Journal of Economic Behavior and Organization 186 (2021) 12–32

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Does appointing team leaders and shaping leadership styles increase effort? Evidence from a field experiment ☆

Matthieu Chemin [a,b]

[a] Department of Economics, McGill University, CIREQ, Canada
[b] CIRANO, Canada

ARTICLE INFO

Article history:
Received 21 September 2020
Revised 18 January 2021
Accepted 10 February 2021
Available online 11 April 2021

JEL classification:
J24
I20

Keywords:
Leaders
Leadership style
Effort in teams

ABSTRACT

This paper presents the results of a randomized experiment varying the method of selection of team leaders (appointment versus self-selection), the characteristics of the appointed leaders and their leadership styles. I find that appointing high-ability and hard-working leaders increases effort, knowledge-sharing and team performance. I find even greater effects for another intervention that attempted to shape leadership style by providing detailed instructions on how to coordinate the team based on previous observation of free-riders' behavior in teams. This indicates that the coordinating abilities of leaders are particularly important in a context of a non-routine cognitive task where knowledge spillovers and the flow of creative ideas are required.

1. Introduction

Good leaders are important. Lazear et al. (2015) shows that boss effects dwarf worker and peer effects. Bloom et al. (2013) shows that teaching modern management techniques to managers of large Indian textile firms causally increase firm performance. Leaders could be especially important in teamwork to resolve free-riding issues and improve knowledge-sharing and technological complementarities among team members. Despite the clear importance of leaders, there is no causal evidence on the most appropriate method of selection of leaders (endogenous or by appointment), the characteristics to be used to select leaders, and the leadership styles that would reduce free-riding and unlock the positive effects of teamwork.

In this paper, I implement the first experiment randomly assigning leaders to teams based on different observable characteristics and randomly shaping their leadership style in order to resolve free-riding, improve knowledge-sharing, and increase team effort and output. The experiment is implemented with 166 randomly-formed teams of four students who are working on an assignment in a real university introductory-level course. The assignment has high stakes in the sense that it counts for 15% of the final grade, which is the difference between an A and a B.

One-third of these teams are randomly assigned to the "Random Leader" (RL) condition, wherein one member is randomly selected to be the team leader. One-third of the teams are randomly assigned to the "Promising Leader" (PL) condition, where a team leader is appointed based on two indicators: grade on a baseline test, and successful track record of cooperation in a previous teamwork task. The individual displaying the best performance on these two indicators from

☆ I gratefully acknowledge financial support from the Social Sciences and Humanities Research Council (SSHRC) of Canada from FRQSC Québec.
E-mail address: matthieu.chemin@mcgill.ca

https://doi.org/10.1016/j.jebo.2021.02.014
0167-2681/

within the team is appointed team leader. One-third of the teams are randomly assigned to the control group. Teams in the control group are asked to designate a "team leader". In all teams, the leaders' task is to is to upload the final slides of the team' s presentation online.

In the two treatment groups RL and PL, the name of the leader is announced to the team by the professor. In other words, his role is established by an authority figure which probably gives it legitimacy in his own eyes and in those of his teammates. In the control group, there are no instructions to organize democratic elections which could have conferred legitimacy upon the leader (as is found in Levy et al. (2011)). In fact, the "designated" leader is essentially non-existent because he is only assigned the trivial task of submitting slides at the end of the assignment. Students in those teams did not have to announce formally the name of their leader and may have proceeded without a leader altogether. We verify in this project whether the control teams actually chose a leader. Followers then respond to these leaders but this response could be quite different than when the leader is officially chosen by an external authority figure. The goal of this empirical project is to study such responses.

In an intervention completely orthogonal to the first one, I also attempt to shape leadership styles. In a randomized third of the teams (which could be either RL, PL, or control, i.e., with no leader announced by the professor), the whole team is told that the leader has the added task of monitoring and reporting free-riding issues to the professor. I call this intervention the "Monitoring" intervention (MON).

In another randomized third of the teams, I use a key insight gleaned from the observation of 191 teams over three years in previous research (Behrens and Chemin, 2018). Careful observation of free-riders revealed that they usually skipped team meetings, in particular the first one (intentionally or not), and later argued that they did not know what to do and were excluded by their teammates, hence their lack of contribution. The goal of the intervention is to avoid this by placing a person in charge of coordinating the team, organizing the first meeting, making sure everybody comes and rescheduling meetings if need be to make sure everybody is included in the teamwork. This corresponds to a reduction in the transaction costs of team meetings, which should be accompanied (at least theoretically) by a reduction in free-riding. I call this intervention the "Coordinating" intervention (COORD).

In the final randomized third of the teams, no indication is given as to what the leader should do (except uploading the final slides of the presentation online). I call this the control group of this intervention on leadership style.

The two interventions on assigning a leader (Control, RL, PL) and on leadership style (Control, MON, COORD) are independent of each other such that all combinations are possible. The experimental design is thus a 3*3 design: 3 (Control, RL, PL) * 3 (Control, MON, COORD), such that the effects of all interventions can be measured separately or in combination. I fully pre-specified the regression to run and the outcomes of interest before the study, and rigorously follow the pre-analysis plan in this paper.[1] To analyze the effect of these interventions, I collect a survey at the end of the teamwork asking if there was a leader in the team and who it was. I also ask each team member to rate themselves and their peers, in order to develop peer-rated measures of effort. I collect 6 measures of effort: whether the student is seen as contributing a fair share to teamwork, as being on time to the team meetings, as respecting the deadlines set by the group, as showing a positive attitude, as teaching others in the team, as well as a continuous peer rating of his contribution. I had not pre-specified how to account for multiple hypothesis testing. However, I use the two standard ways described in Anderson (2008): I build a summary index of effort based on these six variables and I adjust *p*-values for the False Discovery rate (FDR). I also use the Fisher exact test *p*-values (Young, 2018) that do not depend on the assumption of large samples. To understand the mechanisms through which leaders may increase effort, I collect data on leadership styles, i.e., whether the team leader monitored and reported free-riding issues (MON); and whether the team leader followed each precise step suggested to set up the first meeting and improve coordination (COORD).

The first finding of this paper is that the PL intervention increases effort by 0.33 standard deviation, while the RL intervention has no effect. Thus, the method of selecting a leader matters. Having one person of high-ability and hard-working nature being designated as the official leader by someone who exerts authority over the entire team leads to better performance than if he endogenously self-selects. In fact, the control teams have difficulty selecting a leader: the proportion of students saying they have a leader increases significantly by 24% in the PL treatment group.

The second finding of this paper is that the COORD intervention increases effort by an even greater 0.51 standard deviation (more than the MON intervention which increases effort by 0.35 standard deviation). An effect size of 0.51 is essentially moving a student at the mean of the effort distribution to the top third of the class. In the "leadership survey", I find that the COORD leaders make sure everybody comes to the first meeting, exactly in line with the key insight underlying the intervention. As a result, there are more opportunities for knowledge-sharing within the team: I find that more students teach others within their teams. The COORD intervention increases hours spent on the teamwork, and team performance.

The implications of this paper is that the method of selection and the observable characteristics of leaders (such as ability and cooperative attitude) matter. Numerous papers have found that leaders matter (Bertrand and Schoar, 2003; Lazear et al., 2015; Bennedsen et al., 2006; 2012; Schoar and Zuo, 2017; Jones and Olken, 2005; Engelberg et al., 2016; Bloom and Van Reenen, 2007), yet my paper is the first to vary exogenously the selection method of leaders (endogenous or appointment) and the characteristics of leaders. Lab and lab-in-the-field experiments have assigned leaders to teams in

[1] The AEA RCT registry did not exist at the beginning of this study, i.e., January 2014. However, I had fully specified the regression and the outcomes in my ethical approval, which is timestamped by the relevant Ethical Review Board at the end of the Appendix.

Table 1
Descriptive statistics.

	Mean	SD	Obs.
Female	0.63	0.48	658
Baseline ability (out of 3)	1.70	0.90	647
Field of Study			
Humanities	0.41	0.49	658
Business	0.06	0.24	658
Political Science	0.11	0.31	658
Economics	0.23	0.42	658
Science	0.14	0.35	658
Others	0.04	0.20	658
Year of Study			
First year	0.37	0.48	658
Second year	0.37	0.48	658
Third year	0.27	0.44	658

public goods games where players contribute money to a common pot (Gächter et al., 2012; Güth et al., 2007; Jack and Recalde, 2015), yet my paper is the first to look at a situation of real effort provision. I find that appointing as leaders high-ability and hard-working individuals has the strongest effects on team effort.

The second implication is that having a person in the team making sure everybody participates in team meetings is even more critical than having an appointed leader of high-ability or that monitors the rest of the team. This result lends support to the Rotemberg and Saloner (1993) hypothesis, according to which "leaders who adopt a participatory style maximize firm performance when the firm has the potential for exploiting numerous innovative ideas. By contrast, when the environment is poor in new ideas, shareholders benefit from hiring a more selfish leader whose style is more autocratic". This teamwork corresponds to a setting rich with "numerous innovative ideas", thus the COORD style works better than the strictly punitive MON style.

In fact, I argue that the results and this particular setting are of importance for a large section of the population and the workforce. At first glance, one may question the external validity of these findings since they come from a university course, not the actual workplace. However, the teamwork task is intellectually challenging (the presentation of a development economics paper published in a top economics journal by mostly non-economics students), and presents opportunities for knowledge-sharing and technological complementarities (students come from a variety of academic backgrounds such as mathematics, statistics, but also anthropology, history, international development; a diversity of skills that could combine well for the presentation of an empirical development paper). It corresponds very well to the definition of a non-routine cognitive task: "a task demanding flexibility, creativity, intuition, persuasion, problem-solving, and complex communications" according to Autor et al. (2003). Thus, the results of this paper are important considering the large and growing share of the population employed in non-routine cognitive tasks (34% in the US in 2007, Acemoglu and Autor, 2011) and the growing importance of teamwork in academia and the workplace.

The remainder of the paper is organized as follows. Section 2 describes the setting and conceptual framework. Section 3 details the experimental design of the study and explains how the key variables of interest are measured. Section 4 explains the empirical methodology specified in the pre-analysis plan. Section 5 provides an overview of the randomization and shows that the observables are balanced across treatment and control groups. Section 6 provides a manipulation check, i.e., a confirmation that the interventions actually set up a leader and that students followed the experimental design. Section 7 describes the effects of the interventions on various measures of effort including teaching others. Section 8 concludes.

2. Background and conceptual framework

Our experiment was implemented over two years (2014 and 2015) in a three-month undergraduate course on economic development taught by the same professor to 661 students in three different classrooms. Three students dropped out of the course, such that the total sample size is 658 students.

This course is mandatory for a program in international development studies which is offered outside the economics department. As such, the students have diverse backgrounds, with 41% of the students from humanities (this international development program as well as anthropology, english literature, etc.), 6% from business, 11% from political science, 23% from economics, and 14% from science, as shown in Table 1. 63% were female. Average baseline ability on a test given before the teamworks is 1.7 out of 3. The course also includes a broad cross-section of students in terms of their year of study, with 37% being first-year students, 37% second-year, and 27% third-year (students at McGill typically graduate in three or four years).

2.1. Teamwork tasks

Of crucial importance for this paper is the teamwork element. There are two teamworks in the course: the first one is used to detect promising leaders based on their behavior, while the second one is used to analyze the effect of leaders on team effort.

For the first teamwork, students are randomly separated into teams of four. Each team has to download data from the web, perform statistical analysis on Excel, and use concepts from the lectures to test theories which were explained in the course. This assignment is thus called "the statistical teamwork". It lasts for two weeks, counts for 15% of the final grade and provides me with an opportunity to observe the teams and detect promising leaders according to the methodology explained below.

