Gender and leadership in organizations: Promotions, demotions and angry workers*

Priyanka Chakraborty† Allegheny College

Danila Serra‡ Texas A&M University

December 10, 2020

Abstract

Managerial decisions, such as promotions and demotions, please some employees and upset others. We examine whether having to communicate such decisions to employees, and knowing that employees may react badly, have a differential impact on men's and women's self-selection into leadership roles and their performance if they become leaders. In a novel laboratory experiment that simulates corporate decision-making, we find that women are significantly less likely to self-select into a managerial position when employees can send them angry messages. Once in the manager role, there is some evidence of gender differences in decision-making, but no difference in final outcomes, i.e., overall profits. Male and female managers use different language to motivate their employees, yet differences in communication styles emerge only when workers can send angry messages to managers. Finally, low-rank employees send more angry messages to female managers, and are more likely to question their decisions.

JEL Codes: C92, D91, J16

Key words: Gender differences, leadership, backlash, experiment

*We thank Manuela Angelucci, Jeff Carpenter, Catherine Eckel, Lata Gangadharan, Karen Hauge, Alan Love, Elira Kuka, Shanthi Manian and Tim Salmon for useful comments and discussions. We thank seminar participants at the American University, FAIR - The Choice Lab, Georgia State University, Southern Methodist University, Universidad de Los Andes, University of Queensland, University of Texas at Austin, Virginia Tech University, Washington State University.

†Email: pchakraborty@allegheny.edu.

‡Corresponding author. Email: dserra@tamu.edu.

1 Introduction

Individuals in a position of leadership have to make decisions knowing that some people will be negatively affected by their choices and will therefore be unhappy. In organizations, any managerial decision concerning promotions, salary raises, demotions and dismissals will necessarily generate negative judgment and possibly anger among some workers. Numerous studies have shown that, holding performance constant, women in a position of power are judged more negatively than men. This is true in politics, business, academia, as well as laboratory and online experiments (Abel, 2019; Ayalew et al., 2018; Beaman et al., 2009; Boring, 2017; Egan et al., 2017; Grossman et al., 2019; Shurchkov and van Geen, 2017).[1] There is also evidence that women are more sensitive to public scrutiny (Alan et al., 2020; Jones and Linardi, 2014) and negative feedback (Mayo et al., 2012; Johnson and Helgeson, 2002), and are more severely discouraged by such feedback in male-dominated fields or stereotypically male tasks (Ellison and Swanson, 2018; Kugler et al., 2017).

In this paper, we study whether and to what extent the anticipation of (possibly harsher) negative judgment from subordinates induces women, more than men, to select out of leadership positions, and to perform differently when in a leadership role.

We employ a laboratory experiment that simulates a business environment where managerial decisions, such as rank, job title or task allocations, affect the earnings of some workers positively and the earnings of others negatively.[2] This implies that backlash from unhappy employees, in the form of, at the minimum, negative judgment and disapproving messages, is likely. The controlled setting of the experiment allows us to examine whether a gender leadership gap emerges and, if so, whether it is due to gender differences in willingness to expose oneself to the possibility of receiving angry messages from unhappy employees. The experiment is also informative of the extent and nature of gender differences in managers' decision-making, communication styles and overall outcomes, both in

[1] See also: Branton et al. (2018); Rheault et al. (2019); Elsesser and Lever (2011); Hengel (2017); Mengel et al. (2017); Sarsons (2017).

[2] This is a departure from the existing experimental studies of leadership, which have typically employed sequential public goods games or coordination games where leaders can induce followers to increase their contributions through leading by example (e.g., Güth et al., 2007; Grossman et al., 2015; Jack and Recalde, 2015) or through the use of messages suggesting contributions (e.g., Brandts and Cooper, 2007; Reuben and Timko, 2017). Other important studies of leadership have employed minimum-effort games or real effort tasks where leaders incentivize (Shurchkov and van Geen, 2017) or suggest the effort to be put in by followers (Chaudhuri et al., 2018; Erkal et al., 2018), or tasks that require leaders to make decisions on behalf of their group (Alan et al., 2020; Born et al., 2020; Reuben et al., 2012).

the presence and the absence of the possibility of worker backlash.

In the experiment, a manager is matched with the same two employees for multiple decision-making rounds. Following an initial real effort task, the manager's main responsibility is to decide which employee will be high rank and which employee will be low rank, where ranks determine employees' earnings. There can only be one high-rank and one low-rank employee; therefore, rank allocation necessarily creates income inequality between workers. At the beginning of each round, the manager can keep the workers' ranks or switch ranks, i.e., promote the low-rank employee.

In our baseline condition (*No Choice*), we assign the managerial role based on performance in the preceding real effort task. In our treatments, we allow subjects to self-select into the managerial position (*Choice* treatment), and we progressively add the necessity to communicate with employees when assigning ranks (*Choice & Talk* treatment) and the possibility for employees to talk back and send angry emojis to express their disapproval of the rank allocation (*Choice & Backlash* treatment).[3] Our treatment manipulations allow us to test whether women are less likely than men to self-select into the manager role purely due to a differential (un)willingness to create inequalities among workers – this would be observable in the *Choice* treatment – and/or a differential (un)willingness to face employees and provide feedback when promoting or demoting them – this would be observable in the *Choice & Talk* treatment –, and/or a differential (un)willingess to expose themselves to worker backlash – this would be observable in the *Choice & Backlash* treatment. The latter may be caused by expectations of harsher backlash or a higher aversion to backlash, or both.

Our design also allows us to examine gender differences in managerial behavior and outcomes, i.e., the criteria used when assigning ranks, the likelihood of switching worker ranks, the propensity to make mistakes in rank allocation as well as foregone profits due to such mistakes. The main advantage of our experimental setting is that it makes it possible to compare men's and women's managerial behavior and performance in a controlled environment where male and female leaders face the same decision set and incentives.

Through the analysis of the messages that managers send to their employees in the *Choice & Talk* and the *Choice & Backlash* treatments we are also able to examine whether men and women communicate with and motivate employees differently, and whether the

[3] The manager is selected among those who want to be managers based on performance in the preceding real effort task.

threat of worker backlash has a differential impact on men's and women's communication styles. Finally, by comparing the free-form messages and the number of angry emojis sent to male and female managers in the *Choice & Backlash* treatment, we test whether employees have differential attitudes toward male and female managers.

We find evidence of a gender gap in self-selection into leadership only when leaders face the possibility of worker backlash. Neither being responsible for generating income inequality between employees nor having to talk to employees when assigning ranks discourage women, as long as employees cannot send angry messages back. When backlash is possible, men are unaffected, while women are significantly less likely to want to be managers. Once in a leadership position, both men and women assign initial ranks to workers based on their relative productivity. However, when managers have to talk to employees at the rank allocation stage, both male and female managers are more likely to switch workers' ranks in subsequent rounds, therefore giving the initially low-ranked worker a chance. For female managers, the propensity to promote the low-rank employee increases further in the *Choice & Backlash* treatment. An important finding is that managers do not switch ranks to please unhappy employees. In all treatments, male and female managers switch ranks primarily when low-ranking workers are deserving of a promotion, i.e., when they perform equally or better than the other worker. As a consequence, treatment and gender differences in rank-switching behavior do not result in differences in total profits.

The analysis of the messages sent by managers to workers in the *Choice & Backlash* treatment suggests, in line with the existing literature, that men and women have different communication styles. Female managers praise employees more and use cordial words like thank you and sorry, whereas male managers are more likely to explain the ranking decision and to foster competition among workers. However, the comparison of the messages sent under the threat of backlash and in the absence of it, in the *Choice & Talk* treatment, reveals that the observed gender differences in language only emerge when workers can talk back to managers, i.e., they reflect men's and women's different ways to deal with the possibility of worker criticism.

Finally, the analysis of the free-form messages and the angry emojis sent by workers to managers in the *Choice & Backlash* treatment show that high-rank workers send similar messages to male and female managers. On the other hand, low-rank workers have a differential attitude toward male and female managers. They question the ranking decision more if it comes from a female manager, and they send more angry emojis to female than

to male managers.

Our study contributes to the growing literature on the gender leadership gap, which has been documented in all spheres of life. For instance, only 19 percent of firms worldwide have female top managers and only 6 percent of CEOs at S&P 500 companies are women.[4] In politics, women hold only 23 percent of seats in national parliaments worldwide.[5] In academia, averaging across all fields, less than one third of full professors are women. The existing literature has identified behavioral or preference-based constraints to women's self-selection into top leadership roles.[6] These include risk aversion (see, e.g., Eckel and Grossman, 2008), reticence to initiate negotiations (e.g., Bowles et al., 2007; Babcock and Laschever, 2009; Exley et al., 2020), aversion to competitive environments (e.g., Gneezy et al., 2003; Niederle and Vesterlund, 2007; Flory et al., 2014; Preece and Stoddard, 2015),[7] preferences over job attributes (Wiswall and Zafar, 2017), time spent on low promotability tasks (Babcock et al., 2017) and self-stereotyping (Coffman, 2014). A recent study by Born et al. (2020) also shows that being in a male-dominated team reduces women's willingness to lead, due to lower confidence and expectations of team support. We add to this literature by showing that the possibility of backlash from unhappy employees contributes to the leadership gap by inducing women, but not men, to select out of leadership roles.

We also contribute to the literature on gender differences in executives' performance (see, e.g., Adams and Ferreira, 2009; Matsa and Miller, 2014).[8] In line with existing studies, we find that female leaders are more likely to make decisions that benefit low-earning employees, but this holds only if managers have to personally communicate their

[4]For recent statistics on the gender leadership gap in the US, see Warner and Corley (2017)

[5]World Bank DataBank: https://data.worldbank.org/indicator/SG.GEN.PARL.ZS

[6]External demand-side constraints, such as taste-based or statistical discrimination stemming from traditional gender stereotypes concerning men and women's productivities, skills and family constraints, are of course also important. The existing evidence suggests that women are likely to be discriminated against in higher-status jobs, particularly in male-dominated fields. For a review of the literature, see Riach and Rich (2006), Azmat and Petrongolo (2014) and Bertrand and Duflo (2017).

[7]Erkal et al. (2018) show that women are more likely to compete for leadership roles if there is a system in place that, by default, enroll individuals in the competition, while allowing them to opt-out.

[8]The existing studies have generated mixed findings. While female leadership seems to reduce firms' short-term profits, due to fewer workforce layoffs (Matsa and Miller, 2013, 2014), gender diverse boards are more active in monitoring executives (Adams and Ferreira, 2009; Schwartz-Ziv, 2017), improve employees' working conditions (Devicienti et al., 2019), and reduce both the gender pay gap among top executives (Matsa and Miller, 2011) and the gender gap in promotions (Kunze and Miller, 2017). Moreover, Flabbi et al. (2019) find evidence that female CEOs are better at evaluating the productivity of female workers, leading to better allocations of female workers across tasks and to wage distributions that more clearly reflect individual productivities. For a recent review of the literature, see Miller (2017).

decisions to their subordinates, and if they can receive angry messages back as a result. However, since promotions are still primarily driven by worker productivity, the observed gender differences in manager decision-making do not result in significant differences in final outcomes, i.e., total profits.

Our study also relates to the small literature in economics on gender differences in leaders' communication styles (Timko, 2017; Manian and Sheth, 2020), which shows that female leaders send less assertive messages than male leaders (Timko, 2017), even when they know that such messages are more effective in impacting others' behavior (Manian and Sheth, 2020).[9] This is in line with studies in psychology, which provide evidence that men tend to use more assertive language - e.g., through imperative statements - while women use more affiliative language - e.g., statements of support, agreement, and acknowledgment.[10] We contribute to this literature by showing that gender differences in the language used by male and female leaders to motivate employees do not reflect innate gender differences. Rather, they emerge as different strategies used by male and female managers to minimize worker backlash.

Finally, we add to the literature on gender differences in attitudes toward male and female leaders (see, e.g., Abel, 2019; Grossman et al., 2019) by providing evidence of higher criticism and expression of discontent toward female than male managers.[11]

The paper is organized as follows. Section 2 describes the experiment, the treatments and the empirical strategy. Section 3 presents the results on self-selection into leadership (Section 3.1), manager behavior (Section 3.2), manager communication style (Section 3.3), and workers' attitudes toward managers (Section 3.4). Section 4 provides robustness checks and Section 5 concludes.

[9] The operations management literature distinguishes between *transformational* leaders who "transform or change the basic values, beliefs, and attitudes of followers so that they are willing to perform beyond the minimum levels specified by the organization" and *transactional* leaders, who "are founded on an exchange process in which the leader provides rewards in return for the subordinate's effort" (Podsakoff et al., 1990). Based on a meta-analysis of 45 studies, Eagly et al. (2003) conclude that female leaders tend to be more transformational than male leaders, although the difference is small in magnitude.

[10] See, for instance, Leaper and Smith (2004) and Kern et al. (2016).

[11] There is also a large literature on attitudes toward male and female leaders in psychology, sociology and management. These studies typically either provide written description of leadership situations, varying the sex of the leader, or use trained actors to lead, allowing the experimenters to control the degree of success the leader achieves (see, e.g., Swim et al., 1989).

2 The Experiment

2.1 Design

The experiment consists of 6 active stages (Stages 1 to 6), followed by a survey, as shown in Figure 1. An important feature of our design is the method used to reveal subjects' genders to other participants without making gender artificially salient in the game. We achieved this by asking subject to fill in a brief survey at the very beginning of the session, before Stage 1. The survey registers subjects' age, gender, field of study, and previous participation in an experiment. The answer to the gender question leads to a pre-determined list of either male or female names, which we took from Bertrand and Mullainathan (2004)'s correspondence study of race-based discrimination.[12] The male subjects see a list of male names and the female subjects see a list of female names. We inform subjects that for the duration of the experiment they will be identified with a fictitious name, and we invite them to pick a name from the gender-specific list they see on their screen.[13] We do not allow two or more subjects to choose the same name, so each name disappears from the list in real time when picked by another participant.

Stage 1 follows. In this stage, and in the following five stages, subjects engage in a real effort task. Previous studies of leadership have typically used public goods games or coordination games. In order to resemble firm environments where managers are chosen based on qualifications, in our study we wanted a game/task that would allow us to clearly assess participants' relative performance and select the best performing subject in a group as the leader. Specifically, we wanted a gender-neutral task requiring cognitive thinking and focus, where men and women would be equally confident and would perform equally well. We therefore chose a language task, as it has been shown (e.g., Dreber et al., 2014; Niederle, 2016) that language-based tasks are less likely than math-based tasks to generate gender differences in both self-confidence and performance in competitive environments.

In Stage 1, our real-effort task consists in finding a 4-letter word in a 6x6 letter matrix in 5 minutes, for a maximum of 20 matrices.[14] Subjects play individually. They receive an

[12] As our focus is on gender differences, we use Bertrand and Mullainathan (2004)'s list of distinctively white sounding names only. Distinctive names are those that have the highest ratio of frequency in the corresponding racial group.

[13] We did not ask subjects to use their real names as we did not want to lift anonymity nor did we want the potential confounding bias of race, nationality or ethnicity associated with the actual name of the subject to play a role in the experiment.

[14] The decision screen is divided in two halves, as shown in the Online Appendix. On the left, subjects

endowment of 40 Experimental Currency Units (ECU)[15] and earn 2 ECU for each puzzle they solve correctly in 5 minutes. At the end of Stage 1, subjects receive feedback on their performance and are provided instructions on the following 5 stages (Stages 2 to 6) of the experiment. Crucially, they are randomized into groups of 3 and they are shown the fictitious names of their group members. In order to simulate male-dominated environments, the randomization algorithm created groups of 2 men and 1 woman, whenever possible.[16] Since our interest is in work environments where manager and workers interact multiple times, so that a manager is responsible for the long-run income of his or her employees, subjects remain in the same group for the duration of the experiment.

In Stages 2 to 6, two group members play in the role of workers and one in the role of manager. The roles of manager and worker are assigned at the end of Stage 1 according to treatment-specific rules, and retained through Stage 6. The manager gets a fixed wage of 100 ECU and his/her main task is to decide, at the beginning of each stage of the experiment, which worker will be rank A and which worker will be rank B in the following stage. The rank A worker gets a wage of 80 ECU, while the rank B worker gets a wage of 20 ECU. After the rank allocation, all members of the group engage in a similar puzzle task as in Stage 1 of the experiment.[17] Each correctly solved puzzle generates 2 ECU in addition to the initial wage. Moreover, each puzzle solved correctly by the Rank A worker generates 2 ECU also to the manager. This is to incentivize the manager to take the workers' relative productivities into account when making the rank allocation decision. In sum, the earnings from each of the 5 active stages (Stages 2 to 6) of the experiment are determined as follows:

see the matrix and on the right, they see a list of 40 words. Each puzzle contains two words that appear on the list. In order to earn money, subjects have to identify one word per puzzle. We used the website http://tools.atozteacherstuff.com/word-search-maker/wordsearch.php to create the puzzles and the website http://www.thefreedictionary.com to find words of varying lengths. We ran some pilots of the puzzle task with varying levels of difficulty with different sizes of the matrix, different word lengths, and different ways in which words could be identified in the puzzle (forward, backward, up, down, diagonal etc). We found the configuration of finding 4-letter words that appear horizontally or vertically in a 6X6 matrix with a time of 5 minutes in Stage One to be optimal in creating enough heterogeneity in performance among subjects.

