

May 2022

“Motivating Public Sector Employees: Public Good Contributions in Addis Ababa Water and Sewerage Authority”

George Joseph and Josepa Miquel-Florensa

# Motivating Public Sector Employees: Public Good Contributions in Addis Ababa Water and Sewerage Authority \*

George Joseph

World Bank

Josepa Miquel-Florensa

Toulouse School of Economics

May 19, 2022<sup>†</sup>

We present a lab-in-the-field experiment with employees of the Addis Ababa Water and Sanitation Authority with the aim to understand how to improve coordination and collaboration in their daily crew work. Participants play a series of public good games under different rules: standard game, with identifiable set of partners, game with threshold, and game with a randomly selected anonymous leader with the power to punish.

We show that a common goal, in the form of a threshold to be attained for the group success, is significantly more effective than a potentially punishing leader to increase individual effort and ultimately group outcomes. This result advocates for the introduction of team goals as coordination and motivation devices in settings where tasks are performed by groups and are subject to free-riding and coordination challenges.

**Keywords:** Intrinsic motivation, Public utilities, Organizational economics.

**JEL codes:** J45 (Public Sector Labor Markets), M50 (Personnel Economics), O12 (Microeconomic Analyses of Economic Development).

---

\*We thank Cesar Mantilla, Sanjay Pahuja, Erin McDonnell and seminar participants at the ASFEE Virtual seminar and SEEDEC for their comments and suggestions. All errors remain our own. The views presented in this paper are those of the authors and do not represent those of the World Bank or its member countries. The design was reviewed by the TSE-IAST review board for ethical standards in research (2019-08-001).

<sup>†</sup>George Joseph, World Bank [rgjoseph@worldbank.org](mailto:rgjoseph@worldbank.org); Corresponding author: Josepa Miquel-Florensa, Toulouse School of Economics (Université Toulouse 1 Capitole) [pepita.miquel@tse-fr.eu](mailto:pepita.miquel@tse-fr.eu)

## 1. INTRODUCTION

Improving the quality of public service delivery is a challenge in developing countries. Capital investments, institutional reform, and capacity-building are the main ingredients of interventions aiming to improve coverage and use of public utilities. However, even when the projects succeed in implementing the infrastructure and policy reforms, these gains often fail to translate into the expected service delivery improvements for the intended beneficiaries. Physical investments do not improve service delivery without well trained and motivated public agency staff. This staff are responsible for executing the policies and implementing the infrastructure improvements and are the face of the public utility to the consumers. Hence, they are the crucial last mile to ensure take-up and use of the services.

Two challenges often faced by the public utilities are how to attract public service motivated individuals to the utilities and how to ensure that the existing stock of employees remain motivated to deliver their assigned tasks diligently and earnestly (Finan et al., 2017).<sup>1</sup> This paper addresses the second question in the context of the Addis Ababa Water and Sanitation Authority (henceforth AAWSA) - a large water and sanitation service provider in Addis Ababa. To design effective incentives and organizational schemes is key for the utility to reach its service objectives. An additional challenge faced by public utilities comes from the fact that frequently tasks are performed in crews, adding moral hazard and coordination concern on top of motivation challenges (Holmstrom, 1982).<sup>2</sup> This is the case of water and sanitation services provision, where working crews are formed by four to six employees specialized in different roles. Different instruments have been considered, both in the academic literature and in the field, to provide incentives to team effort. Among them, the difficulty to evaluate individual accountability in teams (Marx and Squintani, 2009), the difficulty to monitor heterogeneous teams

---

<sup>1</sup>Social incentives are specially relevant in organizations (see Ashraf and Bandiera (2018) for a survey). This is especially true in sectors where the quality of outcomes depends primarily on the attitude and behaviors of the last-mile service providers (see Ashraf et al. (2014), Ashraf et al. (2018) and Mbiti and Serra (2018) for examples in the health sector). The evidence on the interaction between economic incentives and social preferences is however inconclusive (see Bowles and Polania-Reyes (2012) for a survey).

<sup>2</sup>After the seminal Holmstrom (1982) paper, the literature on monitoring effort in teams is extensive. For example, Gershkov and Winter (2015) study formal versus informal monitoring, Halac et al. (2021) in a theoretical setting look at the difficulty to monitor heterogeneous teams, and Marx and Squintani (2009) look at individual accountability in teams. Herbst and Mas (2015) summarize the experimental and field literature on workers output peer effects, showing that is positive and not statistically significant in the lab versus the field. Hamilton et al. (2003) focuses on team incentives and workers heterogeneity. Villeval (2020) highlights the importance of norms and institutional dynamics on groups performance.

(Halac et al. (2021), Weng and Carlsson (2015)), and the trade-off between formal and informal monitoring schemes (Gershkov and Winter, 2015) have been among the most studied options.

We focus on this paper on the effectiveness of different management policies in increasing crews motivation and hence the team’s performance: Is it about the knowing the other employees in the team? Is it about setting a common goal among employees? Or is it about having supervision (and potential punishments)? To answer these questions, we perform a lab-in-the-field experiment where employees of AAWSA play a series of public good games under different rules: standard game, with identifiable set of partners, game with a threshold, and a game with a randomly selected anonymous leader with the power to punish.

On the first question, we find that knowing the other partners in the team does have a positive and significant impact on contributions. On the second question, we find that that a common goal in form of pre-set threshold does significantly increase contributions, consistently with the literature surveyed in Dannenberg et al. (2015). On the third question, we observe that while players increase their contribution when the leader figure is introduced, this increase is smaller than the one found in the presence of a common goal threshold.

The identification and empowerment of leaders improves group cooperation (Kosfeld and Rustagi (2015)) and the voluntary provision of public goods (Jack and Recalde (2015), d’Adda et al. (2017)). Harnessing the power of personal agency and identifying people who can champion the change initiatives is vital for achieving the desired outcomes in these sectors. Our experiment is consistent with the literature in showing that the leader figure increases contributions, but significantly less than the introduction of a common goal for the team.

The experiment is done with personnel from AAWSA, the public service provider of water supply and sanitation services in Ethiopia’s capital city. There are two facts to be noted about the work structure in water and sanitation branches: (i) there is a vertical distinction across grades of employees, with different training and experience, and (ii) activities are performed in crews of 4-5 employees across different levels. This need of collaboration across levels is key for the provision of good service to customers. Our experimental results shed light on how to design effective motivational incentive schemes in this setting.

The paper is structured as follows. Section 2 provides a description of AAWSA, the setting where the experiment took place. Section 3 describes the experimental sample and procedures. Section 4 presents the empirical analysis, and Section 5 presents the results at the individual and group level, together with the leader

choices. Section 6 concludes.

## 2. SETTING: THE ADDIS ABABA WATER AND SANITATION AUTHORITY

The Addis Ababa Water and Sanitation Authority is the public institution that provides water and sanitation services to Ethiopia's capital (3.5 million inhabitants). It is structured around 8 branches with specific catchment area, covering the whole extension of the city. All branches have an identical internal structure and governance.

The main tasks of the authority's employees are performed in crews of 4-6 employees of different experience. The speed in the reaction to issues on the water/sewerage lines and maintenance of trucks and other materials are key for service delivery. Collaboration within the crew and across levels is crucial for a quick reaction. Hence, the motivation of these crews, and the improvement of their internal coordination, are top concerns for management.

## 3. THE PUBLIC GOOD EXPERIMENT

The literature has considered a variety of experimental games to measure social preferences.<sup>3</sup> In the special case of public employees of the utility, to study social preferences we need to keep in mind that work is performed mainly in crews: groups of 4 to 6 employees, with the usual concerns on free-riding. This is the case in AAWSA, where crews are usually formed by 5 employees with different complementary skills.<sup>4</sup> This peculiarity of the tasks performed makes a public good game a good instrument to measure social preferences and motivations to effort in this setting and to evaluate interventions aiming to improve team outcomes. To mimic the workplace structure, we propose a public good game in teams of 5 members, with a neutral framing.

The main goal of our experimental design is to test how individual contributions to a standard public good game - and ultimately group outcomes - vary with the introduction of two management strategies: (i) a *common goal* motivation in form of a step function, and (ii) with the inclusion of a game leader who has the possibility to punish participants as function of their contribution amounts.

Our benchmark is the standard public good game: in groups of five partic-

---

<sup>3</sup>See surveys of [Charness and Rabin \(2002\)](#), [Camerer and Fehr \(2004\)](#) and [Levitt and List \(2007\)](#).

<sup>4</sup>Teams are composed usually by a driver and a team of engineers of different grades.

ipants, each participant gets an endowment of 10 tokens<sup>5</sup>; contributions to the pool are doubled and divided equally among the team members. In this setting, the payoff of player  $i$  that contributes  $c_i$  (while the other four team members contribute  $c_{-i}$ ) is given by

$$\pi(c_i, c_{-i}) = \underbrace{[10 - c_i]}_{\text{contribution}} + \underbrace{\frac{2}{5} * [\sum_{j=1}^5 c_j]}_{\text{received from box}}$$

where the first part of the payment comes from the endowment not contributed to the pool and the second part comes from the tokens distributed from the pool after amounts contributed by all players have been doubled and split equally among the group members.

The variations on the benchmark public good game played are as follows (see Appendix A for the protocol details). In all games, contributions of each player were private and anonymous, and no information on group return was revealed until payment stage (i.e. after all games had been played). All the games played were paid.