The most important teamwork in this paper however, is the second one where students are randomly re-assigned to new teams of four.[2] They are now tasked with the presentation of an empirical paper related to development economics, more specifically, to the topics covered during lectures. This assignment is thus called "the presentation teamwork". In Fall 2014, 37 papers were selected from Econometrica, the American Economic Review, the Quarterly Journal of Economics, the Journal of Development Economics, and the American Economic Journal: Applied Economics. All papers use a randomized experiment which is essentially a comparison of means between the treatment and control groups, such that students from all backgrounds can more easily understand and interpret the results.[3] Yet, despite being simpler than other econometric techniques, firmly understanding a randomized experiment is still a challenging exercise for some students who have never taken statistical courses.

One goal of the assignment therefore, is to randomly form balanced teams, such that the diversity in the classroom naturally fosters knowledge sharing, building on the unique skill set of each student. This dimension of teamwork was emphasized in class during the explanations given about this assignment.[4]

Students have between 2 to 6 weeks to prepare (depending on the order in which the teams present) before presenting the article in their lab sessions where they are graded by a teaching assistant. The presentation must include the 4 students. The grading system is as follows: 60% for the content of the paper (did the students understand the paper?), 20% for the presentation skills (did the students present in an engaging manner?) and 20% for team cohesion (did the 4 students participate equally?). The assignment grade is the same for all team members which creates opportunities for free-riding.

2.2. Issues of free-riding

In a small number of teams, some members do absolutely nothing[5] and the presence of a free-rider is stressful for the other students.[6]

To understand why free-riding may be occurring despite its detrimental effect on the group, we can look at a simple model of cooperation in teams (see Appendix A for the formal model). Individual effort increases the common team grade, but is also costly. Extra effort by one individual indirectly benefits all other teammates through a higher collective grade. In other words, that individual exerts a positive externality on others. Team members do not internalize the positive externality they bestow on their teammates and thus provide inefficiently low effort in the decentralized equilibrium (see Appendix A for the formal derivation of the effort level in the decentralized equilibrium versus collective efficiency).

The next section explores how leadership may move the equilibrium to the pareto-efficient outcome.

2.3. The effects of assigning a leader on free-riding

Assigning a "Promising Leader" PL, i.e., a high-ability and hard-working individual, versus a "Random Leader" RL, i.e., a randomly selected individual, as the team leader may increase the effort of others in the team through a "leading-by-example" mechanism. There is empirical support for the existence of "leading-by-example" from lab experiments. In Gächter et al. (2012), one individual in a team of agents acts as leader by contributing to a public goods game prior to all others. Gächter et al. (2012) observe that the higher the leader's contribution, the higher the contribution of the other team members. Strikingly, this holds even though there are no monetary incentives that would induce such complementarities. Jack and Recalde (2015) have confirmed these findings by moving the lab to the field. They study voluntary contribution to

[2] When the total number of students was not divisible by 4 and the remainder was 3, I created a team of 3 students. This happened for 3 teams. Additionally, 3 students dropped out of the course. This created another 3 teams of 3 students. The total sample size is thus 661-3=658 students. Overall, this generated 160 teams of 4 and 6 teams of 3 (160*4+6*3=658 students). I control for team size in the regressions below.

[3] For example, Chattopadhyay and Duflo (2004), "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India", Econometrica.

[4] In the instructions for the presentation: "You will probably have in your group students with differing backgrounds (anthropology, political science, economics...). Use the skills of your colleagues adequately to do an excellent presentation, and learn in the mean time from them!". Or in the same presentation: "This work is designed for you to learn from your colleagues with differing backgrounds and skills".

[5] From emails to the professor: "I took on the whole burden of these students who did NO work." (The No written in major case by the student)."It is only now that she has contacted us (late Sunday evening), after we have completed all of the exercises, leaving her having no contribution what so ever to the assignment. There is no work in the final submission that will have been hers."

[6] "We each have two midterm exams this week, in addition to the presentation, and [...]'s absence is adding a lot of extra work and pressure on us."

books for a public library. When leaders are elected from the community versus randomly selected, the leaders give more and followers respond by giving more.

In lab experiments, it is possible to manipulate exactly which individual moves first and therefore who the leader is. In a situation of real effort provision, it is unclear whether appointing a leader will translate into that person "moving first" and followers responding, hence the need for an empirical analysis with the following testable prediction:

Proposition 1. *Assigning a PL reduces free-riding more than assigning a RL.*

The comparison of PL versus the control group is even more unclear. On the one hand, being named "leader" by an external authority figure may give legitimacy to the role. Moreover in the control group, leaders may be altogether absent if choosing a leader is difficult. Güth et al. (2007) document in their public goods experiment that when given the opportunity to endogenously select a leader, only about 40% of groups succeed in appointing a leader, even though groups with a leader outperform groups without a leader by far in their experiment.

On the other hand, in a situation of real effort provision with numerous interactions over several weeks, leaders may naturally emerge in the control group. These endogenous leaders may be even better placed than the externally appointed PL or RL to lead the team. If this is true, appointing a leader may actually be disruptive the proper functioning of a team. Overall, it is quite unclear what the effects of PL could be compared to the control group, hence the need for this empirical investigation.

2.4. The effects of shaping leadership styles on free-riding

The possibility for a leader to impose penalties on those working easy could increase effort levels (see Appendix A). Thus, somehow shaping leaders into adopting a punitive role may increase other's effort.

Proposition 2. *Leaders who monitor and punish will increase others' effort.*

Güth et al. (2007) have found support for this hypothesis in public goods games: leaders who are given the additional power to punish have an even stronger effect on the monetary contribution of followers. Of course, outside of the lab in a situation of real effort provision, punishment may invite retaliation which could dampen any positive effects found. To test this, I implement a "Monitoring" intervention (MON) that instructs leaders to make sure that everybody contributes a fair share to the presentation and to inform the professor if anybody is not contributing to the presentation.

Aside from a punishing role, the leader may also increase others' effort by coordinating the team better. Theoretically, one can simply model this with a reduction in the transaction costs of meeting, which increases equilibrium effort levels (see Appendix A for a formal derivation). The management and psychology literature has focused on the "coordinating" role of leaders in terms of creating a "positive team atmosphere".[7] In contrast, I focus on a new and simple intervention described in greater detail below, which attempts to reduce the transaction costs of meeting by having the leader act as an organizational coordinating device. I thus test the following proposition:

Proposition 3. *Leaders who coordinate will increase others' effort.*

2.5. Interaction between the interventions

The theoretical model in Appendix A makes it clear that the three mechanisms emphasized (leaders-by-example, monitors, coordinators) are essentially similar on a theoretical level. They can all be modeled by an extra cost in the utility function if effort is too low. Concerning leading-by-example, the dislike in the inequality of effort levels decreases with own effort. Concerning monitoring, a pressure is received if effort is too low. Concerning coordinating, the cost of effort is reduced by the leader taking on the coordinating role. Thus, on a theoretical level, it is impossible to differentiate these forces. A field experiment acting independently on each of the three forces is needed to disentangle these forces, and understand which dominates the others.

Moreover, the three forces interact on a theoretical level: if effort goes up because of monitoring or coordinating, then the inequality in effort level is lower, and the force of leading-by-example is weaker. Hence, the following proposition:

[7] In the meta-analysis of Collins and Elwood (2004) on leadership styles, they identified 346 studies that involved a leadership intervention of which only 2 are randomized with a control group and looking at team performance. Both Graen et al. (1982) and Scandura (1984) find significant effects of the leader-member exchange (LMX) theory. This theory emphasizes the quality of leader–member exchange relationships, trust and respect. In their meta-analysis on leadership styles, Morgeson et al. (2010) identify the role of "supporting the social climate" as an important role of leaders. Ten management studies are described where addressing interpersonal issues within the team that may interfere with the team' s performance, showing respect for team members' ideas, exhibiting warmth, facilitating positive working relationships among team members, building cohesion, setting the team climate, demonstrating consideration, and empowering team members were associated with team performance and satisfaction with the leader. In the best identified study, Phillips et al. (2001) conducted a laboratory experiment where the confederate leader was either listening quietly and making a decision with no explanations given, or listening actively, taking notes and providing verbal (e.g., repeating summaries of points made by the speaker) and visual cues (e.g., nodding), showing respect for the staff member and indicating that the points being presented were heard and understood. This study found that a leader' s supportive actions were positively related to satisfaction with the leader and attachment to the team. Finally, in a meta-analysis of team leadership behaviors, Burke et al. (2006) found positive relationships between team leader consideration behaviors and team quantity and quality outcomes.

Proposition 4. *Monitoring and Coordination crowd out leading-by-example.*

In terms of the regression that will be run, this means that the interaction terms between PL and MON or COORD should be negative.[8]

I now describe in more details the experimental design to test these propositions.

3. Experimental design

3.1. Interventions on assigning leaders

In the control group representing a third of the 166 teams, i.e., in 55 teams, the teammates are asked to designate a team leader, but for the sole purpose of uploading the final slides of the presentation online. The exact script is in Appendix 2. I collect data on the identity of the individual who uploads the slides to know the identity of the leader and the followers in each team.

In another set of 55 (randomly selected) teams, one person is randomly selected from within the group to become the leader. I call this intervention the "Random Leader" (RL). No further instructions is given to this RL on how to lead. No extra grade is given to the assigned leader.[9]

In the final 56 teams, I use baseline ability and behavior in the first teamwork (the statistical teamwork) to detect "Promising Leaders" (PL). I use the following formula to develop a leadership score:

$$LeadershipScore_i = \omega_{Ability} \sum_{i=1}^{2} Test_i + \omega_{Motivation} \sum_{j=1}^{4} PeerRatedMotivation_j$$

where i stands for individual i.

The leadership score has two terms. The first term is a measure of ability. $Test_i$ is the grade on individual tests given throughout the year. Tests #1 and #2 take place before the presentation teamwork, and serve as a measure of ability.[10]

The second term is $PeerRatedMotivation_j$, a peer-rated measure of motivation on the first teamwork. After the first teamwork, I collected a survey from the 658 students asking who in the team (1) contributed a fair share to the teamwork, (2) was on time for the team meetings, (3) respected the deadlines set by the group, and (4) showed enthusiasm for the work. For each question and each individual, I calculated the average response given by teammates. This defines four variables that I add together in $\sum_{j=1}^{4} PeerRatedMotivation_j$. Overall, this score reflects the motivation an individual demonstrated for the teamwork, as rated by its peers.

$\omega_{Ability}$ and $\omega_{Motivation}$ are weights associated with ability and motivation. Given the two tests are between 0 and 3 and the four motivation measures are between 0 and 1, I choose $\omega_{Ability} = \frac{2}{24}$ and $\omega_{Motivation} = \frac{3}{24}$, such that the leadership score $LeadershipScore_i$ is a number between 0 and 1.

I then select the individual with the highest $LeadershipScore_i$ to become the leader of the group. The intuition is that high-ability and hard-working individuals are well placed to lead-by-example, as was explained in the theoretical model. PL may also be well-placed to teach others, a hypothesis I will also explore.

3.2. Interventions on leadership style

To test whether monitoring or coordinating leadership styles matter, I give precise instructions to certain leaders explaining how they are expected to behave. This intervention completely orthogonal to the first one.

[8] This is easily seen from the theoretical model in Appendix A. I model "leading-by-example" as an additional cost of $\gamma \times (e_L - e_i)$ where e_L is effort of the leader; "leaders as monitors" as an additional cost of $\max\{\gamma \times (e_{min} - e_i), 0\}$, where $e_{\min}$ as a social norm of "acceptable" effort; and "leaders as coordinators" as a reduction in the costs of effort $C(e_i)$. If monitoring or coordinating is applied, effort increases which reduces by the same token $\gamma \times (e_L - e_i)$, i.e., leading by example.

