Who will lead? Evidence on leadership exposure and the gender gap in aspirations
Preanalysis plan

Sofie Cairo Orsola Garofalo Valentina Tartari*

November 2025

1 Introduction

The EQUALIS diversity barometer (https://equalis.dk/diversitetsbarometer/) shows that women are underrepresented in both private and public leadership positions in Denmark. This is true even in firms with gender parity at entry level, indicating that women are systematically promoted at lower rates than their male peers. The share of women among managers decreases at each step up the career ladder, with the gap particularly pronounced at the C-suite and executive board levels.

While inequality in promotions conflicts with general fairness concerns and the SDG5 of the United Nations Global Compact (https://sdgs.un.org/goals/goal5), the gender gap in promotions may also be an indicator of inefficient allocation of talent to managerial positions (Hsieh et al., 2019), and since managerial quality is crucial for firm productivity (Bloom & Van Reenen, 2007), it can lead to suboptimal growth and

*Cairo: Copenhagen Business School and University of Zürich, sc.si@cbs.dk, Garofalo: Copenhagen Business School og.si@cbs.dk. Tartari: Stockholm School of Economics and Copenhagen Business School, valentina.tartari@hhs.se.
We would like to thank the Carlsberg Foundation for an "Infrastructure Grant" and an "Internationalization Grant" and the Sparnord Foundation for a postdoc grant. We would also like to thank seminar participants in the CNEE network meeting 2023, and at the Arne Ryder Workshop 2024 for "Experiments and Behavioral Economics in Entrepreneurship, Management, and Organization".

productivity at the firm level. Moreover, there is evidence that mixed gender teams with women leaders are in a better position to draw on the skills and knowledge of both male and female team members (Karpowitz et al., 2024), resulting in higher productivity (Adhvaryu et al., 2023), higher quality solutions (Karpowitz et al., 2024), and more breakthrough innovation (Reber et al., 2021) relative to mixed gender teams with male leaders, which tend to underutilize the skills and knowledge of their female team members. The latter result reflects that female members are more reluctant to offer their knowledge input in male dominated groups, male led teams, and in male stereotypical knowledge domains (Coffman, 2014; Born et al., 2022). Hence, mitigating the gender gap in promotions holds a promise to strengthen firm profits, productivity and innovative capabilities.

Addressing the issue of gender gaps in leadership requires an understanding of how firms organize their promotion processes. Most firms either solicit internal and external applications for managerial positions or rely on managers' evaluations of their employees' promotion potential to identify, train (via talent programs), and recruit promising candidates for promotion. Consequently, both gender gaps in individual aspirations for leadership roles, which translate into gender gaps in application rates for managerial positions, and gender gaps in evaluated managerial potential of employees, may act as contributing factors to the underrepresentation of women among candidates for promotion (Benson et al., 2024). That is, if neither a woman herself nor her manager puts her name forward for promotion, she simply does not enter the pool of candidates for promotion. In addition, there may be biases that influence the likelihood of passing from being a potential candidate for promotion to actually being assigned a promotion. Hospido et al. (2022) provide promising evidence that such biases in the promotion process can be addressed institutionally with top-level organizational support. Interestingly, the policy change, which they study, was overall successful in closing the gender gap among managers at the European Central Bank due to a higher success rate for women conditional on applying, but actually resulted in a lower share of applications for promotions from female candidates.

Meanwhile, there is limited prior evidence on how to close the gender gap in leadership aspirations, which emerges already in adolescence (Alan et al., 2020), persists among university graduates and entry-level employees, and deepens further over careers (Azmat et al., 2025; Haegele, 2024).

This motivates our research question and intervention. The question we try to answer is: Can individual leadership aspirations be exogenously stimulated by randomly assigning individuals to a leadership role on a team that has to solve an abstract collaborative task requiring both communication and decision-making. The channel through which this random leadership experience may stimulate willingness to lead is through making own leadership preferences and ability salient to the selected individual, as well as emphasizing typical elements of a leadership role, namely a specific title, certain responsibilities, some private information, and economic incentives.

While leadership tasks in the workplace are typically not randomly assigned, but are rather a strong signal and predictor of future promotions(Bircan et al., 2024), we investigate if these leadership experiences in and of themselves stimulate individuals' willingness to take on leadership roles. We hypothesize that such leadership experiences may be particularly relevant to groups of employees who are underrepresented among managers, and as such have fewer role models, leading to uncertainty and lower confidence in own fit and skills to lead.

Specifically, we conduct an RCT on university students, a large share of which business students, expecting to work in private sector companies after graduation. The RCT is framed as a training session on teamwork and collaboration, centered around playing two rounds of a newly developed multiplayer online game in groups of four at a gaming center in Copenhagen.

In round one of gaming, we randomly assign group leaders who are granted a title, certain responsibilities, access to private information, and individual performance-based economic incentives. After round one, preferences for playing the leader in round two are elicited. Individuals are asked if they are willing to be the leader of the team (1), Would accept to be promoted to leader (2), are indifferent between being a leader or a team member (3), or prefer to be a team member (4). Preference elicitation is incentivized and binding. Participants are informed that teams will be reshuffled and that within each new team, the participant expressing the stronger leadership preference (equivalent to a lower number) will be promoted with ties being settled randomly, and that individuals selecting the fourth option will never be asked to lead. That is, based on their expressed preference, participants are put first, second, or third inline to be the leader of their new team. In the predefined round two groups, the group member with stronger leadership preferences is then promoted to be the leader

in accordance with instructions, and round two is played.

The experimental data are combined with pre-elicited measures of skills and preferences for leadership and gaming. Comparing willingness to lead among randomly selected leaders and randomly selected team members allows us to derive the causal effect of treatment on leadership preferences.

We hypothesize that male team members would already be more likely to express a willingness to lead in the absence of treatment, while women would be more likely to express a willingness to be promoted, as they dislike putting themselves first in line and may be uncertain about their abilities to lead. Next, we hypothesize that all treated individuals will express a greater willingness to lead after having had a leadership experience because they now have a greater understanding of the expected responsibilities and role of the leader. However, we believed that the causal effect will be greater for women, because of their lower priors.

In addition to stimulating individual leadership aspirations, the goal of our intervention is to increase the pool of potential candidates who are interested in taking on leadership responsibility, and thus increase the pool of potential candidates from which firms can choose their next manager. Conditional on a fair promotion process, this is an important avenue towards gender equal representation in management. Equally important, we assess the average quality of leadership-exposed participants who are willing to lead to the average quality of those team members who express a willingness to lead. The quality of promoted treated and non-treated leaders can be proxied by round 2 group performance.

Finally, we examine differences in the performance, risk-taking, decision-making, learning, and communication of randomly assigned male- and female-led teams. Performance is directly measured by units of minerals collected by the team during the game, while communication is measured by number of interruptions, share of time spent on active listening, and the division of talk time between team members. We also assess whether equalized communication predicts higher team performance. Risk-taking is measured by chosen routes to search for mineral, i.e., distance to base camp relative to number of emergencies, such as incidents of hypothermia or running out of oxygen.

From a pilot study that we ran in November 2024 to test the design and technical infrastructure of the experiment, we found indicative evidence that both men and women who were assigned to play the leader in round one, were more likely to express willingness to lead in round two. This indicates that aspirations for leadership are malleable and that random allocation to leadership tasks can be an effective avenue to stimulate leadership aspirations. The effect of a randomly assigned leader role is as strong as the effect of having expressed preferences for being a leader prior to playing the game. Moreover, the results indicate that this may be a viable strategy to increase the pool of candidates for promotion within firms.

2 Prior literature

The literature to which we hope to contribute focuses on individual leadership aspirations, perceived promotion potential, and the allocation of promotable tasks across gender, particularly on tasks involving team or project leadership.

