Political Institutions and Corruption: An Experimental Examination of the "Right to Recall"

Sarah Mansour, Vjollca Sadiraj and Sally Wallace

Abstract

Countries around the world are concerned with corruption as it potentially undermines confidence in government and may reduce the efficiency of public goods provision. While there has been a significant amount of research devoted to identifying the causes of corruption there has been little empirical research on the impact of political institutions on corruption. Given that many nascent governments are establishing new political systems, the time is right for understanding the role that political institutions may play in enhancing or mitigating corruption. This paper uses a series of laboratory experiments to examine the impact of the 'right to recall government officials' on the level of government corruption. We find experimental evidence suggesting that such an institution *can decrease* the level of corruption in government through the increased accountability it imposes on elected politicians, and equity of the system, in terms of income distribution, may also be enhanced.

Keywords: Political economy, corruption, transition economies, experiment, public goods

POLITICAL INSTITUTIONS AND CORRUPTION:

AN EXPERIMENTAL EXAMINATION OF THE "RIGHT TO RECALL"

"There can be no doubt, that if power is granted to a body of men, called representatives, they like any other men will use their power not for the advantage of the community but for their own advantage, *if they can*."

James Mill (1773-1836)

The World Bank (2004) considers corruption to be "…the single greatest obstacle to economic and social development." There are many definitions and versions of corruption, some blatantly illegal, others more nuanced and sometimes even legal. We will confine ourselves in this paper to the definition used by the World Bank (The World Bank 1997) and Bardhan (1997, p.1321) which see corruption as the use of public office for private gains, where an official assigned the authority of making decisions for the group abuses it by delivering decisions that lead to private enrichment.

There is by now an extensive literature that establishes strong association among corruption, the inefficient allocation of public goods, and growth-related outcomes (Bai and Wei 2000; Burki and Perry 1998; Glynn et al. 1997; Kaufmann and Zoido-Lobaton 1999; Mauro 1995).[1] In particular, corruption has been identified as a major source of government failure in public good provision.[2] Understanding what factors breed corruption and whether they can be altered is an important line of inquiry. While there is substantial theoretical literature linking corruption to the type of political institution (Kunicová and Rose-Ackerman 2005; Persson et al. 1997), the empirical literature is limited. This is especially true for economies in transition, which may be able to control corruption to some degree by adopting particular political institutions.

Our study contributes to the emerging empirical literature on the determinants of government corruption with particular attention devoted to the role of citizens' right to recall officials—a political institution that has not been rigorously examined in the literature.[3] The idea is that the right to recall offers a political channel that may increase political accountability of officials compared to those officials who have a fixed term in office. The threat of recall, say due to dissatisfaction with the official's rent-seeking behavior, may reduce the incidence of such behavior. In this context, this paper makes several specific contributions. First, we take a novel approach to understanding the role that the right to recall plays in deterring/enhancing corruption in a controlled environment through a series of laboratory experiments. Secondly, we are the first to study strategic interactions in a stylized game that integrates several games associated with public sector-citizen interaction, previously studied in isolation, such as public good games, tax compliance games, and recall elections. Thirdly, our findings add to the literature on equilibrium selection and behavior in repeated games. A unique feature of our study is conducting experiments in Egypt (with Egyptian students) while the country was experiencing political turmoil; In Egypt the right to recall rulers has been practiced twice in less than three years (removing Mubarak in 2011 and Morsi in July 2013) and in both cases, corruption charges were among the demands of the protesters.[4]

Ex ante, we expect the right to recall an official while in service to be more conducive to the socially intended functioning of officials than simply providing an unchallenged fixed term of service. We test this hypothesis in our paper. Our experimental methodology is particularly relevant given the difficulty of collecting observational data on such events.[5] Specifically, we (i) model corruption that manifests itself through inefficient provision of public goods, (ii) identify two institutions, scheduled replacements and recall-enabled replacements, that may have

different impacts on limiting corruption, (iii) offer an equilibrium analysis of the two and (iv) test the empirical performance of such institutions in the lab.

The need for empirical testing is of a great importance as theoretically the recall option can be predicted to limit corruption, or not, depending on the strategies employed by players; subgame perfect equilibrium predicts no recall-effect on corruption but there are other equilibria in which recall-enabled replacements are more effective in hindering corruption than scheduled replacements. Specifically, we simulate a familiar interaction between public officials and the citizenry (through a tax experiment) to investigate how scheduled replacements of officials versus recall-enabled replacements affect the decisions of officials in charge of a public good provision. In our experiment, subjects earn money by performing a labor task and pay taxes according to their claimed income and face a given probability of getting audited and penalty schedule. Tax proceeds are used to fund a public good that is chosen by the group official. The official has the choice to fund a self-serving (and inefficient in equilibrium) public good or refrain the self-serving behavior by funding a public good that benefits everyone equally and is more efficient (in equilibrium). The frequency of the self-serving public good choice is a stylized measure of corruption. Our data suggest that the recall-enabled replacements of officials has an important impact on limiting corruption behavior but the resulting rate of turnover among officials is high which warrants awareness on other costs of this institution.[6]

The paper proceeds as follows. In the next section, we present the literature review. The theoretical construct and derivation of hypotheses and experimental design are presented in the following sections. A fifth section presents the empirical results and a final section concludes.

Literature Review

Civilized societies have long been known for delegating the power to tax and to provide public goods to representatives. A large fraction of public spending, however, is not devoted to useful public projects, but rather to support projects of self-interested officials and other pork-barrel projects. The political process has been recognized by the economic literature to be a major factor behind this inefficient provision of public goods (Lizzeri and Persico 2001). Representative democracies have been hailed as providing accountability between elected leaders and those who elect. However, democratically elected leaders are not immune to corruption as evidenced in countries as different as Zimbabwe, Venezuela, and the U.S. Accountability of elected officials through the ballot box is thought to curtail the use of public resources for personal gain, but to date, there has been limited empirical analysis of this conjecture.

Public officials are entrusted with decision-making functions, the provision of public goods to the community being an important one. Corruption can be manifested as an allocation of public funds to provision of public goods that provide more benefit to officials at the cost of benefit to the general citizenry. In this paper we focus on corruption as a phenomenon that involves public officials, other citizens, and the allocation of public-sector goods.[7]

The argument that different institutional frameworks can affect levels of corruption follows from opportunities, constraints, and incentives these frameworks provide on strategic interactions among involved agents. One such angle is the "career concerns" approach of Hölmstrom (1999). In this regard, there are numerous studies that argue that leaders who are not eligible for re-election act differently than those who are (Alt et al. 2009; Besley and Case 2003; Besley and Case 1995; Ferraz and Finan 2011). For instance, Alt et al. (2009) find that economic

growth is higher and taxes, spending, and borrowing costs are lower under reelection-eligible incumbents than under term-limited incumbents. Ferraz and Finan (2011) show that, in Brazil, electoral rules that enhance political accountability play a crucial role in constraining politician's corrupt behavior.