[9] Only one student (who is in the in the PL*COORD treatment group) asked about this: "Thank you for choosing me as a group leader. I will make sure to put in my best effort in being one. I was actually wondering if this position will be of any influence in terms of grade, since it seems like it is an added responsibility on top of all other work." The answer was: "There is no influence on your grade. Except through the higher grade you can get thanks to your leadership and coordinating abilities." That same student (a "Promising Leader") wondered about the method of assigning leaders: "I was also curious how the selection process was conducted. Was it completely random?" The answer was: "I chose you because I am sure you will do great as a leader. Please see this as a good exercise of leadership." Answer from the student: "I see, thank you for the insight! I appreciate your words of praise, and I will do my best in leading my group as well as developing my leadership skills." The results are exactly the same when excluding this team from the analysis.

[10] There are a total of five tests, counting for 3 points each (together, 15% of the final grade). Test #1 is undertaken before any teamwork. Test #2 takes place two weeks after Test #1, i.e., immediately after the statistical teamwork; Test #3 immediately before the presentation teamwork; Test #4 two weeks after the start of the presentation teamwork; and Test #5 at the end of the course after all presentations were completed. The questions in the five individual test are related to the course material. For example, the questions of Test #1 in 2010 were: (1) What is the implication in terms of convergence of the Solow model (without technical progress)? (2) Today is the United Nations Millennium Development Goals (MDG) Summit. Cite one of the eight MDGs. (3) A country has no population growth rate, no depreciation, a propensity to save of 10%, and a capital output ratio of 2. (a) According to the Harrod-Domar model, what is the growth rate of capital stock ($\Delta k/k_{t-1}$)? (b) What if the propensity to save goes up to 12%? (c) What if the capital output ratio goes down to 1 (and propensity to save still at 10%)? (d) What is the policy recommendation for this country? (e) How, in practice, could s be increased? (f) Why might these kinds of calculations fail to capture important elements of growth?

In the control group of 55 teams, no indication is given as to what the leader should do (except uploading the final slides of the presentation online).

In another (randomly selected) set of 55 teams, the leader has the additional task of monitoring and reporting free-riding issues. The exact script is:

> Every year, we have problems with "free riders", that is to say, students who let others do the work, and get credited with the group grade. The group leader should make sure that everybody contributes a fair share to the presentation. I count on the group leader to let me know if anybody is not contributing to the presentation.

I call this intervention the "Monitoring" intervention (MON).

For the "coordinating" intervention (COORD), I use a key insight from my observation of 191 teams over three years in another experiment in Behrens and Chemin (2018): despite the apparent simplicity of the task, it is actually very difficult to gather a team of 4 students at a particular meeting place and time.[11] Free-riders usually miss the first meeting (purposefully or not), do not precisely understand the work and their tasks, feel excluded and stop working. A leader who acts as a coordinating device to set up a meeting place, make sure that all students come to the first meeting, cancels the first meeting if all students are not present, and then reschedules another meeting to include everyone, may be key to solve free-riding. The nature of the intervention in this set of teams is to share this insight with the whole team such that the leader is instructed to follow each and every step explained above to set up the first meeting correctly and improve coordination. The exact script is:

> Every year, we have problems with "free riders", that is to say, students who let others do the work, and get credited with the group grade. From my experience as a teacher over the last seven years, I have noticed that it is very difficult to gather a team of 4 students at a particular meeting place and time. Free-riders usually miss the first meeting, do not understand precisely the work and their tasks, feel excluded and stop working. I count on the group leader to act as a central authority to set up a meeting place, make sure that all students come to the first meeting, cancel the first meeting if not all students are present, and then reschedule another meeting to include all. Another important aspect is for the group leader to assign all members to precise and well-defined tasks. This may be very important to include everybody in the team, promote good relations within the group, treat everyone equally and with respect.

The penultimate sentence adds the role of assigning tasks, which may also reduce the transaction costs of working in a team. The last sentence explains that coordinating logistics well is important to promote a positive atmosphere in teams. I test for all these effects (making sure everybody comes to the meetings, assigning tasks, promoting a positive team atmosphere) in the "leadership survey" described below.

Overall, the experiment follows a 3*3 design: 3 (Control, Random leader, Promising leader) * 3 (Control, Monitoring, Coordinating). This allows me to test Propositions 1–4 of the model.

3.3. Measures of effort and leadership style

All students have to fill out a survey at the end of both the statistical and presentation teamworks (the full questionnaire is available in Appendix B). First, the survey contains 9 questions on the general appreciation of the teamwork which is not part of the pre-analysis plan. For the sake of completeness, I still present the effect of the interventions on all these outcomes.

Second, the survey contains 10 questions on effort in the teamwork. I had specified in the pre-analysis plan that these would be the measures used to determine whether the interventions affected effort. In fact, one question (specifically question #14) relates more to the presence of leadership in the team (Was there a leader in your team?). I thus analyze this variable separately in order to establish whether assigning a "Random Leader" (RL) or a "Promising Leader" (PL) actually increases the presence of leaders within the team.

Besides this question about the presence of leaders, six other questions relate to effort.[12] The fact that students not only have to answer for themselves but also for other students in the team allows me to develop peer-rated measures of effort. For each individual, I can measure whether his teammates say that he was actually contributing to the teamwork, was on time at meetings, and so on.

[11] As one students said: "we have tried to contact one of our teammates, [...], since February 23, and she has only replied to our email once, to tell us she was leaving town for the break. We asked her to join the Facebook group on which we are sharing the powerpoint document (because we were not able to find her on Facebook), and she has not done so yet. So far, she has not participated in the work at all. "

[12] (1) Who in the team contributed a fair share to teamwork? (Student 1, 2, 3, or 4, with random identification numbers given to each student in the team); (2) Who was on time to the team meetings?; (3) Who respected the deadlines set by the group?; (4) Who showed a positive attitude?; 5) Who taught others in the team?; (6) What were the contributions of each teammate as a number between 0 and 100 (equal contributions means an answer of 25, 25, 25, 25)?

Third, ten extra questions about leadership styles are included in the survey after the presentation teamwork. Two questions test the adherence to the "Monitoring" script,[13] and eight other questions test the adherence to the "Coordinating" script.[14]

All these outcomes were pre-specified in the ethics certificate.

4. Estimating equation

To identify the causal effect of the different interventions on effort, I estimate the following specification (which was pre-specified in the ethics certificate):

$$\begin{aligned}\text{Effort}_{ig} = \beta_0 &+\beta_1 \text{RL}_g + \beta_2 \text{PL}_g + \beta_3 \text{MON}_g + \beta_4 \text{RL}_g \times \text{MON}_g + \beta_5 \text{PL}_g \times \text{MON}_g \\ &+\beta_6 \text{COORD}_g + \beta_7 \text{RL}_g \times \text{COORD}_g + \beta_8 \text{PL}_g \times \text{COORD}_g + X_i\zeta + \varepsilon_i \end{aligned} \quad (1)$$

where *ig* stands for individual *i* in team (group) *g*. Effort_{ig} is a measure of effort pre-specified in the ethics certificate. RL_g is a dichotomous variable equal to 1 if the individual was in a team where a "Random Leader" had been appointed, and 0 otherwise. The coefficient β_1 tests whether assigning a random individual to become a team leader increases effort. PL_g is a dichotomous variable equal to 1 if the individual was in a team where a "Promising Leader" had been appointed, and 0 otherwise.

MON_g is a dichotomous variable equal to 1 if the individual was in a team where the script instructed the leader to monitor others and report free-riding behavior, 0 otherwise. The coefficient β_3 tests whether delivering this script increases effort. MON_g is interacted with RL_g and PL_g to measure interactions.

Similarly, $COORD_g$ is a dichotomous variable equal to 1 if the individual was in a team receiving the "coordinating" instructions, 0 otherwise. $COORD_g$ is also interacted with RL_g and PL_g.

In the pre-analysis plan, I had specified that I would focus on the followers' effort since the goal is to check whether a leader can increase others' effort. This may create an issue in the PL treatment: if the most able and hard-working student is taken out of the regressions, the rest of the team necessarily appears to be exerting less effort. To make sure this is not a concern, I show results for the whole team and for the followers, which turn out to be very similar.

In the pre-analysis plan, I had also specified a test of heterogeneous effects interacting RL_g with three variables: gender, baseline ability, and effort on statistical teamwork. The reason for interacting gender with RL_g came from Behrens and Chemin (2018). In a similar context, the authors found that the only variable correlating with effort in teams was gender: females were found to exert significantly more effort in teams. Thus, one interesting hypothesis is to test whether assigning females to leadership positions would have a greater effect on others' effort, an hypothesis which I test in this paper.[15] The reason for interacting RL_g with baseline ability and effort was to check which characteristics of leaders mattered. The reason for only interacting RL_g and not PL_g is that promising leaders are already of high baseline ability and effort by design.

X_i is a vector of control variables used in robustness checks. Following the pre-analysis plan, I use: gender, baseline ability, and effort on statistical teamwork. For the sake of completeness, I also include other variables that were not in the pre-analysis plan such as: baseline team ability (to test for peer effects), two semester dummies (to control for cohort effects), field of study, year of study, and team size.

In all regressions, I report robust standard errors clustered at the team level. I also use the exact Fisher test (Young, 2018) which is an exact test regardless of sample size or distribution of error term, as opposed to conventional t-tests which depend on the assumption of large samples.

I address the concern of multiple hypothesis testing in two ways. First, I summarize all effort-related variables into one single index following Anderson (2008). At the most basic level, a summary index is a weighted mean of several standardized outcomes and weights are calculated to maximize the amount of information captured. The optimal weights are the inverse of the correlation matrix. This procedure increases efficiency by ensuring that outcomes which are highly correlated with one another receive less weight, while outcomes that are uncorrelated - and thus represent new information - receive more weight (Anderson, 2008). I also experiment in the appendix with other ways of building summary indices (such as simple averages of measures of effort or dichotomous measures), with similar results. This procedure controls for the increased probability of finding a false positive by reducing the number of tests being conducted.

Second, I use Sharpened False Discovery Rate (FDR) adjusted *q*-values (Anderson, 2008). Intuitively, this method adjusts the *p*-values by dividing the significance level by the number of hypotheses tested in a family of outcomes, taking into account the rank of the variable according to its p-value within the family. For example, with 6 measures of effort (the

[13] (1) Did the group leader make sure that everybody contributes a fair share to the presentation?; (2) Did the group leader inform the professor if anybody was not contributing to the presentation?

[14] (1) Did the group leader set up the first meeting place and time?; (2) Did the group leader make sure that all students were coming to that first meeting?: (3) Did everybody come to that first meeting?; (4) If not, was an alternative meeting set to include everybody?; (5) Did the group leader assign all group members to precise and well-defined tasks?; (6) Did the group leader include everybody in the team?; (7) Did the group leader promote good relations within the group?; (8) Did the group leader treat everyone equally and with respect?

[15] Although not specified in the pre-analysis plan, I also investigate gender effects in greater details in Appendix K by showing the results separately for male or female respondents, as well as for male or female leaders.