- Our *benchmark* is the traditional Public Good game, in groups of 5 players. The goal of this game is to measure contributions in a game that mimics in number of employees the crews that work together daily. Tokens contributed to the common box are doubled and distributed among the team members, independently of the total group contribution. A first round was played with an hypothetical group of five persons in the room, with no information on whether they were sitting at the same table or another. This first round had the goal to familiarize the participants with the game. A second round was played with the participants sitting on the same table. The second round allows to see the impact on contribution choice of identifying the game partners (while keeping in all the games the contributions private).
- *Threshold Public good game*: This situation aims to mimic the effort to be provided to attain many of the tasks performed in the employee's crews on a daily basis, that require a minimum total contribution of effort in order to make any progress towards the competition of the task. Payments in this

---

<sup>5</sup>Each token is valued 1 Ethiopian BIR (0.035 USD at the time of the experiment). Average payment per participant was 66 BIR (2.5 USD), approximately the mean salary for two hours of work. Needs to be noted that the experiment took place in the employee's standard working hours, and hence the payments were adapted accordingly.

case follow a step function: Tokens contributed to the box are doubled when the total contributed amount reaches the threshold of 25 tokens (half of the total tokens of group participants), otherwise all contributions are lost.

- *Public Good game with leader*: Inspired by the protocol presented in [Kosfeld and Rustagi \(2015\)](#), we add to the standard public good game the figure of the game leader - a member of the group with the power of punishing participants when observing their anonymous contributions. Leaders are randomly selected among session participants by picking in front of all participants a card from a bag with all players codes without announcing the outcome. Chosen leaders are informed of their role at the end of the session when all games have been played and no results have been announced to the participants.<sup>6</sup>

### 3.A. *Experimental sample*

The lab-in-the-field sessions took place the first day of the Field Level Leadership workshops between September 2019 and February 2020 at the AAWSA Training facilities located at the outskirts of Addis Ababa.<sup>7</sup> The workshops were performed in groups of 25-30 employees from two of the Authorities branches: Arada and Addis Ketema. The two branches are very similar in terms of employees characteristics and their distribution across departments (see Table A1).<sup>8</sup> All employees of these two AAWSA branches participated in the training and were invited to volunteer for the experiment, with very high participation rates. The employees were distributed across workshops to ensure diversity across and between departments and ranks in the two branches.

---

<sup>6</sup>After all games are played and post-experimental questionnaires have been completed, the randomly selected leaders are discretely called to be told they had been randomly picked for the leader role and are asked to make their choices as leaders.

<sup>7</sup>Field-Level Leadership (FLL) constitutes a set of interventions aimed at identifying and supporting entrepreneurial and motivated employees in public agencies to lead improvements in performance and service delivery outcomes. The identified employees, together with their colleagues from all levels in the organization, are trained to motivate their pairs and improve the branch communication both horizontally (across peers on same rank) and vertically (across ranks). The experimental sessions took place after the introductory session of the training, where participants introduced themselves as per name and branch and had lunch at the training facility. While interaction between participants can not be ruled out, the workshop trainers ensured the research team that no information on the workshop goals and contents was given to the participants in the presentation session.

<sup>8</sup>Figure A1 shows that the employee grade distribution across the two branches is very similar. In the different specifications we control for employee's branch and group's branch composition respectively.

Part A of Table 1 presents the socio-economic characteristics of the participants. We see that the average age is 37 years and that employees are predominantly male (over 80%). In terms of education, it is worth to note that the majority of participants do have over primary education, specially technical diplomas on infrastructure maintenance tasks. In terms of department inside AAWSA, part B of Table 1 shows that the majority of employees work on consumer service tasks both in water and sewerage departments. These departments together with line installations are the ones where crew work is predominant, supporting the choice of public good games in this institutional setting.

### *3.B. Experimental procedures*

At arrival to the experimental room, participants are randomly distributed ID tags. When entering the room participants find the seat that has the number that appears on their tag, with the help of the experiment assistants. That sets participants in tables with five members, distributed in the room so that all participants can see the experiment director and the projection of the (language free) instructions. In front of them, on the table, participants find four closed plastic recipients of different colors with their ID marked on them containing ten white tokens (and some non-valuable black tokens) and a pen for the post-experimental questionnaire. At the center of the table participants find a wooden box that acts as a contribution box.<sup>9</sup> Detailed information on the protocol as well as the experimental materials can be found in Appendix A.

The structure of the experimental session is presented in Figure 1. The randomly selected groups of 5 participants sitting at the same table are the groups playing the public good game from second round on. First game was with partners in the room to let the participants become familiar with the structure of a public good game. Given that participants were not familiar with abstract reasoning, this first round helped them to familiarize with the public good game. The last two games, leader and threshold, were played in different order for different sessions.<sup>10</sup> In the leader game, random selection of the leader was done by one

---

<sup>9</sup>Black tokens have no points value. Their role in the individual boxes and in the common wooden box is to ensure privacy of the individual contributions. On the one hand, when participants pass the box around the table to make the contributions, the tokens in the common box make noise with movement ensuring that nobody can guess the amounts contributed and hence added to the box as it goes around the table. On the other hand, the non-value tokens in the individual boxes also make noise before and after the participant makes the choices, ensuring privacy of decisions.

<sup>10</sup>Session and group fixed effects are included in the analysis to control for potential ordering effects.



member per group picking a card from a box that contained all the participants IDs. The chosen leader was only known by the experiment director, and not by the group members or the leader himself - the "leading by example" mechanism is then blocked. After all games were played and the post-experimental questionnaire was done, experiment assistants would call the leader privately to inform him that his ID had been randomly picked to act as a leader and to ask him for his decision. No information on games outcomes was disclosed until the payment stage.

Figure 1: **Structure of the Experimental Sessions**

| Order 1   | Order 2   |
|---|-----------|
| Test standard PG -Room partners   |           |
| Standard PG - Table partners  |           |
| Threshold   | Leader    |
| Leader  | Threshold |
| Post-experimental survey  |           |
| Leader decisions and payments   |           |
| <i>Notes:</i> Order of the games was randomly allocated to the session. |           |

One important question given that three out of four games are played with the participants sitting on the same table is whether the participants in the table - and in the session - knew each other. Figure A3 in the Appendix shows that around 70% of the participants report to know at least one other participant sitting at the same table, and that on average in each group participants know 2.4 of the table members including themselves. On average, in each table half of the participants came from the same branch (first quartile 40%, third quartile 60%). With respect of the participants in the session, Figure A3 shows that each participant knows between 5 and 15 of the other participants. This number is slightly smaller than the participants in the session they recognize as workers of the same branch.

#### 4. EMPIRICAL STRATEGY

We analyze the experimental data at three levels: individual contributions, group outcomes, and behavior of the randomly chosen leaders.

#### 4.A. Individual contributions: Within-subject analysis

Participants in a given session play four versions of the public good game, as described in Figure 1. Hence, we compare the behavior of each subject across these four decisions. In this analysis, we include individual fixed effects controlling for any individual specific characteristics (ex. pro-sociality, identification with the group, ...) that allow us to identify the changes in behavior given by the difference across games: namely the change of the game partners from room to table, and the introduction of a threshold and of a randomly selected group leader.

We estimate:

$$C_{igt} = \alpha + \beta_g * \sum_g Game_g + \gamma_i + \epsilon_{igt} \quad (1)$$

where  $C_{igt}$  denotes contribution of player  $i$  in game  $g$  sitting at the table of team  $t$ , and  $\gamma_i$  are individual player fixed effects. While this is our preferred specification (and the most conservative one), for robustness we run the same estimation with group fixed effects, group branch composition controls, and individual characteristics. We cluster standard errors at the session level.

#### 4.B. Group-level: Within-group analysis

In our setting, a performance indicator of interest is the outcome at the group level. At this level of aggregation power is limited, since we have 1/5 of the observations at the individual level when considering the group, a total of 87 groups. But is at this level that we are able to observe total contributions to the public good and their relationship with group composition, the threshold and the introduction of a randomly selected group leader. It is also relevant to study the likelihood of a group to fail to attain the threshold as a function of the group composition.

We estimate:

$$G_{gt} = \alpha + \beta_g * \sum_g Game_g + \eta_t + \epsilon_{gt} \quad (2)$$

where  $G_{gt}$  denotes total contribution on game  $g$  of team  $t$ , and  $\eta_t$  denote group fixed effects. As with the case of the individual contributions, this is our main specification since group fixed effects are the more conservative approach, but we present as robustness the specifications with session fixed effects and group characteristics as controls.

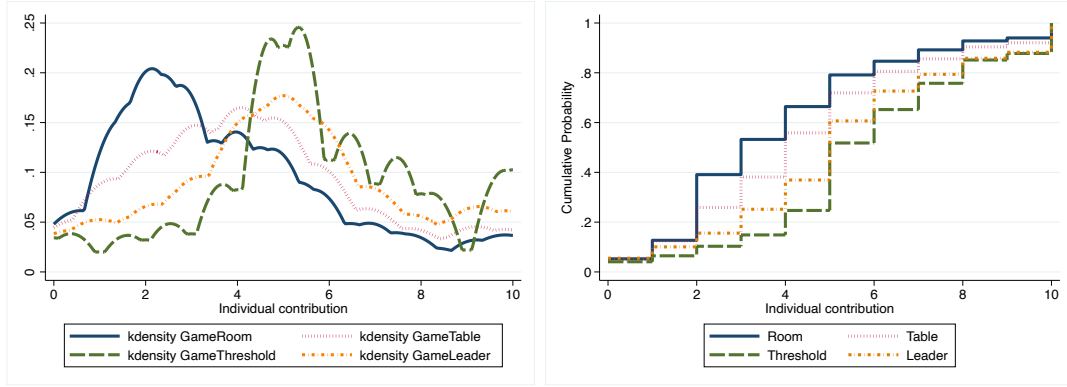


Figure 2: Distribution of individual contributions.

The right figure shows the distribution of individual contributions for the four different games. The left figure shows the cumulative distribution of the individual choices. The four distributions are pairwise significantly different (Wilcoxon distribution test,  $p=0.000$ ).

#### 4.C. Choices of the randomly chosen game leaders

At the leader level, we focus on three variables of interest: the probability that the leader punishes a group, the number of players punished in that case and, within this group, his choice of whom to punish. We look at specifications at the leader level, to analyze the determinants of his decisions, and at specifications at the individual player level.