Lizzeri and Persico (2001) compare the composition of government spending under alternative electoral rules through a political-economy model where the provision of a public good is determined by the electoral incentives of office seeking candidates. When candidates have the option of redistributing resources, public goods will be underprovided relative to the efficient outcome because benefits from the public goods cannot be easily targeted to groups of voters. In the same context, Persson and Tabellini (1999) construct a model of redistributive politics in which a majoritarian system generates less public good provision than a proportional system. Using cross-country data, they find weak support for this prediction. Our paper contributes to this literature by proposing the absence of the right to recall as an explanatory variable for inefficiency of public good provision by government officials.

Concerning the emerging literature on laboratory experiments of corruption[8], Abbink et al. (2002) and Abbink et al. (2000) introduce reciprocity games that mimic situations where corruption arises. Specifically, they separate the influences of the three main characteristics of corruption, namely (i) reciprocity relationships between bribers and public officials, (ii) negative welfare effects, and (iii) high penalties when discovered, in their experiment. In their study, it is a third agent, a sudden death treatment, who may punish corrupt behavior by others. On the contrary, in our study punishment of corrupt behavior can be carried out by the citizens via the recall of the official and/or tax compliance, neither explicitly tested via a political process in Abbink et al. (2002).

Other research assumes that policymakers would act in the interests of those whom they represent simply because of the responsibility to do so by virtue of having been chosen to make decisions for others. Drazen and Ozbay (2014) present experimental evidence that policies chosen by leaders depend on whether they were elected or appointed. They find that elected leaders are significantly more likely to choose a policy not equal to their "type" than leaders who are appointed.

No study—to the best knowledge of the authors—directly examined the effect of citizens' right to recall an official during his/her term in office on corrupt allocations of the public good. This paper bridges a gap within the literature using controlled environments and experimental methods to isolate effects of variables of interest (opportunity to recall in our study) on behavior of decision-makers (prevalence of corruption in our study). We use a series of laboratory experiments to capture the reactions of "citizens" to "officials' behavior," in different political settings, through a stylized official-citizen interaction over tax compliance.

We turn next to the theoretical analysis of our stylized model of corruption in the presence and absence of the recall option.

Theoretical Analysis and Derivation of Hypotheses

To elicit the relationship between the form of political system and corruption, we model it through a natural interaction between citizens and government—that of tax payment and the provision of a public good: all players (citizens and the official) make decisions regarding their tax compliance and the official decides how to use the taxpayer dollars. In the no-Recall game (noR-game) the official is randomly chosen and sits as the incumbent for one fixed term (with a known duration). In the Recall game (R-game) the citizens are allowed to recall the official after

they have observed his/her decision on how to spend the tax proceeds. If a recall is voted for, a new official is chosen among the citizens. To disentangle the effect of recall-enabled replacement on official's behavior (from the election effect), in both games officials are exogenously selected; they can be thrown out of office in the R-game but not in the noR-game.[9]

We model the interaction between citizens and officials with the following sequential stage game. In both political settings, players report their earned income which is used to determine income tax liability. It is common knowledge that with some probability, p_a, any player can get audited. An audited player pays tax on his actual earned (not on declared) income plus a fine; the fine is a convex and increasing function of unreported income. A player who is not audited pays according to his declared (not necessarily true) income, *x*. Tax proceeds[10] are used to finance one of two available public goods; the G-good which favors the official at the expense of other citizens or the C-good, which benefits all players equally. The official makes the decision which public good to fund. Valuation of public goods across players is common knowledge. The valuation of the C-good is identical for citizens and the official whereas the valuation of the G-good is higher than the value of the C-good for the official but lower for the citizens. If we let β_i^j denote the marginal per capita return (mpcr) of *j*-good to *i*-player, valuation of public goods G and C across players is captured by the following set of inequalities,[11]

$$\min\{\beta_o^G, 1\} > \beta_o^C = \beta_c^C > \beta_c^G \geq 1/(n-1)$$

where n is the number of players; subscripts are used for player's type (c for the citizen and o for the official) and superscripts for the type of public goods (G for the G-good and C for the C-good). The social dilemma follows from the lower bound 1/(n-1) and the upper bound 1 on citizen's mpcrs. The official's decision on which public good to fund is made known to all and

payoffs are realized. In the R-game (but not in the noR-game) the stage game continues with the citizens voting on whether to recall the official.

An official who uses office for private benefits would choose to fund the G-good as own return from the G-good (β_o^G) is higher than the return from the C-good (β_o^C) although the funding of the G-good is less preferred by the citizens. This captures in a stylized way some version of legal corruption. The frequency of the G-good being funded will be one of the measures of corruption. To measure the effect of the recall option on economic efficiency and fairness of redistribution of tax proceeds through public good provision we will look at the common measure of efficiency (the ratio between the realized group payoff and the maximum feasible group payoff) and payoff equity (Gini index of the distribution of payoffs) across the two games (R-game and noR-game).

The main question of interest is whether recall-enabled rather than scheduled replacement of officials is a more effective institution in hindering corruption. There is no a priori clear yes/no answer to this question as non-corrupt officials can also be thrown out of office if craving for political power is widespread among voters; if so a recall-enabled institution offers little incentive to officials to behave as socially intended.[12] Theoretically, the level of corruption is expected to be the same in both games if one appeals to subgame perfect equilibria (SPE). However, in other Nash equilibria (not SPE) with players using "maxmin" actions (to punish corruption and tax evasion) out of the equilibrium path, the predictions are more/less corruption in the R-game than in the noR-game depending on whether recalling the official is part of the strategy profile when the game approaches the end. The intuition behind this result is that the official in either game funds the C-good as long as the instantaneous benefits from corruption (funding of the G-good) are smaller than future losses that occur as a result of the corruption

triggering low compliance in both games and recalls in the R-game. But while the instantaneous benefits are the same across the two games the future losses differ as the likelihood of being in the office (and therefore expected payoffs) after funding the G-good are different across the two games. In the two following sections we state the equilibrium analysis of the two games (details in Appendix 1) and use laboratory data to obtain further insights on the empirical validity of theoretical hypotheses.

The following notation will be used: w is the individual's income, τ is the tax rate, p_a and $f(.)$ are the auditing probability and the fine (a convex increasing function) on unreported income. If the likelihood that G-good is funded is p^G then the expected mpcr of the public good to individual i is: $E_i(\beta|p^G) = \beta_i^G p^G + \beta_i^C(1-p^G)$. Letting x_{-i} denote the vector of declared income by others, player i's expected payoff in the stage game from reporting x_i (when the real income is w)[13] is

$$E(\pi_i(x_i, x_{-i}, p^G)) = w - y_i - p_a f(w - x_i) + (T_{-i} + y_i)E_i(\beta|p^G) \qquad (1)$$

where T_{-i} is the expected total tax paid by others and $y_i = \tau(p_a w + (1-p_a)x_i)$ is the expected payment by individual i as income tax.