Table 2
Balance test.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Female	Baseline ability	Effort on Previous Teamwork						Not in PAP
			Contribute fair share	On time at meetings	Respect deadlines	Positive attitude	Taught others	Contrib.	Score Effort
Random Leader (RL)	0.05	−0.12	−0.04	0.02	−0.02	0.01	−0.07	−1.82*	−0.19
	(0.09)	(0.15)	(0.04)	(0.05)	(0.04)	(0.04)	(0.06)	(0.97)	(0.17)
[Fisher p-val]	[0.6]	[0.44]	[0.38]	[0.66]	[0.55]	[0.81]	[0.3]	[0.07]	[0.27]
[FDR q-val]			[1]	[1]	[1]	[1]	[1]	[0.6]	
Promising Leader (PL)	−0.04	−0.00	0.04	0.03	0.03	0.08**	−0.00	−0.45	0.06
	(0.07)	(0.17)	(0.03)	(0.04)	(0.03)	(0.04)	(0.06)	(0.82)	(0.14)
[Fisher p-val]	[0.57]	[0.99]	[0.17]	[0.53]	[0.4]	[0.04]	[0.95]	[0.56]	[0.69]
[FDR q-val]			[1]	[1]	[1]	[0.32]	[1]	[1]	
Monitoring (MON)	0.01	−0.09	−0.01	0.04	−0.03	−0.00	−0.04	−0.57	−0.06
	(0.07)	(0.18)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.86)	(0.19)
MON * RL	−0.01	0.21	0.09	0.01	0.07	0.06	0.10	1.75	0.31
	(0.11)	(0.25)	(0.07)	(0.07)	(0.06)	(0.06)	(0.08)	(1.32)	(0.25)
MON * PL	−0.02	0.17	−0.03	−0.01	0.02	−0.03	0.03	0.47	0.03
	(0.11)	(0.25)	(0.06)	(0.07)	(0.06)	(0.07)	(0.08)	(1.15)	(0.23)
Coordinating (COORD)	0.09	0.10	0.06**	0.09**	0.06*	0.05	−0.00	0.30	0.17
	(0.07)	(0.14)	(0.03)	(0.04)	(0.03)	(0.04)	(0.05)	(0.82)	(0.14)
[Fisher p-val]	[0.18]	[0.48]	[0.026]	[0.017]	[0.081]	[0.26]	[0.96]	[0.71]	[0.24]
[FDR q-val]			[0.11]	[0.11]	[0.12]	[0.24]	[0.63]	[0.63]	
COORD * RL	−0.10	−0.07	−0.01	−0.06	0.02	−0.01	0.04	1.02	0.05
	(0.11)	(0.22)	(0.06)	(0.06)	(0.05)	(0.06)	(0.08)	(1.25)	(0.21)
COORD * PL	−0.06	−0.14	−0.05	−0.05	−0.03	−0.03	0.02	−0.37	−0.09
	(0.11)	(0.22)	(0.04)	(0.05)	(0.05)	(0.05)	(0.08)	(1.19)	(0.19)
Observations	658	647	656	656	656	656	656	627	658
R-squared	0.008	0.006	0.018	0.011	0.014	0.018	0.005	0.012	0.011
Mean control baseline	0.61	1.72	0.85	0.80	0.84	0.81	0.54	25.07	−0.01
(SD)	0.49	0.84	0.23	0.27	0.22	0.24	0.34	5.07	0.94

OLS regressions. Robust standard errors in parentheses, clustered at the team level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. The dependent variable in Column (1) is a dichotomous variable equal to 1 if the individual is a female, 0 otherwise. In Column (2), the outcome is baseline ability, i.e., the grade on Test #3 which is the test immediately before the teamwork. The dependent variable in Column (3) is the peer-rated assessment of fair contribution to the teamwork, i.e., the average number of times teammates say that individual *i* is contributing a fair share. Similarly in the next columns, the dependent variable is the peer-rated assessment of being on time at the meetings in Column (4), respecting the deadlines in Column (5), showing a positive attitude in Column (6), and teaching others in Column (7). Column (8) shows the average quantitative contribution to the teamwork as estimated by peers. In Column (9), the dependent variable is a score of effort which standardizes the six measures of effort in Columns (3) to (8), weighting them by the inverse of the correlation matrix.

family), a 10% significance level becomes a 10%/6 outcomes*1 (first-rank)=1.6% significance threshold for the first-ranked (most significant) variable.

To be fully transparent, I had not pre-specified these two methodologies to address multiple hypothesis testing in 2014. However, they are standard ways of addressing this concern. Throughout this paper, I indicate what was pre-specified or not in every regression.

5. Randomization and balance of observable characteristics

Table 2 presents the balance test. I use specification Eq (1) as described above with outcomes determined before the interventions (gender, baseline ability, and effort on statistical teamwork; as well as baseline team ability, grade on statistical teamwork, field of study, year of study, and team size in the appendix). If well balanced, all the coefficients should be close to zero.

Column (1) shows that the proportion of females is 61% in the control group, with no statistically significant difference in any of the treatment groups. This indicates that there are no more females in the treatment groups versus the control group.

To calculate baseline ability in Column (2), I use the grades on Test #3 since it was done immediately before the presentation teamwork (I show other measures such as the grade on the statistical teamwork in the appendix). Once again, there are no significant differences across all treatment groups.

Columns (3) to (7) look at the 6 questions relating to effort on the statistical teamwork. In Column (3), the dependent variable shows whether an individual contributed a fair share to the teamwork, as rated by her peers.[16] The mean in the

[16] In practice, the question is: "Who contributed a fair share to the teamwork?". The outcome is thus the average number of times that teammates say individual *i* is contributing a fair share. For example, if all 3 teammates say individual *i* is contributing a fair share, then this number is 1. If only 2 teammates out of 3 say individual *i* is contributing a fair share, then this number is 2/3.

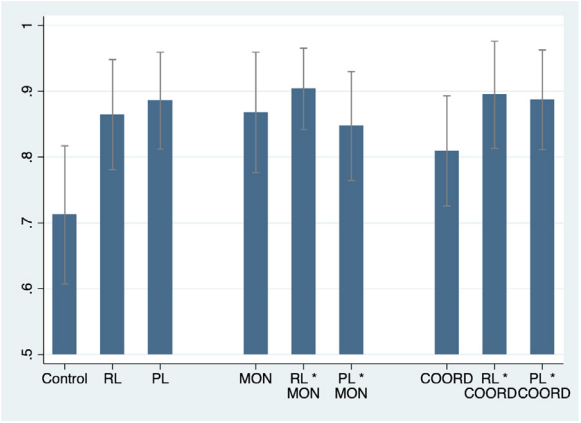

Fig. 1. Effect of Interventions on Presence of a Leader. *Notes:* The variable graphed is a dichotomous variable equal to 1 if the individual answered that any of the individuals in the team was a leader, 0 otherwise.

control group is 0.85, indicating that for the average student, 85% of his teammates say he is contributing a fair share. There are no significant differences across treatment groups, except for the Coordinating intervention, where students are 6 percentage points more likely to be assessed as having contributed a fair share to the statistical teamwork.

Column (4) focuses on the peer-rated dimension of being on time at team meetings. Once again there are no differences at baseline, except for the Coordinating intervention. The result is similar in Column (5) that looks at whether the student respected deadlines set by the team.

In column (6), students in the "Promising Leader" intervention are found to have a significantly greater positive attitude on the statistical teamwork (Column 6). There are no significant differences concerning the likelihood to teach others (Column 7). Finally, students in the "Random Leader" intervention had a significantly lower contribution to the statistical teamwork (Column 8).

None of the few statistically significant differences survive the adjustment of *p*-values for multiple hypothesis testing associated with this family of 6 measures of effort (which are highly correlated with each other as indicated in Table C1). The Sharpened FDR q-values are all above the 10% significance level. Intuitively, for the intervention COORD for example, the best ranked p-value is 0.02, above 10%/6 outcomes*1 (first-rank)=0.016, hence not significant at 10%.

To address multiple hypothesis testing in another way, I also build a summary index of effort based on the 6 measures following Anderson (2008). I compute the weighted average of the six standardized measures of effort. The weights are the inverse of the correlation matrix in Table C1. Column (9) shows no significant different across treatment groups. I experiment in Appendix D with other ways to build summary indices (such as simple averages of measures of effort or dichotomous measures), with similar results, i.e., balance between the control and treatment groups.

In Appendix D, I present other balance tests on variables not specified in the pre-analysis plan. These include baseline team ability, i.e., the average grade of the team excluding individual *i* on Test #3 (the test immediately before the presentation teamwork), the grade on the statistical Teamwork, field of study, year of study, team size, and day of the presentation with few differences across treatment groups.

Overall, this section shows that the control and treatment groups are well balanced. Before moving on to describe the effects on team effort, the next section investigates whether the assigned leader became the actual leader of the team.

6. Manipulation check

6.1. Is there a leader in the treatment group?

To verify this, I use the following question: "Was there a leader in your team? (if yes, indicate who : 1 2 3 4)", I define a dichotomous variable equal to 1 if the individual answered that any member of the team was a leader, 0 otherwise. A first important finding is that not all teams have a leader in the control group: Fig. 1 shows that 70% of students say they have a leader in their team (as visible in the first bar which shows the results for the control group). These results are in line with Güth et al. (2007): in their experiment, only 40% of teams succeed in independently appointing a leader. The prevalence of leaders is higher in my paper, possibly because I use a less conservative measure of leadership. In Güth et al. (2007), all players needed to vote privately and unanimously in favor of one individual for this individual to be selected as a leader. In my paper, I simply use individual answers to the question "Was there a leader in your team?".

The second important finding is that both the RL and PL groups have more leaders as visible in the second and third bars of Fig. 1: in both groups, the proportion of students declaring they have a leader in their team is almost 90%. This large increase shows that merely appoint a leader has the consequence that more students declare they have a leader in their team.

Table 3
Is there a leader in the team?.

	(1) Leader?
Random Leader (RL)	0.15**
	(0.07)
[Fisher p-val]	[0.046]
Promising Leader (PL)	0.17**
	(0.07)
[Fisher p-val]	[0.023]
Monitoring (MON)	0.16**
	(0.08)
[Fisher p-val]	[0.041]
MON * RL	−0.12
	(0.09)
MON * PL	−0.19**
	(0.09)
Coordinating (COORD)	0.10
	(0.07)
[Fisher p-val]	[0.2]
COORD * RL	−0.07
	(0.09)
COORD * PL	−0.10
	(0.09)
Observations	606
R-squared	0.025
Mean control	0.71
(SD)	0.46

OLS regressions. Robust standard errors in parentheses, clustered at the team level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. The dependent variable is a dichotomous variable equal to 1 if the individual answered that any of the individuals in the team was a leader, 0 otherwise.

The MON and COORD groups also have more leaders than the control group, as visible in the fourth bar called MON and the seventh bar called COORD of Fig. 1. But the effect is smaller than RL and PL. This is logical since MON and COORD are not appointing a leader but simply telling the leader what to do. Thus, in the groups MON and COORD not interacted with RL and PL, the leader is the endogenous leader. In contrast, the interactions of RL with MON and COORD (in bars 5 and 6), as well as the interactions of PL with MON and COORD (in bars 8 and 9), have very similar effects on the presence of a leader as RL and PL alone, indicating that telling the leader what to do is not adding on any effect relative to assigning a leader.

Table 3 investigates the statistical significance of these results by running an OLS regression where the dependent variable is the "presence of leader" indicator derived from question #14. The table shows a significant increase in the likelihood of having a leader in the RL and PL and MON groups, but no further increases in the MON*RL and MON*PL groups. As for the COORD intervention, there is no significant effect. These results provide confirmation that assigning a leader (either RL or PL) actually sets one up, but it remains to be seen whether the assigned leader is the actual leader, which is the topic of the next section.

6.2. Who is the leader?

To verify whether the assigned leader becomes the actual leader, I first need to determine who the actual leader is. I use two methods to accomplish this, the first of which is objective and simple: I collect data on the identity of the individual who uploaded the presentation slides. In all teams, the leader was instructed to upload the slides online.

In 64% of the teams in the "Random Leader" group, the person uploading the slides is the "Random Leader".[17] In 67% of the teams in the "Promising Leader" group, the person uploading the slides is the "Promising Leader". This indicates that even though some teams did not follow the experimental design, most teams did.