## 5. RESULTS

Table 2 presents descriptives of the experimental outcomes. In Part A we see that, at the individual level, the threshold game is the one that got higher average contributions (57% of endowment), followed by the leader game (52% of endowment). We see same pattern in Part B, that shows the average group contributions.

#### 5.A. Analysis at the individual level

With respect to individual contributions to the public good, our design allows to answer two questions: Do participants make a difference when they can better identify their game partners? And can the introduction of a threshold or of a randomly assigned game leader increase contributions to the public good game, keeping the game partners constant?

With respect to the first question, in Table 2 Panel A we see that the average contribution to the game with partners in the room is 38% of the endowment,

versus 44% for the game with partners on the participant's table, being the distributions significantly different (Wilcoxon distribution test,  $p=0.00$ ). Figure 2 shows that on the first game (room partners) 40% of the participants contribute less than two tokens, a very high share compared to 20% for the game with partners on the same table.

With respect to the second question, Table 2 Panel A shows that both the threshold and the leader increase the average contribution to the public good game with respect to the benchmark game with partners on the same table to 57% and 52% of their endowment, respectively. Moreover, the four distributions are statistically different, being the threshold game the one that maximizes contributions. Figure 2 shows that in the threshold game, 70% of the participants contribute five or more tokens (25% contribute exactly five) while this share is 50% for the leader game. As Figure 2 shows very clearly, the threshold game does put a reference at contributing half of the endowment and reaching the 25 tokens threshold.<sup>11</sup>

Table 3 shows the results of estimating equation 1. In all columns, the reference game is the game played with table partners, with the goal of comparing this game to the room game - first question- and to the two games with partners in same table but with a twist - second question. Column (1) includes group fixed effects, column (2) adds player branch of origin fixed effects, column (3) is the most conservative estimation including player fixed effects, and last column, column (4) adds to the specification in column (2) individual controls. All specifications show very similar results: with respect to the first question, contributions to the game with partners in the room are significantly smaller than with partners on the same table. With respect to the second question, also consistently across specifications we find that both the threshold and the leader significantly increase individual contributions, with a greater difference for the first. At the bottom of Table 3 we show that the coefficients for the three games being compared with the room game are pairwise significantly different.

We present in Figure A4 some heterogeneity analysis of individual contributions with respect to player's rank in the institution, tenure, and a post-experimental risk-elicitation game. We see that neither of these individual characteristics does have a significant impact on contribution level choices at the different games.

---

<sup>11</sup>The threshold is very likely to set a focal point for the players on posterior games, and for that reason we altered the order of the threshold and leader games as shown in Figure 1 to be able to control for the order effects.

### 5.B. Analysis at the group level

We present in Figure A5 in the Appendix the distribution (and cumulative distributions) of the group aggregate contributions. We see that the patterns as expected mimic the ones of individual contributions presented in Figure 2. Table 2 Panel B shows that average contributions are maximized in the threshold game (27.5 tokens, 78% of the groups reaching the threshold), and are the smaller in the game with room partners (18.3 tokens).

Table 4 shows the estimation of equation 2. Column (1) includes session fixed effects, column (2) adds to that branch composition of the group, and column (3) includes group fixed effects. We observe again that when groups play with room partners they contribute less than when playing with the partners in the table, and that leader and threshold games lead to greater group contributions than playing the benchmark game with partners in the table. The largest contributions are observed on the threshold game. Again coefficients for the different games are pairwise significantly different as shown in the tests on the bottom of Table 4.

### 5.C. Leader's behavior

Our experimental design allows to answer two questions related to including a randomly chosen leader in the game: Whether that increases players contributions, and whether the leader does actually use his authority to punish players. With respect to the first question we see that contributions in the game with leader are significantly higher than in the standard game, both at the individual and at the group level (as shown in Tables 3 and 4). With respect to the second question, Figure 3 shows that leaders do actually choose to punish in around 75% of the groups. As a reminder: leaders are randomly chosen, their identity is not revealed to the groups, not revealed to them as well when contribution decisions are made, and are asked to take their choices as leaders at the end of the experimental session. The choices are taken on a form that shows the anonymous contributions of each member of the group where they were randomly allocated as leaders (See Figure F5 in the Appendix). In terms of payoffs, the leader receives 10 tokens for being chosen for the role, and his choice is whether to punish a player, what implies a cost of one token for the leader and three tokens for the player. Hence, punishing is a costly decision for the leader.

We focus on the leader choice in terms of punishing anyone on the group, and on the number of players punished (columns (1) to (4) in Table 5, where the unit of observation is a leader allocated to a group). We see that the leader's

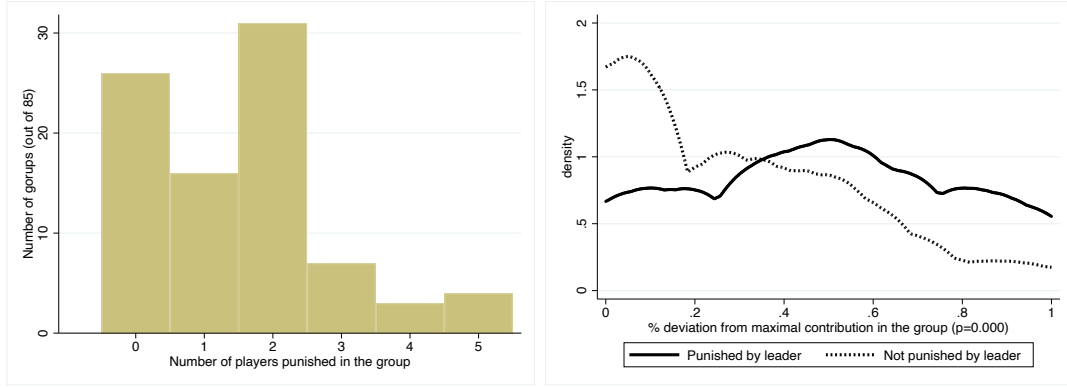


Figure 3: Leader behavior.

The right figure shows the number of players the leader chooses to punish in the group where he is randomly denominated leader. The left figure plots how the deviation from the maximum contributor in the group relates to the probability of being punished. The dotted line shows that the greatest share of punished participants are among those whose contributions represent a small share of the one of the main contributor in the group.

discrete decision to punish - i.e. a dummy variable that takes value one if the leader punishes at least one member of the group - is not correlated to group contributions and neither to the variance of contributions inside the group (Columns (1) and (2)). However, columns (3) and (4) show that the number of players punished in the group is significantly positively correlated with the variance of contributions within the group. We continue our analysis at the individual player level, taking as unit of observation a game participant. Column (5) of Table 5 shows that the likelihood of a player to be punished is significantly negatively correlated with his individual contribution. Given that the number of players punished in a group is positively correlated with the variance of contributions inside a group, we define the percentage difference of each player's contribution with respect to the highest contribution in the group. Figure 3 plots this relative difference for punished and non-punished players, and shows that high deviations (players with contributions smaller than 40% of the highest contributor in the group) are more likely to be among the punished.<sup>12</sup> Column (6) of Table 5 shows, in a regression form, that this negative correlation of deviation with the probability of being punished is positive and significant.

One fair question at this point is how did the leaders behave themselves in the

<sup>12</sup>Figure A7 shows that same pattern can be shown by focusing on absolute contribution to the leader game: the higher the contribution the smaller the likelihood to have been punished by the leader.

games. Table A2 shows that the randomly selected leaders, who did not know they would be leaders at the time they contributed to the different games, do not make significantly different individual contributions to the games than other players in the session. This does show that leaders behavior in the games was not significantly different from other players. Figure A8 looks in more detail at the contribution of leaders versus other players. We see that among the leaders, the ones that do not punish are the ones that make a smaller difference between their contributions to the leader and the table games. While these figures are only per illustration given the size of the sample (we have 87 leaders, of which only 22 do not punish) they do show some correlation between leader decisions and their own behavior in the games. Needs to be noted that leaders make all choices in the games before knowing whether they have been randomly chosen for the role, hence we do not expect their role to affect their behavior in the games.

We also look at the correlation of the leader's choices of whether to punish anyone in the group, and the number of participants punished, and the leader's individual characteristics. Figure A6 shows that there does not seem to be a correlation of the leader's position in the company or his tenure and his choice.

## 6. CONCLUSIONS

This lab-in-the-field experiment shows that a common goal, in the form of a threshold to be attained for the group success, is significantly more effective than the introduction to the game of an anonymous potentially punishing leader to increase individual effort and ultimately group outcomes. This result advocates for the introduction of team goals as coordination and motivation devices in settings where tasks are performed by crews.

The behavior of the randomly chosen team leaders is also interesting as per evaluating their positive impact when comparing outcomes with the standard public good game. We see that leaders do decide to punish participants that free-ride, specially in cases of big dispersion across group contributions. This behavior is illustrative of the expectations of participants on others' behavior, specially in this setting where participants do participate in crews in their daily work.

There are three main insights emerging from the experiment to guide reforms in utilities authorities where crews are key for service provision. First, it is important to encourage interaction among the crews, which will improve identity and cohesion and have implications for expending more effort and not shirking. Second, choosing leaders for the teams (randomly or in some setting-dependent acceptable

fashion) can lead to improvements in individual prosocial behavior. Third, setting a common goal for the crew would encourage prosocial behavior by individual employees and may be more effective than introducing external monitoring. These insights, when used in combination or in isolation by the management, have the potential to improve employee performance standards.

However, the limitations of the extrapolation of lab-in-the-field exercises to actual workplace design need to be acknowledged. First, the (partial) anonymity of the experimental setting abstracts from potential personal incompatibilities among crew and branch members. The history of the employees in the institution is likely to play a key role on their reaction to workplace arrangements. Second, the experiment allows for an anonymous leader, what may likely not be feasible in reality. And third, in the experimental setting is easy to set a threshold for the total contribution. The risk faced in service provision makes it difficult to set thresholds on performance, main argument against piece-rate payments in public utilities.