[15] Earnings from the experiment were later converted to dollars at the exchange rate of 6 ECU for 1 USD, as explained in Section 2.3.

[16] We ended up having 63% of the groups made of one woman and two men, 30% made of two women and one man, and 7% made of men only.

[17] In order to account for learning effects, while in Stage 2 we keep the time limit to solve the 20 matrices equal to 5 minutes, we reduce the time to 4 minutes in Stages 3 and 4, and to 3.5 minutes in Stages 5 and 6.

- The manager gets 100 ECU + 2 ECU per puzzle + 2 ECU per puzzle solved by the rank A worker;
- The rank A worker gets 80 ECU plus 2 ECU per puzzle;
- The rank B worker gets 20 ECU plus 2 ECU per puzzle.

Stages 3 to 6 are identical to Stage 2. However, at the end of each stage of the experiment, the manager is informed about the performances of the current rank A and rank B workers and has to decide whether to keep or reassign ranks before the next stage begins.

Following Stage 6, subjects fill a post-experiment questionnaire,[18] where we elicit demographics, previous leadership experiences, and answers to personality questions that allow us to generate the Big 5 Agreeableness Personality Index, which has been shown to be significantly higher in women than men (Schmitt et al., 2008) and to correlate negatively with leadership ambition (e.g., Ertac and Gurdal, 2012).[19]

One feature of our design requires further discussion. In our setting, the nature of the task is such that the manager is always able to accurately assess the relative performances of the two employees. This allows us to define and clearly measure managerial performance as: 1) the likelihood that the manager makes the first rank allocation based on performance in Stage 1, and 2) the likelihood that the manager makes an ex-post mistake in the rank allocation, by assigning rank A to a worker who ends up not performing as well as the other worker. We acknowledge that, in many settings, workers' performances cannot be objectively or precisely measured, and a manager's rank allocation decisions are at least partly discretionary; it is the lack of transparency and the subjectivity of the decision process that may be especially conducive to worker backlash. Note that even though our task generates objective workers' rankings, we still allow for lack of transparency and perceived subjectivity of the manager's decisions by not disclosing relative performances to the workers. In other words, the workers do not know how they compare to each other, and do not know what criteria the manager followed to allocate ranks.

[18] Due to a computer glitch, we were unable to conduct the post-experiment questionnaire on 12 participants.

[19] The Agreeableness Index measures the tendency to be kind, altruistic, trusting and trustworthy, and cooperative. There is evidence from psychology studies(e.g., Judge and Bono, 2000) that agreebleness predicts transformational leadership, i.e. leadership that operates through inspiration, intellectual stimulation and individual consideration.

2.2 Treatments

In our *No Choice* (NC) treatment, at the end of Stage 1, in each group of three participants the manager is chosen based on performance in Stage 1.[20] Recall that subjects participate in the Stage 1 real effort task individually without knowing anything about Stages 2 to 6, and therefore ignoring the fact that their performance would determine their role in the subsequent stages of the experiment. This prevents competition-driven anxiety from playing a role in determining subjects' performance and subsequent chances of becoming the manager of the group.[21] At the beginning of Stage 2, subjects are informed that the manager was chosen based on performance rather than randomly. This is important, as we aimed to simulate an environment where employees could not doubt the qualifications of their manager. This way, any differences in workers' attitudes toward male versus female managers could not be attributed to differential subjective beliefs about the right of the manager to hold his or her role in the group.

In our *Choice* (C) treatment, we allow subjects to self-select into the leadership position. At the end of Stage 1, after receiving information about the next five stages of the experiment and the fictitious names of their group members, we ask subjects to state whether they would like to be the manager of their group. From the subset of those who want to be manager, we choose the manager based on performance in Stage 1, as in the *No Choice* treatment.[22] Since, by design, the manager is always the highest earner of the group, everybody should want to be manager. Any observed gender difference in self-selection into leadership in this treatment can be attributed to differences in willingness to generate income inequality among workers.[23]

[20] Since, by design, we have more men than women participating in each session of the experiment – due to the objective of having male-dominated groups – we break ties in favor of women. Subjects are unaware of this.

[21] While there is a large literature documenting gender differences in competitiveness and self-confidence (especially in math-based tasks) we wanted to abstract from both factors in our experiment, in order to be able to isolate the role played by the possibility of worker backlash in the origination of gender differences in leadership.

[22] As in our *No Choice* treatment, ties are broken in favor of women, yet subjects are unaware of it.

[23] Individuals may perceive the manager selection as a competitive process, hence, it is possible that gender differences in competitiveness (Niederle and Vesterlund, 2007) may hold women back. However, contrary to standard competition tasks, in our setting not becoming manager, i.e., losing the competition, does not come with a loss, since the consequence would be that the subject plays the game as a worker, i.e., same outcome as if he or she did not self-select into leadership. In other words, there is no risk in wanting to be manager. Still, there may be a gender difference in willingness to receive negative feedback about the outcome of the competition, as it could reveal information about one's relative ability in the real effort

In the *Choice & Talk* (CT) treatment, we inform participants that the manager will have to assign ranks to workers in each stage, as before, and will also have to communicate with each worker when assigning ranks by sending a free-form message to the rank A and the rank B workers before the beginning of the next stage of the real effort task. Subjects are told that the manager can write anything he or she wishes to communicate to each worker. The process is repeated at each rank-allocation stage. As in the *Choice* treatment, after providing information about the next stages of the game, including tasks and payoffs of manager and workers, as well as the fictitious names of their group members, we ask subjects to state whether they want to be the manager of their group. If more than one group member wants to be manager, we select the manager following the same procedure as in the *Choice* treatment. In this treatment, in addition to willingness to generate income inequality among workers, an aversion to having to repeatedly provide feedback to subjects who are affected either positively or negatively by the rank allocation decision, may lead subjects to select out of the manager role[24] and impact men and women differentially.

Finally, in our *Choice & Backlash* (CB) treatment, the manager still has to send a free-form message to the rank A and the rank B workers, yet in this case, each worker can send a free-form message back to the manager. Moreover, the worker assigned rank B can send up to 5 angry emojis to the manager to express disapproval of the rank allocation. The messages sent by the two workers, including the angry emojis, are displayed to the manager before the next real-effort task begins. The process is repeated at each rank-allocation stage. As before, participants receive information about the rules governing Stages 2 to 6 of the experiment, including the presence of two-way communication and the possibility of receiving angry emojis from Rank B workers, before they are asked whether they would like to be the manager of their group.

The *Choice & Backlash* treatment most closely simulates a work environment where managers have to communicate their promotion or task allocation decisions to employees, knowing that employees will be able to talk back and express their discontent. By comparing men's and women's behaviors in this treatment versus the *Choice* and the *Choice & Talk* treatments we are be able to identify the role of anticipation of worker backlash on self-selection into a leadership role. Note that the decision to use angry emojis in the

task. Even if that were true, it would apply to all our self-selection treatments; therefore, it would not act as a confound in our study of the behavioral mechanism of interest.

[24]For instance, subjects may feel uncomfortable at the thought of having to talk to workers assigned to be rank B, possibly multiple times during the course of the experiment.

Choice & Backlash treatment follows the aim to make both workers and manager aware that expression of disapproval of the manager's decision is possible and allowed in the experiment - which is not easy to achieve with the use of free-text messages only, and without framing the instructions in a way that may prime or lead subjects to say specific things to each other. Moreover, emojis have become a widespread form of communication among the younger generations, and the meaning of an angry emoji is universally understood. Therefore, the use of emojis allows us to: 1) hold the boundaries of potential worker discontent constant (between 0 and 5); 2) quantify the attitudes of workers toward managers; and 3) examine how male and female managers respond to angry messages from employees, holding such messages constant.[25]

In addition to self-selection into a manager role, our design allows us to investigate gender differences in managers' decision-making. By design, only the performance of the rank A worker generates additional earnings to the manager. Therefore, the manager should assign rank A to the worker that he or she expects to be the best performer in the upcoming stage. The only information available to the manager is the workers' past performances, so we expect the difference in the workers' past productivities to play a primary role in the rank allocation decisions. However, managers' distributional concerns and/or the desire to avoid the discomfort of having to tell the same worker that he or she has been assigned rank B multiple times, and/or the objective to minimize the receipt of angry emojis, may lead managers to switch ranks between the two workers over the course of the experiment in order to equalize their earnings. Our treatment comparison allows us to examine gender differences in managers' rank switching behavior, and the mechanisms leading to them. We are also able to assess whether differential rank switching propensities lead to different outcomes, i.e., mistakes in rank allocations and foregone profits.

Finally, by examining the free-form messages sent by managers to workers in the *Choice & Talk* and in the *Choice & Backlash* treatments, not only can we study gender differences in communication styles, but we can also examine the extent to which the languages used by male and female managers change under the threat of worker backlash. Furthermore, the analysis of both free-form messages and angry emojis sent by workers to managers provides a comprehensive assessment of differences in workers' attitudes toward male and

[25] We do not allow rank A employees to send happy emojis. Although this could be an interested extension of our study, as we discuss further in the Conclusion, the focus here is the anticipation of negative judgment and hostile messages from unhappy employees, hence the use of angry emojis only.

female managers.

2.3 Implementation

We conducted 29 experimental sessions at the Laboratory for Research In Experimental Economics (LREE) at Southern Methodist University. We involved a total of 417 participants, of which 41% are women, as shown in Table 1. We employed a between-subject design, with each subject participating in only one session and one treatment. In each session, we had multiple independent groups of one manager and two workers. Groups were fixed for the duration of the experiment, and members of each group made decisions independently from all the other groups participating in a session.

As described in Section 2.2, the experiment consisted of an initial brief survey and name-assignment stage, followed by six active stages plus a post-experiment survey. Subjects were presented with the instructions for each stage on their computer screen immediately before that stage began. Additionally, at the beginning of Stage 1, they received hand-outs and verbal instructions about the puzzle-solving task, and at the end of Stage 1 they received further verbal and written instructions (hand-outs) about the rules applying to Stages 2 to 6 of the experiment.[26]

Only one randomly selected active stage of the experiment was used for actual payments. Experimental earnings were converted from ECU to dollars at the exchange rate of $1 for 6 ECU. The experiment was programmed in z-Tree (Fischbacher, 2007) and subjects were recruited among pre-registered LREE students.

In order to guarantee anonymity, at the beginning of each session subjects were randomly assigned an identification number, which they kept for the duration of the experiment. At no point during the experiment did we ask subjects to reveal their names and, although actual names were used during the payment process for accounting purposes, we informed subjects that we would not register their names and therefore would not be able to link them to the choices made in the experiment. Each session lasted between 60 and 90 minutes, with average earnings of $28 per subject, including a $10 show-up fee.

2.4 Empirical Strategy

We estimate self-selection into the manager role using equation 1 below,

[26] The experiment instructions are provided in the Online Appendix.

$$Y_i = \beta_0 + \beta_1 dF_i + \beta_2 dT_{3i} + \beta_3 dT_{4i} + \beta_4 dF_i * dT_{3i} + \beta_5 dF * dT_{4i} + \beta_6 P_i + \delta \mathbf{X}_i + u_i \quad (1)$$

where Y_i is participant i's stated desire to be the manager of the group; dF_i is a dummy equal to 1 if subject i is a woman; dT_{3i} is a dummy equal to 1 if i is in the *Choice & Chat* treatment and dT_{4i} is a dummy equal to 1 if i is in the *Choice & Backlash* treatment. The *Choice* treatment is the excluded category. We also control for the performance in Stage 1 of the experiment, P_i, and include the interactions between each treatment dummy and the female dummy. We report p-values generated by Wald tests for linear combinations of the estimated coefficients of interest, e.g. for $\beta_2 + \beta_4 = 0$. In our most comprehensive specification, we include X, a vector of demographic controls, i.e., age, field of study, being a native English speaker, having held a leadership position outside the lab and the Big 5 agreeableness measure.[27] We estimate equation (1) using linear probability models, since interpreting interaction terms in non-linear models is not straightforward, as the marginal effect of an interaction term may not be the same as the estimated coefficient, and further the standard t-test is inaccurate (Norton et al., 2004). Even though we employ a between-subject design and groups are independent from each other, the gender composition of the group may affect self-selection into leadership, as recently shown by Born et al. (2020); therefore, we cluster the standard errors at the group level. In our most comprehensive specification, we also control for the number of women in the group.

We adopt the same estimation strategy when examining gender difference in managers' decision-making, except that the analysis now includes the *No Choice* treatment and explores the five stages of the experiment in which managers made active decisions affecting the workers in their group. In these regressions, we still keep the *Choice* treatment as our benchmark category, to be able to directly identify the impact of self-selection, manager feedback and possibility of backlash on leaders' choices. We estimate equation 2 below:

[27] Note that, while age and field of study where registered in Stage 0 of the experiment and are available for all subjects, the other control variables were collected at the end of experiment, and due to a software glitch with one session, these variables we missing for 12 experimental participants. In all regressions, we present results with and without these additional controls.

$$Y_{it} = \beta_0 + \beta_1 dF_i + \beta_2 dT_{1i} + \beta_3 dT_{3i} + \beta_4 dT_{4i} + \beta_5 dF_i * dT_{1i} + \beta_6 dF_i * dT_{3i} \\ + \beta_7 dF_i * dT_{4i} + \beta_8 DP_{t-1} + \delta \mathbf{X}_i + \lambda_t + u_{it} \tag{2}$$

where Y_{it} is manager i's decision in Stage t, dF_i is a dummy equal to 1 if the manager is a woman; dT_{1i} is a dummy equal to 1 if the manager is in the *No Choice* treatment, dT_{3i} is a dummy equal to 1 if the manager is in the *Choice & Talk* treatment, dT_{4i} is a dummy equal to 1 if the manager is in the *Choice & Backlash* treatment. We control for the difference in the performances of the two workers in the previous Stage, DP_{t-1}, as it is likely to affect manager decision-making. We also include the interactions between each treatment dummy and the female dummy. As before, we report p-values generated by Wald tests for linear combinations of the estimated coefficients of interest, e.g. for $\beta_2 + \beta_5 = 0$. In our most comprehensive specification, we employ the same set of controls as in equation 1. Finally, we include stage fixed effects in all regressions and, as before, we cluster the standard errors at the group level, which in this case it is the equivalent of clustering the standard errors at the individual level, as we have the same manager per group for the duration of the experiment. We employ four measures of managerial decision-making: 1) whether in Stage 2 (i.e., the first rank allocation), the manager assigns rank A to the best performing worker in stage 1; 2) whether the manager decides to subsequently switch ranks and promote the rank B worker; 3) Whether the manager makes an ex-post mistake in the rank allocation by promoting a worker who ends up performing worse than the other worker, and 4) the amount of foregone profits caused by mistaken rank allocations.

In order to test whether male and female managers are more or less likely to promote a rank B worker after receiving angry emojis, we restrict the analysis to the *Choice & Backlash* treatment and estimate equation 3 below:

$$Y_{it} = \beta_0 + \beta_1 dF_i + \beta_2 A_{it-1} + \beta_3 dFM_i * A_{it-1} +_i + \beta_4 DP_{t-1} + \delta \mathbf{X}_i + \lambda_t + u_{it} \tag{3}$$

where Y_{it} is a dummy equal to 1 if manager i, at the beginning of Stage t promotes the worker who was rank B in Stage $t-1$, dF_i is a dummy equal to 1 if manager i is a woman, A_{it-1} is the number of angry emojis that i received from the rank B worker when allocating ranks at at $t-1$, DP_{t-1} is the difference between the performance of the rank A

and the rank B workers at $t-1$. Importantly, to test whether male and female managers respond differently to the receipt of angry emojis, we include the interaction between the gender of the manager and the number of angry emojis received from the rank B worker at $t-1$. As before, we include stage fixed effects and cluster the standard errors at the group level.

Finally, when analyzing the number of angry emojis that rank B workers send to male versus female managers in the *Choice & Backlash* treatment, we estimate equation 4 below,

$$Y_{jit} = \beta_0 + \beta_1 dF_i + \beta_2 P_{jt-1} + \beta_3 dB_{jt-1} + \delta \mathbf{X}_j + \lambda_t + u_{jit} \tag{4}$$

where Y_{jit} is the number of angry emojis, including 0, that the rank B worker j sent to manager i at the beginning of Stage t, dF_i is a dummy equal to 1 if manager i is a woman, P_{jt-1} is the performance of worker j in the task in Stage $t-1$, and dB_{jt-1} is a dummy equal to 1 if worker j was rank B also in Stage $t-1$. We include our set of controls, stage fixed effects and we cluster the standard errors at the group level. To account for the fact that rank B workers can send angry emojis *after* reading the free-form message sent by their manager, we categorize the manager messages based on their content and in our most comprehensive specification we also control for the type of message sent by the manager.