Would the "Promising Leader" have ended up being the leader anyway in the control group. because he/she is "promising"? I calculate the Leadership Score in the control group and identify who would have been the "Promising Leader" in the control group had the PL intervention been implemented. In only 39% of the teams, the person uploading the slides is the person who would have been designated PL. Thus, the intervention PL clearly affected the identity of the leader.

Of course, one cannot discard the possibility that RL and PL merely upload the slides for the sake of appearances, but are not real leaders. To address this issue, I use the more subjective question "Was there a leader in your team? (if yes, indicate

[17] This is more than 25%, the rate at which one would expect a randomly selected individual out of a team of 4 students to end up being the leader of a team (in fact, the true figure should be less than 25% considering only 70% of the teams have a leader as evidenced above).

Table 4
Effects of interventions on effort.

	(1) Contribute fair share	(2) On time at meetings	(3) Respect deadlines	(4) Positive attitude	(5) Teach others	(6) Contrib.	(7) Not in PAP Score Effort
Random Leader (RL)	−0.02	−0.06	−0.05	−0.07**	0.10	−0.12	0.15
	(0.05)	(0.04)	(0.05)	(0.03)	(0.06)	(0.43)	(0.18)
Promising Leader (PL)	0.01	−0.02	−0.02	−0.04	0.14**	0.08	0.33**
	(0.04)	(0.04)	(0.04)	(0.03)	(0.06)	(0.32)	(0.16)
[Fisher p-val]	[0.77]	[0.61]	[0.61]	[0.13]	[0.02]	[0.84]	[0.045]
[FDR q-val]	[1]	[1]	[1]	[0.51]	[0.12]	[1]	
Monitoring (MON)	0.00	−0.03	−0.04	−0.06	0.15**	0.41	0.35**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)	(0.27)	(0.15)
[Fisher p-val]	[0.99]	[0.42]	[0.41]	[0.17]	[0.014]	[0.12]	[0.015]
[FDR q-val]	[0.99]	[0.57]	[0.57]	[0.44]	[0.11]	[0.44]	
MON * RL	0.04	0.08	0.08	0.10*	−0.08	−0.07	−0.10
	(0.06)	(0.05)	(0.06)	(0.06)	(0.08)	(0.53)	(0.23)
MON * PL	0.03	0.01	0.06	0.06	−0.15*	−0.20	−0.23
	(0.05)	(0.06)	(0.05)	(0.05)	(0.09)	(0.36)	(0.22)
Coordinating (COORD)	0.05	0.01	0.01	−0.03	0.15**	0.41	0.51***
	(0.04)	(0.03)	(0.03)	(0.03)	(0.06)	(0.26)	(0.15)
[Fisher p-val]	[0.13]	[0.74]	[0.75]	[0.26]	[0.017]	[0.12]	[0.001]
[FDR q-val]	[0.26]	[0.58]	[0.58]	[0.33]	[0.079]	[0.26]	
COORD * RL	−0.04	0.00	0.04	0.03	−0.11	−0.02	−0.28
	(0.06)	(0.05)	(0.06)	(0.05)	(0.09)	(0.48)	(0.25)
COORD * PL	−0.10*	−0.02	−0.02	−0.02	−0.12	−0.26	−0.45**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.08)	(0.38)	(0.22)
Observations	658	658	658	658	658	643	658
R-squared	0.022	0.010	0.012	0.019	0.030	0.002	0.023
Mean control	0.91	0.93	0.93	0.97	0.50	24.49	0.00
(SD)	0.20	0.16	0.18	0.12	0.32	4.14	1.01

OLS regressions. Robust standard errors in parentheses, clustered at the team level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. The dependent variable in Column (1) is the peer-rated assessment of fair contribution to the teamwork, i.e., the average number of times teammates say that individual i is contributing a fair share. Similarly in the next columns, the dependent variable is the peer-rated assessment of being on time at the meetings in Column (2), respecting the deadlines in Column (3), showing a positive attitude in Column (4), and teaching others in Column (5). Column (6) shows the average quantitative contribution to the teamwork as estimated by peers. In Column (7), the dependent variable is a score of effort equal to the standardized six measures of effort in Columns (1) to (6) weighted by the inverse of the correlation matrix.

who : 1 2 3 4)". I had not pre-specified exactly how to determine the leader from this question, but the methodology is as follows. First, I calculate the average answer for each individual (e.g., if 2 out of 3 teammates say an individual was a leader, the score for that individual is 2/3). Second, I identify as leader the individual with the highest such average in the team.

There are two issues with this methodology. First, it is unclear who the leader is if the maximum score is shared by multiple individuals. This is a distinct possibility in practice since the average can only take four values in teams of 4 (1, 2/3, 1/3 or 0),[18] and it actually happens in 33% of the teams. Second, does the team have an actual leader if the highest score in the team is 1/3, i.e., only 1 teammate out of 3 thinks the individual is a leader? Observing a maximum average of 0 or 1/3[19] actually happens in 27% of the teams.

Despite the limitations of this methodology, the results are similar to when I use the presentation slides: 51% of individuals designated as Random Leaders end up being the actual leader according to their peers (59% in the Promising Leader group). In the control group, only 43% of individuals who would have been designated as Promising Leaders end up being the actual leader according to their peers.

However, considering these limitations, I use the first methodology (the person uploading the slides) to identify the actual leader in the rest of the paper. This method has the merit of simplicity and objectivity.

Overall, the results in this section showed that the interventions RL and PL increased the likelihood of having a leader; and that in most of the cases, the assigned leader ended up being the actual leader. The next section looks at the effect of the interventions on effort.

7. Results

7.1. Effects on effort

Table 4 shows the effect of the interventions on the 6 dimensions of effort using the regression specified in the pre-analysis plan. Column (1) shows peer-rated assessments of whether the individual contributed a fair share to the teamwork.

[18] and three values in teams of 3 (1, 1/2 or 0).
[19] and 1/2 in teams of 3.

The mean in the control group is 0.91, indicating that students are seen as contributing a fair share to the teamwork by 91% of their peers. This propensity to contribute a fair share is not significantly affected by the interventions. In fact, Columns (2), (3) and (4) do not show much effect of the interventions on the likelihood of being on time at meetings, respecting deadlines, and showing a positive attitude.

Column (5) shows that the strongest effect of the interventions is on teaching others in the team, i.e., knowledge-sharing. The dependent variable is an average "score" similar to the one in Section 6.2: for example, if all 3 teammates say individual *i* is teaching others, then this number is 1. If only 2 teammates out of 3 say individual *i* is teaching others, then this number is 2/3. The mean control is 0.5, indicating that students are seen as teaching others by 50% of their peers. This variable is not significantly affected by the "Random Leader" intervention, but increases significantly with the other interventions: the propensity to teach others increases by 14 percentage points in a team with a "Promising Leader", 15 percentage points in a team where the leader was instructed to monitor others, and 15 percentage points in a team where the leader was instructed to coordinate the team compared to the control group where no intervention is taking place. These results are robust to using Fisher *p*-values instead of t-tests.

This is an important and new finding of this paper. Because some students have trouble understanding their paper,[20] other students have to teach them otherwise the quality of the presentation and the team's grade are negatively affected.[21] Yet, teaching others is costly[22] and subject to the classic public goods dilemma: if two students both understand the paper and must decide whether or not to teach others in the team or not, they will both prefer to let the other teach, benefit from a higher grade without bearing the costs of effort. If everybody thinks this way, no teaching occurs and the team grade is low. Thus, teaching others may be under-provided in teams. If a student is a "Promising Leader" who moves first and teaches others in the team, this may entice other students to teach as well by a conformity effect. A monitoring leader who reports others who do not attempt to learn or to teach others may also solve the situation. A coordinating leader could also encourage knowledge spillovers by making sure everybody participates in team meetings. I find support for these three hypotheses in Column (5).

Interaction terms between MON and RL or PL, as well as interaction terms between COORD and RL or PL are not significant if anything they are negative and of the same magnitude as the level terms (around 15 percentage points), in line with the theoretical prediction about the interaction between treatments. This is intuitive when considering the leading-by-example theory: if students already teach others because of MON or COORD, there is less inequality in effort levels, thus the positive effects of leading-by-example is crowded out.

Column (6) shows the quantitative contribution to the teamwork, as rated by peers. The mean control is 24.49, showing that students contribute on average 24.49% of the work according to their peers. This figure is increased in PL, MON, and COORD teams, but not significantly.

To address multiple hypothesis testing, I once again build a summary index of effort, following Anderson (2008). Table C2 shows very high correlations among all outcomes, but less so for "teaching others". This is understandable since contributing a fair share, being on times at meetings, respecting the deadlines and showing a positive attitude can be seen as the minimum requirements of teamwork. Teaching others however, goes above and beyond such requirements and requires extra effort. This variable therefore represents new information and since weights are attributed using the inverse of the correlation matrix, it will have a greater weight than the other variables which are more correlated with each others. Column (7) shows that this score of effort is significantly increased by the PL, MON, and COORD interventions while the interaction terms remain negative. I also experiment in Appendix G with other ways to build summary indices (such as simple averages of measures of effort or dichotomous measures) with similar results.

When using sharpened FDR *q*-values, the only intervention to survive the test is COORD and only the outcome of teaching others remains significant. This provides a first indication that the COORD intervention has the strongest effects overall.

Graphically, these results can be seen in Fig. 2. Effort is increased by 0.33 standard deviations in the PL group (the standard deviation is 1 by construction in the control group), 0.35 in the MON group, and 0.51 in the COORD group, which provides a second indication that the COORD intervention was the most impactful overall.

Table 5 presents robustness checks with additional controls where the dependent variable is the summary index score of effort. Column (1) adds gender of the student which is found not to be significantly associated with effort. Most importantly, the main coefficients of interest, i.e., PL, MON and COORD, remain of similar value and significant.

In fact, the coefficients of PL, MON and COORD do not change much if one controls for baseline individual ability (Column (2)) and the score of effort on the previous assignment (Column (3)). These control variables were specified in the pre-analysis plan but out of completeness, I also show robustness checks for other variables which were not. Once again, the coefficients of PL, MON and COORD are not affected when controlling for baseline team ability, i.e., the average grade of the team excluding individual *i* on the baseline test (Column (4)). In fact, baseline team ability is not significantly related with effort at all. This shows that peer effects are weaker than leader effects, in line with Lazear et al. (2015).

[20] From one email to the professor: "i'm (sic) really not good in math and i (sic) have difficulties to understand the tables in general".

[21] "One of them told us she couldn't do her part just two hours before the presentation, so I had to take over. The other one just forgot what to say in the middle of the presentation so I took over. I made the group as cohesive as possible, explained all their parts to them, but it was not my job to police them and make them work hard. I feel so helpless because even the TA said it was very clear that I knew my material but they did not. Every person who I spoke to in the room was horrified that the three girls in my group were so unprepared."

[22] "I have been too busy explaining things"

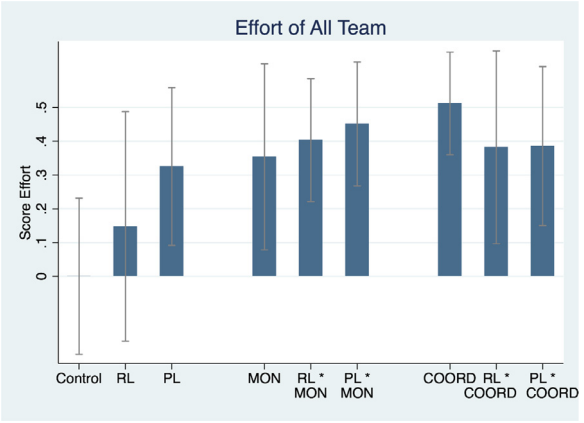

Fig. 2. Effect of Interventions on Effort of Entire Team. *Notes:* The variable graphed is the score of effort equal to the standardized six measures of effort in Columns (1) to (6) weighted by the inverse of the correlation matrix.