It follows from the linearity in the public good payoff specification that the optimal declared income is in dominant strategies. But unlike in linear public good games, full free riding (that is, declaring zero income) is not optimal for penalty functions that are sufficiently convex. It follows from partial free-riding (through partial tax compliance) and statement above that in equilibrium under provision of public goods is expected. As the optimal free-riding decreases in the mpcr in our games, the higher the corruption the lower public good provision.

The outcomes of the subgame perfect equilibrium in either game (see Proposition 1.1 and 2.1 in Appendix 1) are underfunding of the G-good which is the only public good being funded; in addition, in the Recall game the officials are always thrown out.

The subgame perfect equilibrium builds on the Nash equilibrium of the stage game. However, in our games players' payoffs in the Nash equilibrium of the stage game are larger than the minmax payoff.[14] Hence, there are Nash equilibria (not SPE) in which players' payoffs are close to any strictly enforceable payoff profile if the game is played long enough. In such equilibrium with grim punishing actions being triggered by corruption or free-riding, there exists an r* such that C-good is funded during the first r* rounds of the game and the G-good is funded during R-r* end rounds, R is the total number of rounds the stage game is played. The number of rounds without corruption, r*, varies with the length of the service term (i.e., the value of R), but the number of end rounds with corruption, (i.e., the value of R-r*) does not. The number of end rounds in which the G-good is funded is determined by the ratio between the instantaneous benefit (the round additional payoff) that the official earns by making self-serving decisions and the future losses (the difference between the payoff he gets by not defecting and the minmax payoff in any following round): the larger the future losses (for e.g., the larger the fraction of voters that engage in punishing the official) the smaller the number of end rounds with corruption, R-r* (See Appendix 1). In the R-game, there are similar Nash equilibria in which the official is recalled if he funds the G-good during the no-corruption rounds. In such equilibria the number of end rounds with corruption is (weakly) smaller than in the noR game if the recall option is not exercised during the end rounds (with corruption). If the recall option is exercised during the end rounds then the recall-enabled replacement institution cannot be superior to the scheduled replacement in delaying corruption. There are many such Nash equilibria; which one

is played out is an empirical question. Nevertheless, theoretical analysis shows that whether recall-enabled replacement (compared to the scheduled replacement) is a better institution in hampering corruption depends on strategies used during the end of the game. This may explain why we see both institutions across countries.

Theoretical predictions of SPE and equilibria with trigger strategies across the two games are summarized in the following proposition. Assume risk neutrality and that players' preferences on the payoff space are represented by (1). Then

1. SPE predict full corruption in both games.
2. There are Nash equilibria with punishing (grim) strategies out of the equilibrium path that predict no corruption in all but the end rounds. The predicted number of end rounds with corruption is:
    a. *lower* in the noR-game than in the R-game for strategy profiles that *recall* officials during the end rounds.
    b. *higher* in the noR-game than in the R-game for strategy profiles that *do not recall* officials during the end rounds (i.e. when recall is exercised only after defections during the non-end rounds).
3. Inefficiency of public good provision increases with corruption.

Proof: See Appendix 1.

Our first hypotheses that follow from the outcomes of SPE stated above and Part 3 of the proposition are:

H1o: Corruption level is the same across the two institutions.

H2o: Inefficiency of public good provision is similar across the two institutions.

In the R-game, "always recall" the official is part of a SPE (see Appendix 1, result P2.1) so we have the third null hypothesis

H3o: The likelihood of a recall does not depend on official's behavior.

The alternative hypotheses that follow from Part 2 and 3 of the Proposition are:

H1a: Corruption level differs across the two games.

H2a: Inefficiency of public good provision is different across the two games.

The one-sided alternative hypotheses for the Nash equilibria of type 2.a (2.b) are higher (lower) corruption and inefficiency of the public good provision in the R-game than in the no-R game.

In the R-game, in (both types 2.a and 2.b) Nash equilibria with trigger strategies, a funding of the G-project by the official during the non-end rounds of the game triggers recalls. Hence, the one-sided alternative hypothesis to H3o is

H3a: Funding of the G-good affects positively the likelihood of recall.

We turn now to an explicit discussion of the experimental design.

Experimental Design

The experimental design that we report here is a 2x1 design implemented across subjects.[15] In both settings, subjects are randomly matched into groups of five at the beginning of the experiment; groups remain fixed during the entire experiment. At the beginning of round 1, an initial official is randomly selected. Subjects earn experimental pounds[16] according to the performance in a simple task of correcting spelling mistakes in the Arabic language. After the earning money task is completed, subjects decide how much income to report; the reported income is taxable at the rate of 25 percent.[17] No taxes are paid on unreported income unless a

subject is audited; an audited subject, in addition to paying taxes on earned income, pays a penalty on any undeclared income determined by a known penalty structure; one out of the five members is randomly selected to be audited.[18] Total taxes paid by all subjects are used to fund one of two feasible public goods as being decided by the official of the group. The mpcr of the C-good is 0.6 for any member of the group, whereas the G-good is valued most by the official (mpcr=1.5) but less so by citizens (mpcr=0.375).

To capture non-excludability and non-rivalry characteristics of a public good, we follow a standard implementation in the experimental literature that distributes some multiple (3 for us) of the total individual investments in the public good (i.e., total tax revenue in our games) among group members. In case of the C-good, tripled tax revenues are equally distributed among group members; in case of the G-good, half of the amount goes to the official whereas the remaining half is distributed equally among the other group members, i.e., the other four citizens. Thus while the C-good is valued the same across group members, the G-good provides more benefit to the official at a cost of reduced benefit to the citizens—what we call a "corrupt" decision. The net of the round's earnings for each subject is then calculated (earned income minus taxes less penalties (if audited) plus the payoff from the public good chosen by the official). One full term in the office lasts for seven rounds and the whole experiment consists of 14 rounds.

The two experimental settings differ as follows. In the no-recall treatment the official remains in power for seven rounds whereas in the recall treatment the group members are allowed to vote for a recall-replacement at any round out of the seven rounds. If the majority of members (including the official)[19] vote for a recall, the computer randomly chooses a new official from eligible members.[20] After the first seven rounds, a new official is randomly selected in the no-recall treatment, and the experiment continues for seven more rounds (i.e., until the 14th

round); in the recall treatment a random selection of an official takes place only if the initial official was never recalled for seven rounds. Several studies (e.g., Blume and Sobel 1995; Crawford and Sobel 1982; Farrell and Gibbons 1989) find that communication can affect behavior. Therefore, after the tenth round in both treatments we allow subjects for a "cheap talk" chat via text messages within the group members. Subjects were not allowed to communicate with one another during the experiment other than the chat allowed after round 10.