3 Results

We start by describing our subject pool and conducting balance tests across treatments. A total of 245 men and 172 women participated in the experiment. In Table A1, in the Appendix, we report descriptive statistics for our male and female sample pools, i.e. their average age, whether they were majoring in STEM, Business or Economics or in a different field, whether they were native English speakers, whether they reported having held a leadership position, and their average Big 5 Agreeableness Index. The average age is 22.9, with no significant gender differences across treatments and genders. Most of our participants are STEM, Business or Economics majors, although the percentage of men majoring in these fields (83%) is significantly higher than the percentage of women (64%, $p = 0.000$). The percentage of men studying STEM or Economics is higher in the *Choice* treatment than in the *Choice & Talk* ($p = 0.044$) treatment. There are no significant differences across treatments for female participants. Most subjects reported having held

a leadership role in the past, with no significant differences across treatments and genders. About 41% of men and 55% of women are native English speakers, ($p = 0.008$). We do not see any significant differences in the percentages of native speakers across treatments for women, whereas we see fewer native speakers among men in *Choice* than in the *Choice & Talk* ($p = 0.011$) and in the *Choice & Backlash* ($p = 0.048$) treatments. In line with the existing literature, our female participants score significantly higher than men in the Big 5 Agreeableness Index ($p = 0.004$), with no significant differences across treatments for both the male and the female samples.

Before presenting and discussing our main findings, we assess possible gender differences in the performance in the real effort task employed in the study. Recall that we aimed to design a real effort task that would be as much as possible gender-neutral. On average, subjects solved 14 out of 20 puzzles correctly in Stage 1. The only difference we see across genders and treatments is that women are better at the task in the *Choice & Talk* treatment than in the other treatments ($p = 0.04$). Table A2 in Appendix shows estimates from regression analysis where the dependent variable is the number of correctly solved puzzles per stage. The table shows that there are no statistically significant gender differences within and across treatments, which suggests that we were successful in employing a gender-neutral task.

We also wanted our real effort task to lead to stable performance rankings within a group. In other words, we wanted to employ a task whereby being the best(worst) performer in the task in Stage 1 would be a good predictor of the likelihood of being at the top(bottom) of the group in the subsequent stages. This is what we see in the data. Descriptive statistics[28] show that if a subject is the top(bottom) performer in Stage 1, he or she is the top(bottom) performer in about 80% of the following stages. When restricting the analysis to the two workers, the best performer in Stage 1 is the best performer in 77% of the following stages, i.e., about 4 of the following 5 stages.

In what follows, we present and discuss the core results of the paper, i.e., the effects of our treatments on men's and women's self-selection into the manager position (Section 3.1), and on their decisions once in the manager role (Section 3.2). We then present our findings on male and female managers' communication styles (Section 3.3). We conclude by reporting on the attitudes of workers toward male and female managers, as measured by

[28] Not presented here but available upon request.

the free-form messages and the number of angry messages sent to managers in the *Choice & Backlash* treatment (Section 3.4).

3.1 Self-selection into the manager position

Almost all subjects wanted to be a manager in both our *Choice* and our *Choice & Talk* treatments, with no significant differences between men and women, as shown in Table 2 and Figure 2. However, a large and statistically significant gender gap emerges in our *Choice & Backlash* treatment, where we see 78% of women self-select into the manager role as opposed to 95% of men ($p = 0.007$). The decline in managerial ambition among women is significant compared to both the *Choice* ($p = 0.055$) and the *Choice & Talk* treatments ($p = 0.011$). On the other hand, we do not see a decline in self-selection into the manager role in the *Choice & Talk* as compared to the *Choice* treatment. This suggests that, absent the possibility of worker backlash, women are not less willing than men to assume the leadership role in our setting. In other words, the gender leadership gap that we observe is due purely to gender differences in the reaction to the possibility of receiving negative messages from low-rank workers.

This is confirmed by the regression analysis displayed in Table 3. We start by testing for treatment effects by including our treatment variables only. We then gradually add interaction terms and controls. Note that we lose a few observations when controlling for past leadership positions, being a native English speaker and the Big 5 Agreeableness index. This is due to a software glitch that prevented us from conducting the post-experiment survey for 12 participants in the *Choice* treatment. Among the control variables, we also include individuals' performance in the real effort task in Stage 1, as this may affect subjects' perceived probability of becoming the manager. Finally, following Born et al. (2018), which show that the gender composition of a group may affect women's self-selection into leadership, we include a dummy equal to 1 if there were two women in the group. Recall that by design we aimed to have groups with one woman and two men to simulate male-dominated environments. However, we ended up having two women and one man in 30% of the groups, and three men in 6% of the groups. In Section 4, we conduct robustness checks where we restrict the analysis to groups made of two men and one woman.

The estimates in Table 3 confirm that women, but not men, are less likely to self-select into the manager role only when facing the possibility of worker backlash (*p-value*=0.037,

Wald test for the sum of coefficients of *Choice & Backlash* and its interaction with the female dummy in Column 3, and *p-value*=0.082 in column 4, where we lose a few observations). Importantly, the likelihood of self-selection is also significantly lower in the *Choice & Backlash* treatment when compared to the *Choice & Talk* treatment (*p-value*=0.008 in Column 3 and *p-value*=0.012 in Column 4), suggesting that it is the possibility of backlash from employees and not the necessity to give feedback to employees that prevents women from wanting to be managers. In particular, women are 17 percentage point less likely to self-select into the manager role in the *Choice & Backlash* treatment as compared to the *Choice&Talk* treatment.[29] We report the treatment effects estimated for the female sample in Table A4 in Appendix. Table 3 also provides evidence of a 16 percentage point gender leadership gap in the *Choice & Backlash* treatment only.

Among the controls, we find that performance in the task is a strong predictor of women's managerial ambition.[30] We therefore examine whether gender differences in the decision to volunteer are more extreme for bottom performers than for the middle and top performers. Figure A1 in Appendix reports the percentages of volunteers among men and women in all treatments by performance terciles.[31] In the *Choice* and *Choice & Talk* treatments, we do not see evidence of significant gender differences in self-selection into the manager role for any tercile. In contrast, under *Choice & Backlash*, significant gender differences in volunteering exist among both bottom ($p = 0.065$) and middle performers ($p = 0.009$). Importantly, the gender gap closes among the very top performers ($p = 0.883$), i.e. the men and women who solved 20 out of 20 puzzles in Stage 1. The figure also shows that while there are no significant differences in self-selction among performance terciles in the male sample for either treatment, in the *Choice & Backlash* treatment, female top performers are significantly more likely to volunteer than female bottom performers, treatment ($p = 0.058$). This suggests that, although in our setting the performance of the

[29] They are 14 percentage point less likely to self-select in *Choice & Backlash* than in the *Choice* treatment.

[30] For women, the only robust determinant of volunteering decision is the field of study, with women majoring in STEM, Business and Economics being more likely to want to be managers. This is essentially indicating that women who have already self-selected into fields that are typically male-dominated and conducive to competitive high-paying jobs, are also more likely to self-select into leadership roles in the experiment. In the male sample, the only significant variable is the Big 5 Agreeableness score, which, in line with the existing studies, appears with a negative sign, suggesting that less agreeable, hence more competitive and aggressive men, are more likely to self-select into the manager role.

[31] The bottom tercile is made of subjects who completed 12 or less puzzles correctly in Stage 1 of the experiment. The middle tercile consists of students who completed more than 12 puzzles but less than 20 puzzled in Stage 1. The top tercile is made of students who completed 20 out of 20 puzzles in Stage 1.

manager is never disclosed to the employees and the manager is never in competition with his or her workers, women feel that, in order to be managers of their group in a setting when employees can express negative feedback, they need to be extremely good at the task. Men do not seem to have the same concerns.

3.2 Managers' decision-making

When comparing the behaviors of male and female managers in stages 2 to 6 of the experiment, we consider a number of outcome variables. First we examine whether at the time of the first rank allocation, in stage 2, the manager assigns rank A to the most productive worker in stage 1 (Section 3.2.1). We then assess whether, in the following stages, the manager switches ranks and promotes the rank B worker, while controlling for the relative performance of the two workers in the previous stage (Section 3.2.2).

Ideally, we would want a measure of managerial success. A measure of performance that can be compared among managers is the likelihood of making a mistake when assigning ranks to workers. The mistake is revealed ex-post, i.e., if the worker chosen to be rank A in a given round ends up performing worse than the other worker. In Section 3.2.3 we examine male and female managers' likelihood of making such mistakes across treatments. Ranking mistakes cause losses in profits, given that only the rank A worker generates money to the manager. In Section 3.2.4 we analyze the foregone profits caused by ex-post mistakes. Finally, we examine the total profits that male and female managers made from the productivity of the (chosen) rank A workers over the five stages of the experiment, and assess the inequality in the total workers' earnings.

3.2.1 First rank assignment

We start by examining the likelihood that the first rank assignment, at the beginning of Stage 2, is based on workers' relative productivity, i.e., the number of correctly solved puzzles in the previous stage. Specifically, we ask whether the manager assigned rank A to the worker who solved more puzzles in Stage 1. Overall, 94% of the managers assigned rank A to the best performing worker in Stage 1.[32] Self-selected managers seem more likely to assign rank A to the best worker in stage 1 as compared to exogenously chosen

[32] If the two workers solved the same number of puzzles, we code our outcome variable as a 1, no matter the worker who was chosen to be rank A.

managers. The difference is statistically significant for women at the 10 percent level (98.36% versus 89.89%, Chi-Square test *p-value*=0.065), and statistically insignificant for men (93.18% versus 81.25% Chi-Square test *p-value*=0.173). We do not see any significant differences across the *Choice* treatments for either male or female sample.[33] We do not see any significant gender or treatment differences in the likelihood of assigning ranks based on past performance, as shown in Table 4. In the regressions, the strongest predictor of the likelihood of allocating rank A to the best performing worker is the difference in the numbers of puzzles that each worker solved correctly in stage 1.

3.2.2 Rank-switching and promotions

We have previously noted that it is generally the case that the best performing worker in stage 1 performs better than the other worker in the following stages. On average, the best worker in stage 1 outperforms the other worker in nearly 4 of the 5 subsequent stages. Therefore, once ranks are established based on past performance at the beginning of Stage 2, switching ranks and promoting the rank B worker is likely to result in a mistake, i.e. to generate less profits than if the ranks had been kept constant. However, a desire to equalize workers' payoffs, or to reward the rank B worker if he or she performs equally or better than the rank A worker in a given stage, or to placate the worker's anger and reduce the likelihood of backlash, may induce managers to switch ranks across stages. In this section we analyze treatment-specific gender differences in rank-switching behavior.

Overall, averaging all treatments, female managers switch ranks 1.44 times, which is higher than the 1.03 times observed for male managers (*p-value*=0.053). Figure 3 displays the average number of rank switches by male and female managers over the five stages of the experiment by treatment. It shows that self-selected male and female managers tend to switch ranks more when having to repeatedly provide feedback to the two employees. Such tendency seems to increase further for female managers under the threat of backlash, but not significantly so. Figure A2 in the Appendix distinguishes between rank switches that led to the promotion of a rank B worker who performed worse than the rank A worker, and therefore did not deserve to be promoted, and rank switches that were justified by the fact that the rank B worker performed equally or better than the rank A worker. The likelihood of promoting a rank B worker when he or she underperformed is low no matter

[33] The statistics are not displayed in a table, but available upon request.

the gender of the manager. On average, low performing rank B workers were promoted in only 12 percent of all cases, with no significant differences across treatments for either male or female managers. Promotions, i.e., rank switching, are significantly more likely in all treatments if the rank B worker performed equally or better than the rank A worker.[34]

We estimate equation 2 of Section 2.4 in Table 5. By doing so, we are able to control for the difference in the performances of the two workers, which is a strong predictor of manager ranking decisions. While the estimates provide no evidence of gender differences in the likelihood to switch workers' ranks over time, they confirm that having to talk to workers after assigning ranks induces both male and female managers to switch ranks more often. Moreover, the possibility of worker backlash further increases female managers' likelihood to promote the rank B worker, although the difference with the *Choice & Talk* treatment is only significant at the 10 percent level ($p = 0.087$ based on the estimates in column 3, and $p = 0.098$ based on the estimates in column 4). The estimated treatment effects, which, for the female sample, are displayed in column 2 of Table A3 in the Appendix, reveal that having to communicate with workers when assigning ranks increases male and female managers' likelihood of switching ranks by about 16 and 13 percentage points, respectively. Such likelihood increases further for female managers by about 15 percentage points under the possibility of worker backlash.

3.2.3 Ex-post mistakes in rank allocation

In every stage, workers are first assigned rank A or rank B and then they perform in the real effort task. Here, we assess whether managers assigned rank A to a worker who ended up performing worse than the rank B worker, leading to a loss in profits.[35] Figure 4 shows the percentage of ex-post mistakes in rank allocations, aggregated over Stages 2 to 6, by treatment and gender of the manager. The figure suggests that self-selected female managers are less likely to make mistakes in rank allocations as compared to exogenously chosen female managers (Chi-square test *p-value*=0.003), everything being equal, i.e., when

[34] Overall, male and female managers promote deserving rank B workers respectively 56% and 65% of the times (*p-value*=0.215). Having to talk to workers induces male managers to switch ranks significantly more frequently (*p-value*=0.014). There is an increase also for female managers in the *Choice & Talk* treatment, but it is not statistically significant. The possibility of angry messages from employees induces women to switch ranks more often, as compared to the Choice and the *Choice & Talk* treatments, although only the former difference is statistically significant at the conventional level (*p-value*=0.008 when comparing to *Choice* and *p-value*=0.092 when comparing to *Choice & Talk*).

[35] Recall that the manager earns 2 ECU for each puzzle that is correctly solved by the rank A worker.

comparing the *No Choice* and the *Choice* treatments. Self-selection does not matter for male managers, most likely because virtually all men self-select into leadership anyway.

Table 6 presents the results generated by regression analysis. The estimates confirm that female managers in the *No Choice* treatment are more likely to make mistakes than self-selected managers in the *Choice* treatment (*p-value*=0.028 in column 3 and *p-value*=0.024 in column 4). Both having to repeatedly communicate with workers and the possibility of receiving angry messages back, have no impact on female managers' likelihood of making a mistake when allocating ranks to workers. On the other hand, there is some evidence, albeit weak, that the possibility of worker backlash induces male managers to make fewer mistakes, resulting in a significant gender difference in the likelihood of ex-post mistakes in the *Choice & Backlash* treatment only.

3.2.4 Manager's profits

As a final assessment of managerial performance, we compute foregone profits caused by possible mistakes in rank allocations, as well as the total profits generated by the the workers assigned rank A to the manager over the five stages of the experiment.

Foregone profits capture the economic consequences of any treatment- or gender-driven difference in managers' decision making highlighted in the previous subsections. Foregone profits are equal to zero if the manager assigns rank A to the worker who ends up being the best performer in the current stage. If this is not the case, foregone profits are defined as the difference between the profit that the manager would have made by allocating rank A to the best performing worker (i.e., the counterfactual) and the profits that he or she actually made. Recall that each worker (and manager) could solve up to 20 puzzles. Each puzzle solved correctly by the rank A worker generates 2 Experimental Currency Units (ECU) to the manager.

Figure 5 shows average foregone profits, expressed in ECU, by treatment and gender. Foregone profits are low on average, i.e., below 2 ECU. This is both because mistakes only happen 22 percent of the times on average, and because, when they do happen, the difference in the performances of the two workers tends to be small, i.e., less than 4 puzzles on average. Results from regression analysis, displayed in Table 7, show no evidence of gender or treatment effects on foregone profits. In other words, neither having to talk to employees when assigning ranks, or facing the possibility of receiving angry messages from

low-ranked workers leads managers to make decisions that ultimately result in financial losses. This is true for both male and female managers.

Finally, in Table 8 we compare the total profits that male and female managers made over the 5 stages of the experiment from the productivity of the worker assigned rank A in any given stage. We find no evidence of gender differences in such profits. The only significant treatment effect is the positive impact of the *Choice & Backlash* treatment on the profits generated by rank A workers to male managers, albeit significant only at the 10 percent level.

Table 8 displays also the difference in the total earnings of the two workers matched with a male versus a female manager. This is to assess whether treatment-induced differences in managers' behaviors result in differences in the level of income inequality between the two workers. While we do not find evidence of gender differences in inequality between workers in any of the treatments, Table 8 shows that the difference in the workers' total earnings declines significantly when female managers have to repeatedly talk to workers and face the possibility of backlash.

3.2.5 Reaction to angry emojis

As a final assessment of gender differences in managers' decision-making, we examine how male and female managers react to the receipt of angry emojis in the *Choice & Backlash* treatment. Rank B workers could send up to 5 angry emojis to their manager after learning their rank, at the beginning of each stage (Stages 2 to 6). We analyze workers' likelihood of sending angry emojis, and the number of emojis sent, in Section 3.4. Here, we assess whether managers are more likely to promote a rank B worker after receiving angry emojis from him or her, and whether the answer depends on the manager's gender.

We report our estimates of equation 3, presented in Section 2.4, in Table 9. We employ the same specification as in Table 5, except that the analysis is restricted to the *Choice & Backlash* treatment. The dependent variable is still the decision to switch ranks and promote the rank B worker at the beginning of stage t. However, we now include the number of angry emojis that the rank B worker sent to the manager when he or she was notified of the assigned rank in the previous stage. Importantly, we interact the number of angry emojis received with the gender of the manager.