Similarly, the coefficients of PL, MON and COORD do not change much after including the 2 semester dummies (for Fall 2014 and Winter 2015, Winter 2014 being the omitted category) to control for cohort effects (Column (5)), the field of study (Column (6)), the year of study (Column (7)), team size (Column (8)) or all control variables at once (Column (9)).

Results are similar when using the question about teaching others as the main outcome rather than the summary score of effort: the coefficients of PL, MON and COORD remain significant when including the control variables.

To summarize, the interventions PL, MON, and COORD increase effort. In particular, the effect of COORD remain significant even after correcting for multiple hypothesis testing. Furthermore, the effect is being driven by knowledge sharing, i.e., students teaching each other. One may wonder however, whether this effect comes from leaders teaching others or rather from the followers. The next section explores this by splitting the sample into followers and leaders.

7.2. Do Followers and leaders experience different effects?

The summary finding is that results are very similar when restricting the sample to either followers or leaders.

I first restrict the sample to followers. To do so, I focus on students *not* uploading the slides online. This methodology allows me to determine who the leader is even in the control group, as well as in the MON and COORD treatment groups where no leaders are designated.

Table 6 shows the effect for followers. The results are very similar to those in the pooled sample: followers in the PL, MON and COORD are found to teach others more and the results remain significant for the PL and COORD interventions when using the Sharpened FDR q-values adjusting for multiple hypothesis testing. The score of effort is increasing in all interventions. Graphically, these results can be seen in Fig. E1 in the Appendix.

In Table E1 in Appendix E, I use another way to define followers. In the RL and PL treatment groups, I restrict the sample to the students not designated RL or PL. The problem with this approach is that there are no designated leaders in either the control group or the MON and COORD treatment groups. In these groups, I thus keep all observations in the sample and results turn out to be very similar with this methodology.

In the Appendix F, I also restrict the sample to leaders by focusing on students who are uploading the slides online. Table F1 and Fig. F1 show the effects of the intervention on the effort of leaders themselves. Leaders teach more others and in general exert more effort in PL and MON and COORD. The coefficients are similar in magnitude to the results on followers, but not significant, possibly due to the reduced sample size for leaders.

Overall, the positive effect was experienced in the same way by all members of the team regardless of their role. The next section investigates the mechanisms for this positive effect, in other words, what leaders do exactly.

7.3. What do leaders do?

In the pre-analysis plan, I specified 10 questions that exactly followed the script which was sent to the students to verify whether leaders adhered to the instructions being given. Columns (1) and (2) test adherence to the Monitoring script, while Columns (3) to (10) test adherence to the Coordinating script.

Results are shown in Table 7. Column (1) shows the answers to the question “Did the group leader make sure that everybody contributes a fair share to the presentation?”, referring to the first sentence of the Monitoring script. The outcome is a dichotomous variable equal to 1 if the student answered yes, 0 otherwise. The sample includes the whole team. The average in the control group is 0.94, indicating that 94% of students say that their leader makes sure everybody contributes a fair share to the presentation. Considering the very high average answer to that question, there may be a “ceiling effect”,

Table 5
Robustness checks.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
						Not specified in PAP				
RL	0.15	0.24	0.19	0.15	0.14	0.12	0.13	0.17	0.15	0.24
	(0.18)	(0.19)	(0.18)	(0.18)	(0.18)	(0.19)	(0.18)	(0.18)	(0.19)	(0.19)
PL	0.33**	0.30*	0.31*	0.33*	0.31*	0.32**	0.33**	0.33**	0.33**	0.29*
	(0.16)	(0.16)	(0.17)	(0.17)	(0.16)	(0.16)	(0.17)	(0.16)	(0.16)	(0.17)
MON	0.35**	0.35**	0.37**	0.35**	0.35**	0.34**	0.36**	0.37**	0.35**	0.35**
	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.16)
MON * RL	−0.10	−0.21	−0.16	−0.10	−0.09	−0.06	−0.09	−0.13	−0.10	−0.21
	(0.23)	(0.24)	(0.22)	(0.23)	(0.24)	(0.23)	(0.23)	(0.23)	(0.24)	(0.23)
MON * PL	−0.23	−0.23	−0.23	−0.22	−0.21	−0.23	−0.24	−0.23	−0.23	−0.22
	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)
COORD	0.51***	0.46***	0.48***	0.52***	0.50***	0.50***	0.51***	0.53***	0.51***	0.44***
	(0.15)	(0.14)	(0.16)	(0.15)	(0.15)	(0.16)	(0.15)	(0.15)	(0.15)	(0.16)
COORD * RL	−0.27	−0.33	−0.29	−0.28	−0.28	−0.25	−0.26	−0.35	−0.28	−0.35
	(0.25)	(0.25)	(0.25)	(0.25)	(0.25)	(0.26)	(0.25)	(0.25)	(0.25)	(0.27)
COORD * PL	−0.45**	−0.39*	−0.43*	−0.46**	−0.43*	−0.45**	−0.46**	−0.47**	−0.45**	−0.39*
	(0.22)	(0.22)	(0.23)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)
Female	0.05									0.01
	(0.09)									(0.09)
Baseline		0.15***								0.12**
Ability		(0.05)								(0.06)
Score Effort			0.28***							0.20**
Past			(0.09)							(0.08)
Baseline Team				−0.03						−0.04
Ability				(0.07)						(0.08)
Fall2014					0.08					0.08
					(0.08)					(0.11)
Winter2015					0.03					0.02
					(0.10)					(0.14)
Business						−0.12				−0.05
						(0.20)				(0.22)
Political						−0.17				−0.09
Science						(0.12)				(0.12)
Economics						−0.11				−0.05
						(0.11)				(0.11)
Science						0.12				0.13
						(0.11)				(0.11)
Others						−0.17				−0.20
						(0.21)				(0.24)
First year							0.05			0.07
							(0.10)			(0.12)
Second year							−0.05			−0.06
							(0.10)			(0.12)
Team size								−0.49**		−0.41**
								(0.20)		(0.19)
Day of									0.01	0.01
Presentation									(0.04)	(0.04)
Observations	658	647	658	658	658	658	658	658	658	647
R-squared	0.024	0.038	0.060	0.024	0.024	0.032	0.025	0.030	0.023	0.075

OLS regressions. Robust standard errors in parentheses, clustered at the team level. *** $p \int < 0.01$; ** $p \int < 0.05$; * $\int p < 0.1$. The dependent variable in all columns is the score of effort equal to the standardized six measures of effort weighted by the inverse of the correlation matrix.

whereby it would be hard for any intervention to further raise this figure. In fact, Column (1) indicates that no intervention does so.

The RL intervention is negatively related with this outcome, indicating that "Random Leaders" may not be the best placed to address free-riding issues. More generally, I find in this paper that RL is not positively related with any outcome. This is in line with anecdotal evidence from students who stated that the RL was not the best leader.[23]

Column (2) tests adherence to the second statement in the Monitoring script with the question: "Did the group leader inform the professor if anybody was not contributing to the presentation?". For this variable, the average in the control group is 60%. This figure increases in the PL, MON and COORD treatment group by 8, 5 and 11 percentage points, but only

[23] In one team, the RL "has showed little skills of a group leader and does not show motivation for this assignment". One message from this RL shared by the previous student: "Hey guys I won't be able to make it because my flight is tomorrow morning to Vancouver. :(Can someone brief me on the meeting and what tasks I am assigned?". This was predictable: leadership score of this RL is only 0.3 on the statistical teamwork, with a score of 0 on "contributed a fair share", 0 on time at meetings, 0.33 on respected deadlines, and 0 on positive attitude.

Table 6
Effects of interventions on effort of followers.

	(1) Contribute fair share	(2) On time at meetings	(3) Respect deadlines	(4) Positive attitude	(5) Teach others	(6) Contrib.	(7) Not in PAP Score Effort
Random Leader (RL)	−0.02	−0.06	−0.04	−0.08*	0.11*	0.24	0.24
	(0.05)	(0.04)	(0.06)	(0.04)	(0.06)	(0.47)	(0.19)
Promising Leader (PL)	0.02	−0.02	0.00	−0.04	0.15**	0.32	0.42**
	(0.04)	(0.04)	(0.05)	(0.03)	(0.06)	(0.49)	(0.19)
[Fisher p-val]	[0.66]	[0.56]	[0.93]	[0.16]	[0.016]	[0.55]	[0.024]
[FDR q-val]	[1]	[1]	[1]	[0.65]	[0.098]	[1]	
Monitoring (MON)	−0.00	−0.03	−0.03	−0.09	0.14**	0.80*	0.42**
	(0.05)	(0.04)	(0.06)	(0.06)	(0.07)	(0.47)	(0.20)
[Fisher p-val]	[0.94]	[0.39]	[0.62]	[0.11]	[0.049]	[0.097]	[0.035]
[FDR q-val]	[0.88]	[0.43]	[0.56]	[0.28]	[0.28]	[0.28]	
MON * RL	0.03	0.08	0.07	0.13*	−0.07	−0.76	−0.24
	(0.07)	(0.06)	(0.08)	(0.07)	(0.09)	(0.65)	(0.27)
MON * PL	0.03	0.01	0.06	0.08	−0.15	−0.83	−0.34
	(0.06)	(0.06)	(0.07)	(0.07)	(0.10)	(0.58)	(0.27)
Coordinating (COORD)	0.06	−0.00	0.03	−0.03	0.16**	0.53	0.59***
	(0.04)	(0.03)	(0.05)	(0.03)	(0.06)	(0.45)	(0.17)
[Fisher p-val]	[0.11]	[0.98]	[0.51]	[0.26]	[0.014]	[0.25]	[0.001]
[FDR q-val]	[0.35]	[0.56]	[0.56]	[0.43]	[0.066]	[0.43]	
COORD * RL	−0.06	0.05	0.04	0.05	−0.14	−0.65	−0.46
	(0.07)	(0.05)	(0.07)	(0.06)	(0.09)	(0.69)	(0.28)
COORD * PL	−0.13**	−0.02	−0.01	−0.02	−0.14	−0.21	−0.54**
	(0.06)	(0.05)	(0.06)	(0.05)	(0.09)	(0.61)	(0.25)
Observations	503	503	503	503	503	492	503
R-squared	0.027	0.009	0.016	0.019	0.029	0.003	0.025
Mean control	0.91	0.93	0.91	0.97	0.49	24.30	−0.09
(SD)	0.20	0.16	0.20	0.11	0.30	4.47	1.04

OLS regressions. Robust standard errors in parentheses, clustered at the team level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. The sample only includes the followers, identified as the students not uploading the presentation slides. The dependent variable in Column (1) is the peer-rated assessment of fair contribution to the teamwork, i.e., the average number of times teammates say that individual i is contributing a fair share. Similarly in the next columns, the dependent variable is the peer-rated assessment of being on time at the meetings in Column (2), respecting the deadlines in Column (3), showing a positive attitude in Column (4), and teaching others in Column (5). Column (6) shows the average quantitative contribution to the teamwork as estimated by peers. In Column (7), the dependent variable is a score of effort equal to the six standardized measures of effort in Columns (1) to (6) weighted by the inverse of the correlation matrix.

significantly so for the COORD intervention. This may appear surprising at first since the instruction was in the MON, not COORD, script. Yet, the logical next step for a COORD leader who tries to coordinate but faces resistance from teammates - despite the COORD script sent to the whole team - is to inform the professor, an action available to all students at all times in this course.