The prior literature has identified differential leadership aspirations at graduation and career entry as a major culprit of the gender gap in promotions because career aspirations directly affect the level of self-promotion on the job, effort and thus performance on the job, and likelihood of applying for managerial positions (Azmat & Ferrer, 2017; Haegele, 2024).

Moreover, individuals who aspire to leadership are likely to engage in private and professional opportunities to develop their leadership skills, e.g., by taking on leadership roles in student organizations or by volunteering for heading teams or tasks in their workplace to collect relevant experience and demonstrate their ability to lead. As such career aspirations are highly predictive of promotions. None-the-less, preferences for leadership do not predict leadership talent or skill per se. In a recent lab experiment using a game that required leaders to allocate tasks to workers based on the principle of comparative advantage, the authors found that the quality of leaders who self-selected into leadership roles was lower than that of individuals who were randomly selected to act as leaders (Weidmann et al., 2024). This supports the notion that personal interest in leadership may not predict leadership quality, and that promoting decisions based solely on individual preferences for leadership may not be

in the interest of firms aiming to identify the most talented leaders. Weidmann et al. (2024) found that teams, which were headed by individuals who scored high on pre-elicited 'economic decision-making', a measure of individual understanding of how to allocate scarce resources in the most efficient way, added more value to their teams on average. 'Economic decision making' has previously been found to correlate with higher wage earnings (Caplin et al., 2023), particularly in occupations with a high decision-making intensity. Other contributions underline that individuals with relatively high line or solo performance (and field experts) are not necessarily the best team leaders. Benson et al. (2019) analyzed observational data from a US-based sales firm and found that the quality of leaders (leader-value-added as proxied by team performance), who were promoted based on high individual sales performance, was lower than that of leaders with lower solo performance, but greater experience in coordinating team sales. They drew the conclusion, that incentivizing high performance with pecuniary bonuses rather than promotions was preferable, while promotion decisions should evaluate demonstrated team work ability.

A related literature documents that managers assess the leadership potential of men and women differently. When performance is held fixed, male employees receive higher managerial potential ratings than their female peers, and this translates into a higher probability of promotion for men at all levels of the firm hierarchy (Benson et al., 2024). In firms that rely on managers to identify candidates for talent programs and promotions, men and women employees are then promoted at different rates because managers systematically evaluate the leadership potential of men as higher than that of women conditional on performance.

The question is why managers perceive men to have higher leadership potential, and how this perception may affect their allocation of promotable tasks in the workplace. Bircan et al. (2024) find evidence that managers' gender homogeneous assignment of employees to visible and promotable entry-level tasks including project leadership roles (on investment projects or sales), rather than differential leadership aspirations, almost entirely explain the gender promotion gap at the Turkish bank, which they study.

Our research project extends their findings by asking how the allocation to and experience of being a team leader affects individual willingness to take on team leadership responsibility. We are not claiming to change individual aspirations for leader-

ship more generally, but rather aiming to demonstrate that within firms, a relatively low-cost policy to increase entry of candidates into the pool for promotion, is to systematically expose more employees to leadership tasks that inform them about their own leadership potential. [1]

Beyond individual career aspirations and evaluated leadership potential, structural barriers such as homophilic sponsorship and mentorship between male managers and employees, may contribute to the unequal representation of women in leadership positions. Cullen & Perez-Truglia (2023) find that smoking male workers with smoking male managers in a large Asian Bank are promoted faster and at higher rates than their female and non-smoking male colleagues.

Finally, there is a large literature on discrimination of women who aspire to be leaders or actually pursue leadership – both from above and from below (Koenig et al., 2011; Blau & Kahn, 2017). Women who show ambition for leadership, risk repercussions from their current line manager, cf. talent hoarding (Haegele, 2022), and employees of female leaders tend to deliver less effort than they would for a male leader (Kacperczyk et al., 2023), while particularly male employees are more likely to evaluate their female leaders as unqualified (Chakraborty & Serra, 2024).

Moreover, there is evidence that men and women conditional on applying are hired and promoted at different rates due to biases (Neumark, 1999), especially in male-dominated domains. However, Hospido et al. (2022) find that biased decision-making in hiring and promotion can be turned around by top-down diversity policies, cf. observational data around a policy change at the European Central Bank. However, for such policy changes to be effective, individuals typically will need to apply or express interest in a position or promotion.

[1] In future experiments, we aim to further examine whether a randomly assigned leadership role, affects the perceived promotion potential of the individual relative to randomly assigned team members. By varying the knowledge managers have on who was randomly selected for leader and whether the team that they led had a high performance, we hope to separate the effect of allocation to leader and the effect of demonstrated leadership ability (as proxied by group performance) on promotion potential.

3 Design

3.1 Formalities

The student experiments are planned for the fall of 2025 and the spring of 2026. The experiments take place at the gaming center Let's Play in Copenhagen, Denmark, which offers a suitable field setting for running our experiment based on computer gaming in teams.

Ethical approval has been obtained from Copenhagen Business School, and the experiments will be preregistered with The American Economic Association's registry for randomized controlled trials (https://www.socialscienceregistry.org/).

The project is financed by three grants obtained from the Carlsberg Foundation and the SparNord Foundation.

3.2 The game by Intraction

The RCTs that we run rely on a newly developed multi-player computer game by the Swedish start-up company *Intraction*, which offers an artificial setting for training collaboration skills similar to those required in workplace collaborations. The game is designed for teams of four players. The players communicate by voice (online) throughout the game.

The setting of the game is a mission to save mankind by collecting as much of the rare mineral uridium as possible in a cold and barren landscape and carry it back to base. The collection of mineral requires collaboration and communication between team members outside and inside of the base. Outside the base, team members must carry their own oxygen, heaters, and med kits to avoid suffocation and hypothermia. There is a high risk of depleting one's resources, and when that happens team members depend on other team members coming to their rescue.

At different stages of the game, group members can make use of personal abilities (superpowers), activate "Entanglement" to allow for sharing of abilities, and use their

creative and analytical skills to build drones that can be used to transport mineral and save lives. All the latter elements serve to make the collection of mineral more efficient.

Each game consists of two 15-minutes game loops for mineral collection, interrupted by a short debrief, where prompts induce shared learning and reflection. After the completion of the second loop, participants complete a 360-degrees survey, and receive feedback on team and individual performance, as well as on team collaboration, communication and decision-making metrics.

The game does not have built-in economic incentives, but we incentivize team collection of mineral that is team performance[2].

Performance incentives are similar for leaders and other group members in round one, where leaders are randomly allocated. Team members and leaders alike receive a bonus of USD10 if their group 1) finishes with top-three uridium collection and 2) is drawn in a lottery among the top-three. In round two, where individuals are selected for leaders based on elicited aspirations, leaders receive a higher bonus than team members (USD 20 vs. USD 10), if their group 1) finishes in the top-three of uridium collection, and 2) is drawn in a lottery among the top-three.

We selected this particular game due to its unique data collection, behavioral science foundation, and because the game storyline centers around a team mission that requires communication and collaboration to be successful.

The game is unique in its collection of rich and detailed data at the individual and team level. Registration of avatar movements and transcription of team conversation allow for deriving rich behavioral metrics on collaboration, decision-making, communication, risk-taking and assisting others.

3.3 Student participants

Participants are recruited among students from Copenhagen Business School (CBS) and among social science students from the University of Copenhagen (UCPH). We

[2]One could have chosen to incentivize e.g. helping and assisting others or obtaining equalized talk time etc.

cannot rule out that some UCPH students have participated in other experiments at CLEE in the past, nor that they may know each other outside of the experiment. CBS students have not participated in previous CLEE experiments.

Both the research team at CBS and student assistants from CLEE were involved in recruitment efforts, visiting lectures, approaching students on break, and writing to the existing database of CLEE.