As noted when looking at managers' rank-switching behavior, female managers in the

Choice & Backlash treatment are more likely to promote the rank B worker, controlling for the difference in performance between the two workers. While the receipt of angry emojis does not affect male managers' ranking decisions, it actually lowers the likelihood that a female manager would promote the rank B worker by 7 percentage points, everything else being equal ($p = 0.002$, Wald test of the linear combination of angry emojis and its interaction with the female dummy). This indicates that the observed higher propensity of female managers to promote the rank B worker in the *Choice & Backlash* treatment as compared to the other *Choice* treatments is not a response to *received* angry emojis and is not due to a desire to appease employees who have expressed disapproval of the ranking decision.

3.3 Managers' communication styles

In the *Choice & Talk* and the *Choice & Backlash* treatments, at the beginning of each stage, the manager had to send free a form message to each worker following the rank allocation decision and before the real effort task began. Crucially, in the *Choice & Talk* treatment, workers could not send messages back - either through text or through angry emojis - to the manager. This treatment variation allows us to examine gender differences in communication styles with and without the possibility of worker backlash. Specifically, it makes it possible to assess whether any gender differences in communication styles possibly observed in the *Choice & Backlash* treatment are the result of intrinsic differences in the way men and women provide feedback and motivate employees, or if they are the result of gender-specific strategies aimed at minimizing worker backlash.

We conducted text analysis of the messages sent by managers over the 5 stages of the experiment. We identified seven primary types of messages: 1) praising the worker for past performance, e.g., "amazing work!"; 2) using motivating words, e.g., "You got this!"; 3) providing an explanation for the rank allocation, e.g., "the other worker did better"; 4) inducing competition among workers, e.g., "the other worker is catching up so keep it up"; 5) mentioning fairness, e.g., "trying to be fair"; 6) using team building words, e.g., "let's go team!"; and 7) using cordial words like "thank you" and "sorry." Table A4 in Appendix provides additional examples of the messages of each kind sent by managers to workers. Table 10 and 11 show the percentages of messages within each category sent by male and female managers to rank A and rank B workers, respectively, over the five stages of the

experiment, in the *Choice & Talk* and the *Choice & Backlash* treatments.

Both male and female managers sent primarily praising and motivating messages to rank A workers. In the *Choice & Talk* treatment, where workers could not send messages back, we do not see any gender difference in the type of sent messages. Gender differences however emerge when worker backlash is possible, as male managers praise less, explain more and become more likely to induce competition between workers. The only change we see among female managers in the *Choice & Backlash* treatment is in that they now also explain the rank allocation more, but still to a less extent than male managers.

The comparison of the messages sent to rank B workers generates similar insights. We see a gender gap in the propensity to provide explanations at the rank allocation stage in the absence of worker backlash. When workers can send messages back, male managers send fewer praising and motivating messages, and increase explanations and messages fostering competition between workers. Female managers respond to the possibility of backlash by also praising employees less and explaining themselves more. Contrary to male managers, they do not send messages inducing competition. Instead, they tend to use more cordial words, such thank you and sorry. Notably, the resulting gender difference in the propensity to send cordial messages to the rank B workers is purely due to the possibility of worker backlash. We do not find any gender difference, or treatment effects, in the propensity to use team building words, or to mention fairness.

In sum, male and female managers have similar communication styles when workers cannot talk back. However, when workers can send messages to managers, we see that female managers are more likely than male managers to praise workers and use cordial words, whereas male managers are more likely to explain the ranking decision and foster competition between workers. These gender differences are in line with previous studies of differences in the language used by men and women, both in psychology (e.g., Kern et al., 2016) and economics (Timko, 2017). However, our treatments reveals that such gender differences are generated by differential responses to the possibility of worker backlash, rather than by innate gender differences in communication styles.

3.4 The attitude of workers toward male and female managers

In the *Choice & Backlash* treatment, rank A and rank B workers sent free-form messages to their manager at the rank allocation stage, after receiving the manager's free-form message.

Moreover, the rank B workers could send up to five angry emojis to the manager. We conducted text analysis of the messages sent by the Rank A workers and by Rank B workers. The most common messages can be categorized as follows: 1) thanking the manager or joking; 2) expressing approval of the ranking decision; 3) expressing commitment to work hard; 4) apologetic; 5) containing negative feelings, i.e., upset; 6) questioning the ranking decision. Table A5 in Appendix provides examples of the messages belonging to each category.

We report the percentage of messages of each kind sent to male and female managers by rank A and rank B workers in Tables 12 and 13, respectively. Rank A workers primarily thanked the managers and committed to work hard, with no differences based on the gender of the manager. We also do not see significant differences in the messages sent by rank B workers, with one exception. Female managers are more likely to receive messages in which rank B workers question the ranking decision. This happens 17 percent of the times for female managers, versus 6 percent for male managers (p=0.032).

Next, we examine the angry emojis that Rank B workers sent to male versus female managers in the *Choice & Backlash* treatment. A first look at the data shows that male and female managers were equally likely to receive at least one angry emoticon from rank B workers. Over the 5 stages of the experiment, both male and female managers received at least one angry emoticon about 45 percent of the times.[36] Panel b of Figure 6, however, shows that, conditional on receiving at least one emoticon, female managers tend to receive more angry emoji on average.

In Table 14, we estimate equation 4 of Section 2.4. In particular, we conduct regression analysis of the decision of Rank B workers to send one or more angry emojis to their manager over the 5 stages of the experiment. We control for the performance of the worker in the previous stage, which is likely to affect the worker's belief about his or her right to be rank A in the current stage, and whether the worker was rank B also in the previous stage, which may increase his or her frustration with the ranking decision. We then include demographics, and finally, in column 4, we control for the content of the free-form message (see Table 11) sent by the manager,[37] since the rank B worker sees such

[36] Figure A3, in Appendix, shows the percentages of male and female managers who received at least one angry emoticon in each stage of the experiment. While female managers started off by being slightly more likely to receive at least one angry message from their rank B worker, the difference is not statistically significant.

[37] We include dummy variables for each category of message described in Section 3.3 and displayed in

message immediately before he or she is given a chance to send angry emojis. The estimates show that female managers receive more angry emojis from rank B workers. Controlling for the type of message sent by the manager strengthens this result. Rank B workers tend to send 1 more angry emoji to a female manager than to a male manager.

Overall, the analysis suggests that while rank A workers do not display differential attitudes toward male or female managers, as revealed by the content of the free-form messages they send to them, rank B workers are both more likely to verbally question the ranking decision when it comes from a female manager, and more likely to express disapproval of the choice made by a female manager by sending more angry emojis. This is despite the fact that female managers switch ranks more often and that the propensity to send angry emojis declines when workers' ranks are switched, as suggested by the positive and significant sign of the dummy indicating that the worker was not switched, i.e., he or she was rank B also in the previous round.

4 Robustness checks: The gender composition of the group

When recruiting subjects to participate in the experiment, we aimed to have groups of two men and one woman to simulate male-dominated environments. About 63 percent of our groups ended up being of the desired gender composition, with the remaining groups consisting primarily of two women and one man. When examining self-selection into leadership in Section 3.1, we controlled for the number of women in the group, and we clustered the standard errors at the group level. In the regression analysis, the number of women in the group did not seem to impact the decision to become the manager of the group.

In Figure A4 in Appendix, we reproduce Figure 2 for the restricted sample of groups composed of two men and one woman only. The pattern we saw for the full sample remains, with women less likely than men to self-select into leadership in the *Choice & Backlash* treatment only (p=0.009). Among women, we see a lower percentage of women wanting to be manager in the *Choice & Backlash* treatment than in the *Choice & Talk* treatment, although the difference is not statistically significant in the restricted smaller sample (p=0.123). We do not see treatment-driven significant differences among male participants.

Table 11. Note that the dummies are not mutually exclusive, since the same message can for example be both praising and motivating.

The gender composition of the group, combined with the gender of the manager, has immediate consequences for the genders of the workers a manager is matched with. In our setting, averaging across all treatments, 77 percent of male managers are matched with one male and one female worker, versus 47 percent of female managers (p=0.000). We conduct robustness checks of managers' decision-making, restricting the analysis to groups where both male and female managers are matched with one male and one female worker. The results obtained for rank-switching, ex-post mistakes and foregone profits are displayed in Table A6 in Appendix. Even though the sample is smaller by about one third, the signs of the estimates are consistent with those obtained in Tables 5 to 7, with male and female managers more willing to switch ranks when having to talk with employees, and overall no significant differences in ex-post mistakes and foregone profits.[38]

5 Conclusion

The literature on gender differences in leadership is still in its nascent stage. In this paper, we asked whether women are less likely to self-select into managerial positions that require decision-making that generates inequalities among employees, possibly leading to worker disapproval. We find strong evidence of gender difference in willingness to assume a managerial role only when worker backlash is possible.

By employing a controlled experiment, we were also able to clearly assess the performances of male and female managers under identical incentive systems and decision sets. We find some evidence of treatment and gender differences in the propensity to switch workers' ranks and promote the low-rank employee. Such propensity is higher for both male and female managers when they have to verbally interact with employees; for female managers it increases further when workers can express their disapproval of the managerial decision. However, managers do not switch ranks to please unhappy employees. On the contrary, their ranking decision is either unaffected by the received angry emojis, as it is the case for male managers, or it is impacted negatively by it, as it is the case for female managers. In all treatments and regardless of the manager gender, the probability of a promotion is always strongly higher for deserving employees, i.e. those who performed equally or better than the other worker, and mistakes in rank-allocations are infrequent.

[38] The higher likelihood of female managers to switch ranks in the *Choice & Backlash* treatment disappears in the restricted sample, but we may be underpowered to detect the treatment effect.

Therefore, any observed difference in rank-switching behavior does not lead to significant differences in final outcomes, i.e., profits, across treatments and manager genders.

Male and female managers differ in the language used when communicating with workers. Female managers seem nicer, praising and thanking employees more, while male managers are more direct and more likely to foster competition among employees. However, the observed differences emerge only when workers can talk back to managers. This suggest that male and female workers adopt different communication strategy to cope with, and possibly minimize, the possibility of worker backlash.

Our final set of results concerns workers' attitudes toward male and female managers. Such attitudes are biased against female managers, who are more often questioned about their ranking choices, and receive more angry emojis.

In evaluating the relevance and implications of our findings, it is important to note that the analysis is based on a laboratory setting where decisions are anonymous and worker disapproval takes the form of angry emojis sent by only one other subject via a computer terminal, rather than via personal face to face interaction. We may therefore be underestimating the role that the possibility of worker backlash plays in contributing to the gender leadership gap in field settings, where managers have to face more direct and personal expressions of anger from multiple unhappy subordinates.

Our study could be extended in many interesting ways. First, more work is needed to assess to what extent the observed gender difference in self-selection into leadership is due to women's greater aversion to disapproval and negative feedback, or to the correct anticipation of harsher negative judgment from employees. Second, it would be interesting to examine the relationship between the possibility of worker backlash and leadership decisions in a setting where managers can also receive approval messages from happy employees. In particular, future work could assess how strong or numerous the expected approval messages should be to compensate for the expected negative messages, and induce more women to self-select into top leadership roles.

References

Abel, M. (2019). Do workers discriminate against female bosses?

Adams, R. B. and D. Ferreira (2009). Women in the boardroom and their impact on governance and performance. *Journal of financial economics 94*(2), 291–309.

Alan, S., S. Ertac, E. Kubilay, and G. Loranth (2020). Understanding gender differences in leadership. *The Economic Journal 130*(626), 263–289.

Ayalew, S., S. Manian, and K. Sheth (2018). Discrimination from below: Experimental evidence on female leadership in ethiopia. Technical report, mimeo.

Azmat, G. and B. Petrongolo (2014). Gender and the labor market: What have we learned from field and lab experiments? *Labour Economics 30*, 32–40.

Babcock, L. and S. Laschever (2009). *Women don't ask: Negotiation and the gender divide*. Princeton University Press.

Babcock, L., M. P. Recalde, L. Vesterlund, and L. Weingart (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review 107*(3), 714–47.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: female leadership and gender bias. *Quarterly Journal of Economics 124*(4), 1497–1540.

Bertrand, M. and E. Duflo (2017). Field experiments on discriminationa. In *Handbook of Economic Field Experiments*, Volume 1, pp. 309–393. Elsevier.

Bertrand, M. and S. Mullainathan (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *The American Economic Review 94*(4), 991–1013.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics 145*, 27–41.

Born, A., E. Ranehill, and A. Sandberg (2020). Gender and willingness to lead does the gender composition of teams matter?

Bowles, H. R., L. Babcock, and L. Lai (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational*

Behavior and Human Decision Processes 103(1), 84–103.

Brandts, J. and D. 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), 1223–1268.

Branton, R., A. English, S. Pettey, and T. D. Barnes (2018). The impact of gender and quality opposition on the relative assessment of candidate competency. *Electoral Studies 54*, 35–43.

Chaudhuri, A., Y. Li, and E. Sbai (2018). (un)willingness to lead? men, women and the leadership gap.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics 129*(4), 1625–1660.

Devicienti, F., E. Grinza, A. Manello, and D. Vannoni (2019). What are the benefits of having more female leaders? evidence from the use of part-time work in italy. *ILR Review 72*(4), 897–926.

Dreber, A., E. von Essen, and E. Ranehill (2014). Gender and competition in adolescence: task matters. *Experimental Economics 17*(1), 154–172.

Eagly, A. H., M. C. Johannesen-Schmidt, and M. L. Van Engen (2003, July). Transformational, transactional, and laissez-faire leadership styles: a meta-analysis comparing women and men. *Psychological Bulletin 129*(4), 569–591.

Eckel, C. C. and P. J. Grossman (2008). Men, women and risk aversion: Experimental evidence. *Handbook of experimental economics results 1*, 1061–1073.

Egan, M. L., G. Matvos, and A. Seru (2017). When harry fired sally: The double standard in punishing misconduct. Technical report, National Bureau of Economic Research.

Ellison, G. and A. Swanson (2018). Dynamics of the gender gap in high math achievement. Technical report, National Bureau of Economic Research.

Elsesser, K. M. and J. Lever (2011). Does gender bias against female leaders persist? quantitative and qualitative data from a large-scale survey. *Human Relations 64*(12), 1555–1578.

Erkal, N., L. Gangadharan, and E. Xiao (2018). Leadership selection: Can changing the default break the glass ceiling?

Ertac, S. and M. Y. Gurdal (2012). Deciding to decide: Gender, leadership and risk-taking in groups. *Journal of Economic Behavior & Organization 83*(1), 24–30.

Exley, C. L., M. Niederle, and L. Vesterlund (2020). Knowing when to ask: The cost of leaning in. *Journal of Political Economy 128*(3), 816–854.

Flabbi, L., M. Macis, A. Moro, and F. Schivardi (2019). Do female executives make a difference? the impact of female leadership on gender gaps and firm performance. *The Economic Journal 129*(622), 2390–2423.

Flory, J. A., A. Leibbrandt, and J. A. List (2014). Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *The Review of Economic Studies 82*(1), 122–155.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

Grossman, P. J., C. Eckel, M. Komai, and W. Zhan (2019). It pays to be a man: Rewards for leaders in a coordination game. *Journal of Economic Behavior & Organization 161*, 197–215.

Grossman, P. J., M. Komai, and J. E. Jensen (2015). Leadership and gender in groups: An experiment. *Canadian Journal of Economics/Revue canadienne d'économique 48*(1), 368–388.

Güth, W., M. V. Levati, M. Sutter, and E. Van Der Heijden (2007). Leading by example with and without exclusion power in voluntary contribution experiments. *Journal of Public Economics 91*(5), 1023–1042.

Hengel, E. (2017). Publishing while female. are women held to higher standards? evidence from peer review.

Jack, B. K. and M. P. Recalde (2015). Leadership and the voluntary provision of public goods: Field evidence from bolivia. *Journal of Public Economics 122*, 80–93.

Johnson, M. and V. S. Helgeson (2002). Sex differences in response to evaluative feedback: A field study. *Psychology of Women Quarterly 26*(3), 242–251.

Jones, D. and S. Linardi (2014). Wallflowers: Experimental evidence of an aversion to standing out. *Management Science 60*(7), 1757–1771.

Judge, T. A. and J. E. Bono (2000). Five-factor model of personality and transformational leadership. *Journal of applied psychology 85*(5), 751.

Kern, M. L., G. Park, J. C. Eichstaedt, H. A. Schwartz, M. Sap, L. K. Smith, and L. H. Ungar (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological methods 21*(4), 507.

Kugler, A. D., C. H. Tinsley, and O. Ukhaneva (2017). Choice of majors: Are women really different from men? Technical report, National Bureau of Economic Research.

Kunze, A. and A. R. Miller (2017). Women helping women? evidence from private sector data on workplace hierarchies. *Review of Economics and Statistics 99*(5), 769–775.

Leaper, C. and T. E. Smith (2004). A meta-analytic review of gender variations in children's language use: talkativeness, affiliative speech, and assertive speech. *Developmental psychology 40*(6), 993.

Manian, S. and K. Sheth (2020). Follow my lead: Assertive cheap talk and the gender gap.

Matsa, D. A. and A. R. Miller (2011). Chipping away at the glass ceiling: Gender spillovers in corporate leadership. *American Economic Review 101*(3), 635–39.

Matsa, D. A. and A. R. Miller (2013). A female style in corporate leadership? evidence from quotas. *American Economic Journal: Applied Economics 5*(3), 136–69.