Columns (3) to (10) test adherence to the Coordinating script. Once again, averages of these outcomes are very high in the control group: 84% of students say that their leader sets up the first meeting (Column (3)), and makes sure that all students come to that first meeting (Column (4)). There is not much room for improvement and none of the interventions significantly increases these outcomes - although COORD show positive coefficients as opposed to RL, PL, and MON where students did not receive the COORD script.

The difference comes from Column (5) with the question: “Did everybody come to that first meeting?”. This is the key insight of the COORD intervention: free-riders usually miss the first meeting (voluntarily or not), feel excluded or argue they did not know what to do, and stop working altogether. An intervention making sure that all students participate in the first meeting would address this issue and as Column (5) shows, only 65% of students in the control group say that everybody came to the first meeting. This is a decidedly low number and the COORD intervention significantly raises the figure by 16 percentage points. The other interventions RL, PL and MON are not associated with a significant increase because none of these groups received the Coordinating script. Thus, by placing an emphasis on the organization and attendance of that first meeting, the COORD script increases the likelihood that everybody indeed participates.

There is no effect of the COORD intervention on setting up an alternative meeting if not everybody came to the first one (Column (6)). This however is understandable since there is no need to schedule an alternative meeting if everybody was there the first time.

While the first part of the Coordinating script strictly relates to organizational aspects of teamwork, the second part relates to assigning tasks and building a positive and cooperative atmosphere in teams. However, I find no effect of the COORD intervention on: assigning all group members to precise and well-defined tasks (Column (7)), including everybody in the team (Column (8)), a small and insignificant effect on promoting good relations within the group (Column (9)), and no effect on treating everyone equally and with respect (Column (10)). Thus, the COORD intervention seems to be improving organizational aspects of teams, not their positive atmosphere, although one cannot exclude that no effect is found since

Table 7
Effects on leadership style.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Monitoring Script		Coordinating Script								Not in PAP		
	Fair Contrib.?	Inform Prof.?	Set up First Meet?	Make Sure First Meet?	Everybody First Meet?	Alternative Meet?	Assign Tasks?	Include?	Promote Good Rel.?	Respect?	Score Leader	Score Leader MON	Score Leader COORD
												(1)-(2)	(3)-(10)
RL	−0.05*	0.06	−0.04	−0.07	0.06	0.04	−0.08**	−0.05**	−0.06*	−0.02	0.16	0.25	0.08
	(0.03)	(0.06)	(0.08)	(0.08)	(0.10)	(0.07)	(0.04)	(0.02)	(0.03)	(0.02)	(0.20)	(0.17)	(0.20)
PL	−0.03	0.08	−0.03	−0.07	0.10	−0.14	−0.05	−0.03	−0.04	−0.02	0.01	0.27	−0.09
	(0.03)	(0.07)	(0.06)	(0.06)	(0.11)	(0.09)	(0.03)	(0.02)	(0.03)	(0.02)	(0.19)	(0.19)	(0.20)
MON	−0.04	0.05	−0.05	−0.07	0.10	−0.03	−0.01	−0.01	−0.03	−0.00	0.14	0.19	0.09
	(0.03)	(0.07)	(0.07)	(0.07)	(0.11)	(0.09)	(0.04)	(0.01)	(0.03)	(0.02)	(0.23)	(0.19)	(0.21)
MON * RL	0.06	−0.04	0.08	0.14	−0.09	0.04	0.06	0.03	0.07	0.01	−0.06	−0.20	0.01
	(0.04)	(0.09)	(0.10)	(0.11)	(0.15)	(0.12)	(0.06)	(0.03)	(0.04)	(0.03)	(0.32)	(0.25)	(0.30)
MON * PL	0.05	−0.07	−0.01	0.15	−0.08	0.21	0.01	0.02	0.06*	0.03	0.12	−0.26	0.23
	(0.04)	(0.09)	(0.10)	(0.09)	(0.15)	(0.13)	(0.05)	(0.02)	(0.04)	(0.03)	(0.29)	(0.26)	(0.29)
COORD	0.03	0.11*	0.03	0.07	0.16*	−0.01	0.00	0.00	0.04	0.01	0.38**	0.26	0.32*
	(0.02)	(0.06)	(0.06)	(0.06)	(0.10)	(0.06)	(0.03)	(0.02)	(0.02)	(0.02)	(0.17)	(0.16)	(0.17)
[Fisher]	[0.17]	[0.07]	[0.54]	[0.19]	[0.087]	[0.88]	[0.93]	[0.89]	[0.12]	[0.45]	[0.026]	[0.11]	[0.075]
[FDR]	[0.15]	[0.15]	[1]	[0.72]	[0.72]	[1]	[1]	[1]	[0.72]	[1]			
COORD	0.02	−0.09	0.04	0.05	−0.03	−0.14	0.03	0.04	0.01	0.02	−0.26	−0.27	−0.19
* RL	(0.04)	(0.08)	(0.10)	(0.10)	(0.13)	(0.12)	(0.06)	(0.03)	(0.04)	(0.03)	(0.26)	(0.24)	(0.26)
COORD	−0.03	−0.13	0.07	0.10	−0.12	0.06	0.01	0.01	−0.03	−0.01	−0.20	−0.33	−0.10
* PL	(0.04)	(0.09)	(0.08)	(0.08)	(0.15)	(0.12)	(0.05)	(0.03)	(0.04)	(0.03)	(0.26)	(0.26)	(0.26)
Obs.	589	554	585	587	605	500	589	588	587	587	605	589	605
R-squared	0.025	0.007	0.016	0.026	0.015	0.022	0.013	0.015	0.028	0.012	0.015	0.006	0.015
Mean cont.	0.94	0.60	0.84	0.84	0.65	0.77	0.85	0.98	0.93	0.97	0.00	0.00	0.00
(SD)	0.12	0.35	0.37	0.37	0.48	0.43	0.22	0.07	0.15	0.11	1.00	1.00	1.00

OLS regressions. Robust standard errors in parentheses, clustered at the team level. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. The outcome in Column (1) is a dichotomous variable equal to 1 if the student answered yes to the question "Did group leader make sure everybody contributes a fair share?", 0 otherwise. The next questions are: in Column (2): "Did group leader inform the professor if anybody was not contributing?", Column (3): "Did group leader set up the first meeting place and time?", Column (4): "Did group leader make sure that all students were coming to that first meeting?", Column (5): "Did everybody come to that first meeting?", Column (6): "If not was an alternative meeting set to include everybody?", Column (7): "Did group leader assign all group members to precise and well-defined tasks?", Column (8): "Did group leader include everybody in the team?", Column (9): "Did group leader promote good relations within the group?", Column (10): "Did group leader treat everyone equally and with respect?"

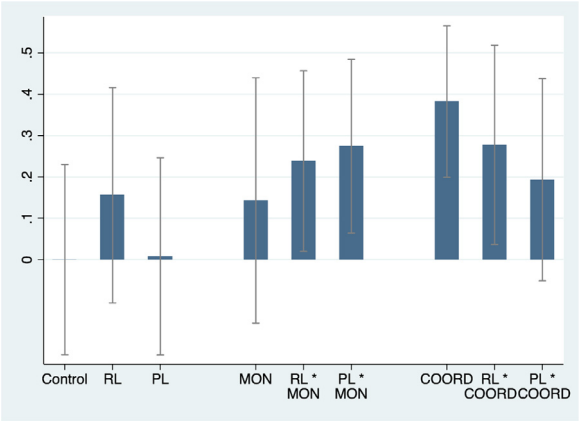

Fig. 3. Effect of Interventions on Leadership Style. *Notes:* The variable graphed is the score of leadership equal to the standardized ten measures of leadership in Columns (1) to (10) weighted by the inverse of the correlation matrix.

the averages of these variables are high. For example, almost all students in the control group say their leaders assigned tasks (85%), included everyone (98%), promoted good relations (93), and treated everyone with respect (97 percent), making it difficult for any intervention to further raise these figures.

Overall, the COORD intervention is found to improve the organizational skills of leaders by informing them of how important it is to make sure that everybody comes to the first meeting. It also increases the propensity of leaders to inform the professor if anybody is not contributing to the presentation. These two results remain significant when using Fisher *p*-values.

To address multiple hypothesis testing, I build a summary index by standardizing all 10 outcomes into a single leadership score, again following Anderson (2008). Results are in Column (11) which shows that COORD leaders do better than the control group and Fig. 3 shows these results in a graph where it is clear that only COORD is significantly associated with the score of leadership.

I then separate the variables and build two summary indices: one for the Monitoring script (that includes outcomes from Columns (1) and (2)) and one for the Coordinating script (that includes outcomes of Columns (3) to (10)). I find that COORD leaders have both higher Monitoring (Column (12)) and Coordinating scores (Column (13)), although the former is not significant

Using FDR *q*-values, the effect of COORD on informing the professor is almost significant at 10%, but the effect on everybody being at the first meeting is not. This is because the unadjusted *p*-values are too large to pass a test of 2 hypotheses in the family of the monitoring outcomes, and 8 hypotheses in the family of the coordinating outcomes.

Nonetheless, it is somewhat reassuring that the results remain significant in three other summary indices of leadership which also address the multiple hypothesis concern. In Appendix G, I use the simple average of the 10 measures of leadership and a more dichotomous measure of leadership. The results are very similar with these additional measures that address multiple hypothesis testing: only COORD is significantly related with the leadership scores. In Appendix I, Table I1 shows the results do not differ whether the respondents are followers or leaders. They are thus not merely driven by the leaders saying they acted this way while the followers disagree.

To summarize, these results indicate that COORD leaders make everybody come to the first meeting (Column (5) of Table 7) which in turn provides good conditions for knowledge-sharing in teams (Column (5) of Table 4). In the next section, I investigate the effects of the interventions on team performance, hours spent working on the teamwork, and benefits gained from the teamwork.

7.4. Heterogeneous effects by gender

The pre-analysis plan also specified to look at heterogeneous effects by interacting “Random Leader” with gender. Table K1 in Appendix K shows the results: the leader’s gender makes no difference in itself. A more detailed analysis of results by gender is available in the appendix in Table K. Essentially, female leaders increase effort by the same amount as male leaders.

7.5. Other outcomes not specified in the pre-analysis plan

In Appendix L, I use all the other questions in the questionnaire. Overall, I find that the RL, PL, MON and COORD interventions increased grades on the teamwork (Table L1). These interventions also increased hours of work spent on the teamwork (Table L2), and not at the detriment of hours spent on the rest of the course.

Moreover, these interventions increased the grade of the leaders on an endline test score not related to the teamwork (Table L3). This is in line with Anderson and Lu (2017) who find that becoming a leader significantly increases one's performance. One explanation for this finding may be that RL, PL, MON and COORD leaders teach more others and are more involved in the course as shown by the extra hours spent on both the teamwork as well as on the rest of the course, thus their academic performance increases.

7.6. Why would leaders not emerge organically?

Considering the large effects of these three interventions (PL, MON and COORD), one may wonder why leaders are not organically emerging in the control group to take on these leading-by-example, monitoring, and coordinating roles.

One explanation for PL not emerging organically in the control group is that there are transaction costs to designating a leader. Students must discuss, agree, resolve competing claims in case of multiple candidates, which can all be awkward phases. I find in this paper that the PL intervention significantly raises the proportion of students who say they have a leader in their team, thereby reducing the transaction costs of designating a leader.