After completion of the main experiment, subjects completed a post-experimental online questionnaire (see Appendix 3) that included questions designed to get information about idiosyncratic individual characteristics such as attitudes toward risk, views regarding the performance of political institutions, gender, religion, academic performance, among others.

All 120 subjects (60 subjects in each treatment; each session was run with 30 subjects) who participated in the experiment were volunteers from undergraduate classes at Cairo University. The experiment was conducted in March 2013, three months before the ousting of Egypt's first democratically elected president, Morsi. Each subject participated only once in the experiment. At the end of the experiment, subjects were paid for all 14 rounds and the total earnings were typically between $26.00 and $60.00[21]. The experiment typically approximately two hours, and it was conducted in Arabic. Both treatments were conducted in the Laboratory of the Faculty of Economics and Political Sciences at Cairo University.

Empirical Results

Before we report subjects' behavior, it will be helpful to look at incentives for corruption across the two games given parameters used in the experiment.

We begin by noting that the maximum feasible payoff for a group is EP150. For optimal claims of income given the type of public good funding, if the tax proceeds go to fund the C-

good then the round payoff is EP27 for everyone, whereas under corruption (G-good funding) the round payoffs are EP32 and EP21 for the official and the citizen, respectively. Thus, funding C-good offers not only more fair redistribution of tax proceeds but also higher efficiency as optimal claims are higher: economic efficiency is 90 percent (=135/150) in case of the C-good and down to 77 percent (=116/150) if the G-project is funded. Yet, the round payoff is EP5 higher for the official if the official decides to fund the G-good.

Next, in a two-terms of service situation (two parts in our experiment, each with seven rounds), a norm of corruption might be tempting in the scheduled replacement treatment as the expected payoff to an official is EP386 which exceeds the expected payoff of EP378 from no corruption (C-good always funded). A 'norm' of corruption is less tempting in the recall-enabled replacement treatment as a recalled official needs to wait for three elections before he becomes an eligible member for office, so subject's expected payoff under a 'norm of corruption' is bounded from above by EP338 which is smaller than expected payoff of EP378 if the norm of no corruption is in place.[22]

The SPE predicts full corruption in either institution. Given the parameters used in our experiment, the Nash equilibria of type 2a reported in the Main Result also predict full corruption (in both games) as seven rounds are not sufficient to support rounds with no corruption: the round gain is 22.5 whereas any future round comes with a loss of 3.5, hence the predicted number of rounds with corruption is seven. Nevertheless, for equilibria of type 2b (that restrain from recalls during the end of the game) seven rounds are enough to support rounds with no corruption in the recall-enabled replacement treatment as the round gain remains the same, 22.5, but the future round loss is almost tripled, 9.34; so the predicted number of end rounds with corruption in the R-game is three. If so, then the frequency of corruption is between 43 percent

and 100 percent (depending on whether recalls are utilized during the end-rounds) in the R-game and 100 percent in the noR-game. So, given the parameters used in our experiment, the alternative hypothesis H1a and H2a are one sided:

H1b: Corruption level is higher in the scheduled replacement than recall-enabled replacement treatment.

H2b: Inefficiency of public good provision is higher in the scheduled replacement than recall-enabled replacement treatment.

Finally, if the C-good is funded more often in the R treatment (H1b) then payoffs are more equal across group members because they all get equal returns for the public good. In addition if corrupt officials are often recalled then in case of G-good provision each player earns (the high) official's payoffs in some rounds and (the low) citizen's payoffs in others. Thus, it follows that:

H4 (inequity): Inequality of final earnings is negatively affected by the recall option.

Experimental Results

Of the 60 subjects who participated in the no-recall treatment, 40 (67 percent) subjects never served as a group official, 16 subjects (27 percent) served as group officials for seven rounds, and four subjects (7 percent) served for 14 rounds. In the recall treatment, on the other hand, of the 60 subjects, there were six subjects (10 percent) who never served as officials, five subjects (8 percent) served for seven or eight rounds, and no subject served for 14 rounds; half of the subjects (30) served as group officials for two or three rounds.

FIGURE 1 Histogram of Recalls

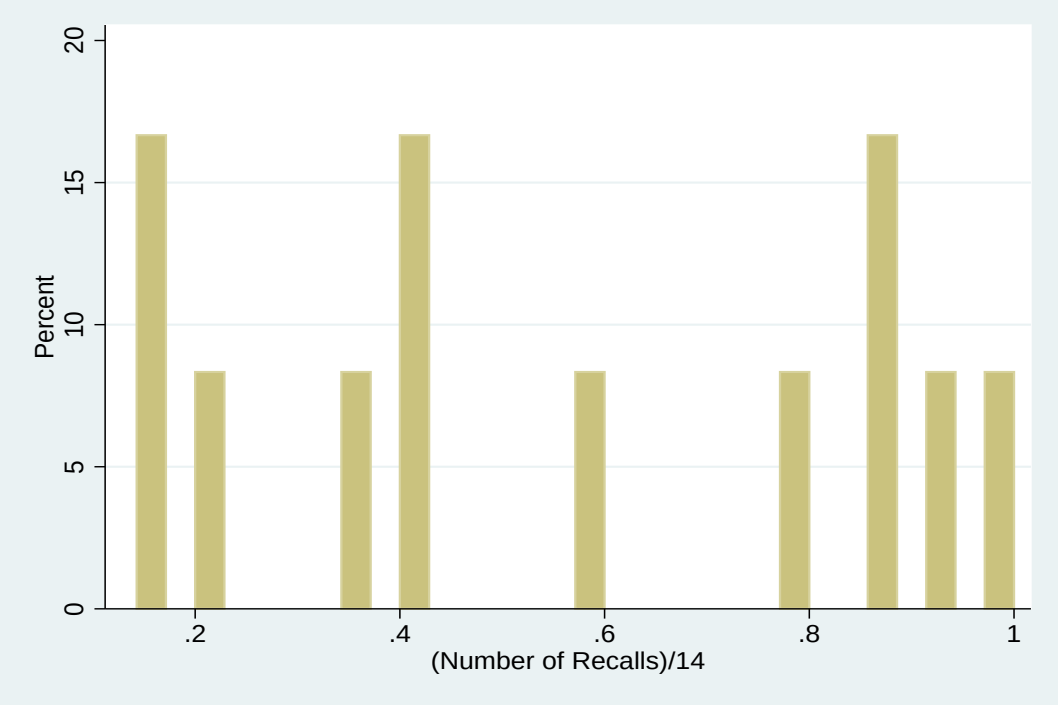


Figure 1 shows empirical distribution of recalls across 12 groups; the mean likelihood of a recall is 56 percent (standard deviation=0.498). Clearly our subjects weren't shy of exercising the option to recall the group official but the 56 percent is a far cry from 100 percent rate of recall predicted by a SPE (hypothesis H3o). So what determines the likelihood of a recall?[23]