Matsa, D. A. and A. R. Miller (2014). Workforce reductions at women-owned businesses in the united states. *ILR Review 67*(2), 422–452.

Mayo, M., M. Kakarika, J. C. Pastor, and S. Brutus (2012). Aligning or inflating your leadership self-image? a longitudinal study of responses to peer feedback in mba teams. *Academy of Management Learning & Education 11*(4), 631–652.

Mengel, F., J. Sauermann, and U. Zölitz (2017). Gender bias in teaching evaluations. *Journal of the European Economic Association*.

Miller, A. R. (2017). Women and leadership.

Niederle, M. (2016). gender effects in economic experiments. *Handbook of experimental economics. Princeton: Princeton university press. Google Scholar*.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 1067–1101.

Norton, E. C., H. Wang, C. Ai, et al. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal 4*, 154–167.

Podsakoff, P. M., S. B. MacKenzie, R. H. Moorman, and R. Fetter (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *The leadership quarterly 1*(2), 107–142.

Preece, J. and O. Stoddard (2015). Why women don't run: Experimental evidence on gender differences in political competition aversion. *Journal of Economic Behavior & Organization 117*, 296–308.

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

Reuben, E. and K. Timko (2017). On the effectiveness of elected male and female leaders and team coordination.

Rheault, L., E. Rayment, and A. Musulan (2019). Politicians in the line of fire: Incivility and the treatment of women on social media. *Research & Politics 6*(1), 2053168018816228.

Riach, P. A. and J. Rich (2006). An experimental investigation of sexual discrimination in hiring in the english labor market. *Advances in Economic Analysis & Policy 5*(2).

Sarsons, H. (2017). Recognition for group work: Gender differences in academia. *American Economic Review 107*(5), 141–45.

Schmitt, D. P., A. Realo, M. Voracek, and J. Allik (2008). Why can't a man be more like a woman? sex differences in big five personality traits across 55 cultures. *Journal of personality and social psychology 94*(1), 168.

Schwartz-Ziv, M. (2017). Gender and board activeness: the role of a critical mass. *Journal of Financial and Quantitative Analysis 52*(2), 751–780.

Shurchkov, O. and A. van Geen (2017). Why female decision-makers shy away from promoting competition.

Swim, J., E. Borgida, G. Maruyama, and D. G. Myers (1989). Joan mckay versus john mckay: Do gender stereotypes bias evaluations? *Psychological Bulletin 105*(3), 409.

Timko, K. (2017). Gender, communication styles, and leader effectiveness.

Warner, J. and D. Corley (2017). Womens leadership by the numbers.

Wiswall, M. and B. Zafar (2017). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics 133*(1), 457–507.

Tables and figures

	Independent Groups	Participants		
		Men	Women	Total
No Choice (NC)	34	61	41	102
Choice (C)	33	57	42	99
Choice&Talk (CT)	37	63	48	111
Choice&Backlash (CB)	35	64	41	105
Total	139	245	172	417

Table 1: Groups and treatments

	Men %	Women %	H_0: M=W p-value
Choice (C)	94.74	92.86	0.698
Choice&Talk (CT)	92.06	95.83	0.418
Choice&Backlash (CB)	95.31	78.05	0.007
H_0: C = CT (p-value)	0.558	0.539	
H_0: C = CB (p-value)	0.884	0.055	
H_0: CT = CB (p-value)	0.451	0.011	

Note: We report the percentages of participants who wanted to be managers. P-values are generated by Chi-square tests.

Table 2: Percentages of subjects who want to be managers

	Dep. Var: Wants to be manager			
	(1)	(2)	(3)	(4)
Choice&Talk (CT)	-0.002	-0.027	-0.009	-0.005
	(0.037)	(0.055)	(0.051)	(0.052)
Choice&Backlash (CB)	-0.055	0.006	0.010	0.023
	(0.038)	(0.047)	(0.046)	(0.048)
Female	-0.049	-0.019	-0.007	-0.008
	(0.036)	(0.057)	(0.055)	(0.064)
Female x CT		0.056	0.045	0.046
		(0.077)	(0.074)	(0.079)
Female x CB		-0.154*	-0.160*	-0.157*
		(0.092)	(0.090)	(0.095)
Constant	0.960***	0.947***	0.907***	0.810***
	(0.032)	(0.038)	(0.131)	(0.147)
H_0: CT = CB	0.158	0.499	0.639	0.505
H_0: CT + Fem. x CT=0		0.548	0.455	0.441
H_0: CB + Fem. x CB=0		0.047**	0.037**	0.084*
H_0: CT + Fem. x CT=CB + Fem. x CB		0.011**	0.008***	0.013**
Controls	No	No	Yes	Yes^{+}
Observations	315	315	315	303
Clusters	105	105	105	101

Linear probability models. Robust standard errors, clustered at the group level, in parentheses. Controls are: performance in the task, age, and STEM or economics field of study. Additional controls in column 4 are: native English speaker, having held a leadership position, Big5 agreeableness measure and number of women in the group. The number of observations is lower in column 4 because a computer glitch prevented us from collecting post-experiment survey data from 12 participants in the *Choice* treatment. p<0.01, ** p<0.05, * p<0.1.

Table 3: Self-selection into the manager position

	Dep. Var: Assigned Rank A to best performer			
	(1)	(2)	(3)	(4)
No Choice (NC)	-0.085	-0.087	-0.087	-0.091
	(0.074)	(0.100)	(0.099)	(0.103)
Choice&Talk (CT)	0.000	0.039	0.056	0.043
	(0.055)	(0.095)	(0.088)	(0.090)
Choice&Backlash (CB)	0.061	0.074	0.066	0.071
	(0.042)	(0.070)	(0.071)	(0.075)
Female	0.060	0.066	0.062	0.045
	(0.045)	(0.070)	(0.070)	(0.074)
ΔPerformance$_{t-1}$		0.023***	0.023***	0.023***
		(0.006)	(0.005)	(0.005)
Female x NC		0.020	0.023	0.045
		(0.122)	(0.123)	(0.123)
Female x CT		-0.072	-0.084	-0.057
		(0.106)	(0.102)	(0.101)
Female x CB		-0.015	-0.012	-0.006
		(0.081)	(0.081)	(0.086)
Constant	0.907***	0.789***	0.969***	0.814***
	(0.057)	(0.080)	(0.134)	(0.259)
H_0: NC = CT	0.236	0.186	0.112	0.143
H_0: NC = CB	0.019**	0.026**	0.042**	0.036**
H_0: CT = CB	0.124	0.603	0.867	0.650
H_0: NC + Fem. x NC=0		0.344	0.374	0.520
H_0: CT + Fem. x CT=0		0.499	0.581	0.778
H_0: CB + Fem. x CB=0		0.164	0.195	0.159
H_0: CT + Fem. x CT=CB + Fem. x CB		0.094*	0.137	0.155
Controls	No	No	Yes	Yes$^+$
Observations	139	139	139	135
Clusters	139	139	139	135

Linear probability models. Robust standard errors, clustered at the group level (same as individual), in parentheses. The dependent variable is a dummy equal to 1 if the Manager, in Stage 2, assigned rank A to the best performer in Stage 1. Δperformance$_{t-1}$ is the difference in rank A's and rank B's performances in Stage 1. Controls are: age and STEM or economics field of study. Additional controls in column 4 are: native English speaker, having held a leadership position, Big5 agreeableness measure and number of women in the group. The number of observations is lower in column 4 because a computer glitch prevented us from collecting post-experiment survey data from 12 participants in the *Choice* treatment. p<0.01, ** p<0.05, * p<0.1.

Table 4: First rank allocation

	Dep. Var: Manager switched workers' ranks			
	(1)	(2)	(3)	(4)
No Choice (NC)	0.104*	0.104	0.101	0.081
	(0.053)	(0.067)	(0.069)	(0.070)
Choice & Talk (CT)	0.147***	0.172**	0.182**	0.165**
	(0.053)	(0.068)	(0.074)	(0.078)
Choice & Backlash (CB)	0.224***	0.149*	0.129	0.108
	(0.059)	(0.084)	(0.083)	(0.082)
Female	0.069*	0.042	0.024	-0.025
	(0.040)	(0.069)	(0.072)	(0.074)
Δperformance$_{t-1}$	-0.037***	-0.037***	-0.037***	-0.037***
	(0.003)	(0.003)	(0.003)	(0.003)
Female x NC		0.000	0.008	0.060
		(0.104)	(0.104)	(0.108)
Female x CT		-0.033	-0.047	-0.001
		(0.101)	(0.103)	(0.109)
Female x CB		0.137	0.155	0.204*
		(0.117)	(0.116)	(0.116)
Constant	0.214***	0.228***	0.492***	0.484**
	(0.052)	(0.056)	(0.130)	(0.214)
H_0: NC = CT	0.441	0.336	0.277	0.272
H_0: NC = CB	0.057*	0.597	0.737	0.740
H_0: CT = CB	0.218	0.799	0.799	0.518
H_0: NC + Fem. x NC=0		0.198	0.166	0.078*
H_0: CT + Fem. x CT=0		0.066*	0.080*	0.031**
H_0: CB + Fem. x CB=0		0.001***	0.001***	0.000***
H_0: CT + Fem. x CT=CB + Fem. x CB		0.077*	0.087*	0.096*
Stage FE	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes^{+}
Observations	556	556	556	540
Clusters	139	139	139	135

Linear probability models. Robust standard errors, clustered at the group level (same as individual), in parentheses. The dependent variable is a dummy equal to 1 if the Manager switched ranks and promoted the worker who was Rank B in the previous Stage. Δperformance$_{t-1}$ is the difference in the performances of the rank A and the rank B workers in previous Stage. The analysis is restricted to Stages 3 to 6. Controls are: age and STEM or economics field of study. Additional controls in column 4 are: native English speaker, having held a leadership position, and Big5 agreeableness measure. The number of observations is lower in column 4 because a computer glitch prevented us from collecting post-experiment survey data from 12 participants in the *Choice* treatment. p<0.01, ** p<0.05, * p<0.1.

Table 5: Managers' decision to switch ranks and promote the Rank B worker

	Dep. Var: Ex-post mistake in rank allocation			
	(1)	(2)	(3)	(4)
No Choice (NC)	0.069	-0.004	-0.002	-0.033
	(0.045)	(0.069)	(0.069)	(0.066)
Choice & Talk (CT)	0.030	0.006	-0.013	-0.052
	(0.048)	(0.077)	(0.073)	(0.072)
Choice & Backlash (CB)	-0.008	-0.114*	-0.099	-0.125*
	(0.044)	(0.064)	(0.064)	(0.064)
Female	0.012	-0.085	-0.074	-0.106
	(0.033)	(0.063)	(0.065)	(0.066)
Δperformance$_{t-1}$	-0.024***	-0.024***	-0.023***	-0.025***
	(0.003)	(0.003)	(0.003)	(0.003)
Female x NC		0.136	0.129	0.165*
		(0.090)	(0.090)	(0.090)
Female x CT		0.053	0.069	0.104
		(0.097)	(0.092)	(0.094)
Female x CB		0.195**	0.183**	0.204**
		(0.086)	(0.086)	(0.086)
Constant	0.303***	0.354***	0.040	-0.126
	(0.052)	(0.064)	(0.112)	(0.161)
H_0: NC = CT	0.418	0.893*	0.876	0.778
H_0: NC = CB	0.082*	0.071*	0.105	0.117
H_0: CT = CB	0.418	0.087*	0.177	0.251
H_0: NC + Fem. x NC=0		0.023**	0.028**	0.024**
H_0: CT + Fem. x CT=0		0.312	0.339	0.404
H_0: CB + Fem. x CB=0		0.156	0.156	0.188
H_0: CT + Fem. x CT=CB + Fem. x CB		0.733	0.654	0.677
Stage FE	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes^{+}
Observations	695	695	695	675
Clusters	139	139	139	135

Linear probability models. Robust standard errors, clustered at the group (same as individual) level in parentheses. The dependent variable is a dummy equal to 1 if the worker who was assigned rank A ends up performing worse than the worker who was assigned rank B in the current Stage. Δperformance$_{t-1}$ is the difference in the performances of the rank A and the rank B workers in previous Stage. The analysis is for Stages 2 to 6. Controls are: age and STEM or economics field of study. Additional controls in column 4 are: native English speaker, having held a leadership position, and Big5 agreeableness measure. The number of observations is lower in column 4 because a computer glitch preventing us from collecting post-experiment survey data from 12 participants in the *Choice* treatment. p<0.01, ** p<0.05, * p<0.1.

Table 6: Likelihood of ex-post mistakes in rank allocations

	Dep. Var: Foregone profit due to mistakes in ranking			
	(1)	(2)	(3)	(4)
No Choice (NC)	0.836	1.024	1.025	0.747
	(0.525)	(0.882)	(0.884)	(0.908)
Choice & Talk (CT)	0.160	0.101	-0.041	-0.394
	(0.465)	(0.746)	(0.710)	(0.739)
Choice & Backlash (CB)	0.251	-0.173	-0.117	-0.316
	(0.499)	(0.761)	(0.762)	(0.747)
Female	-0.243	-0.382	-0.358	-0.682
	(0.378)	(0.652)	(0.672)	(0.719)
Δperformance$_{t-1}$	-0.278***	-0.279***	-0.277***	-0.285***
	(0.048)	(0.049)	(0.050)	(0.050)
Female x NC		-0.360	-0.385	-0.006
		(1.140)	(1.149)	(1.195)
Female x CT		0.113	0.202	0.615
		(0.960)	(0.939)	(0.970)
Female x CB		0.780	0.750	0.941
		(1.015)	(1.034)	(1.056)
Constant	2.186***	2.264***	1.448	-0.367
	(0.476)	(0.537)	(1.279)	(1.883)
H_0: NC = CT	0.205	0.333	0.255	0.221
H_0: NC = CB	0.306	0.208	0.223	0.258
H_0: CT = CB	0.857	0.744	0.923	0.924
H_0: NC + Fem. x NC=0		0.320	0.340	0.294
H_0: CT + Fem. x CT=0		0.725	0.789	0.739
H_0: CB + Fem. x CB=0		0.361	0.356	0.397
H_0: CT + Fem. x CT=CB + Fem. x CB		0.530	0.463	0.533
Stage FE	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes^{+}
Observations	695	695	695	675
Clusters	139	139	139	135

OLS. Robust standard errors, clustered at the group level, in parentheses. The dependent variable is equal to 0 if the rank A worker performed better than the rank B worker. If the rank B worker performed better than the rank A worker, the dependent variable is the difference between the profits that the manager could have made by assigning rank A to the best performing worker, and the profits actually made in the current Stage. Δperformance$_{t-1}$ is the difference in the performances of the rank A and the rank B worker in previous Stage. The analysis is for Stages 2 to 6. Controls are: performance in the task, age, and STEM or economics field of study. Additional controls in column 4 are: native English speaker, having held a leadership position, and Big5 agreeableness measure. The number of observations is lower in column 4 because a computer glitch prevented us from collecting post-experiment survey data from 12 participants in the *Choice* treatment. p<0.01, ** p<0.05, * p<0.1.

Table 7: Foregone Profits

	Total profits			Total difference in workers' earnings		
	Male Manager	Female Manager	H_0: M=W pvalue	Male Manager	Female Manager	H_0: M=W pvalue
No Choice (NC)	166.25	172.44	0.486	261.75	193.89	0.112
Choice (C)	163.33	178.44	0.159	246.40	278.44	0.441
Choice & Talk (CT)	177.69	179.83	0.842	176.77	212.75	0.424
Choice & Backlash (CB)	181.37	168.53	0.201	240.62	180.10	0.155
H_0: NC=C	0.794	0.478		0.717	0.043*	
H_0: NC=CT	0.278	0.419		0.091*	0.623	
H_0: NC=CB	0.048**	0.712		0.621	0.737	
H_0: C=CT	0.269	0.876		0.161	0.090*	
H_0: C=CB	0.075*	0.341		0.891	0.023**	
H_0: CT=CB	0.677	0.280		0.198	0.395	

We report the total profits generated by rank A workers to the manager, and the difference in the total earnings of the highest and the lowest earning worker within each group. Both total profits and differences in workers' earnings are computed over the 5 stages of the experiment. p-values are generated by test of equality of means. p<0.01, ** p<0.05, * p<0.1.

Table 8: Total profits and inequality in workers' total earnings

	Dep Var: Manager promotes the Rank B worker			
	(1)	(2)	(3)	(4)
Female	0.182*	0.263**	0.234**	0.225**
	(0.093)	(0.105)	(0.099)	(0.098)
Angry Emojis$_{t-1}$	-0.055***	-0.019	-0.037	-0.026
	(0.020)	(0.024)	(0.025)	(0.029)
Δperformance$_{t-1}$	-0.047***	-0.047***	-0.045***	-0.045***
	(0.006)	(0.006)	(0.006)	(0.006)
Female x Angry Emojis$_{t-1}$		-0.061*	-0.034	-0.043
		(0.035)	(0.033)	(0.040)
Constant	0.442***	0.396***	1.284***	1.221**
	(0.097)	(0.097)	(0.366)	(0.481)
H_0: Emojis$_{t-1}$+ Fem. x Emojis$_{t-1}$=0		0.003***	0.001***	0.002***
Stage FE	Yes	Yes	Yes	Yes
Controls	No	No	Yes	Yes^{+}
Observations	140	140	140	140
Clusters	35	35	35	35

Linear probability models. Robust standard errors, clustered at the group level in parentheses. The dependent variable is a dummy equal to 1 if the Manager switched ranks and promoted the rank B worker at the beginning of the current Stage in the *Choice & backlash* treatment. Angry Emojis$_{t-1}$ is the number of angry emojis sent by the rank B worker in the previous Stage. Δperformance$_{t-1}$ is the difference in the performances of the rank A and rank B workers in the previous Stage. Controls are: age and STEM or economics field of study. Additional controls in column 4 are: native English speaker, having held a leadership position, and Big5 agreeableness measure. p<0.01, ** p<0.05, * p<0.1.