Students from CBS are typically hired by private industry firms, banks, and consultancies after graduation, and they are known to have a high potential for becoming private sector executives and leaders. The students from UCPH, particularly in the economics and political science program, are typically hired by the central administration, by NGOs, and by public financial institutions such as the central bank, the export credit agency, and the pension funds. As such they have a high potential for becoming public sector leaders.

The recruitment flyer invites students to participate in an event on teamwork and collaboration based on playing a newly developed multiplayer computer game. The invitation email does not mention leadership aspirations nor gender gaps in leadership, promotions, or willingess to lead.

If a student sign up, they are assigned a username to be used during the experiment. That is, they are told to identify as participant pxx with email-address pxx.syy@cbs2025.dk during the experiment. The provided username/email-address is used to create a gaming profile, and generally functions as a link between the survey and gaming data.

We enroll groups of 40 students at a time, equivalent to 10 teams of four players. We repeat the RCT 13 times, such that the total number of participants is 520 students, resulting in 130 teams, headed by 65 female leaders and 65 male leaders. Each team of four consists of a leader and three mixed gender team mates.

Students are compensated economically for their participation, including the filling of a presurvey prior to the event and for showing up for 2.5 hours at the gaming center to participate in the event. As a first step in the presurvey, students are asked to give consent to make their survey responses and collected experimental data available to the research team at CBS. They also provide consent that the game supplier may use the gaming data generated during the experiment to improve their product.

3.4 Design

The intervention takes place during the first round of gaming in teams of four. Essentially, we randomly allocate one participant to play a leader role, and three participants to play team members. The former is then our treated participant, and the latter are our controls.

The intervention aims to strengthen willingness to lead among treated participants by randomly exposing them to a leadership experience. After the first round of gaming, we elicit willingness to lead in the second round of gaming for all participants, and measure any difference between treated and controls.

Concretely, playing the leader consists of having a leader title, responsibility for some simple tasks and access to some private information during the game. The treatment is therefore a bundle of elements, and we do not attempt to separate the effect of different elements.

While both leaders and team members become familiar with the functioning and mission of the game during the first round of gaming, random exposure to holding a leadership role may enhance individual taste for leadership, enhance salience of the content of the leadership role, and provide personal information to the individual on her own leadership skills.

This is likely to influence perceived payoffs to playing the leader in a future round, and as such randomly exposed individuals may be more willing to stand first in line for taking on a leadership role. Indeed, marginal candidates, who are apriori uncertain about their talent or taste for leadership may perceive that their expected returns to applying increase with leadership exposure.

We hypothesize a positive average treatment effect on willingness to lead among the treated relative to other team members. Though the treatment is clearly strongest for the individual assigned to leadership, it may also affect other team members' taste for leadership. If the willingness to lead is stimulated among other, particularly same-gender, team members, the estimated effect of randomly assigning a leadership role on willingness to lead may be biased downwards.

3.5 Procedure

The experiment proceeds as follows.

Presurvey: Prior to the event, participants are asked to fill a presurvey in Qualtrics. The questions of the presurvey are shown in the appendix. The main purpose of the presurvey is to elicit socio-demographic characteristics such as age, gender, origin, and high school GPA of participants, as well as to anchor their leadership preferences and expectations and their prior experience with computer gaming. The survey is also used to elicit risk preferences, competitive preferences, and patience.

Arrival: On the day of the experiment, participants arrive at the lab and are allocated to pre-defined teams of four people. The four team members are seated on a line, each at their own computer. Three of the participants receive team member instructions, while one participant who has been randomly selected to play the leader, receives team leader instructions.

The instruction sheets explain the mission of the game, the assigned role in the game which is either team member or leader, the road map of the gaming session, and the economic incentives. The leader instructions in addition, instruct the individual to use the title "Team leader" for their game avatar, explains the leader's tasks, e.g., to start the game, and provides some private information, i.e., private hints on how to solve the game. The team members are instructed to follow the cues of the leader.

Tutorial and game round 1: Both leaders and team members are instructed to create a gaming profile and put on their headphones. Next, they are instructed to complete an online gaming tutorial that prepares them for the game they are about to play. The hands-on tutorial involves instructions and training on how to move one's avatar in the game, how to pick up oxygen and medicine kits and store them in one's bag-pack, and how to empty one's backpack. After completing the individual tutorial, all players meet in the game lobby, and from then on start communicating via the game. The leader starts the game, which involves the completion of two game loops interrupted by a short debrief. After the completion of the two game loops the game

ends, and participants receive team and individual level feedback within the universe of the game.

Midsurvey and allocation of leaders in game round 2: Next, participants are asked to fill an ultrashort online survey in Qualtrics that elicits their willingness to play the leader in round two of gaming, and then take a break of 15 minutes. Every participant is asked whether 1) they want to play leader, 2) they will accept to be promoted to leader, 3) they are indifferent, or 4) they prefer to play a team member.

Before eliciting preference, we explain the economic incentives for taking on leadership, and we explain that the leader will gain access to some private information to help them solve the game.

We also explain that the groups for round two have already been made, and that we will select a leader randomly among anyone who has expressed a willingness to lead. If there are none, that we will select randomly among those who are willing to be promoted, and finally if there are none, we will select randomly among those who are indifferent. In the situation, where everyone on a predefined team has expressed a preference for being a team member, no leader will be selected.

During the break, the experiment assistant, assigns leader roles within each of the predefined (round two) teams based on elicited willingness to lead from the midsurvey.

Game round 2: After the break, students return to their computer, and they are no longer sitting close to their group members. They receive new instructions to complete a game session consisting of two loops. On each team three of the participants receive team member instructions and the last team member who has been promoted receives team leader instructions. The game proceeds as before.

End of experiment: After the second round of the gaming, the experiment ends, and student participants are informed about the schedule for payment of participation fees and prizes. The payments are facilitated by CLEE.

Data We rely on survey data and data collected within the game universe. The data from the presurvey and midsurvey are collected from Qualtrics and the data from the game is collected by the game supplier, Intraction, and shared with the research team.

The datasets are linkable through the assigned username (email-address) provided to enrolled students by CLEE, and used to create gaming profiles.

The presurvey elicits sociodemographics, namely, age, gender, origin, field of study, high school GPA, expected education, and parents' level of education. Next, the presurvey elicits general preferences, such as risk appetite, appetite for competition and patience, and leadership preferences and expectations for becoming a leader, perception of own proficiency in team working skills, as well as prior experience and preferences for playing computer games. Finally, participants are asked about career aspirations, facing the choice between pairs of hypothetical jobs offering, e.g., high pay or flexible hours etc.

The midsurvey elicits preferences for leading in round two, but also asks participants how they perceive the contribution of the team leader and themselves to their team's performance,

The game delivers three types of data, survey-like data (see questions in the appendix), gaming data on performance and collaboration organized by teams, individuals, and by game loop, and finally transcribed conversation data from the teams gaming.

The data on performance and innovation/creativity includes collection of mineral, opening of droids, activation of entanglement, construction of drones, use of drones, lives saved, and time spent doing nothing, while collaboration data records incidents of helping and assisting others, sharing of talk time, number of interruptions, time spent on active listening and more.

Finally, the transcribed conversation data, is used to proxy informal leadership and psychological safety on teams. These measures are derived from specific language and text constructs.

4 Analysis

The overall aim of the intervention is to increase individual appetite for taking a leadership role by random exposure to a leadership experience.

Hence, we aim to answer the research question: Does leadership exposure causally affect willingness to lead, i.e. to put oneself first in line for taking the leader role, among the treated individuals relative to other team members?