## REFERENCES

- Ashraf, N. and Bandiera, O. (2018). Social incentives in organizations. *Annual Review of Economics*, 10(1):439–463.
- Ashraf, N., Bandiera, O., and Jack, B. K. (2014). No margin, no mission? a field experiment on incentives for public service delivery. *Journal of Public Economics*, 120:1 – 17.
- Ashraf, N., Bandiera, O., and Lee, S. (2018). Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services.
- Bowles, S. and Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements? *Journal of Economic Literature*, 50(2):368–425.
- Camerer, C. F. and Fehr, E. (2004). Measuring social norms and preferences using experimental games: A guide for social scientists. *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*, 97:55–95.
- Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.



- Dannenberg, A., Löschel, A., Paolacci, G., Reif, C., and Tavoni, A. (2015). On the provision of public goods with probabilistic and ambiguous thresholds. *Environmental and Resource economics*, 61(3):365–383.
- d’Adda, G., Darai, D., Pavanini, N., and Weber, R. A. (2017). Do leaders affect ethical conduct? *Journal of the European Economic Association*, 15(6):1177–1213.
- Finan, F., Olken, B., and Pande, R. (2017). Chapter 6 - the personnel economics of the developing state. In Banerjee, A. V. and Duflo, E., editors, *Handbook of Economic Field Experiments*, volume 2 of *Handbook of Economic Field Experiments*, pages 467 – 514. North-Holland.
- Gershkov, A. and Winter, E. (2015). Formal versus informal monitoring in teams. *American Economic Journal: Microeconomics*, 7(2):27–44.
- Halac, M., Kremer, I., and Winter, E. (2021). Monitoring teams.
- Hamilton, B. H., Nickerson, J. A., and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of political Economy*, 111(3):465–497.
- Herbst, D. and Mas, A. (2015). Peer effects on worker output in the laboratory generalize to the field. *Science*, 350(6260):545–549.
- Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, pages 324–340.
- Jack, B. K. and Recalde, M. P. (2015). Leadership and the voluntary provision of public goods: Field evidence from bolivia. *Journal of Public Economics*, 122:80–93.
- Kosfeld, M. and Rustagi, D. (2015). Leader punishment and cooperation in groups: Experimental field evidence from commons management in ethiopia. *American Economic Review*, 105(2):747–83.
- Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives*, 21(2):153–174.
- Marx, L. M. and Squintani, F. (2009). Individual accountability in teams. *Journal of Economic Behavior & Organization*, 72(1):260–273.

- Mbiti, I. and Serra, D. (2018). Health workers' behavior, patient reporting and reputational concerns: Lab-in-the-field experimental evidence from kenya.
- Villeval, M. C. (2020). Public goods, norms and cooperation. In *Handbook of Experimental Game Theory*. Edward Elgar Publishing.
- Weng, Q. and Carlsson, F. (2015). Cooperation in teams: The role of identity, punishment, and endowment distribution. *Journal of Public Economics*, 126:25–38.

Table 1: Descriptives of the experiment participants

|   | N. obs. | Mean   | sd    |
|---|---------|--------|-------|
| <b>Part A: Socio-Economic characteristics</b> |         |        |       |
| Age   | 364     | 37.659 | 9.951 |
| Male  | 364     | .813   | .390  |
| Married                                       | 364     | .653   | .476  |
| Never married                                 | 364     | .291   | .454  |
| Primary                                       | 364     | .156   | .363  |
| Secondary                                     | 364     | .203   | .403  |
| Technical diploma                             | 364     | .376   | .485  |
| First degree and above                        | 364     | .255   | .436  |
| Language Amharic                              | 364     | .755   | .430  |
| Lives in Addis                                | 364     | .840   | .366  |
| Family breadwinner                            | 364     | .862   | .344  |
| Owns dwelling                                 | 364     | .239   | .427  |
| Public rent                                   | 364     | .159   | .366  |
| Private rent                                  | 364     | .527   | .499  |
| <b>Part B: Employment – AAWSA departments</b> |         |        |       |
| <i>Water services:</i>                        |         |        |       |
| Line installation                             | 364     | .167   | .374  |
| Non-revenue water                             | 364     | .060   | .238  |
| Water consumer service                        | 364     | .252   | .435  |
| <i>Sewerage:</i>                              |         |        |       |
| Line installation                             | 364     | .079   | .271  |
| Sewer connection                              | 364     | .013   | .116  |
| Sewerage consumer service                     | 364     | .208   | .407  |
| <i>Support department:</i>                    |         |        |       |
| Human resources                               | 364     | .013   | .116  |
| Finance                                       | 364     | .027   | .163  |
| Procurement                                   | 364     | .013   | .116  |
| General service                               | 364     | .065   | .248  |
| Planning and budgeting                        | 364     | .008   | .090  |
| Monitoring and evaluation                     | 364     | .008   | .090  |
| <i>Non specified:</i>                         |         |        |       |
| Other   | 364     | .079   | .271  |

*Notes:* Descriptive statistics are presented for the participants (87.3%) that participated in a paralel employee survey with detailed information on job history.

Table 2: **Summary of experiment results**

**Part A: Results at the individual level**

|                | N   | mean     | sd       |
|----------------|-----|----------|----------|
| Game Room      | 417 | 3.834532 | 2.547426 |
| Game Table     | 417 | 4.422062 | 2.636521 |
| Game Threshold | 417 | 5.738609 | 2.525118 |
| Game Leader    | 417 | 5.199041 | 2.715907 |

**Part B: Results at the group level**

|                | N  | mean     | sd       |
|----------------|----|----------|----------|
| Game Room      | 87 | 18.37931 | 8.828128 |
| Game Table     | 87 | 21.1954  | 9.048756 |
| Game Threshold | 87 | 27.50575 | 8.72163  |
| Game Leader    | 87 | 24.91954 | 8.352995 |

Table 3: **Individual contributions**

|                              | (1)                     | (2)                     | (3)                     | (4)                     |
|------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| <b>Reference game: Table</b> |                         |                         |                         |                         |
| Room game                    | -0.5875 ***<br>(0.1620) | -0.6010 ***<br>(0.1632) | -0.5875 ***<br>(0.1327) | -0.5658 ***<br>(0.1654) |
| Leader game                  | 0.7770 ***<br>(0.1620)  | 0.7762 ***<br>(0.1632)  | 0.7770 ***<br>(0.1327)  | 0.8179 ***<br>(0.1654)  |
| Threshold game               | 1.3165 ***<br>(0.1620)  | 1.3090 ***<br>(0.1632)  | 1.3165 ***<br>(0.1327)  | 1.3417 ***<br>(0.1654)  |
| Constant                     | 4.4221 ***<br>(0.1146)  | 4.4619 ***<br>(0.1307)  | 4.4221 ***<br>(0.0938)  | 2.8661 ***<br>(0.4678)  |
| Test Room=Leader             | p=0.000<br>F=70.94      | p=0.000<br>F=71.24      | p=0.000<br>F=105.70     | p=0.000<br>F=69.95      |
| Test Room=Threshold          | p=0.000<br>F=138.13     | p=0.000<br>F=137.04     | p=0.000<br>F=205.82     | p=0.000<br>F=132.94     |
| Test Leader=Threshold        | p=0.000<br>F=11.09      | p=0.001<br>F=10.67      | p=0.000<br>F=16.53      | p=0.001<br>F=10.02      |
| Observations                 | 1668                    | 1644                    | 1668                    | 1428                    |
| Player ID                    | 416                     | 416                     | 416                     | 357                     |
| $R^2$                        | 0.2909                  | 0.2908                  | 0.6241                  | 0.3699                  |
| Group FE                     | Yes                     | Yes                     | No                      | Yes                     |
| Player FE                    | No                      | No                      | Yes                     | No                      |
| Branch of origin FE          | No                      | Yes                     | No                      | Yes                     |
| Individual controls          | No                      | No                      | No                      | Yes                     |

*Notes:* Robust standard errors, cluster session. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level. Unit of observation is the choice of a participant in a given game. Individual controls included in column (4) are age, dummies of education level, department, grade and branch. The smaller sample is due to the participants that missed the parallel employee survey.

Table 4: **Group contributions**

|                              | (1)                    | (2)                    | (3)                    |
|------------------------------|------------------------|------------------------|------------------------|
| <b>Reference game: Table</b> |                        |                        |                        |
| Room Game                    | -2.8161 **<br>(1.2443) | -2.8161 **<br>(1.2267) | -2.8161***<br>(0.7659) |
| Leader Game                  | 3.7241 ***<br>(1.2443) | 3.7241***<br>(1.2267)  | 3.7241***<br>(0.7659)  |
| Threshold Game               | 6.3103***<br>(1.2443)  | 6.3103***<br>(1.2267)  | 6.3103***<br>(0.7659)  |
| Constant                     | 21.195***<br>(0.8798)  | 24.411***<br>(1.3173)  | 21.195 ***<br>(0.5416) |
| Test Room=Leader             | p=0.000<br>F=27.63     | p=0.000<br>F=28.43     | p=0.000<br>F=72.92     |
| Test Room=Threshold          | p=0.000<br>F=53.80     | p=0.000<br>F=55.35     | p=0.000<br>F=141.99    |
| Test Leader=Threshold        | p=0.038<br>F=4.32      | p=0.035<br>F=4.44      | p=0.000<br>F=11.40     |
| Observations                 | 348                    | 348                    | 348                    |
| Groups ID                    | 87                     | 87                     | 87                     |
| R2                           | 0.2716                 | 0.2942                 | 0.7843                 |
| Group FE                     | No                     | No                     | Yes                    |
| Session FE                   | Yes                    | Yes                    | No                     |
| Branch composition           | No                     | Yes                    | No                     |

*Notes:* Robust standard errors. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level. Unit of observation is the total contribution of a group in a given game. Branch composition is a variable with the share of participants belonging to the Arada branch.