Table 9: Likelihood of promotion conditional on angry emojis

	Praise	Motivate	Explain	Compete	Fairness	Team	Cordial
Choice & Talk (CT)							
Male Manager	0.74	0.46	0.12	0.08	0.06	0.01	0.11
Female Manager	0.73	0.35	0.07	0.04	0.03	0.04	0.10
H_0: M=F (p-values)	(0.751)	(0.137)	(0.193)	(0.311)	(0.353)	(0.335)	(0.869)
Choice & Backlash (CB)							
Male Manager	0.51	0.31	0.29	0.18	0.05	0.03	0.08
Female Manager	0.67	0.29	0.18	0.06	0.04	0.06	0.14
H_0: M=F (p-values)	(0.030)	(0.799)	(0.088)	(0.021)	(0.357)	(0.229)	(0.190)
M: H_0 : CT=CB (p-values)	(0.001)	(0.234)	(0.000)	(0.002)	(0.770)	(0.745)	(0.479)
F: H_0 : CT=CB (p-values)	(0.378)	(0.118)	(0.023)	(0.756)	(0.415)	(0.229)	(0.395)

We report the percentage of messages of each type sent by male and female managers over the 5 stages of the experiment in the *Choice & Talk* and the *Choice & Backlash* treatments. Since the categories are not mutually exclusive, the percentages do not sum up to 1. P-values are generated by Chi-square tests.

Table 10: Messages sent by managers to rank A workers

	Praise	Motivate	Explain	Compete	Fairness	Team	Cordial
Choice & Talk (CT)							
Male Manager	0.52	0.29	0.31	0.09	0.06	0.00	0.25
Female Manager	0.57	0.36	0.13	0.04	0.06	0.03	0.15
H_0: M=F (p-values)	(0.569)	(0.364)	(0.004)	(0.164)	(0.930)	(0.137)	(0.107)
Choice & Backlash (CB)							
Male Manager	0.27	0.21	0.51	0.19	0.07	0.03	0.20
Female Manager	0.42	0.28	0.30	0.06	0.08	0.04	0.33
H_0: M=F (p-values)	(0.044)	(0.276)	(0.005)	(0.012)	(0.823)	(0.536)	(0.060)
M: H_0 : CT=CB (p-values)	(0.000)	(0.045)	(0.000)	(0.001)	(0.636)	(0.865)	(0.757)
F: H_0 : CT=CB (p-values)	(0.039)	(0.386)	(0.038)	(0.902)	(0.436)	(0.330)	(0.007)

We report the percentage of messages of each type sent by male and female managers to the the rank B workers over 5 stages of the experiment in the *Choice & Talk* and the *Choice & Backlash* treatments. Since the categories are not mutually exclusive, the percentages do not sum up to 1. P-values are generated by Chi-square tests.

Table 11: Messages sent by managers to rank B workers

		Messages sent by Rank A		
		Thanks or jokes	Approves rank	Commits to work hard
Male Manager		0.54	0.07	0.26
Female Manager		0.58	0.08	0.20
H_0: M=F	*p-value*	0.582	0.823	0.327

Note: We report the percentage of messages of each types sent to managers by rank A workers over the 5 stages of the experiment. P-values are generated by Chi-square tests. *** p<0.01, ** p<0.05, * p<0.1.

Table 12: Messages sent by Rank A workers to managers in *Choice & Backlash*

		Messages sent by Rank B					
		Thanks or jokes	Accepts rank	Commits to work hard	Apologetic	Upset	Questions rank
Male Manager		0.16	0.59	0.16	0.06	0.09	0.06
Female Manager		0.17	0.50	0.17	0.09	0.04	0.17
H_0: M=F	*p-value*	0.916	0.277	0.916	0.434	0.218	0.032**

Note: We report the percentage of messages of each types sent to managers by rank B workers over the 5 stages of the experiment. P-values are generated by Chi-square tests. *** p<0.01, ** p<0.05, * p<0.1.

Table 13: Messages sent by Rank B workers to managers in *Choice & Backlash*

	Dep. Var: Angry emojis sent to Manager			
To Female Manager	0.823*	0.744*	0.767*	1.038**
	(0.452)	(0.440)	(0.431)	(0.498)
Performance$_{t-1}$	0.097**	0.101**	0.084**	0.071*
	(0.040)	(0.038)	(0.039)	(0.040)
RankB$_{t-1}$	0.918**	0.777*	0.757*	1.075**
	(0.430)	(0.391)	(0.381)	(0.415)
Constant	-0.982	-0.346	-1.001	-1.026
	(0.816)	(1.616)	(1.915)	(2.064)
Stage FE	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes^{+}	Yes^{+}
Manager's message type	No	No	No	Yes
Observations	140	140	140	140
Clusters	35	35	35	35

OLS. Robust standard errors, clustered at the group level in parentheses. The dependent variable is the number of angry emojis, including zeros, that the rank B worker sent to his or her Manager in the *Choice & backlash* treatment. Performance$_{t-1}$ is the number of correct puzzles solved by the rank B worker in the previous Stage. RankB$_{t-1}$ is a dummy equal to 1 if the rank B worker was rank B also in the previous Stage. Controls are: age, gender and STEM or economics field of study. Additional controls in columns 3 and 4 are: native English speaker, having held a leadership position, and Big5 agreeableness measure. In column 4, we add the manager's message dummies presented in Table 11, i.e. seven dummy variables indicating whether the manager sent a message praising the worker, motivating the worker, explaining the ranking decision, inducing competition, mentioning fairness, team building or using cordial words. p<0.01, ** p<0.05, * p<0.1.

Table 14: Angry emojis sent to managers

Figure 1: Stages of the experiment

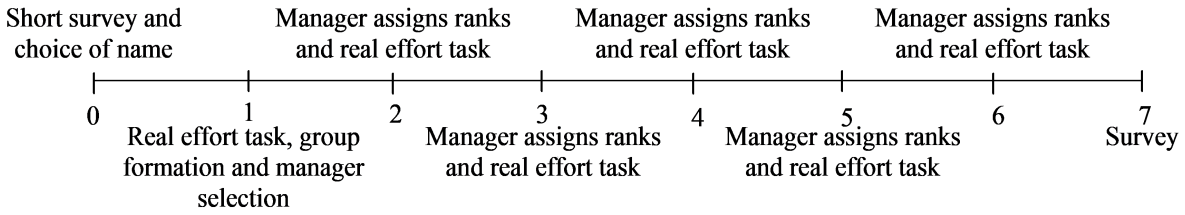
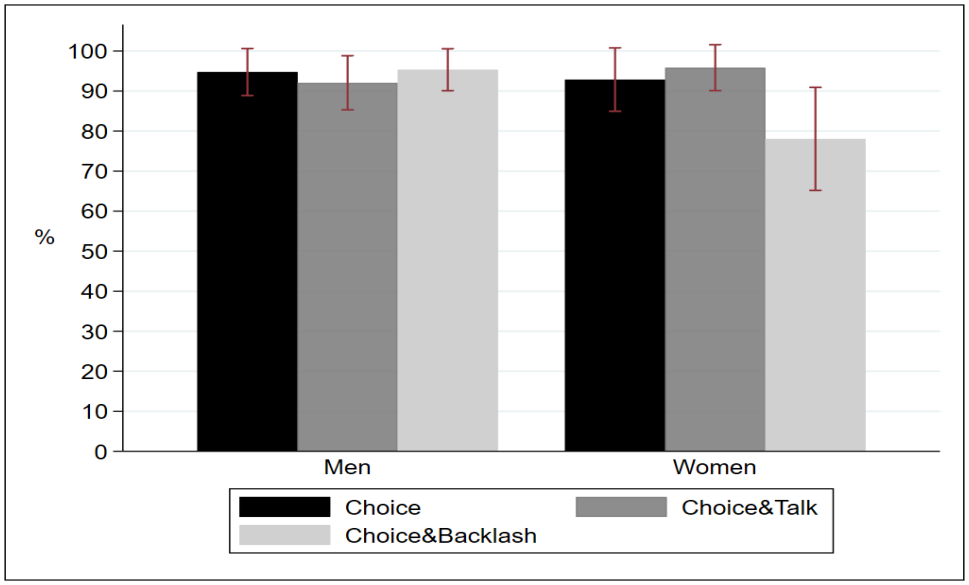


Figure 2: Self-selection into the manager position

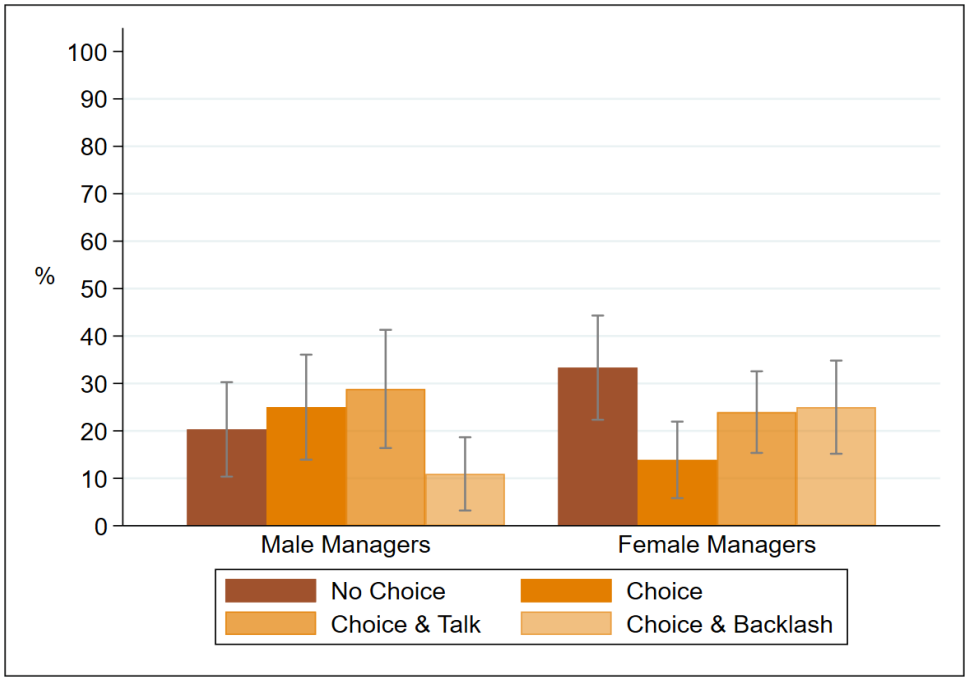


Note: The figure shows the percentages of men and women who wanted to be the Manager of their group in each treatment.

Figure 3: Average number of rank switches by treatment

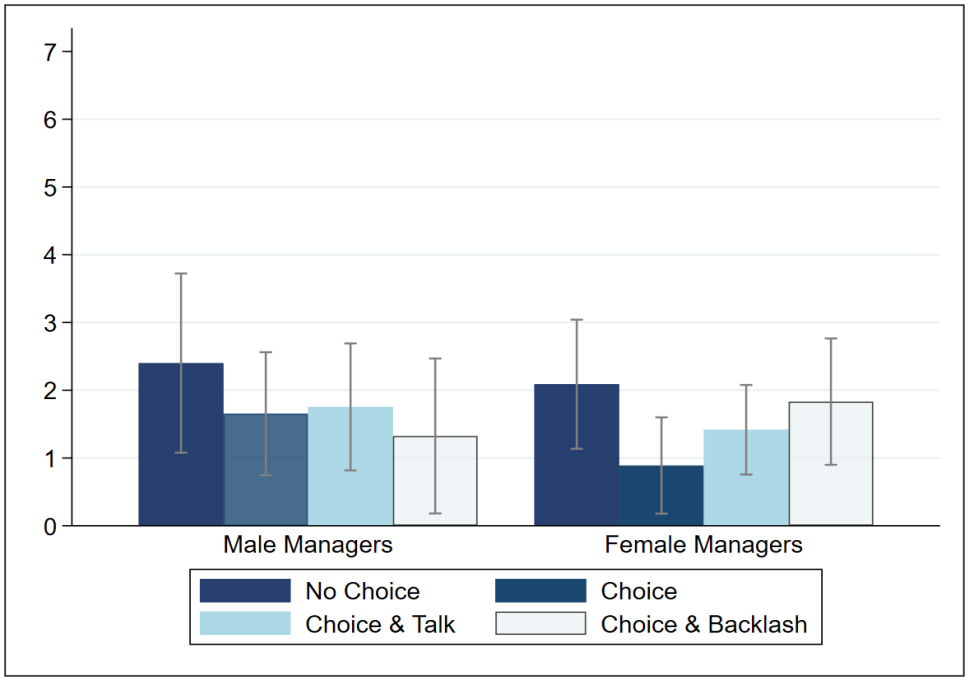
Note: The figure shows the average number of times a manager switched workers' ranks over the 5 stages of the experiment, by treatment and gender of the manager.

Figure 4: Ex-post mistakes in rank allocations



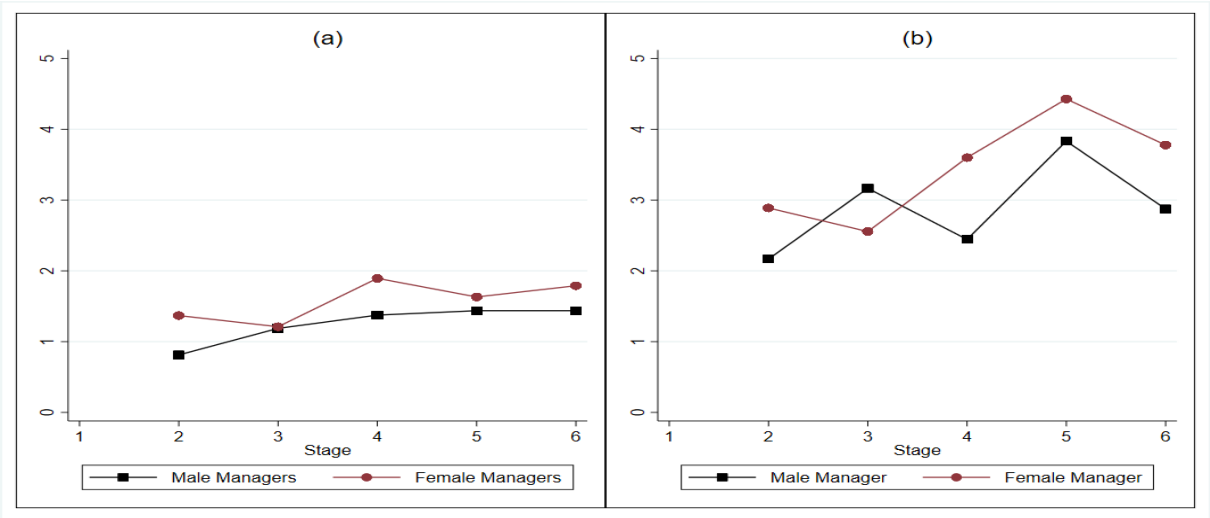
Note: The figure shows the percentages of male and female managers who assigned rank A to a worker who ended up performing worse than the rank B worker, across stages 3 to 6.

Figure 5: Foregone profits



Note: The figure shows the difference - averaged across stages - between the highest profit that the manager could have made in a Stage by hiring the worker who ended up performing the best and the profits the manager actually made. Foregone profits are greater than 0 if the worker who was assigned Rank A performed worse than the worker who was assigned rank B. Foregone profits are expressed in Experimental Currency Units (ECU). Note that in any give stage, the manager could make at most 40 ECU in profits from the work of the Rank A worker.

Figure 6: Angry emojis sent to managers



Note: The figure in panel (a) shows the average number of angry emojis sent to male and female managers in the *Choice & Backlash* treatment. The figure in panel (b) shows the emojis sent, conditional on sending at least one.

