel, and Mana Komai. 2015. "An Experimental Study of Leadership Institutions in Collective Action Games." *Journal of the Economic Science Association* 1(1): 100–113.

Timko, Krisztina. 2017. "The Selection Process and Not Gender Matters for Effective Leadership." University of Helsinki.

Weber, Roberto A, Colin F Camerer, Yuval Rottenstreich, and Marc J Knez. 2001. "The Illusion of Leadership: Misattribution of Cause in Coordination Games." *Organization Science* 12(5): 582–98.

Zizzo, Daniel J. 2010. "Experimenter Demand Effects in Economic Experiments." *Experimental Economics* 13(1): 75–98.