Recall and corruption. If an intrinsic need for power is the main driver of a recall then we expect to see that the likelihood of recall does not depend on an official's choice (C or G); a result that would be consistent with the SPE. On the other hand, and according to the NE that are not SPE, a corrupt official can trigger recalls in the R-game; meaning that we should expect no recalls of non-corrupt officials during the non-end rounds of the game (end-rounds as well for type 2.b equilibrium). Examining the data at the aggregated level, however, we found that the empirical likelihood of a recall is 16.28 percent following a C-good and almost six times as high, 97.56 percent following a G-good funding. Moreover, by classifying groups into two categories: committed (six groups that recalled fewer than half of their officials) and volatile (six groups that

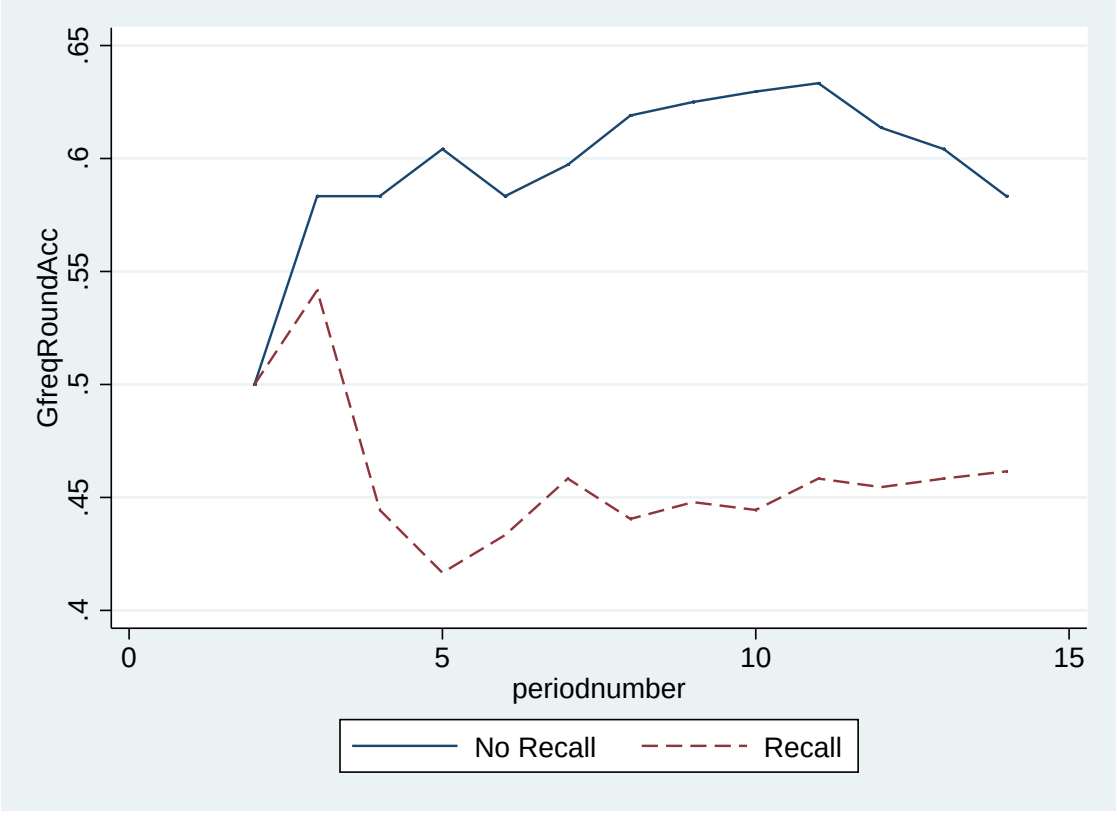
recalled more than half of their officials), we found that: (i) the likelihood of recalling non-corrupt officials is a high 41.67 percent (100 percent for the corrupt officials) among volatile groups and a low 6.45 percent (90.91 percent for corrupt officials) for committed groups, and (ii) the corruption level is 2.7 times higher in the volatile category: 71.43 percent (volatile groups) and 26.19 percent (committed groups). So, although there seems to be some intrinsic need for power exhibited among volatile groups, the effect of corruption seems to be stronger. Indeed probit regression (with clusters at group level) supports the hypothesis that corruption is the main cause of recalls: the estimated effect of a dummy variable that takes the value of 1 (0) when the G-good (C-good) is funded increases the likelihood of a recall by 81 percent (robust standard error = 0.057, p-value=0.000).[24] We conclude that our data reject the null hypothesis H3o in favour of the alternative hypothesis H3a.

Result 1: *Corruption, and not an intrinsic need for power, is the main cause of recalls.*

The strong effect of corruption on the likelihood of a recall brings to the forefront the question of the interaction between political institutions and the level of corruption. The data from the end game (round 14) supports the hypothesis that it is the threat of recall followed by ineligibility to serve as an official for at least three rounds that may sway officials to fund the G-good less often in the recall treatment. In the last round of the experiment (when the recall comes with no consequences) we observe that 83 percent of the officials fund the G-good which is not statistically different (Pearson chi2(1)=0.25, p=0.615) from behavior of officials in the noR-treatment: 75 percent choose to fund the G-good in round 7, the last round before a scheduled official replacement in the no-recall takes place.[25] On the contrary, for round 7 (which is not the end of life in office for 83 percent (10 out of 12) officials in the recall treatment) only 33.33 percent of the officials decided to fund the G-good (Pearson chi2(1)=4.20, p=0.041).

Overall there is less corruption in the recall treatment. With groups as the unit of observation, we find that the mean of the distribution of frequencies of G-good until the chatting event is 63 percent (95 percent confidence interval is (0.41, 0.86)) in the No-Recall treatment and down to 46 percent (95 percent confidence interval is (0.27, 0.64)) in the Recall treatment. To capture the evolution of "corruption" at a group level, we construct a new variable, "Time Frequency of G-good (TFG)." The value of the new variable at round *t* for group *i* is the ratio of the total number of times that the G-good is funded up to round *t* and the value of *t*. This variable will be used in the statistical analysis of the determinants of corruption reported below, but to get an overall impression of the effect of the Recall option on the frequency of corruption over time Figure 2 shows TFG values for the two treatments.[26]

FIGURE 2 Time Frequencies (at Group Level) of G-Good Being Funded



A visual inspection of Figure 2 suggests that: (i) the prevalence of corruption is negatively affected by the Recall option (as the Recall [dotted] line is everywhere (but at the very beginning rounds) by at least 10 percent below the no Recall [solid] line), (ii) subjects in the recall treatment learn quickly (as early as round 3) to fund the G-good less often but after that behavior seems stabilized, (iii) chatting (after round 10) seems to have a positive effect on reducing corruption in the No-Recall treatment but not in the Recall treatment and (iv) there is a persistent upward trend in the frequency of corruption in the No-Recall treatment until the chatting event.