APPENDIX TABLES AND FIGURES

	Male				Female				M=F (All)
	NC	C	CT	CB	NC	C	CT	CB	p-value
Age	22.61	22.84	24.38	22.55	23.32	22.19	23.27	22.00	0.414
	(2.49)	(2.87)	(6.98)	(4.63)	(4.51)	(3.09)	(6.59)	(4.27)	
STEM-Bus-Econ (frequency)	0.85	0.91	0.78	0.80	0.71	0.67	0.54	0.66	0.000
	(0.36)	(0.29	(0.42)	(0.41)	(0.46)	(0.48)	(0.50)	(0.48)	
Native speaker (frequency)	0.34	0.29	0.52	0.47	0.41	0.51	0.67	0.56	0.008
	(0.48)	(0.46)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	
Past Leadership (frequency)	0.80	0.85	0.89	0.83	0.90	0.83	0.85	0.83	0.724
	(0.40)	(0.36)	(0.32)	(0.38)	(0.30)	(0.38)	(0.36)	(0.38)	
Big 5 Agreeableness Index	27.80	26.67	27.92	27.70	29.44	28.34	29.35	29.95	0.004
	(5.05)	(5.53)	(6.98)	(5.18)	(7.26)	(6.78)	(5.94)	(5.21)	

Table A1: Subjects' characteristics

	Number of correctly solved puzzles		
	(1)	(2)	(3)
No Choice (NC)	0.396	0.269	0.130
	(0.739)	(0.706)	(0.704)
Choice&Talk (CT)	0.735	0.894	0.172
	(0.746)	(0.696)	(0.696)
Choice&Backlash (CB)	0.230	0.019	-0.441
	(0.762)	(0.737)	(0.711)
Female	0.574	0.119	-0.500
	(0.858)	(0.851)	(0.863)
Female x NC	-0.620	-0.198	0.171
	(1.193)	(1.115)	(1.138)
Female x CT	0.707	0.620	1.074
	(1.101)	(1.068)	(1.027)
Female x CB	0.855	1.015	1.322
	(1.153)	(1.129)	(1.073)
Constant	13.172***	19.253***	14.558***
	(0.585)	(1.174)	(1.717)
H_0: NC = CT	0.642	0.359	0.948
H_0: NC = CB	0.824	0.729	0.390
H_0: CT = CB	0.502	0.221	0.330
H_0: NC + Fem. x NC=0	0.811	0.935	0.736
H_0: CT + Fem. x CT=0	0.075*	0.057*	0.098*
H_0: CB + Fem. x CB=0	0.210	0.229	0.272
Controls	No	Yes	Yes$^+$
Stage FE	Yes	Yes	Yes
Observations	2,502	2,502	2,430
Clusters	417	417	405

OLS estimates. Note: each subject solved up to 20 puzzles in each of 6 rounds. Robust standard errors, clustered at the individual level, in parentheses. Controls are: performance in the task, age, and STEM or economics field of study. Additional controls in column 4 are: native English speaker, having held a leadership position, Big5 agreeableness measure. The number of observations is lower in column 4 because a computer glitch preventing us from collecting post-experiment survey data from 12 participants in the *Choice* treatment. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Performance in the real effort task

	Wants to be manager	First rank allocation	Switch ranks	Ex-post mistake	Lost profits
No Choice		-0.064	0.109	0.127**	0.640
Choice & Talk	0.036	-0.028	0.134*	0.055	0.161
Choice & Backlash	-0.150**	0.055	0.284***	0.084	0.633

Note: The coefficients displayed in each column are the linear combinations of the coefficients estimated in column 3 of Tables 3, 4, 5, and 6. The *Choice* treatment is the benchmark treatment. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Estimated treatment effects for the female sample

Examples of messages sent by managers
Praising
"you are a powerhouse", "Amazing work, you are doing great"
"You're a star. Keep shining.", "Outstanding work as always!"
"you are KILLING IT!", "Manager material! I like the consistency!"
Motivating
"Keep it up!", "You got this dude"
"Keep up the good work!", "I believe in you Jill! You got this."
"I know you can do better than that.", "we are almost there! finish strong!!"
Explaining the rank allocation
"Greg did better in general which is why he is A do better and you will be"
"You are B because Emily did better in 1. If you beat her, I will switch your rank.",
"This is not a tactic: If you do better than Worker A, you will be rewarded as such."
Competition-inducing
"Your coworker outscored you! Earn your spot back!"
"Good job, Carly. Roger is catching up so keep it up"
"your percentage increase was higher than worker A, so beat her this time and Rank A is yours"
Fairness
"Trying to be fair and switch it up"
"I just want to distribute the ranks equally"
"I'll alternate A and B so you have a better chance to be paid more"
Team building
"lets do this together"
"Let's go, team!", "The team needs your best effort, John."
"Let's finish strong here team!"
Cordial
"Sorry", "sorry had to give him a chance"
"Thank you for your hard work!", "Good luck! Hope you make lots of $ today :)"
"Great job! Thank you so much!"

Table A4: Examples of messages sent by managers to workers

Examples of messages sent by workers
Thanks or jokes
"Thanks. Don't agree, but thanks."
"I understand. Thank you."
"haha what?"
Accepts rank
"no problems"
"Fine."
"Fair enough"
Commits to work hard
"I will be better and become rank A"
"I will try my best."
"I will make you proud."
Apologetic
"I apologize for slack last stage and hope to make it up"
"I am disappointed in myself. Need to concentrate more."
"sorry....got stucked..lol"
Upset
"I am mad."
"eat dirt, im out here grinding and u missin out"
"i hate u"
Questions the ranking decision
"why??? i did 20 correct in first and 19 in second?"
"you really can't do better can you"
"are u kidding me... u changed to mine after the failed then u go back to them"

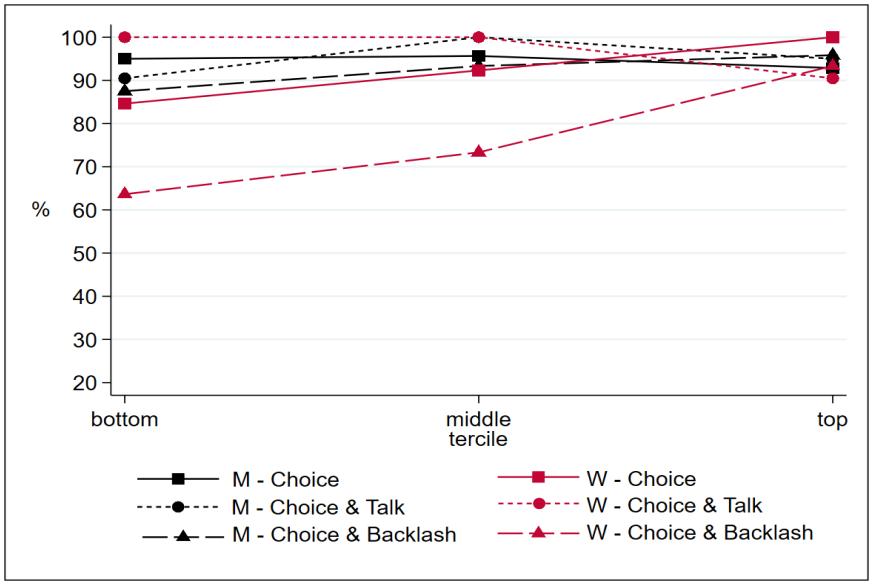
Table A5: Examples of messages sent by rank B workers

	Switched ranks			Ex-post mistake			Foregone profit		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
No Choice (NC)	0.141*	0.147*	0.126	0.066	0.060	0.004	1.523	1.467	0.925
	(0.080)	(0.083)	(0.078)	(0.078)	(0.078)	(0.074)	(1.075)	(1.066)	(1.174)
Choice & Talk (CT)	0.212**	0.204**	0.189*	-0.067	-0.080	-0.147	-1.035	-1.293	-1.906**
	(0.086)	(0.092)	(0.098)	(0.087)	(0.091)	(0.090)	(0.848)	(0.905)	(0.952)
Choice & Backlash (CB)	0.152*	0.119	0.092	-0.089	-0.080	-0.109	-0.153	-0.185	-0.523
	(0.088)	(0.088)	(0.092)	(0.069)	(0.068)	(0.069)	(0.846)	(0.858)	(0.843)
Female	0.081	0.060	-0.007	-0.026	-0.023	-0.046	0.198	0.138	-0.127
	(0.098)	(0.103)	(0.113)	(0.078)	(0.080)	(0.082)	(0.921)	(0.956)	(1.140)
Female x NC	0.029	0.017	0.098	0.010	0.007	0.042	-1.637	-1.743	-1.296
	(0.140)	(0.141)	(0.158)	(0.114)	(0.114)	(0.116)	(1.564)	(1.557)	(1.775)
Female x CT	-0.106	-0.109	-0.056	0.061	0.059	0.111	0.217	0.177	0.562
	(0.149)	(0.151)	(0.166)	(0.119)	(0.119)	(0.119)	(1.250)	(1.250)	(1.363)
Female x CB	0.021	0.039	0.114	0.132	0.135	0.133	-0.289	-0.142	-0.084
	(0.165)	(0.164)	(0.169)	(0.096)	(0.097)	(0.094)	(1.177)	(1.236)	(1.358)
ΔPerformance$_{t-1}$	-0.032***	-0.032***	-0.033***	-0.025***	-0.025***	-0.026***	-0.317***	-0.320***	-0.333***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.068)	(0.069)	(0.071)
Constant	0.226***	0.511***	0.498**	0.264***	0.060	-0.144	3.358***	1.655	-0.901
	(0.071)	(0.150)	(0.212)	(0.063)	(0.125)	(0.186)	(0.928)	(1.399)	(2.301)
H_0: NC = CT	0.448	0.537	0.524	0.130	0.122	0.084*	0.029**	0.023**	0.024**
H_0: NC = CB	0.915	0.751	0.709	0.026**	0.038	0.077*	0.127	0.127	0.192
H_0: CT = CB	0.548	0.353	0.320	0.776	0.994	0.651	0.332	0.243	0.159
H_0: NC + Fem.x NC=0	0.145	0.155	0.073*	0.361	0.420	0.584	0.915	0.793	0.760
H_0: CT + Fem.x CT=0	0.392	0.467	0.336	0.948	0.800	0.660	0.407	0.232	0.232
H_0: CB + Fem.x CB=0	0.222	0.237	0.151	0.522	0.298	0.713	0.587	0.704	0.569
H_0: CB + Fem.x CB = CB + Fem.x CB	0.640	0.682	0.636	0.516	0.297	0.384	0.325	0.306	0.393
Stage FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes$^+$	No	Yes	Yes$^+$	No	Yes	Yes$^+$
Observations	332	332	316	415	415	395	415	415	395
Clusters	83	83	79	83	83	79	83	83	79

Linear probability models in columns 1 to 6. OLS in columns 7 to 9. The analysis is restricted to groups with one male worker and one female worker. Robust standard errors, clustered at the group level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

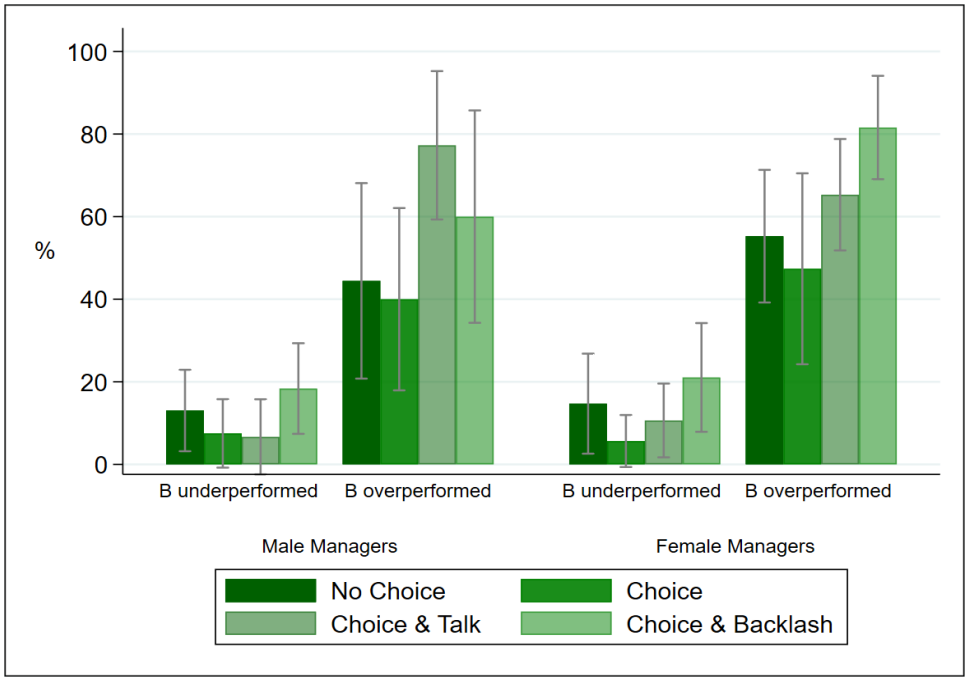
Table A6: Robustness: Rank-switching, ex-post mistakes and foregone profits

Figure A1: The decision to be Manager by performance terciles



Note: The figure shows the percentages of bottom-, middle- and top-performers who wanted to be manager by gender and treatment.

Figure A2: Rank switching by Rank B performance and treatment



Note: The figure shows the percentage of times the rank B worker was promoted to rank A when he or she performed worse than the rank A worker, and when he or she performed the same or better, by the gender of the manager.

Figure A3: Percentage of managers receiving at least one angry emoticon

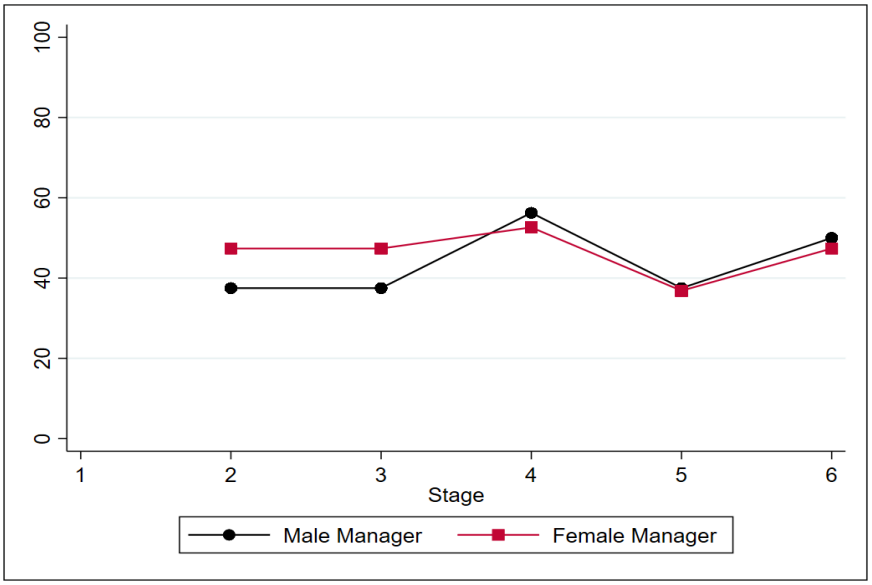
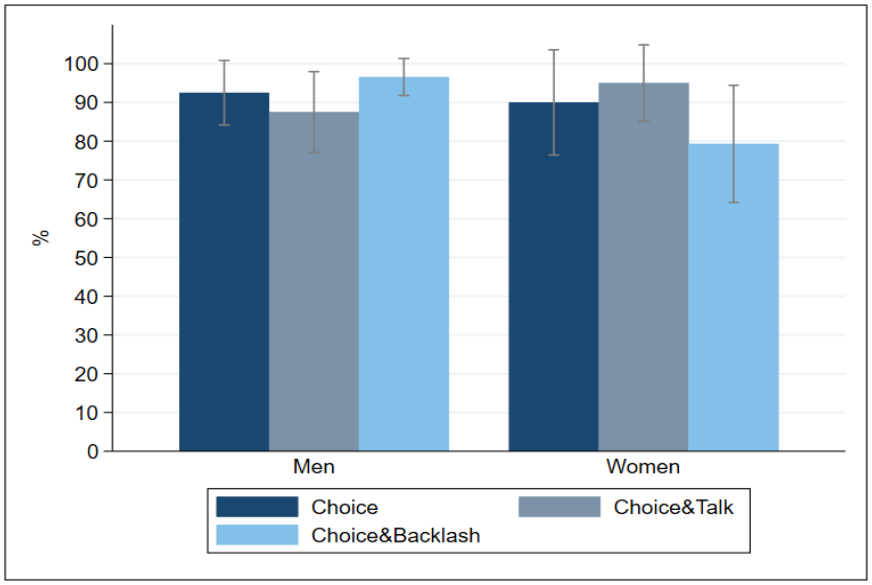


Figure A4: Self-selection into leadership in groups of 2 men and 1 woman



ONLINE APPENDIX

EXPERIMENTAL INSTRUCTIONS

General instructions

Thank you all for coming today. You are here to participate in an experiment. In addition to a $10 participation fee, you will be paid any money you accumulate from the experiment. You will be paid privately, by check, at the conclusion of the experiment.

The experiment will consist of six stages and the instructions will be provided separately on your screen at the beginning of each stage. You will have the chance to earn money in each stage of the experiment. Earnings during the experiment will be denominated in Experimental Currency Units, or ECU. At the end of the session one stage of the experiment will be randomly selected for payment and your earnings in that stage will be converted to dollars at the exchange rate of $1 for 6 ECU. After participating in all the stages of the experiment you will be asked to complete a brief questionnaire. You will then be paid the money your earned in the selected stage of experiment.