To answer this question we will run the following regression:

$$\text{Willing to lead} = \beta_0 + \beta_1 \text{female} + \beta_2 \text{treated} + \beta_3 (\text{treated} \times \text{female}) + \varepsilon$$

In this regression, we evaluate the treatment effect of leadership exposure by gender. Our hypothesis is that females overall will be less likely to express a willingness to lead equivalent to standing first in line to be the leader relative to males, because standing first in line to lead is gender-incongruent behavior - and even more so for a team task in a male-stereotypical domain such as computer gaming. Moreover, women may be uncertain of what is required from them if they step into the role of leader, and they may expect resistance from their team members if they were to demonstrate leadership.

Next, we expect that the interaction between treatment and female will be positive, because the experience of taking a leader role will stimulate willingness to lead by informing individuals of their own ability to lead, their own taste for leadership, and generally make it more salient what being a leader requires and offers. Our hypothesis is that the treatment effect will be smaller for men than for women, because standing first in line would a priori be gender role congruent.

The leader role that we are exposing individuals to involves a title, some tasks, some incentives and access to some private information. As such it does not involve any direct costs. However, exposed individuals risk to incur indirect psychological costs - e.g., if facing team member resistance.

After running the simple regression, in a second step, we control for prior leadership and gaming preferences based on measures elicited in the presurvey.

The idea is to isolate the effect of treatment, when holding leadership preferences and domain-preferences fixed. An alternative is to interact treatment with leadership preferences. This would allow us to test the hypothesis that individuals who already have leadership preferences respond less to treatment, because of lower uncertainty of own leadership preferences.

$$\begin{aligned}\text{Willing to lead} = \beta_0 + \beta_1 \text{preference for leading} + \beta_2 \text{preference for gaming} \\ + \beta_3 \text{female} + \beta_4 \text{treated} + \beta_5 (\text{treated} \times \text{female}) + \varepsilon\end{aligned}$$

Finally, we ad a control for prior gaming experience to explore if the treatment effect is primarily driven by individuals with prior domain-specific experience.

$$\begin{aligned}\text{Willing to lead} = \beta_0 + \beta_1 \text{preference for leading} + \beta_2 \text{preference for gaming} + \beta_3 \text{gaming experience} \\ + \beta_3 \text{female} + \beta_4 \text{treated} + \beta_5 (\text{treated} \times \text{female}) + \varepsilon\end{aligned}$$

We will also run the regressions with sociodemographic controls.

Overall, we are interested in knowing the effect of treatment on standing first in line for a promotion by gender. This will inform diversity policies in the work place. Particularly, the potential for stimulating willingness to lead among women.

4.1 Pool size and average quality of candidates for promotion

The second set of question we ask concern the size and quality of the pool of candidates for promotion. In this part of the analysis, we measure share of treated who are willing

to lead and/or willing to accept a promotion relative to the share among controls to evaluate whether firms implementing this type of intervention would have a greater pool of leadership candidates to choose from.

Finally, we ask whether the average quality and the quality distribution among treated individuals who are either willing to lead and/or willing to be promoted is higher than among controls. As a proxy for leader quality, we consider the team performance, that is aggregate collection of uridium, and team communication, in round two of gaming.

Assuming that not all candidates would be promoted, it is particularly important to compare the upper tails of the quality distributions, and evaluate whether the firm would have access to a similar set of top candidates with and without a policy that randomly exposes entry level employees to some leadership responsibility.

We will consider both team performance with and without the contribution of the actual leader as a proxy for leader quality. If the leader was a real expert himself or just highly motivated by economic incentives, then high team performance could be primarily driven by solo player performance, rather than by team effort and collaboration, which would then imply that the leader quality would vanish as the team task at hand changed.

One important factor to control for when comparing the performance of the teams of treated (and promoted) leaders relative to control (and promoted) leaders is the cumulative prior leadership experience of the team.

When considering the team composition, the share of individuals with prior leadership experience from round one of gaming may correlate with team performance, because prior leaders have had access to some private information.

4.2 Quality of leader by gender

We are also interested in knowing, whether random exposure to a male or a female leader changes team performance and communication in round one of the game.

4.3 Extensions of the analysis

We plan to do exploratory analysis beyond the main analysis to understand the team mechanisms at play.

Informal leadership and perceived leader contribution One question is whether team members are as compliant and attentive when facing a female relative to a male leader. Let's imagine that we find that random exposure to a male leader in round one enhances team performance relative to random exposure to a female leader. Then we would like to be able to understand whether the team members systematically behave differently with a male relative to a female leader.

In the midsurvey, we do ask participants to rate the performance contribution of their leader. The perceived contribution can then be compared to the actual contribution on communication and performance of the individual leaders, cf. game metrics. This will allow us to explore whether female leaders are judged differently than male leaders.

We also hope to explore the transcription of the team communication to explore whether the assigned formal leader is also the person with greatest informal leadership as expressed in dialogue and team exchanges. For instance, one may study whether male and female led teams are equally likely to experience another team member than the formal leader taking on informal leadership.

We hypothesize that a female leader may face a higher resistance, or questioning of her decisions than a male leader due to male stereotypes around computer gaming.

Collaboration, communication, and psychological safety on teams: We aim to shed light on the correlation between team collaboration and team performance. Relying on transcribed conversation, we will derive team psychological safety, and investigate if it is correlated with sharing of talk time, interruptions, active listening, that is whether it is correlated with team performance via team collaboration.

Another aspect of collaboration that could enhance team performance, is the use of learning opportunities, e.g. during the debrief session between the game loops.

We hypothesize that teams with higher psychological safety are more likely to discuss mistakes and diffuse knowledge during the debrief session, leading perhaps to greater performance increases in the second loop of gaming.

5 Pilot

We ran a pilot experiment with 44 students equivalent to 11 groups on November 25, 2024. Participants were recruited and paid by the Laboratory for Experimental Economics at University of Copenhagen (CLEE). The venue was the gaming center Let's Play on Nordre Fasanvej. Participants received approximately 50DKK for filling the presurvey and 450DDK for their participation. Students who showed up and were sent home were paid DKK100. At the experiment, we were two researchers, two representatives from Intraction, two assistants from CLEE, and two clerks from Let's Play.

Before the experiment, we had asked students who were interested to sign up and fill a presurvey. The survey took 15 minutes to complete and was filled by 60 people (November 20-23). On the day of the experiment, participants arrived and were assigned a seat and a group. They received instructions according to the role they would play. They then signed into the game and completed a preassessment survey (15min) and a gaming tutorial (15min). During the experiment groups of four played the game for two rounds of each 45 minutes (3 game loops), and answered one survey between Round 1 and Round 2 of gaming (midsurvey 10 minutes), and a final survey after the experiment (endsurvey 10 minutes). They also had a break of 15 minutes between round 1 and 2.

In the pilot we enrolled eight teams with leaders and three groups without leaders. The analysis indicated that the effect of treatment when comparing the treated to team members in groups with a leader was similar to the effect when comparing treated and team members from groups with no leader. To increase power, we have therefore decided to enroll only groups with leaders in the main experiment.

5.1 Pilot descriptives

The following section contains some descriptives from the pilot. We had a total of 58 students who filled the presurvey, of which a majority went on to participate in the RCT.

From the descriptives, we see that the women who participated, on average, were older, had higher GPAs, and slightly higher preferences for leadership than the participating men. However, the men were more likely to expect to become leaders in their future jobs and had greater preferences for competition.

It is worth noting that we had the students play the decision-making game from Weidmann et al.(2024), which tests individual ability to allocate workers to tasks according to the principle of comparative advantage. We found no correlation between the rank they achieved in the decision-making game, and their individual GPA. Neither did we find a correlation between their rank in the decision-making game and their performance in the present game by Intraction. One obvious difference is that the present game requires continuous communication and collaboration to maximize team performance.