One explanation for why monitoring and punishing is not naturally done in the control group is the fear of retaliation.[24] The fear of retaliation is reduced with the MON intervention since all the students receive the precise instruction that the team leader must inform the professor. The intervention thus provides a justification for informing the professor. On an anecdotal level, multiple messages from the MON treatment groups were sent to the professor.[25]

One explanation for COORD is (lack of) information: the idea for the COORD intervention only became evident after three years of observation of free-riders in teams in Behrens and Chemin (2018). It is very difficult for students to meet.[26] Hard working students, for whom it is most usually the first experience working in teams, usually assume that other students will show the same level of dedication and hard work and will come to meetings. The intuition for the COORD intervention is to share the insight that free-riders usually miss the first meeting such that leaders can tackle free-riding at an early stage before it is too late.

8. Conclusion

When measuring the causal effect of leaders on team performance, the empirical difficulty is the nonrandom assignment of workers to bosses. Rigorous identification strategies of the causal effect of bosses have been developed using: the switch of CEOs across firms (Bertrand and Schoar, 2003); the switch of workers across bosses (Lazear et al., 2015); CEO deaths (Bennedsen et al., 2006); CEO hospitalizations (Bennedsen et al., 2012); economic conditions when managers enter the labor market (Schoar and Zuo, 2017); the deaths of national leaders (Jones and Olken, 2005); and religious leaders' s performance in his or her first church (Engelberg et al., 2016).

In a general sense, the first contribution of my paper is to provide an experiment where some teams have an assigned leader (treatment group) while other teams do not (control group). However, to test economic theories of leadership, I separated the treatment group into subgroups which differ in how the leader was chosen and how he or she was instructed to behave. Along these lines, the second contribution I make is using a unique design to exogenously vary the skills of the person in charge (Promising Leader PL or Random Leader RL). This constitutes a test of the "leading-by-example theory" because -under the assumption that higher skilled leaders exert more effort - the presence of a Promising Leader should cause others to work harder to reduce effort inequality. Finally, the last contribution of my paper is to teach different leadership styles (Monitoring MON or Coordinating COORD) following Bloom et al. (2013) who have found that teaching management practices has large effects on performance. Together, these contributions should help determine which method of selection (endogenous versus by appointment) and which leadership style are most relevant for the effective execution of intellectually challenging tasks.

The most important finding from this study is that the COORD intervention produced the strongest effect when compared to RL, PL and MON. This intervention shared a key insight with leaders and their team: free-riders usually miss the first (also subsequent) meetings, and use this excuse later on to justify their free-riding. The intervention was aimed at making sure leaders organize the first meeting to include all students. This coordinating role was well done in practice: students in the COORD treatment group have significantly better-attended first meetings. I also find that teams in the COORD treatment group meet longer hours, teach each others more, and have better grades on the presentation; indicating better

[24] "Please do not contact the group members directly, as they will know I complained.""I would highly like for this to be confidential as I do not want their grades lowered because of me."

[25] Examples are: "I am having problems coordinating with my team and was wondering if you had time to speak with me about how to improve this issue. As of now, I feel as though I am carrying the majority of the burden and I feel as though I am being taken advantage of."

[26] "The day of the meeting X showed up two hours late, claiming that he is "not a morning person" and Y never showed up, which disappointed us as she had showed willingness but then made no effort to follow up on this.""Unfortunately only one of my 3 other teammates came, the other two informed about an hour before the meeting that they were sick and were going to be late (and then never came).""I have been completely frustrated with the process, as I have been forced to be responsible for coordinating ALL of their activities in relation to this project and they are also being highly unresponsive and unaccountable when it comes to meeting.""We made a doodle earlier in the week and scheduled to meet up and do most of the work this weekend (starting Friday) during which [...] has been completely MIA (even though she said she was available on the doodle), unreachable and unresponsive to our emails.""It's hard to work in team work since lot of members are out of town for the reading week."

team performance overall. These results support the coordinating theory of leadership, at least the organizational aspect of it. There was no evidence suggesting that the COORD intervention led to a more positive work environment.

Despite COORD standing out as the most effective treatment, other interventions were also successful. I find that simply assigning a leader to teams (especially a high-achieving one) or instructing leaders to monitor also increases effort. Once again, this increased effort manifested itself especially in terms of students taking the time to teach and learn from each others but hours of labor increased and team performance also improved. These results are easily reconcilable with a simple model of free-riding and leading-by-example in teams. Effort is under-provided in the control group so assigning leaders, especially those who work hard, increases effort of others as predicted by the leading-by-example theory. Other mechanisms are less credible: "Promising Leaders" are not found to monitor more or coordinate their teams better. The fact that interaction terms between PL and MON (or PL and COORD) are negative is also easily understandable: if MON and COORD already increase effort, there is less inequality in effort levels, thus less leading-by-example.

Taken together, the results suggest that a coordinating leadership style is the most impactful in this context of a non-routine cognitive task where knowledge spillovers and technological complementarities are important. These results are important given the growing extent of non-routine cognitive tasks in the economy (Acemoglu and Autor, 2011), as well as the growing prevalence of teamwork.

Appendix A. Intervention scripts

Control group: No appointed leader

Hello,

Please designate a group leader who will upload your presentations slides on MyCourses in the assignment section.

[Here copy and paste either (1) nothing, OR (2) the coordinating leader script, OR (3) the monitoring leader script].

Please come visit me at my office hours Tuesdays 4–5pm or Friday 2:30–3:30pm for clarifications, or to discuss any problem that may happen in your group.

Sincerely,

Matthieu Chemin

Treatment group: Random leader or Promising leader

Hello,

I designate XX to be the group leader who will upload your presentation slides on MyCourses in the assignment section.

[Here copy and paste either (1) nothing, OR (2) the coordinating leader script, OR (3) the monitoring leader script].

Please come visit me at my office hours Tuesdays 4–5pm or Friday 2:30–3:30pm for clarifications, or to discuss any problem that may happen in your group.

Sincerely,

Matthieu Chemin

Script for coordinating leader:

Every year, we have problems with "free riders", that is to say, students who let others do the work, and get credited with the group grade. From my experience as a teacher over the last seven years, I have noticed that it is very difficult to gather a team of 4 students at a particular meeting place and time. Free-riders usually miss the first meeting, do not understand precisely the work and their tasks, feel excluded and stop working. I count on the group leader to act as a central authority to set up a meeting place, make sure that all students come to the first meeting, cancel the first meeting if not all students are present, and then reschedule another meeting to include all. Another important aspect is for the group leader to assign all members to precise and well-defined tasks. This may be very important to include everybody in the team, promote good relations within the group, treat everyone equally and with respect.

Script for monitoring leader:

Every year, we have problems with "free riders", that is to say, students who let others do the work, and get credited with the group grade. The group leader should make sure that everybody contributes a fair share to the presentation. I count on the group leader to let me know if anybody is not contributing to the presentation.[27]

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2021.02.014.

[27] In order to provide a justification for these messages, the script started with: "Every year, we have problems with "free riders", that is to say, students who let others do the work, and get credited with the group grade." Only one student commented on this within the COORD intervention: "I'm a little confused as to why our group was singled out for this comment, not that it was not appreciated. Did we fail to do something?" The answer was "Your group was not singled out. There were many problems on the statistical teamwork, and I am trying to solve them."

References

Acemoglu, D., Autor, D., 2011. Handbook of labor economics, 4b, pp. 1043–1171.

Anderson, M., 2008. Multiple inference and gender differences in the effects of early intervention: a reevaluation of the abecedarian, perry preschool, and early training projects. J. Am. Stat. Assoc. 103 (484), 1481–1495.

Anderson, M., Lu, F., 2017. Learning to manage and managing to learn: the effects of student leadership service. Manag. Sci. 63 (10), 3246–3261.

Autor, D., Levy, F., Murnane, R., 2003. The skill content of recent technological change: an empirical exploration. Q. J. Econ. 1279–1333. http://jhr.uwpress.org/content/early/2019/09/10/jhr.55.4.0717-8907R2.full.pdf+html.

Behrens, K., Chemin, M., 2018. Confidential peer review and effort in teams: Evidence from a field experiment. J. Human Resour. http://jhr.uwpress.org/content/early/2019/09/10/jhr.55.4.0717-8907R2.full.pdf+html.

Bennedsen, M., Perez-Gonzalez, F., Daniel, W., 2006. Do CEOS matter?NYU Working Paper No. FIN-06-032.

Bennedsen, M., Perez-Gonzalez, F., Daniel, W., 2012. Evaluating the impact of the boss: Evidence from ceo hospitalization events. http://conference.iza.org/conference_files/Leadership_2012/perez-gonzalez_f7878.pdf.

Bertrand, M., Schoar, A., 2003. Managing with style: The effects of managers on firm policies. Q. J. Econ. 118 (4), 1169–1208.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., Roberts, J., 2013. Does management matter? evidence from india. Q. J. Econ. 128 (1), 1–51.

Bloom, N., Van Reenen, J., 2007. Measuring and explaining managament practices across firms and countries. Q. J. Econ. 122 (4), 1351–1408.

Burke, S., Stagl, K., Klein, C., Goodwin, G., Salas, E., Halpin, S., 2006. What type of leadership behaviors are functional in teams? a meta-analysis. Leadersh. Q. 17, 288–307.

Collins, D., Elwood, H., 2004. The effectiveness of managerial leadership development programs: a meta analysis of studies from 1982 to 2001. Human Resour. Devel. Q. 15, 217–248.

Engelberg, J., Fisman, R., Hartzell, J., Parsons, C., 2016. Human capital and the supply of religion. Rev. Econ. Stat. 98 (3).

Gächter, S., Nosenzo, D., Renner, E., Sefton, M., 2012. Who makes a good leader? cooperativeness, optimism, and leading-by-example. Econ. Inqu. 50 (4), 953–967.

Graen, G., Novak, M., Sommerkamp, P., 1982. The effects of leader-member exchange and job design on productivity and satisfaction: Testing a dual attachment model. Organ. Behav. Human Perform. 30, 109–131.

Güth, W., Levati, M.V., Sutter, M., Van Der Heijden, E., 2007. Leading by example with and without exclusion power in voluntary contribution experiments. J. Publ. Econ. 91 (5–6), 1023–1042.

Jack, B.K., Recalde, M.P., 2015. Leadership and the voluntary provision of public goods: field evidence from Bolivia. J. Publ. Econ. 122 (C), 80–93. doi:10.1016/j.jpubeco.2014.10.

Jones, B.F., Olken, B.A., 2005. Do leaders matter? National leadership and growth since World War II. Q. J. Econ. 120 (3), 835–864. doi:10.1093/qje/120.3.835.

Lazear, E., Shaw, K., Stanton, C., 2015. The value of bosses. J. Labor Econ. 33 (4).

Levy, D., Padgitta, K., Peart, S., Houser, D., Xiaoc, E., 2011. Leadership, cheap talk and really cheap talk. J. Econ. Behav. Organ. 77, 40–52.

Morgeson, F., DeRue, s., Karam, E., 2010. Leadership in teams: A functional approach to understanding leadership structures and processes. J. Manag. 36 (1), 5–39.

Phillips, J., Douthitt, E., Hyland, M., 2001. The role of justice in team member satisfaction with the leader and attachment to the team. J. Appl. Psychol. 86, 316–325.

Rotemberg, J.J., Saloner, G., 1993. Leadership style and incentives. Manag. Sci. 39 (11), 1299–1318. doi:10.1287/mnsc.39.11.1299.

Scandura, T.A.G., 1984. Moderating effects of initial leader-member exchange status on the effects of a leadership intervention. J. Appl. Psychol. 69 (3), 428–436.

Schoar, A., Zuo, L., 2017. Shaped by booms and busts: How the economy impacts ceo careers and management styles. Rev. Financ. Stud. 30 (5), 1425–1456. doi:10.1093/rfs/hhw111.

Young, A., 2019. Channelling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. Q. J. Econ. 134 (2), 557–598. doi:10.1093/qje/qjy029.