Table 5: **Leader behavior**

|   | (1)                  | (2)                   | (3)                 | (4)                   | (5)                      | (6)                   |
|---|----------------------|-----------------------|---------------------|-----------------------|--------------------------|-----------------------|
|   | <b>Group Level</b>   |                       |                     |                       | <b>Player Level</b>      |                       |
|   | Any punishment       |                       | Number punished     |                       | Player punished          |                       |
| Group contribution                      | 0.006486<br>(0.0096) | 0.0002284<br>(0.0114) | 0.01290<br>(0.0207) | -0.007984<br>(0.0278) |                          |                       |
| Std. deviation group contributions      |                      | 0.08628<br>(0.0600)   |                     | 0.2770 *<br>(0.1400)  |                          |                       |
| Individual contribution                 |                      |                       |                     |                       | -0.04905 ***<br>(0.0145) | -0.009554<br>(0.0169) |
| % deviation from max. group contributor |                      |                       |                     |                       |                          | 0.3998*<br>(0.1886)   |
| Observations                            | 87                   | 85                    | 87                  | 85                    | 417                      | 417                   |
| R2                                      | 0.1540               | 0.1620                | 0.1794              | 0.2248                | 0.1438                   | 0.1630                |
| Session FE                              | Yes                  | Yes                   | Yes                 | Yes                   | Yes                      | Yes                   |

*Notes:* Robust standard errors, cluster session. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level. For columns (1) to (4) the unit of observation is a group, the dependent variable is the behavior of the leader randomly and anonymously allocated to this group. For columns (5) and (6) the unit of observation is a player.

Table A1: Descriptives of the experiment participants by branch

| Variable                  | Addis Ketema |                  | Arada |                  | T-test Difference<br>(1)-(2) |
|---------------------------|--------------|------------------|-------|------------------|------------------------------|
|                           | N            | Mean/SE          | N     | Mean/SE          |                              |
| <b>Water services</b>     |              |                  |       |                  |                              |
| Line installation         | 181          | 0.160<br>(0.027) | 181   | 0.166<br>(0.028) | -0.006                       |
| Non revenue water         | 181          | 0.083<br>(0.021) | 181   | 0.039<br>(0.014) | 0.044*                       |
| Water Customer Service    | 181          | 0.276<br>(0.033) | 181   | 0.232<br>(0.031) | 0.044                        |
| <b>Sewerage</b>           |              |                  |       |                  |                              |
| Line installation         | 181          | 0.072<br>(0.019) | 181   | 0.088<br>(0.021) | -0.017                       |
| Sewer connection          | 181          | 0.006<br>(0.006) | 181   | 0.022<br>(0.011) | -0.017                       |
| Sewerage Customer service | 181          | 0.188<br>(0.029) | 181   | 0.232<br>(0.031) | -0.044                       |
| <b>Support department</b> |              |                  |       |                  |                              |
| Human resources           | 181          | 0.011<br>(0.008) | 181   | 0.017<br>(0.010) | -0.006                       |
| Finance                   | 181          | 0.022<br>(0.011) | 181   | 0.033<br>(0.013) | -0.011                       |
| Procurement               | 181          | 0.011<br>(0.008) | 181   | 0.017<br>(0.010) | -0.006                       |
| General service           | 181          | 0.061<br>(0.018) | 181   | 0.072<br>(0.019) | -0.011                       |
| Planning and Budgeting    | 181          | 0.017<br>(0.010) | 181   | 0.000<br>(0.000) | 0.017*                       |
| Monitoring and Evaluation | 181          | 0.006<br>(0.006) | 181   | 0.011<br>(0.008) | -0.006                       |
| <b>Non specified</b>      |              |                  |       |                  |                              |
| Other                     | 181          | 0.088<br>(0.021) | 181   | 0.072<br>(0.019) | 0.017                        |

*Notes:* Descriptive statistics are presented for the participants (87.3%) that participated in a parallel employee survey with detailed information on job history. The value displayed for t-tests are the differences in the means across the groups. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level.

Table A2: Leaders contributions in the games

|              | (1)<br>Room            | (2)<br>Table           | (3)<br>Threshold       | (4)<br>Leader          |
|--------------|------------------------|------------------------|------------------------|------------------------|
| Leader       | 0.6045 *<br>(0.3030)   | 0.7666 *<br>(0.3742)   | 0.1301<br>(0.3742)     | 0.1557<br>(0.4792)     |
| Constant     | 3.7113 ***<br>(0.0618) | 4.2658 ***<br>(0.0763) | 5.1725 ***<br>(0.0763) | 5.7069 ***<br>(0.0977) |
| Session FE   | Yes                    | Yes                    | Yes                    | Yes                    |
| Observations | 417                    | 417                    | 417                    | 417                    |
| R2           | 0.0859                 | 0.0894                 | 0.0667                 | 0.0538                 |

*Notes:* Robust standard errors. \*\*\*, \*\*, and \* indicate significance at the 1, 5, and 10 percent critical level. The unit of observation is a player.

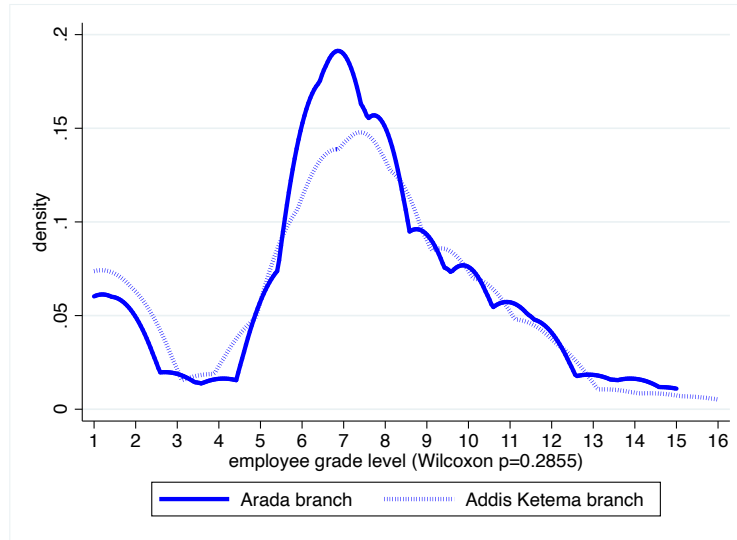


Figure A1: Distributions of employee grades in each branch.

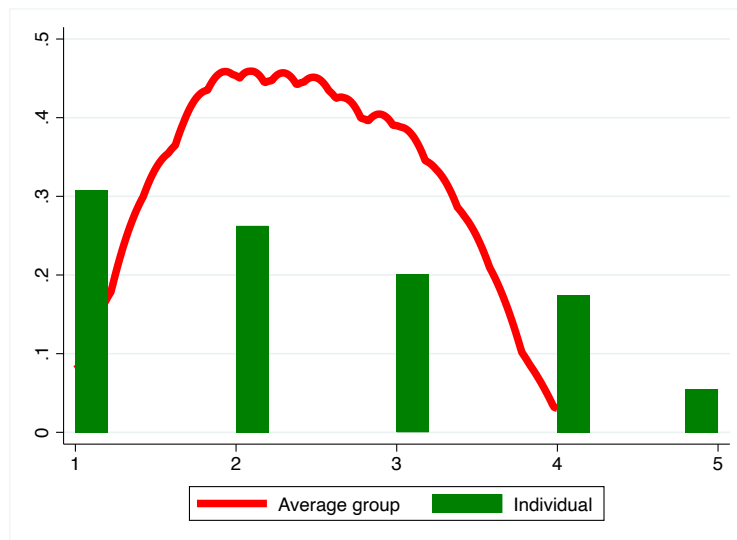


Figure A2: Knowledge of other participants in the group.

The histogram shows the answer of the participants to the post-experimental survey question *"Including yourself, how many participants in your group do you know?"*. The line represents the kernel density of the average of this question by group.



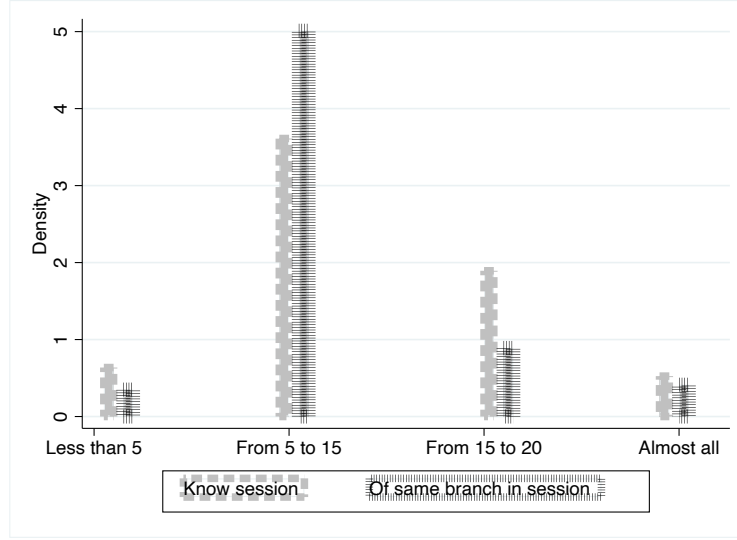


Figure A3: Knowledge of other participants in the session.

The histogram shows the answer of the participants to the post-experimental survey question *"How many of the participants in the session do you know?"* (lighter bar) and *"How many participants in the session work at the same AAWSA branch?"* (darker bar).

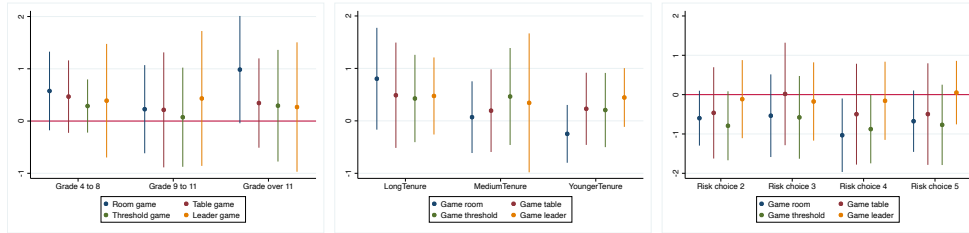


Figure A4: Heterogeneity of individual contributions.