Table 1: Descriptive Statistics and T-tests by Gender

	Men		Women		T-test
Variable	Mean	SD	Mean	SD	p-value
Danish origin	0.72	0.45	0.45	0.51	0.050
Age	26.00	5.30	30.32	8.77	0.045
High school GPA	9.40	1.94	10.32	1.64	0.062
Social science field	0.78	0.42	0.41	0.50	0.007
STEM field	0.06	0.23	0.05	0.21	0.866
Other field	0.17	0.38	0.55	0.51	0.005
Rank, cf. Weidmann et al(2024)	82.56	12.89	85.24	9.09	0.404
Game ability (self-rated)	5.97	2.08	6.36	2.13	0.496
Game preference	6.08	2.55	6.00	2.73	0.908
Expects to be a leader	0.47	0.51	0.41	0.50	0.646
Likes leadership	0.14	0.35	0.18	0.39	0.677
Willingness to take risks	5.78	2.40	5.68	2.19	0.877
Patience	5.36	3.43	5.73	3.27	0.687
Competitiveness	7.36	2.47	6.18	2.02	0.053
Observations	36		22		58

5.2 Analysis

The pilot we ran was hugely underpowered, however, we used the data to test our main hypotheses.

The first question we aim to answer is, does leadership exposure causally affect willingness to lead, i.e. to put oneself first in line for taking the leader role, among the treated individuals relative to other team members?

To answer this question we ran the following regression:

$$\begin{aligned}\text{Willing to lead} = \beta_0 + \beta_1 \text{preference for leading} + \beta_2 \text{preference for gaming} \\ + \beta_3 \text{treated} + \beta_4 (\text{treated} \times \text{female}) + \varepsilon\end{aligned}$$

Using the pilot data, we compare 1) treated to all other team members, 2) treated to other team members controlling for assignment to a leader-team 3) treated to team members on non-leader teams, and 4) treated to other team members on leader-teams.

We find that only the treated women are significantly more likely to select "Willing to lead", cf. results in Table 2. Our preferred specification in column(4) compares the treated to team members from teams with a leader.

The coefficient on the interaction term $Treatedxfemale$ is positive and significant across all specifications. It is worth noting that the coefficient on female is insignificant, because we control also for initial leadership preferences, which are lower among women. We do not find any significant treatment effect on "Accept to be promoted" nor on the joint category "Willing to lead OR accept to be promoted", cf. Tables 10 and 11 in the appendix.

The pilot results indicate that in firms, which rely on individuals to come forward to enter the pool of potential candidates, our treatment may contribute to a higher number of women coming forward and entering the pool of potential candidates. Whether our hypothesis will find support when we run the main experiment is of course not a given.

Table 2

	Dependent variable:			
	Willing to lead			
	(1)	(2)	(3)	(4)
Preference for leading	0.131	0.126	0.096	0.118
	(0.262)	(0.267)	(0.310)	(0.268)
Preference for gaming	−0.099**	−0.097**	−0.097*	−0.099**
	(0.044)	(0.045)	(0.050)	(0.045)
Treated	−0.017	0.086	−0.062	−0.042
	(0.205)	(0.322)	(0.226)	(0.211)
Member, leaderteam		0.114		
		(0.269)		
Female	−0.231	−0.230	−0.299	−0.266
	(0.160)	(0.163)	(0.188)	(0.169)
Treated x female	0.783**	0.785**	0.856*	0.819**
	(0.374)	(0.380)	(0.411)	(0.384)
Constant	1.008***	0.893**	1.039***	1.033***
	(0.311)	(0.417)	(0.355)	(0.319)
Observations	31	31	26	29
R^2	0.506	0.510	0.471	0.511
Adjusted R^2	0.407	0.387	0.338	0.405
Note:				*p<0.1; **p<0.05; ***p<0.01

The second question we ask is, does the pool of candidates for promotion increase with treatment, such that firms have access to at least the same number of candidates, when deciding on whom to promote? The pool of candidates could either be defined as the number of individuals, who are willing to lead, or as the number of individuals, who are either willing to lead or willing to accept a promotion.

For the pilot, we found that the share of potential leaders was higher among the treated than among the controls, both when focusing on the narrow and the broad definition. The share of individuals in the pool for promotion was approximately 20 ppts. higher among the treated (random leaders) relative to the untreated (random team members), and it increased both for women and men, though most for women on "Willingness to lead", and most for men on "Accept to be promoted"

Table 3: Descriptive Statistics and T-tests by Treatment

	Untreated		Treated		T-test
Variable	Mean	SD	Mean	SD	p-value
Willing	0.23	0.43	0.43	0.53	0.394
Promo	0.27	0.45	0.29	0.49	0.927
Willing × Promo	0.50	0.51	0.71	0.49	0.326
Observations	30		7		37

Table 4: Descriptive Statistics and T-tests by Treatment

	Untreated women		Treated women		T-test
Variable	Mean	SD	Mean	SD	p-value
Willing	0.00	0.00	0.67	0.58	0.184
Promo	0.33	0.49	0.00	0.00	0.039
Willing × Promo	0.33	0.49	0.67	0.58	0.430
Observations	12		3		15

Table 5: Descriptive Statistics and T-tests by Treatment

	Untreated men		Treated men		T-test
Variable	Mean	SD	Mean	SD	p-value
Willing	0.39	0.50	0.25	0.50	0.639
Promo	0.22	0.43	0.50	0.58	0.418
Willing × Promo	0.61	0.50	0.75	0.50	0.639
Observations	18		4		22

Table 6: Group performance by treated leader

	Untreated leaders		Treated leaders		T-test
Variable	Mean	SD	Mean	SD	p-value
Mineral collection	71.94	22.70	44.83	18.94	0.110
Variation in airtime	39929.79	20730.47	39188.46	8820.03	0.936
Total interruptions	41.50	20.85	70.83	22.43	0.133
Observations	8		3		11

Now, we have seen that a higher share of treated participants enter the promotion pool. The question is then, if the average quality of individuals in the promotion pool is at least as high among the treated as the untreated. To make a preliminary assessment we can compare the performance of round two groups with a treated leader and round two groups with a non-treated leader.

Specifically, we focus on performance as measured by mineral collection in the game, and communication as proxied by division of airtime and total number of interruptions. From Table 6, we see that groups with an untreated leader have a better performance and interrupt each other at a lower frequency, than groups with a treated leader. This could potentially be a concern if this holds true in the main experiment.

We can also compare the outcomes of groups with a male and a female leader. When considering randomly allocated leaders, we see that groups with a female leader have a higher performance, more equal talk time and they interrupt each other at lower rates. In round two, the communication is still superior in female-led teams, but the performance is higher in male-led teams, cf. Table 8.

Table 7: Group performance by gender, round 1

	Male leader		Female leader		T-test
Variable	Mean	SD	Mean	SD	p-value
	mean	sd	mean	sd	p
Mineral collection	32.45	25.41	56.89	6.19	0.000
Variation in airtime	49046.83	48484.94	15613.21	10411.05	0.005
Total interruptions	71.15	27.64	38.28	16.94	0.000
Observations	22		12		34

Table 8: Group performance by gender, round 2

Variable	Male leader Mean mean	Male leader SD sd	Female leader Mean mean	Female leader SD sd	T-test p-value p
Mineral collection	69.81	24.42	46.71	7.22	0.000
Variation in airtime	39169.78	18590.44	42443.29	4703.04	0.360
Total interruptions	52.42	22.92	33.00	20.58	0.051
Observations	37		7		44

6 Planned manager pilot

In parallel with the main student experiments, we will run one experiment with a group of 40 managers from the engineering union IDA. The group of managers is not large enough to derive causal evidence, but the results from the manager pilot will still allow us to indicatively address whether students (entrants to the labor market) behave differently than experienced managers when exposed to a leadership experience. We hypothesize that managers may be more wiling to take on leadership responsibility. A direct comparison is complicated by the fact that managers know each other already, and as such are influenced by existing work place dynamics. Also, they may find the domain of computer gaming more foreign due to age differences.