We turn our attention to finding out which of the features above survive statistical significance criteria. The use of probit model with clusters at the subject level is warranted as we have more than one observation per subject serving as a group official. The dependent variable is a dummy that takes the value of 1 (0) if the group official decides to fund G-good (C-good). In the list of regressors that are expected to affect the official's decision are the frequency of G-good being funded in the official's own group at the time of decision (TFG), whether in the preceding round the official funded C-good and remained in office (C&In), the opportunity to communicate via a chatting event (Chatting) and, in model 2, we add additional regressors that control for individual idiosyncratic characteristics such as gender and religion, for example.

Table 1 reports estimated marginal effects (p-values in brackets) of the regressors. Estimates reveal that previous level of corruption is positively associated with officials' self-serving choice of funding G-good: 31.6 percent increase in the likelihood of corruption (p=0.057). Variables that are negatively correlated with corruption include retaining a non-corrupt official (the event of a C-good provision and the official not being recalled) in the preceding round (-26.3 percent),

being a Muslim in the recall treatment (-23.7 percent) and good academic performance (-25.1 percent).

TABLE 1 (Marginal Effects) Probit Regression of G-Good Funding

G-good Funding (D)	**Recall Data**		**All Data**		
	(1)	(2)	(1)	(2)	(3)
C-good & no Recall	-0.263**	-0.245**		-0.243**	-0.236**
(C & In preceding round)	(0.034)	(0.049)		(0.029)	(0.030)
Time Frequency of G-good	0.316*	0.361*	0.641***	0.407***	0.369***
(lagged)	(0.057)	(0.057)	(0.000)	(0.002)	(0.009)
Chatting (D)	0.109	0.119	0.122	0.112	0.102
	(0.127)	(0.111)	(0.121)	(0.122)	(0.168)
Demographics					
Female (D)		-0.138			-0.085
		(0.197)			(0.334)
Muslim (D)		-0.237**			0.019
		(0.041)			(0.886)
Single (D)		-0.018			0.086
		(0.919)			(0.643)
Junior and up (D)		-0.135			-0.125
		(0.313)			(0.167)
High GPA (D)		-0.251***			-0.223***
		(0.008)			(0.003)
Treatment Effects					
Chatting No Recall (D)			-0.386**	-0.344**	-0.387***
			(0.014)	(0.017)	(0.009)
No Recall (D)			0.147*	0.154*	0.137*
			(0.085)	(0.053)	(0.081)
Nr. Of Observations	156	156	312	312	312
Nr of Clusters	54	54	74	74	74
R^2	0.127	0.189		0.168	0.200
Log-likelihood	-94.40	-87.67	-184.4	-179.4	-172.4
Obs. P	0.487	0.487	0.532	0.532	0.532
Predicted P	0.483	0.484	0.539	0.538	0.538

To get some information on the association between the two types of institutions and the level of corruption we used the pooled sample and added two dummy variables for the No-Recall treatment: one captures additional chatting effect whereas the other measures the overall effect of

taking away the Recall option on the observed likelihood of G-good funding. Our data support the conclusion that in the absence of Recall option the likelihood of corruption goes up 13.7 percent (one-sided p=0.04). Our data reject the null hypothesis H1o in favor of the alternative hypothesis H1b.

Result 2: *The level of corruption is lower if citizens are allowed to recall the official.*

Allowing chatting among group members has no significant effect on the likelihood of corruption, however, in the absence of Recall, communication lowers the likelihood of corruption by almost 38.7 percent.[27]

Finally, we look at the inefficiency of the public good provision across the two treatments. The average inefficiency of public good provision is 15.22 percent in the no Recall treatment and 22.63 percent in the Recall treatment. According to the Kolmogorov-Smirnov test, data rejects the null hypothesis, H2o of similar inefficiencies of public good provision across the two treatments.

Result 3: *Inefficiency of public good provision is higher if citizens are allowed to recall the official.*

As the inefficiency of the public good provision is determined by tax compliance, we'll have more to say about this in the following section.

Economic consequences of the Recall Option. An expected economic consequence of a lower level of corruption in our experiment is lower income inequality.[28] The range of payoffs in the Recall treatment is [239, 405] which is a strict subset of the range of payoffs in the No-Recall treatment, [213, 550]. The Gini index in the No-Recall treatment is twice the index in the Recall treatment: 10.3 percent (No-Recall) and 5.1 percent (Recall). Figure 2 shows estimated kernel densities of the distributions of final earnings in the two treatments (the solid line shows data

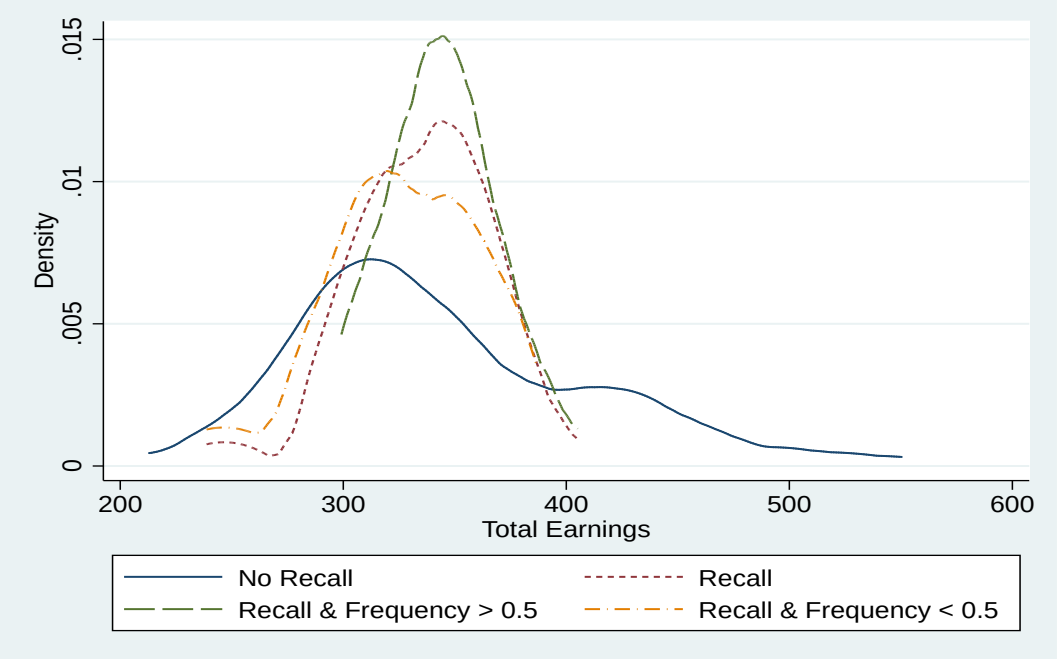
from No-Recall treatment whereas the dashed line correspond to data from Recall treatment). The null hypothesis of final earnings in the two treatments coming from the same distribution is rejected by Kolmogorov-Smirnov test (p=0.028).