The figures show coefficients of estimating  $C_{igt} = \alpha + \beta_i * \sum_i Variable_i + \gamma_t + \epsilon_{igt}$ , where  $Variable_i$  is a set of dummies on the individual characteristic of interest, and  $C_{igt}$  is contribution of player  $i$  in group  $t$  in game  $g$ . Standard errors clustered at the session level. Tenure dummies are defined as per tenure quartiles being the reference the shortest tenure quartile. Risk choices come from post-experimental risk-elicitation game, being the reference the safest alternative offered (choice 1), and with alternatives increasing in risk.

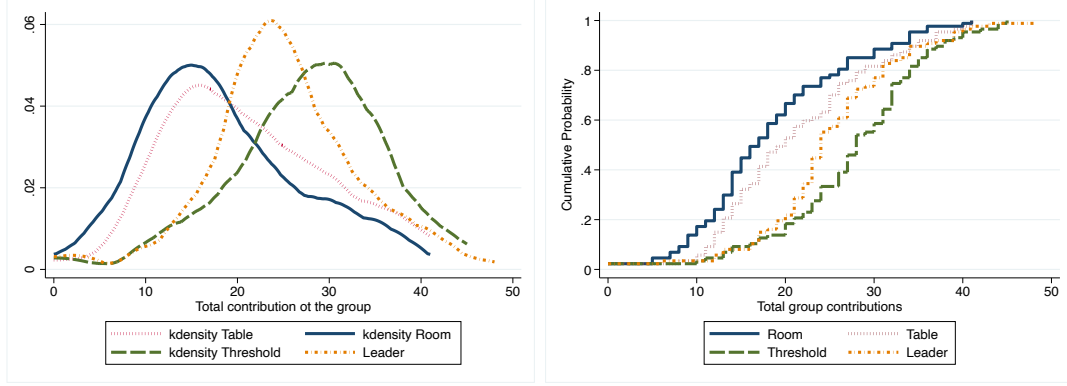


Figure A5: Distribution of group aggregate contributions.

The right figure shows the distribution of total group contributions (sum of contributions of all members in the group) at the four different games. The left figure shows the cumulative distribution of the groups totals. The four distributions are pairwise significantly different (Wilcoxon distribution test,  $p=0.000$ ).

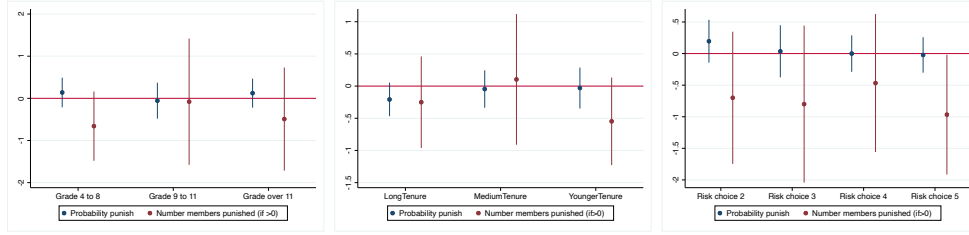


Figure A6: Heterogeneity of leader's likelihood of punishment

The figures show coefficients of estimating  $P_{is} = \alpha + \beta_i * \sum_i Variable_i + \epsilon_{is}$ , where  $Variable_i$  is a set of dummies on the individual characteristic of interest, and  $P_{igt}$  is choice of leader  $i$  in session  $s$ . Standard errors clustered at the session level. Tenure dummies are defined as per tenure quartiles being the reference the shortest tenure quartile. Risk choices come from post-experimental risk-elicitation game, being the reference the safest alternative offered (choice 1), and with alternatives increasing in risk.

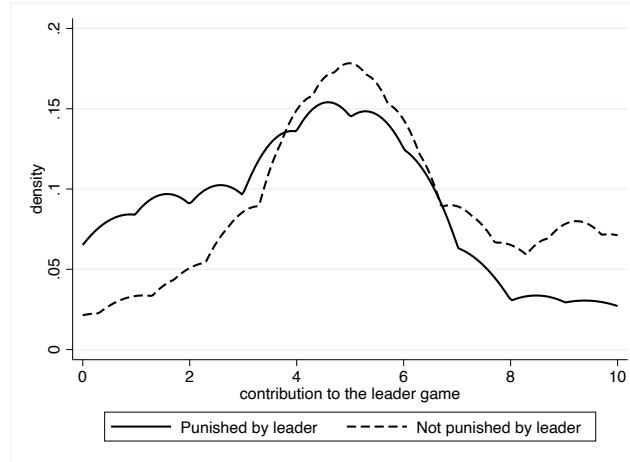


Figure A7: Probability of punishment as function of contribution

The Figure shows the distribution of contributions in the leader game for players punished and not punished by the leader. The pattern is similar to the one showed in Figure 3 for levels instead of deviation from main contributor in the group.

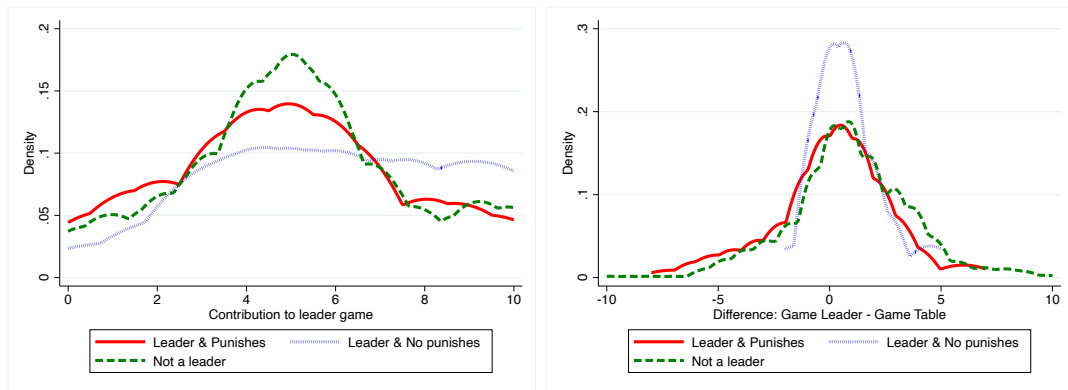


Figure A8: Behavior of leaders compared to other participants.

The figures show the contribution to the leader game and the difference in contribution between the leader and the table games for non-leaders, and among the leaders for the ones that decide to punish and not to punish separately.

## APPENDIX A: LAB-IN-THE-FIELD EXPERIMENTAL PROTOCOL

This Appendix presents the instructions for the experimental protocol. Instructions were provided orally in Amharic by the experiment director, with help of the experimental materials (slides projected).

### *1.A. Before the start of the session:*

Prepare the experimental room and materials: In the room there need to be 6 tables with 5 persons around each table. Each table is identified with a number (groups 1 to 5) and each participant is identified by a letter (player A to E).

In front of each participant there is:

- Consent form and a pen.
- Four pots of different colors with the player ID written on them. Inside each box there need to be 10 tokens of the color of the participant and some white tokens.
- For each group, a wooden box with some transparent tokens (to ensure noise when moved) and a whole to allow players to introduce tokens in the box without seeing the content.

### *1.B. General structure of the session:*

1. Welcome and consent
2. Presentation of experimental materials and payment determination
3. Public good games: Standard (room and table), and with random order with threshold and with leader
4. Post-experimental survey
5. Risk-elicitation game and payment

\*\*\*\*\*

#### *Part 1: Welcome and consent*

Good morning /Good afternoon. We are researchers from the World Bank and from Toulouse School of Economics. First of all, we will like to thank you for voluntarily assisting to this activity.

Today we will do a decision-making activity. We will explain each step of every activity, and we will make all the pauses needed to answer all the questions you may have. We will spend about one hour between explaining the exercise, playing it and finishing with a short survey at the exit.

Before we start, we want to explain three very important points (*write the 3 points at blackboard as they are presented*).

Anonymity: All the decisions and all the answers will be completely anonymous. We will never tell anyone what you did or answered. Each one of you got an ID at entry: from this moment on all of what you do will be registered to that ID, never with your name; therefore, there will be no connection between your name and the decisions you take.

Privacy: All participants will avoid turning around and watch the movements of other participants in the activity or to speak to them.

Yours decisions are worth money: During this morning/afternoon, you will have to make decisions and you will win points. Be a little patient, we will explain later how you can win points. The important thing for you to know is that each point is worth 1 BIR.

Clear for everyone?

Your participation is completely voluntary. You may leave the room at any time you want. Unfortunately, if you decide to leave the activity before the end of the session, we won't be able to calculate your points and, therefore, your payment. The payments will be made at the end of the session after you answer a short questionnaire.

You will be playing in groups of 5 participants. Hence, if you leave the room your team will need to stop playing: we will compensate with YYY Bir the rest of participants.

Clear for everyone? Any questions?

You have in front of you a form that clarifies all the points I just presented. Let us read it together. If you agree to participate, please sign the form and give it to my colleague. Thank you. (*collect consent forms*).

Perfect, thank you so much. Now we are ready to start the activity.

You have in front of you four pots of different colors with your ID number written on them. True for everyone? You all have the 4 pots? Great. So let's take the pink pot. Can you please all take the tokens and count them? Everyone has

10 black tokens of your color and a few white ones ?

Perfect, thank you.

So now we will start a sequence of 4 games, and for each of them we will use one of the four pots. We start with the pink one.

## *Part 2: Description of the games*

### *Game A and Game B : Standard public good game*

Note: This game will have two versions. The order of the versions will be randomized across sessions.