7 Power calculation

Our main outcome of interest is the probability of selecting the category ”Willing to lead”, which puts the individual first in line to be the team leader. We aim to be able to detect a statistically significant effect at the 5-% significance level with a power of 80%. We saw from the pilot that the baseline proportion of individuals willing to be a leader is close to 20%, while it is 50% when pooling individuals willing to be a leader and individuals willing to be promoted. Within women the pooled share is 1/3 and within men 2/3. The treatment effect varies between 15% and 30%, and it is higher for women than it is for men on ”Willing to lead”.

Power Calculations for Difference in Proportions

Assumptions. Unless otherwise noted, calculations assume:

- Two-sided test with significance level $\alpha = 0.05$ and power $1-\beta = 0.80$ (so $z_{1-\alpha/2} = 1.96$ and $z_{1-\beta} = 0.84$).
- Independent samples with simple random assignment (no clustering or spillovers).
- Outcome is a binary indicator (e.g., willingness to lead), with baseline control proportion p_0.
- Normal (large-sample) approximation to the difference in proportions.

MDE formula. For treated group size n_T and control group size n_C, the minimum detectable effect (absolute difference in proportions) is approximated by:

$$\text{MDE} \approx \left(z_{1-\alpha/2} + z_{1-\beta}\right) \sqrt{p_0(1-p_0)\left(\frac{1}{n_T} + \frac{1}{n_C}\right)}. \tag{1}$$

We plug in p_0 (the baseline in the control group) for the variance term.

Table 9: Minimum Detectable Effects (Difference in Proportions)

Scenario	n_T	n_C	p_0	MDE (abs. diff.)
1) 130 treated vs. 390 controls, baseline 0.23	130	390	0.23	0.119 (11.9 pp)
2) 130 treated vs. 390 controls, baseline 0.50	130	390	0.50	0.142 (14.2 pp)
3) 65 treated vs. 65 controls, baseline 0.33	65	65	0.33	0.231 (23.1 pp)
4) 65 treated vs. 65 controls, baseline 0.61	65	65	0.61	0.240 (24.0 pp)

Notes: MDEs computed using Eq. (1) with $z_{1-\alpha/2} + z_{1-\beta} = 1.96 + 0.84 = 2.80$. Values shown are absolute differences; for a two-sided test the detectable increase is $p_0 + \text{MDE}$ and the detectable decrease is $p_0 - \text{MDE}$ (bounded in $[0,1]$).

Given the baseline descriptives from the student pilot, we would need at least 65 women and 65 men in order to detect minimum treatment effects of around 25% on the joint outcome of being willing to lead or accepting to be promoted. This in turn

would imply that we have a total of 130 treated and 390 controls allowing us to detect minimum treatment effects of 12-14% for the full sample. This is equivalent to enrolling 520 students split into 130 teams, consisting of 130 leaders and 390 team members. Therefore, we aim to run 13 RCTs with 40 students or 10 teams in each session. However, there is a risk that our budget will not suffice.

It is worth noting that the baseline probability of women selecting "Willing to lead" (stand first in line to lead) is likely to be lower than 20% as was found for the joint sample of men and women. In that scenario, we may be able to detect MDE's as low as 5-10% for the given sample size.

8 Conclusion

In this draft of a preanalysis plan, we have provided an idea of the types of questions our research project aims to answer and the mechanisms we plan to explore.

References

Adhvaryu, A., Howell, P., Nyshadham, A., & Tamayo, J. A. (2023). *Rapport in organizations: Evidence from fast food*. Harvard Business School.

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

Azmat, G., Cuñat, V., & Henry, E. (2025). Gender promotion gaps and career aspirations. *Management Science*, *71*(3), 2127–2141.

Azmat, G., & Ferrer, R. (2017). Gender gaps in performance: Evidence from young lawyers. *Journal of Political Economy*, *125*(5), 1306–1355.

Benson, A., Li, D., & Shue, K. (2019). Promotions and the peter principle. *The Quarterly Journal of Economics*, *134*(4), 2085–2134.

Benson, A., Li, D., & Shue, K. (2024). Potential and the gender promotions gap. *Available at SSRN*.

Bircan, Ç., Friebel, G., & Stahl, T. (2024). Gender promotion gaps in knowledge work: The role of task assignment in teams.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, *55*(3), 789–865.

Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The quarterly journal of Economics*, *122*(4), 1351–1408.

Born, A., Ranehill, E., & Sandberg, A. (2022). Gender and willingness to lead: Does the gender composition of teams matter? *Review of Economics and Statistics*, *104*(2), 259–275.

Caplin, A., Deming, D. J., Leth-Petersen, S., & Weidmann, B. (2023). *Economic decision-making skill predicts income in two countries* (Tech. Rep.). National Bureau of Economic Research.

Chakraborty, P., & Serra, D. (2024). Gender and leadership in organisations: The threat of backlash. *The Economic Journal*, *134*(660), 1401–1430.

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

Cullen, Z., & Perez-Truglia, R. (2023). The old boys' club: Schmoozing and the gender gap. *American Economic Review*, *113*(7), 1703–1740.

Haegele, I. (2022). Talent hoarding in organizations. *arXiv preprint arXiv:2206.15098*.

Haegele, I. (2024). The broken rung: Gender and the leadership gap. *arXiv preprint arXiv:2404.07750*.

Hospido, L., Laeven, L., & Lamo, A. (2022, 09). The Gender Promotion Gap: Evidence from Central Banking. *The Review of Economics and Statistics*, *104*(5), 981-996. Retrieved from https://doi.org/10.1162/rest_a_00988 doi: 10.1162/rest_a_00988

Hsieh, C.-T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The allocation of talent and us economic growth. *Econometrica*, *87*(5), 1439–1474.

Kacperczyk, O., Younkin, P., & Rocha, V. (2023). Do employees work less for female leaders? a multi-method study of entrepreneurial firms. *Organization Science*, *34*(3), 1111–1133.

Karpowitz, C. F., O'Connell, S. D., Preece, J., & Stoddard, O. (2024). Strength in numbers? gender composition, leadership, and women's influence in teams. *Journal of Political Economy*, *132*(9), 3077–3114.

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? a meta-analysis of three research paradigms. *Psychological bulletin*, *137*(4), 616.

Neumark, D. (1999). Wage differentials by race and sex: The roles of taste discrimination and labor market information. *Industrial Relations: A Journal of Economy and Society*, *38*(3), 414–445.

Reber, M., Schmallenbach, L., & Singh, H. (2021). The role of female versus male stars in team knowledge recombination. In *Academy of management proceedings* (Vol. 2021, p. 14369).

Weidmann, B., Vecci, J., Said, F., Deming, D. J., & Bhalotra, S. R. (2024). *How do you find a good manager?* (Tech. Rep.). National Bureau of Economic Research.

9 Appendix

Pre-survey

Student Surveys

Surveys are implemented in Qualtrics and distributed via email. Students were asked to insert their ad-hoc email, which then functions as an identifier across the datasets.

Survey Invitation to Students

Thank you for your interest in participating in our experiment, which consists of a presurvey (approximately 30 minutes) and a gaming session of 3 hours on November 25 from 13:00 to 16:00. Participants will play a multiplayer computer game for groups of four developed by the Swedish start-up Intraction.

The gaming session will take place on November 25 from 13:00 to 16:00. It will be hosted at Let's Play, Nordre Fasanvej 113, 2000 Frederiksberg, which is located a 15-minute walk from the metro stop Fasanvej.