Result 4: *The Recall option has a positive effect on income equality*.

Lower income inequality can also result from a high frequency of recalls; if so then lower inequality might not be that desirable as frequent recalls signal cabinet instability. To test whether the high recall is the main cause of the low earning inequality, we look at data from the Recall treatment and compare equality of earnings' distribution of subjects from groups with Recall frequencies below 50 percent (dash-dot line in Figure 3) and above 50 percent (short-dot line in Figure 3). The null hypothesis of earnings in these two categories being drawn from the same distribution is not rejected by Kolmogorov-Smirnov test (p=0.236). We conclude that

Result 5: *Low corruption and not high rate of recalls is the main cause of low inequality of earning distributions.*

FIGURE 3 Estimated Kernel Densities of Final Earnings



A remaining question is whether the lower inequality of income distribution in the Recall treatment comes at a cost of lower economic efficiency, which is measured as the ratio of the realized earnings and the maximum possible earnings. We find that realized efficiency is 82.22 percent in the no-recall treatment and 79.91 percent in the recall treatment; Kolmogorov-Smirnov test doesn't reject the null hypothesis of equal distributions of efficiency across the two treatments (p=0.403).

Result 6: *Economic efficiency is similar across the two institutions, with and without the Recall option.*

Our data do however reveal higher wasted resources through higher penalties in the Recall treatment (26.79, 95% CI (12.63, 24.51)) than in the No-Recall (18.57, 95% CI (19.50, 34.08)); the null hypothesis of equal penalties across treatments is rejected in favor of the alternative hypothesis of (p=0.067 (t-test)).

Result 7: *Recall option has a positive effect on wasted resources through penalties.*

Allowing citizens to recall the official seems to provide incentives for officials to limit the use of office for private benefits but, paradoxically, at the same time it encourages lower compliance rates. What could explain these perverse patterns? Some thought exercises are in order: (i) a thrown out of office official can retaliate by contributing less to the public fund, and she can do so through low compliance, and (ii) low level of corruption can be sustained by the use of punishing strategies (such as minmax actions). One might expect that type (ii) reasoning should be more pronounced in the No-Recall treatment as there low compliance is the only tool to discipline the official but the problem is that it is also more costly to do so than in the Recall treatment.[29] To see whether type (i) reasoning has any validity we looked at our data: the mean compliance rate of a citizen who was previously an official and was recalled is 70.9 percent

whereas it is 79.7 percent for officials in office (not recalled). The difference is significant at the 5 percent level. This suggests support for the retaliation hypothesis.

Conclusion

Can political institutions impact corruption? In this research, we find experimental evidence to suggest that indeed, the type of institution *can* have a limiting effect on the level of corruption in government. Through a laboratory experiment run in Cairo, we differentiate officials' behavior regarding the use of public funds in situations that allow "citizens" to recall the official or not. In cases where officials take a "corrupt" decision (by choosing to use public funds in a public good that benefits themselves and reduces the benefits to other group members—G-good), there is a substantially larger chance of a recall than when the official has taken the more equitable decision (C-good). In addition, we find evidence that the culture of corruption is quickly affected by the recall treatment. We find that the prevalence of corrupt decisions falls early and significantly in the experiment in the recall treatment. In the no-recall treatment, there is an upward drift in the prevalence of corruption until the subjects chat. The recall environment also shows promise for increased social stability by the correlation we find with equity in the post-experiment distribution of income.

There is significant policy relevance of these findings especially considering the number of nascent democracies resulting from the Arab Spring. Corruption has been hailed as one of the primary reasons for toppling these governments. The practical limitations and complications of establishing new representative democracies have become painfully obvious in countries including Egypt, Tunisia, Libya, and Yemen, among others. In many countries the definition and practice of representative governance will be years in the making. In the meantime, the results of

this research strongly suggest that citizens' ability to censure their leaders (through recall) can reduce corruption, and is therefore an important lever to consider in the development of new political processes.

References

Abbink, Klaus. 2006. "Laboratory Experiments on Corruption." In *International Handbook on the Economics of Corruption*, ed. Susan Rose-Ackerman. Cheltenham, UK: Edward Elgar.

Abbink, Klaus, Bernd Irlenbusch, and Elke Renner. 2002. "An Experimental Bribery Game." Journal of Law, Economics, and Organization 18 (2): 428-54.

Abbink, Klaus, Bernd Irlenbusch, and Elke Renner. 2000. "The Moonlighting Game: An Experimental Study on Reciprocity and Retribution." Journal of Economic Behavior and Organization 42 (2): 265-77.

Alt, James, Ethan Bueno de Mesquita, and Shanna Rose. 2009. "Disentangling Accountability and Competence in Elections: Evidence from U.S. Term Limits." Journal of Politics 73 (1): 171-86.

Bai, Chong En, and Shang-Jin Wei. 2000. "Quality of Bureaucracy and Open-Economy Macro Policies." NBER Working Paper #7766. Cambridge, MA: National Bureau of Economic Research.

Bardhan, Pranab. 1997. "Corruption and Development: A Review of Issues." Journal of Economic Literature 35 (3): 1320-46.

Besley, Thomas, and Anne Case. 2003. "Political Institutions and Policy Choices: Evidence from the United States." Journal of Economic Literature 41 (1): 7-73.

Besley, Thomas, and Anne Case. 1995. "Does Electoral Accountability Affect Economic Policy Choices? Evidence from Gubernatorial Term Limits." Quarterly Journal of Economics 110 (3): 769-98.

Blume, Andreas, and Joel Sobel. 1995. "Communication-Proof Equilibria in Cheap Talk Games." Journal of Economic Theory 65 (2): 359-82.

Burki, Shahid Javed, and Guillermo E. Perry. 1998. *Beyond the Washington Consensus: Institutions Matter, World Bank Latin American and Caribbean Studies : Viewpoints*. Washington, DC: World Bank.

Central Agency for Public Mobilization and Statistics (CAPMAS). 2013. *Annual Statistical Yearbook for the Arab Republic of Egypt*. Egypt: Central Agency for Public Mobilization and Statistics.

Crawford, Vincent, and Joel Sobel. 1982. "Strategic Information Transmission." Econometrica 50 (6): 1431-51.