- Version 1: Group are the 5 persons sitting together.
- Version 2: Group are 5 persons in the room but not necessarily the ones sitting together. Since will be 5 groups of 5 participants, we will construct the alternative group by putting together all participants with same letter but different group number.

Ok, now that everyone has his space and the materials let me present the first game. For this game we will use the 10 black tokens of the pink box.

Why do you have some white tokens in the pot also? This is to make sure that the box always makes some noise, so that nobody can guess your decisions. The white tokens have ZERO value (*write on a white board that white tokens are worth zero BIRR and that black tokens are worth one BIRR to make sure is at the participants view during the whole session*).

As you can see, you are sitting in groups of 5 persons. You all have the same number, that determines the group, and different colors. Is everyone sitting in groups of 5?

Version 1: You are playing with the group you are sitting together with numbers of same color as you // Version 2: You are playing with 4 other persons in this room, not necessarily the ones having same color as you.

(Wait for answer) Great !

Ok, so the 10 black tokens are yours. In this game, you just have to take one decision: how to split these 10 tokens between yourself and the group magic box.

You can allocate any number of tokens you want between the two alternatives.

The first alternative is keep the tokens for yourself – and remember, at the end of the game each token will be exchanged for 1 BIR.

The second alternative is to put tokens in the magic box. There they will be combined with all the tokens allocated to the common box by everyone in your group. (*show the magic box/bag for each group*). That is, the total number of tokens in the box will be the sum of the tokens each person in your group decides to put there. We will then take the total number of tokens there, double the total, and then divide it equally among the 5 people in your group. That means that each person in your group will earn the exact same number of tokens from the box regardless of how much each person has contributed. Similarly, you will receive tokens back from the box regardless of whether you put tokens into the box or not. Is it clear for everyone?

Let me put an example – Visual materials in Figures F1 - F2.

Clear for everyone? Ok.

Now we ask you to make your choice for this game: Keep on your box the black tokens you want to keep for yourself, and put in the magic box – that the assistant is passing - the ones you decide to allocate to the common box. Remember: what you keep is yours, what goes to the magic box is multiplied and shared equally among group participants.

One detail before playing for the first time: The magic box has some white tokens with no value, and the assistant will make sure they make noise when collecting everyone's contribution. Like this, nobody from the noise can infer anything about your contribution and your choices remain totally private.

Everyone agrees? Great! Then let's make the choice. Keep in your hand what you want to put in your group magic box and the assistant is passing to collect. (*Assistant: make sure to make noise with white tokens inside the box so that nobody can hear noise of new tokens*)

*Game C: Public good game with a threshold.*

Thank you so much. Now we proceed to play the third game, and to do so please take the green box.

Everyone has the green box at hand? Everyone is counting the black tokens inside the box, and has 10 ? Perfect, so now let me explain the next game.

Your decision in this game is the same as in the first ones: how to split your 10 black tokens between your box and the common box. But now there is a difference: tokens in the common box will not be multiplied and will be lost if the

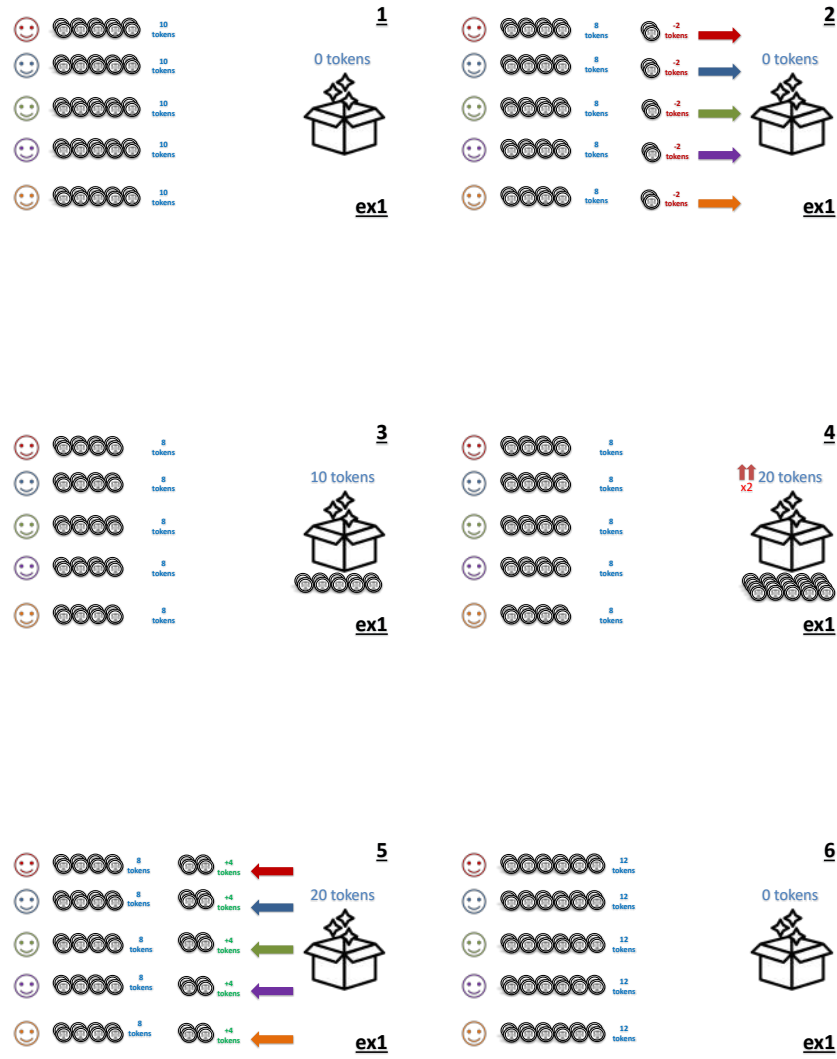


Figure F1: Experimental materials - Example 1/4

The Figure presents 1/4 examples presented for the understanding of the functioning of the magic box.



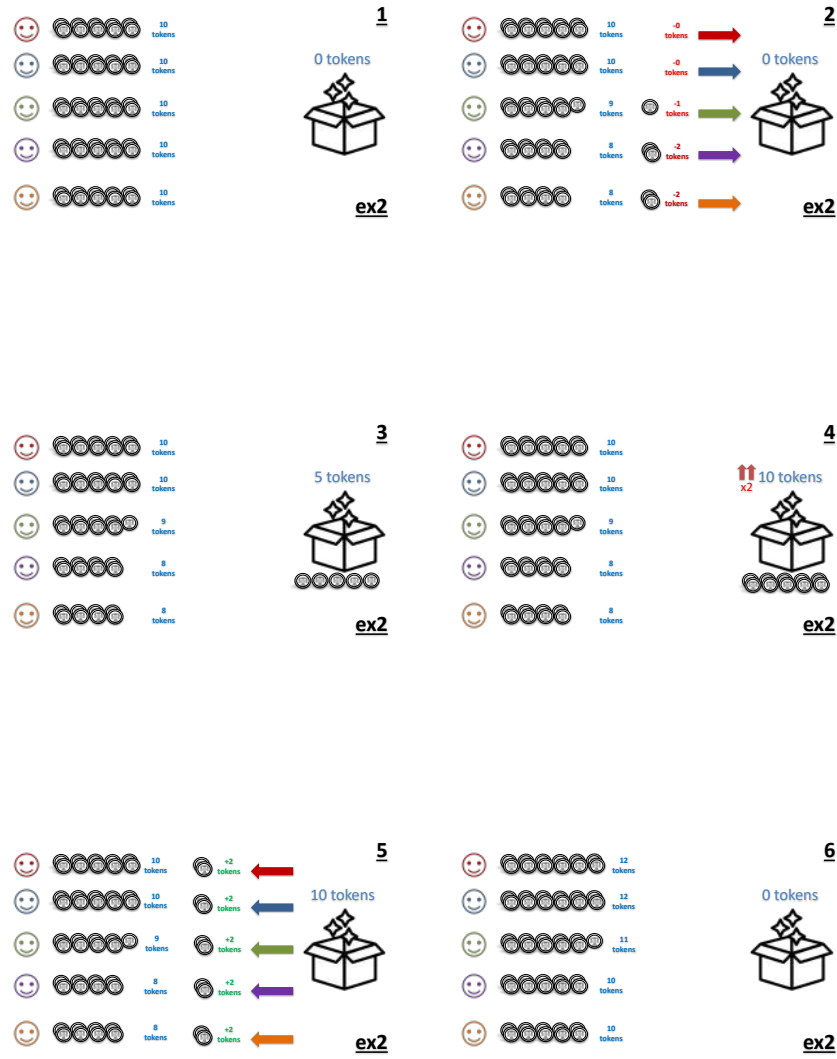


Figure F2: Experimental materials - Example 2/4

The Figure presents 2/4 examples presented for the understanding of the functioning of the magic box.