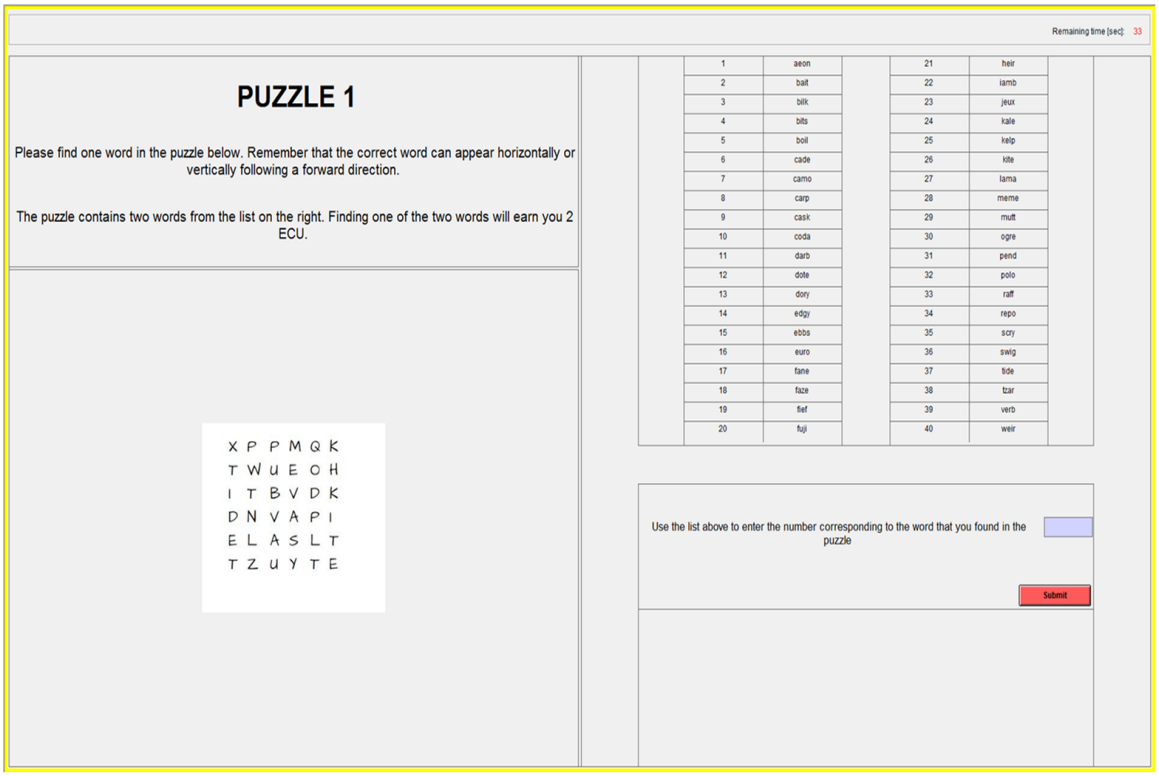
This study has been reviewed and approved by the SMU Human Subjects Committee. If you have any questions during the experiment, please raise your hand and wait for an experimenter to come to you. Please do not talk, exclaim, or try to communicate with other participants during the experiment. Participants intentionally violating these rules may be asked to leave the experiment and may not be paid.

Please read and sign the Consent Form that you found on your desk. Please raise your hand if you have any question about any of the information on the Consent form. We will proceed with the experiment once we have collected all signed consent forms.

[Collect consent forms. Start program. When everybody is on Screen 3, distribute Puzzle Example]

PUZZLE EXAMPLE (*Handout 1*)

During the experiment, you will engage in multiple rounds of a puzzle-solving task. Please refer to the paper you have been given to see an example of the task. Each task consists of finding a 4-letter word in a 6x6 matrix. As you can see on the example you have been given, the screen will be divided in two halves. On the left, you will see the matrix and on the right, you will see a list of 40 words. Each puzzle has two words that appear on the list. In order to earn money, you will have to identify one word per puzzle. Once you identify the word, you will have to enter the number next to that word in the list. You will then have to press "submit" to move to the next puzzle.



Please note that the word you are looking for can appear horizontally or vertically in the matrix, following a forward direction. You should ignore words that are read backward or diagonally. You should also ignore words that do not appear in the list.

Look at the example you have been given. In order to earn points, you would have to

find either the word “tide” or the word “kite” and enter the corresponding number. The word “tide” appears vertically on the first column. The word “kite” appears vertically on the sixth column.

Can you all see the two words in the puzzle? Raise your hand if you cannot see them.

Note that there are other words that you may identify in the matrix. For instance, the word “sale” [appears horizontally on the fifth row, but it reads backward] and the word “bale” [appears diagonally]. These words would not be valid entries, since they either appear backward or are not on the list. Remember that for an entry to be valid, it MUST be on the list to the right of the matrix.

Can you all see the two words in the puzzle? Raise your hand if you cannot.

Do you have any questions or doubts about the puzzle-solving task?

Instructions for Stages 2 to 6 *(Handout 2)*

Stage 2 of the experiment is about to begin.

- In this stage and in the following 5 stages of the experiment you will be part of a group, together with two other participants. One group member will assume the role of Manager and the other two group members will assume the role of Worker.
- The manager gets a wage of 100 ECU. The main role of the manager is to decide which worker will be rank A and which worker will be rank B in the group, in this stage of the experiment.
- A Rank A worker gets a wage of 80 ECU. A Rank B worker gets 20 ECU.
- After the rank allocation, all members of the group will engage in a similar puzzle task as in Stage 1 of the experiment. Each correctly solved puzzle generates 2 ECU in addition to the initial wage. Moreover, each puzzle solved correctly by the Rank A worker generated 2 ECU also to the manager.
- Therefore the earnings from this stage of the experiment are determined as follows:
  - The Manager gets 100 ECU + 2 ECU per puzzle + 2 ECU per puzzle solved by Rank A worker

  - Rank A worker gets 80 ECU plus 2 ECU per puzzle
  - Rank B worker gets 20 ECU plus 2 ECU per puzzle

- Stages 3 to 6 will be identical to Stage 2. However, at the end of each stage of the experiment, the Manager will be informed about the performance of Rank A and Rank B workers and will have the chance to reassign ranks before the next stage begins, at his or her discretion.

Please turn this page around to have a look at the screen that the Manager will see when making the rank allocation decision.

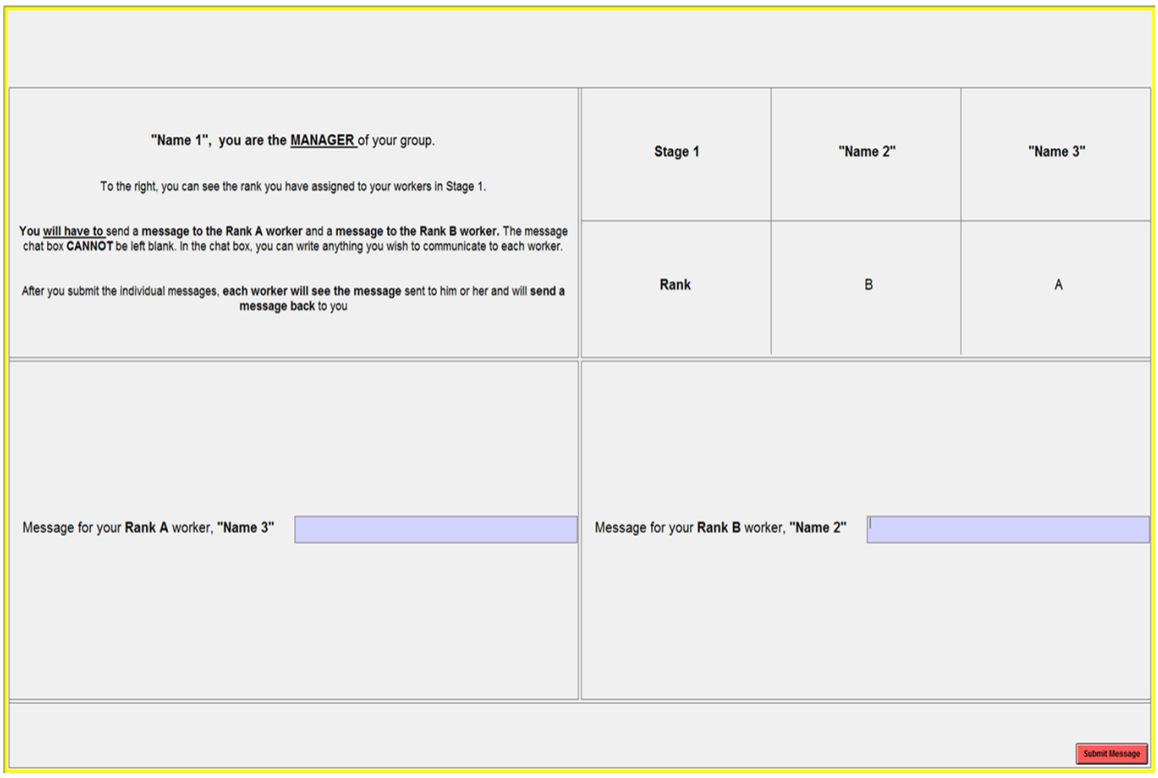
The Manager will have to decide which worker will be Rank A and which worker will be Rank B. After the Manager makes the allocation decision, each worker will be informed about the Rank they have been assigned, either rank A or rank B.

[**CHOICE & TALK**:*After assigning ranks A and B to the workers, the Manager will have to send a message to the Rank A worker and a message to the Rank B worker. The message chat box CANNOT be left blank. In the chat box, the Manager can write anything he or she wishes to communicate to each worker. After the Manager submits the individual messages, each worker will see the message sent to him or her.*]

[**CHOICE & BACKLASH**: *After assigning ranks A and B to the workers, the Manager will have to send a message to the Rank A worker and a message to the Rank B worker. The message chat box CANNOT be left blank. In the chat box, the Manager can write anything he or she wishes to communicate to each worker. After the Manager submits the individual messages, each worker will see the message sent to him or her and will have to send a message back to the Manager.*]

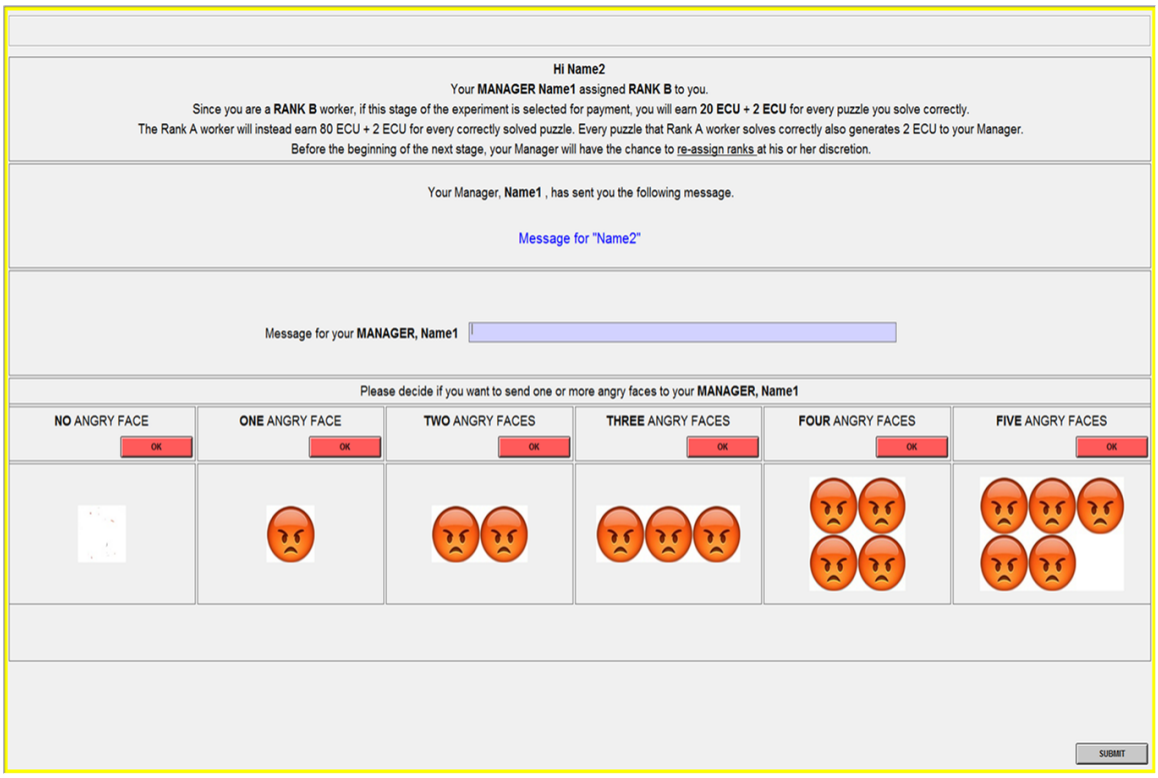
[**CHOICE & TALK and CHOICE & BACKLASH**:

The Manager will see the following screen and will have to send a message to each of the two workers. As before, we are referring to the Manager as "Name 1" and the workers as "Name 2" and "Name 3" but in the actual experiment the names of the three group members will be displayed.

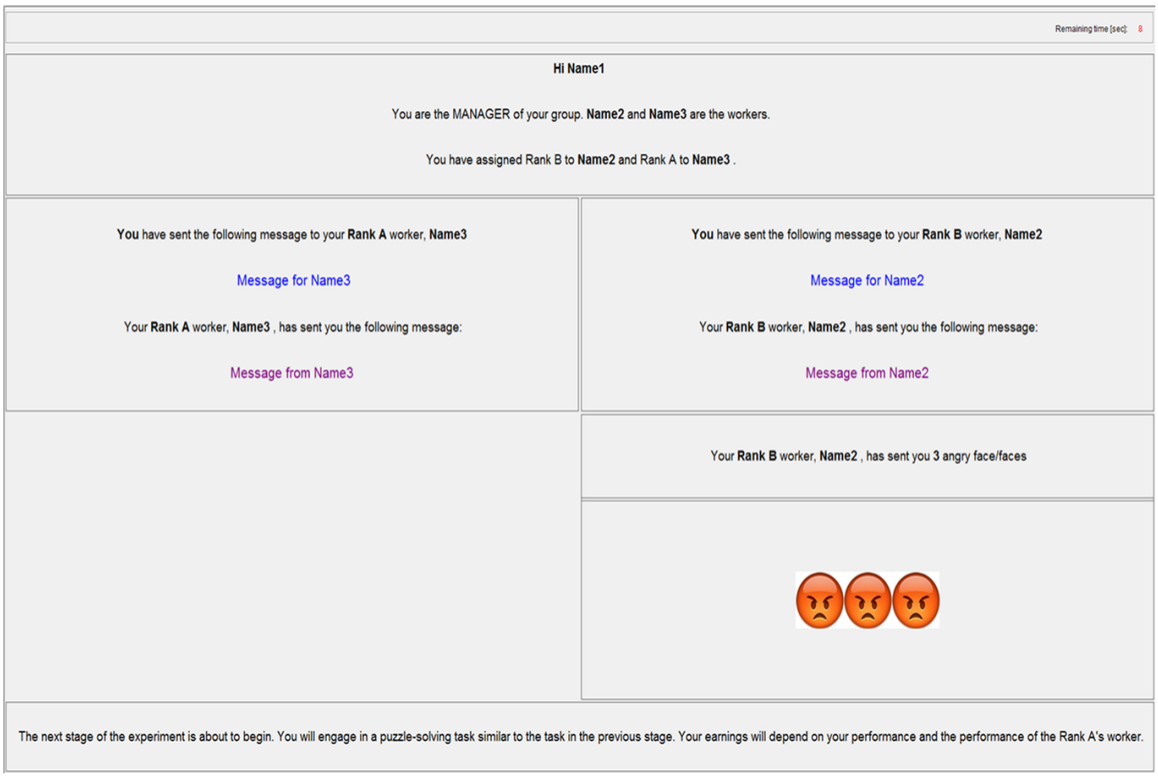


[**CHOICE & BACKLASH:** *Each worker will see the message sent to him or her and will have to send a message back to the Manager. In addition, the Rank B worker can send one or more angry faces to the Manager to express their disapproval of the ranking decision. In particular, the Rank B's worker can send up to 5 angry faces to the Manager, as shown below.*

Please turn this page to see the screen that Worker Bs and the Manager will see. Rank B worker will see the following screen and will have to decide how many angry faces, if any, he or she will want to send to the Manager. Rank B worker will also have to write a message in the chat box, in response to the Manager's message. Rank A worker will see a similar screen, except that he or she will not be able to send angry faces to the Manager.



The messages sent by the workers will be displayed to the Manager as shown in the screen below before the next stage begins. [Please note that in this example there is no actual text displayed in the Message Box.] In the example below, Rank B worker has sent 3 angry faces. Remember that Rank B can send between 0 and 5 angry faces.]



ALL TREATMENTS: At the end of each of the next 5 stages of the experiment, the Manager will have to decide whether to keep or re-allocate ranks A and B to the two workers, at his or her discretion.

[**CHOICE & TALK**: *The Manager will also have to send messages to Rank A and Rank B worker before the beginning of each stage of the experiment.*]

[**CHOICE & BACKLASH**: *The Manager will also have to send messages to Rank A and Rank B worker before the beginning of each stage of the experiment, and the workers will have to reply to those messages. At the beginning of each stage, the Rank B workers will also have to decide whether to send angry faces to the Manager.*]

ALL TREATMENTS: Is the role of the Manager clear? Please raise your hand if you have any questions about the next 5 Stages of the experiment.

[**ALL CHOICE TREATMENTS**: *In the next screen, you will be asked* ***whether you want to be the Manager of your group.***]