Remember to Bring the Following Items on the Day of the Experiment

- Bring your cell phone, and make sure you can access your private email on your phone. During the experiment, we will distribute short surveys via your private email.
- Bring the following email and password to the venue, as you will need to use the email to create a user profile for gaming.
  - **Mail:** px@kuleemail.com
  - **Password:** cbsku778

- On the day of the event, you will need to open the email inbox and retrieve your code to activate your user profile for gaming. The inbox can be accessed at `https://login.one.com/mail`.

Payment

Payment for filling the current survey is DKK 50, which we will pay you upon arrival at the gaming session. The payment for the gaming session is 3 × DKK 140. The total payment for filling the survey and completing the experiment will be DKK 470. In case of oversubscription, participants who have filled the survey and show up for the experiment but are sent home will receive DKK 100.

To participate in the experiment, the presurvey must be completed no later than November 22, 2024, at CET 23:00.

We thank you in advance for your participation!

Press here to Start the survey: `[SURVEYLINK]`

Survey

Consent Form

Before we start, we ask for your consent for the following items:

Do you allow the research team at Copenhagen Business School to access and analyze the following data?

- Survey data from the present survey and surveys during the experiment.
- Experimental data as collected in the game.
- Video recording of the gaming session, including group conversation.

Response options: YES, I do consent/NO, I do not consent and I will not come to the experiment.

10 Sociodemographics

Next, we would like to ask you some questions about your background.

- [[**birth**]] What year were you born? [Scroll wheel 1960-2010][3]
- [[**gender**]] What is your gender? [Male, Female, non-binary, other/prefer not to say]
- [[**origin**]] What is your nationality? [Danish, other]
- [[**origin**$_{o}ther$]]$IFOTHER: Pleasespecify$[[**hs**]]$Didyoucompletehighschool(HHX,HTX,STX)$?{
- [[**grade**]] IF YES: What was your GPA from high school? {-3,0,2,4,7,10,12}
- [[**exp_educ**]] What is the highest education you expect to complete? {high school, BA, MA , PhD, other, please specify}
- [[**educ_ongoing**]] Are you currently a student? {YES/NO}
- [[**field**]] What do you study? [Free text]
- [[**dad_educ**]] What is the highest completed education of your father? {Primary school, Secondary school, BA, MA (Kandidat), PhD, other}
- [[**mom_educ**]] What is the highest completed education of your mother? {Primary school, Secondary school, BA, MA (Kandidat), PhD, other}
- [[**parent_manager**]] Did any of your parents ever hold a position as a manager with responsibility over other employees? {YES/NO}

[3] We construct [[**age**]]

11 Prior Leadership Experience

Now, we invite you to think back to your adolescence (age 12 to 20).

- **[[sport]]** Did you practice competitive sports? {Yes/No}

Prior Game Experience

[[game_intro]]Now we would like to ask you about your previous experience with playing games.

- **[[game_skill]]** How good are you at solving games, such as for instance Word Puzzle or Sudoku? Please move the slider to a number between 1 and 10 to indicate your level of proficiency, when 1 is "Extremely bad" and 10 is "Extremely good".
- **[[game_pref]]** How much do you like to play and solve games, such as for instance Word Puzzle or Sudoku? Please move the slider to a number between 1 and 10, when 1 equals "I don't like it at all" and 10 equals "I like it very much."
- **[[pcgame_any]]** Have you ever played any computer games, such as FIFA, Super Mario or similar games? {Yes, No}
- **[[pcgame_freq]]** IF YES: How many times have you played? {A few times, several times, I play on a regular basis}
- **[[mpgame_any]]** Have you ever played multiplayer computer games such as Fortnite, CS:GO, or World of Warcraft? {Yes, No}
- **[[mpgame_freq]]** IF YES: How many times have you played? {A few times, several times, I play on a regular basis}

Self-Perception

Now we would like to ask you some questions on how you perceive yourself.

- **[[compete]]** How competitive do you consider yourself to be? Please choose a value on the scale below, where 0 means 'not competitive at all' and 10 means 'very competitive'. {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
- **[[leader_enhance_perform]]** Do you think a team leader can enhance group performance in creative and collaborative tasks? {Definitely not, Probably not, Might or might not, Probably yes, Definitely yes}
- **[[risk]]** Are you generally a person who is willing to take risks or do you try to avoid taking risks? {0 (not at all willing), ..., 10 (very willing)}
- **[[skill_enhance_perform]]** Which leadership skills do you think are key to enhance team performance in a creative and collaborative task. Please select the five most important leadership skills. {Ability to motivate a team, Communication skills, Taking responsibility, Cooperation skills, Emotional intelligence, Collaboration skills, Problem solving, Decision-making, Setting a strategy, Supporting and assisting others, Creating a safe environment}
- **[[patient]]** On a scale from 1 to 10, how willing are you to wait for a larger reward in the future rather than receive a smaller reward right now? {0, 1, 2, ..., 10}

On a scale from 1 to 5, please rate your proficiency in each of the following skills, where 1 = 'Very Little' and 5 = 'Very Much':

- Ability to motivate a team: [1] [2] [3] [4] [5]
- Communication skills: [1] [2] [3] [4] [5]
- Setting a strategy: [1] [2] [3] [4] [5]
- Decision making: [1] [2] [3] [4] [5]

- Problem and crisis solving: [1] [2] [3] [4] [5]
- Cooperation skills: [1] [2] [3] [4] [5]
- Taking responsibility: [1] [2] [3] [4] [5]
- Emotional intelligence: [1] [2] [3] [4] [5]
- Supporting and assisting others: [1] [2] [3] [4] [5]
- Collaboration skills: [1] [2] [3] [4] [5]

Career Aspirations

We would now like to ask some questions about your career aspirations after university graduation.

[[career_pref]]What best describes where you would like to see yourself five years after graduation with respect to your career? I would like to see myself in a position that ...
Please move the element you care more about to position 1 (the top position), and the element you care second most about to position 2, and so on.

- High pay
- Responsibility over a team
- Responsibility over a project
- Intellectual challenge
- Flexible hours

[[enjoy_project_lead]] How much do you think you would enjoy being a project leader in your future career?

- Not at all

- Not a lot
- A lot
- Very much

[[**exp_leader**]] Do you expect to become a team leader in your future career?

- Definitely not
- Probably not
- Not sure
- Probably yes
- Definitely yes

[[**lead_challenge**]] Next, we would like to know which type of position is most appealing to you. If you were to apply for a new position five years after graduation, which position would you prefer?

- Position A: Offers leadership of a small team
- Position B: Offers intellectual challenge.

[[**presence_flex**]] If you were to apply for a new position five years after graduation, which position would you prefer?

- Position A: Pays 20% more than B, but requires presence on at least 5 days of the week.
- Position B: Requires presence on at least 3 days a week

[[**team_small_large**]] If you were to apply for a new position five years after graduation, which position would you prefer?

- Position A: Offers leadership of a small team of max 3 people.

- Position B: Pays 20% more than A and offers leadership of a large team of 10 people.

[[**overtime_regular**]] If you were to apply for a new position five years after graduation, which position would you prefer?

- Position A: Pays 20% more than B, but often requires overtime
- Position B: Has very predictable hours

[[**collegial_unknown**]] If you were to apply for a new position five years after graduation, which position would you prefer?

- Position A: Involves leading a team that is known to be very collegial.
- Position B: Pays 20% more, but involves leading a team that you do not know.

[[**transport**]] If you were to apply for a new position five years after graduation, which position would you prefer?

- Position A: Involves 15 minutes of transportation time each way.
- Position B: Pays 30% more, but involves transportation time of 1 hour each way.

[[**exp_leader_project**]] Do you expect to become a project leader in your future career?

- Definitely not
- Probably not
- Not sure
- Probably yes
- Definitely yes

[[hours_unpredict]] In a typical working week, how often do you believe that project leaders experience an unpredictable event that causes them to work longer or different hours than usual?