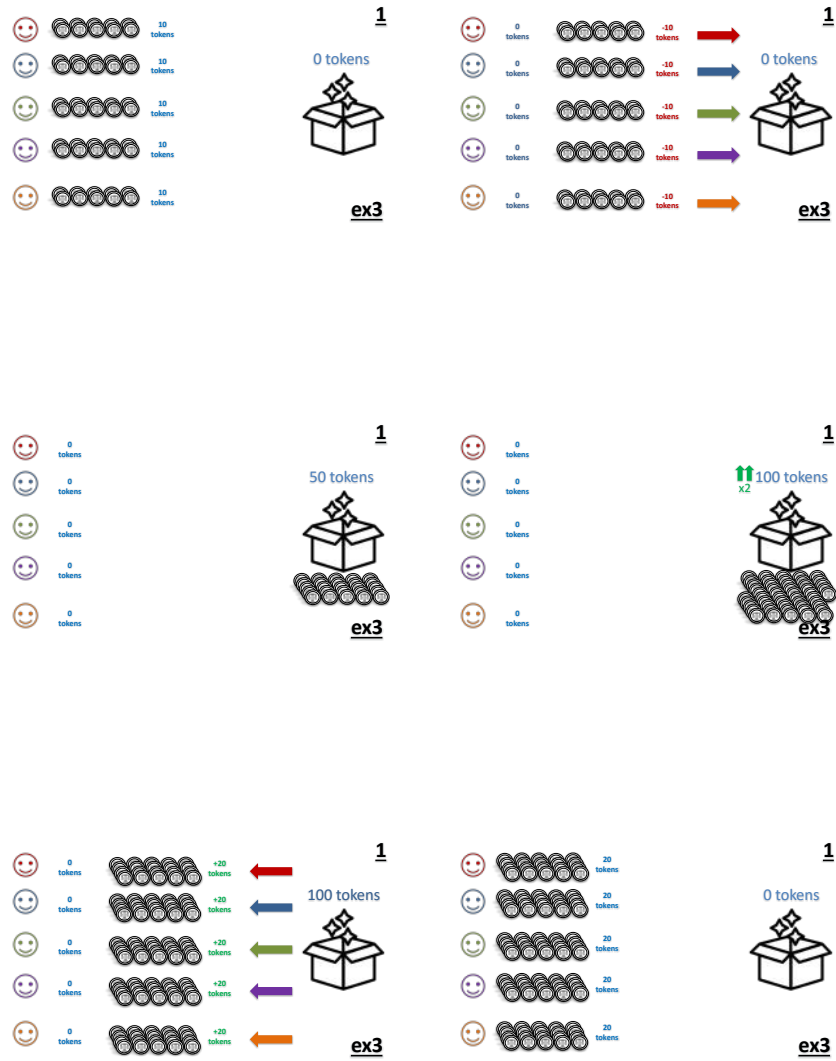


Figure F3: Experimental materials - Example 3/4

The Figure presents 3/4 examples presented for the understanding of the functioning of the magic box.

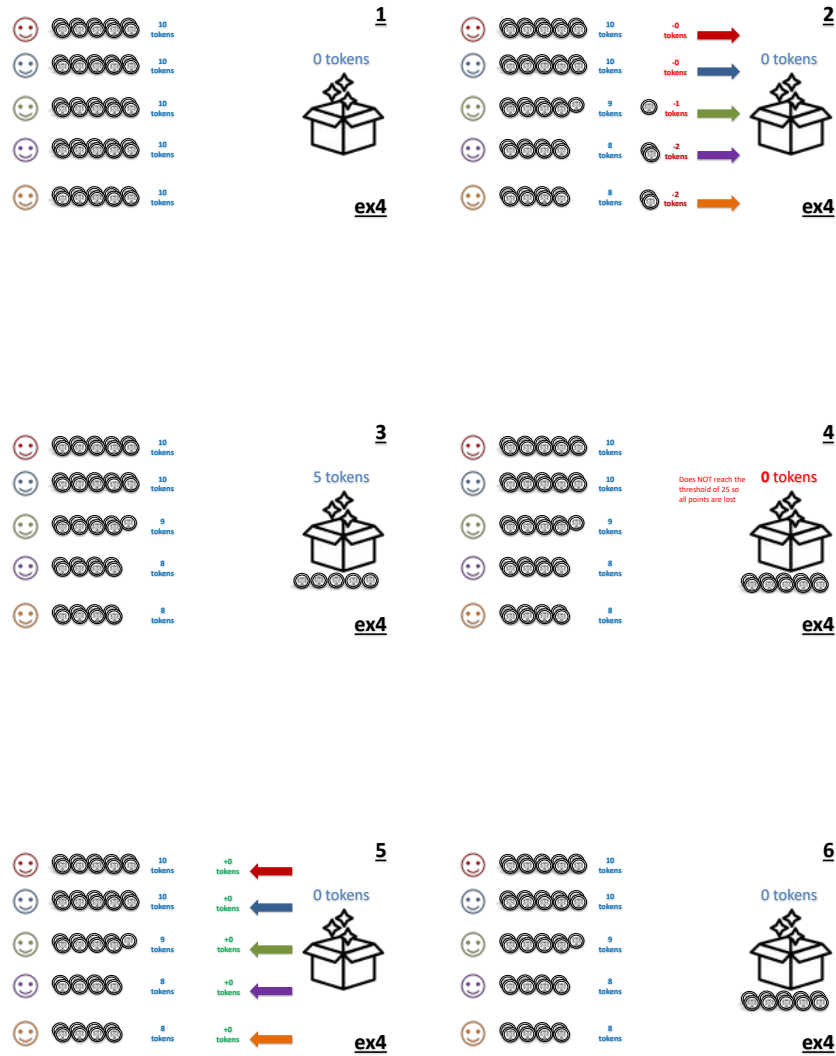


Figure F4: Experimental materials - Example 4/4

The Figure presents 4/4 examples presented for the understanding of the functioning of the magic box.

total contribution from the group to the box is not over 25 points.

For example: Figures F3 - F4.

Is it clear? Good ! Then, let's take the decision for this game. Assistants will pass to collect the common boxes and the individual boxes.

*Game D: Public good game with third party punishment*

Thank you so much. Ok, now let's move to the last game. Please take the red box you have in front.

Before playing the game, that will have the same structure of the first one played, let us introduce the "Game leader". The game leader is one participant – now we will talk about who will have the leader role in a few minutes – that has the role to check the amounts contributed and decide whether to punish participants as function of their contribution.

So, what we will do now is first play the game as we did in the first game: everyone contributes, what is put in the common box is doubled and shared equally among the group members.

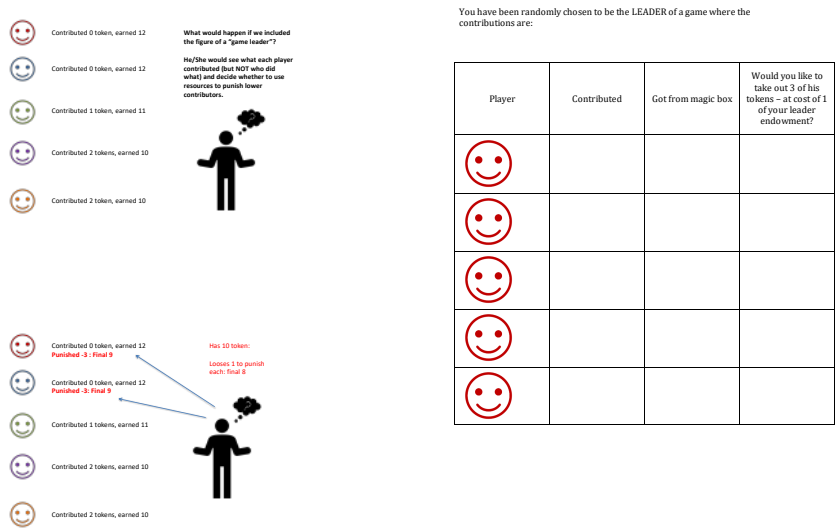


Figure F5: Experimental materials - Leader example

The Figure presents the example of leader and the form given to the leaders to make their choice.

After that has happened, we will call the leader of your group and show him the amount each participant (including himself if he is in that group) did put in the common box. Leader is randomly chosen among the room participants and revealed after all games are played – to ask him to take the decision. But the leader will not come to know of your identity. We will give the leader 10 tokens (See Figure F5). After the leader has seen the amount that you put in the project, he has an option to reduce the income of any of the players. If the leader decides to reduce the income of a player, he has to spend one token, but the player loses 3 tokens. However, if the leader decides not to reduce, he gets to keep all 10 tokens in his pocket and the players don't lose any income either. Your final income from the game will be calculated after subtracting any amount that was reduced by the leader. (*Explanation with visual material*)

Please note that you will not come to know if the leader reduced your income or not. We will not tell you this. We will now give you two examples to show how this works.

OK. Since we do not know who the leader will be, let's think on which is the leader's choice. So let's imagine the contributions of players are – use visual material. Then, the leader sees that. I will give the leader 10 tokens and he may use this money to reduce one or more than one player's income.

What happens if the leader says he does not want to reduce? The money goes to his pocket.

Let us now see what happens if the leader reduces players' income. Example, if leader decides to 'punish' each player by one token – then, he needs to spend 4 tokens, and players lose 3 tokens each. How much does each player earn? Will earn the share from the common box ( $15 \times 2 = 30/5 = 6$ ), minus the 3 points of punishment, plus the tokens in their personal box. And the leader? The 10 tokens minus the 4 used to punish the four players.

One very important thing: before you put any money in the project, please think if the leader will reduce your income because of the money you put in the project or not. We will ask you on this.

Ok, now time to tell you who will have the leader role. For each of the groups, we will pick from this bag where we have all your name tags one of them. This will be the person playing the leader role in your group. Note that you will not know the leader until all games are played and the leader is selected to privately make his/her choice.

*Part 3: Instructions for the leader*

We hope that you have understood the game. In this game you have to take a decision on whether or not you would like to reduce any one or both player's income. This decision is entirely yours and we will not help you on this. You are not obliged to reduce. We will not inform anyone of the decision you take.

You will receive 10 tokens from us, which you may use to reduce a player's income.

In case you decide to reduce a player's income, you will have to spend 1 token for each reduction, but the player's income will be reduced by 3 tokens.

Have you understood the game? Would you like to play a mock round?

Now, we would like to show you the contribution decisions of four members from your group (See Figure F5) Can you tell me where are the payoffs of player D on this display? How much did player D put in the project?

Good. Here is your endowment of 10 tokens. You may now take a reduction decision nothing on the paper.

*Part 4: Post-experimental questionnaire and thanks.*

Thank you so much! That was the last box and the last game. Now we will proceed to call the participants that got the role of leader to ask them to make their choices.

In the meantime, while we do that and calculate your individual payments, we will ask you to please fill in a small questionnaire. The payments will be distributed after lunch in closed envelopes with information on the game chosen for payment and where the points you earned come from.