- Almost never
- 1-2 days a week
- 3 or more days a week

[[leader_resist]] How often would you say that team leaders encounter resistance from employees (e.g., because they have to make an unpopular decision)?

- A few times a year
- Several times a year
- Several times a month
- Several times a week

[[enjoy_team_lead]] How much do you think you would enjoy being a team leader in your future career?

- Not at all
- Not a lot
- A lot
- Very much

MID SURVEY [Distributed after Round 1 of gaming]

Intro paragraph *You have now played 3 loops of the game developed by Intraction. The overall team collection of uridium was written in the feedback that you received after the*

game. The three teams that collected the most uridium during round 1 of gaming will participate in a lottery for DKK400, which will be paid out as DKK100 for each team member. We will ask you about your team's Uridium collection after verifying your identity.

Question 1

- Please insert the email that you used to create your gaming profile. The email has been shared with you by the Experimental Lab. It should look something like px@kuleemail.com

Question 2

- Please insert the email that you used to create your gaming profile. The email has been shared with you by the Experimental Lab. It should look something like px@kuleemail.com

Question 3

- How much uridium did your team collect in round 1 (3 loops) of gaming?

Mid paragraph *After a short break of 15 minutes you will be matched with a new group to play the game once more. Before composing the teams, we will recruit team leaders for the new teams based on personal preferences. The three groups that collect the most uridium will participate in a lottery for DKK400. The leader of the winning group will receive DKK100 extra. We will shortly ask you about your preferred role on the team for the second round of gaming. Please note that the assigned Team Leader will receive specific instructions to guide them and private information to help them solve the game.*

Question 4 Now, we would like to ask you, what role you would like to play in round 2 of gaming.

Note that, we will first make a draw of team leaders among individuals who would like to be promoted. Next, we will make a draw among individuals who would accept to be promoted. Finally, we will draw leaders among individuals who are indifferent. Individuals who prefer to be team members will not be promoted.

1. I would like to be promoted to team leader.
2. I would accept to be promoted to team leader, but I prefer to be second in line.
3. I am indifferent
4. I would prefer to be a team member.

Question 5 Did your group have a team leader in round 1 of gaming?

[r1_team] Yes, but I was a team member

[r1_lead] Yes, I was the leader

[r1_nolead] No, my group had no leader

Question 5a [[lead_qual]] IF "Yes, but I was a team member": Do you think that the leader enhanced the overall performance of the team?

- Definitely not
- Probably not
- Might or might not
- Probably yes
- Definitely yes

Question 5b [[lead_qual2]] IF "Yes, I was the leader": Do you think that you enhanced the overall performance of your team?

- Definitely not
- Probably not
- Might or might not
- Probably yes
- Definitely yes

You can now take a 15 minutes' break, while we create new groups. You are welcome to buy a drink, use the restroom, use your phone or step outside for a bit of fresh air. After the break, we will provide you with instructions for round 2 of gaming.

Conclusion

[[**outro3**]] Thanks a lot for your participation. It is a great help to our research project.

Questionnaires distributed within the game

Onboarding Questionnaire (1-7 Likert scale)

- I prioritize the needs of my colleagues.
- I quickly grasp new situations at work.
- I neglect my work responsibilities.
- I bring energy and enthusiasm to my workplace.
- I worry a lot about work outcomes.
- I treat all my colleagues with equal respect.

- I enjoy brainstorming and problem-solving at work.
- I am thorough and avoid errors in my work.
- I quickly connect with my coworkers.
- I easily become discouraged by work challenges.
- I show concern for my coworkers' well-being.
- I enjoy handling complex work-related problems.
- I strictly follow work protocols.
- I handle workplace social dynamics effectively.
- My mood changes often at work.
- I always forgive and move past conflicts at work.
- I am open to new work processes and ideas.
- I ensure accuracy in my work tasks.
- I tend to take lead at work.
- I get upset by sudden changes at work.
- I always have positive feedback for my team.
- I appreciate the importance of creativity in my work.
- I execute my work plans effectively.
- I tend to be reserved with new colleagues.
- I get stressed out easily by work demands.

Self-Perceived Level of Competencies (1-8 Unorthodox Likert Scale)

- I often take informed risks in my work.
- I share resources at work in an effective way.
- I am an equally good listener and speaker in dialogues at work.
- Depending on the situation I take different types of risk in my work.
- I take part in planning and strategizing at work to the best of my abilities.
- I communicate effectively in a leadership position.
- I make decisions at a balanced speed.
- I promote synergy and teamwork.
- I share relevant and required information with my colleagues.
- I make decisions with high quality.
- I adapt easily to new work roles and activities.
- I am in touch with my emotions.

Psychological Safety (1-7 Likert Scale) - Filled every third game loop

- When I make a mistake, the other players occasionally make me feel uneasy about it.
- I feel that it is easy to bring up things that don't work with this team.
- I feel that this team lets people know when they make mistakes or perform poorly.
- I feel that we encourage each other to try new things and take risks.
- I find it harder than usual or I feel embarrassed to ask for help in this team.
- No one in this team deliberately acts in a way that negatively affects my efforts.
- I feel that I can use my strengths in this team and that I am appreciated for it.

360 Evaluation (1-7 Likert Scale)

- I can tell that player N is working on or has improved their [skill goal].
- I can tell that player N+1 is working on or has improved their [skill goal].
- I can tell that player N+2 is working on or has improved their [skill goal].

Overview of variables in gaming data

- Uridium collect:
- Robots unlocked:
- Abilities used:
- Hypothermia events:
- Entaglement activated:
- Drones built:
- Lives saved:
- Airtime:
- Relative airtime:
- Interruptions:
- Active listening:
- Abilities shared:
- Group index:

12 Additional tables and figures

Table 10

	Dependent variable:			
	Accept to be promoted			
	(1)	(2)	(3)	(4)
Leader preferences	−0.236	−0.228	−0.314	−0.243
	(0.347)	(0.352)	(0.379)	(0.348)
Gaming preferences	0.021	0.018	0.024	0.015
	(0.058)	(0.059)	(0.062)	(0.058)
Treated	0.244	0.087	0.283	0.232
	(0.271)	(0.425)	(0.276)	(0.274)
Member, leader-group		−0.173		
		(0.356)		
Female	0.158	0.156	0.316	0.087
	(0.212)	(0.216)	(0.230)	(0.220)
Treated x female	−0.617	−0.620	−0.767	−0.557
	(0.495)	(0.503)	(0.503)	(0.499)
Constant	0.101	0.276	0.034	0.153
	(0.411)	(0.552)	(0.435)	(0.415)
Observations	31	31	26	29
R^2	0.134	0.143	0.208	0.116
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 11

	Dependent variable:			
	Willing to lead or accept to be promoted			
	(1)	(2)	(3)	(4)
Leader preferences	−0.105	−0.103	−0.218	−0.126
	(0.369)	(0.377)	(0.419)	(0.359)
Gaming preferences	−0.078	−0.079	−0.072	−0.084
	(0.062)	(0.063)	(0.068)	(0.060)
Treated	0.227	0.174	0.221	0.191
	(0.289)	(0.455)	(0.305)	(0.283)
Member, leader-group		−0.059		
		(0.381)		
Female	−0.073	−0.074	0.016	−0.179
	(0.226)	(0.231)	(0.254)	(0.227)
Treated x female	0.167	0.166	0.089	0.262
	(0.528)	(0.538)	(0.556)	(0.515)
Constant	1.109**	1.169*	1.073**	1.187**
	(0.439)	(0.591)	(0.480)	(0.428)
Observations	31	31	26	29
R^2	0.167	0.168	0.129	0.225
Note:				*p<0.1; **p<0.05; ***p<0.01