Drazen, Allen and Erkut Y. Ozbay. 2014. *How Do Elected and Appointed Policymakers Act When In Office? Experimental Evidence on Citizens, Candidates, and Leaders*. College Park, MD: University of Maryland.

Farrell, Joseph, and Robert Gibbons. 1989. "Cheap Talk with Two Audiences." American Economic Review 79 (5): 1214-23.

Ferraz, Claudio, and Frederico Finan. 2011. "Electoral Accountability and Corruption in Local Governments: Evidence from Audit Reports." American Economic Review 101 (4): 1274-311.

Fuest, Clemens, Giorgia Maffini, and Nadine Riedel. 2013. "Do Corruption and Taxation Affect Corporate Investment in Developing Countries?" In *Critical Issues in Taxation and Development*, eds. Clemens Fuest and George R. Zodrow. Cambridge, MA: MIT Press.

Glynn, Patrick, Stephen J. Kobrin, and Moises Naim. 1997. "The Globalization of Corruption." In *Corruption and the Global Economy*, ed. Kimberly Ann Elliott. Washington, DC: Peter G. Petersen Institute for International Economics.

Hölmstrom, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." Review of Economic Studies 66 (1): 169-82.

Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobaton. 1999. *Governance Matters*. Washington, DC: World Bank.

Kaul, Inge, Isabelle Grunberg, and Marc Stern, eds. 1999. *Global Public Goods: International Cooperation in the 21st Century*. Published for the United Nations Development Program. New York, NY: Oxford.

Kunicová, Jana, and Susan Rose-Ackerman. 2005. "Electoral Rules and Constitutional Structures as Constraints on Corruption." British Journal of Political Science 35 (4): 573-606.

Lizzeri, Alessandro, and Nicola Persico. 2001. "The Provision of Public Goods Under Alternative Electoral Incentives." American Economic Review 91 (4): 225-39.

Mauro, Paolo. 1995. "Corruption and Growth." Quarterly Journal of Economics 110 (3): 681-712.

Persson, Torsten, Gerard Roland, and Guido Tabellini. 1997. “Separation of Powers and Political Accountability.” Quarterly Journal of Economics 112 (4): 1163-202.

Persson, Torsten, and Guido Tabellini. 1999. “The Size and Scope of Government: Comparative Politics with Rational Politicians.” European Economic Review 43 (4-6): 699-735.

Powell, G. Bingham. 2000. *Elections as Instruments of Democracy: Majoritarian and Proportional Visions*. New Haven, CT: Yale University Press.

The World Bank. 2004. *The World Development Report: Making Services Work for the Poor*. Washington, DC: World Bank.

The World Bank. 1997. *Helping Countries Combat Corruption: The Role of the World Bank, Poverty Reduction and Economic Management*. Washington, DC: World Bank.

Supporting Information

Appendix 1

Equilibrium Analysis

PROOF of the Main Result

We provide equilibrium analysis separately for the two games below. Part 1 and 2 of the Main result follows from part 1 and 2 (respectively) of Propositions 1 and 2 below. Part 3 of the main result is a straightforward implication of statements (*) and (A.1) below.

Let *w* and *x* denote the earned and the claimed income by an individual. Let the penalty function $f(.)$ defined on underreported income, $(w - x \in [0, w])$, be an increasing and convex function and $f(0) = 0, f'(0) = 0$. Let the valuation of the C-good be identical for citizens and the official whereas the valuation of the G-good be asymmetric: it is valued more than the C-good by the official but less by the citizens. This is captured by the following order of the marginal per capita return, β of the public goods G and C across players,

(*) $$\min\{1, \beta_o^G\} > \beta_o^C = \beta_c^C > \beta_c^G \geq 1/(n-1)$$

where n is the number of players, player type in subscripts and public good type in superscripts. The lower bound 1/(n-1) is a sufficient condition for funding of each public good to be socially efficient whereas the upper bound $\min\{1, \beta_o^G\}$ provides incentives for free riding. As we are mainly interested in cases for which the corruption is costly to the citizens as a population we will assume that *n* is large enough to satisfy,

(**) $$n > (1 - \beta_c^C)/(\beta_c^C - \beta_c^G)$$

We use R to denote the total number of rounds the game is played, i.e., the full term of the official in the office. Assume selfish preferences and risk-neutrality.

Proposition 1 (No-Recall Game)

1. The outcomes of the SPE are: under provision of the G-good, the only public good being funded.
2. There exist Nash equilibria that are Pareto improvement of the SPE. The outcomes of one such equilibria are: C-good being funded during the first r* rounds and G-good being funded during the remaining rounds, R-r*, for some r*. The number of rounds, r* during which the C-good is funded increases with the number of citizens using trigger strategies to punish corruption.

PROOF. First note that if public good *j* (*j* from {G, C}) is funded then it is optimal for player *i* to report income, x_i from (0, *w*) given by

$$x_i = x_i^*, if\ \beta_i^j < 1$$

$$= w, if\ \beta_i^j \geq 1 \qquad \text{(A.1)}$$

where x_i^* solves $f'(w - x_i^*) = \tau(1 - \beta_i^j)(1/p_a - 1)$, and it is 0 if at x=w the left hand side of the last equation is smaller than the right hand side expression, i.e., $f'(w) < \tau(1 - \beta_i^j)(1/p_a - 1)$.

Note also that (A.1) and statement (*) imply that $f'(w - x_c^*) \geq f'(w - x_o^*)$ and by convexity of the penalty function *f(.)* we get

$$x_c^* \leq x_o^* \qquad \text{(A.2)}$$

for a public good j.

Next, let T denote the total tax revenue. At the end of the stage game, it follows from statement (*) that funding the G-good is optimal for the official as:

$$\pi_o(x_i, x_{-i}, 1) - \pi_o(x_i, x_{-i}, 0) = (\beta_o^G - \beta_o^C)T \geq 0$$

Given that the G-good is funded, player *i* declares income, x^* that maximizes his expected payoff:

$$\max_{x\in[0,w]} E\big(\pi_i(x_i, x_{-i}, 1)\big) = w - (1-p_a)\tau x(1-\beta_i^G) - p_a[f(w-x) + \tau w(1-\beta_i^G)] + \beta_i^G T_{-i}$$

where the second and the third terms correspond to *i*'s payoff in two possible states of audition. As the penalty function, *f(.)* is convex and increasing, the optimal claimed income, x^* is determined by f.o.c., hence the specifications on the optimal x_i as stated above follow.

Part 1. The SPE Nash equilibrium of the stage game is a SPE of the R-round game. Thus, G-good is funded in every round. Under provision of the G-good in the SPE follows from the observation that under full compliance, an amount of $T^e = \tau nw$ goes to fund the G-good which is a Pareto improvement. Indeed, the difference between T^e and the expected total tax revenue in the SPE is

$$T^e - T^* = \tau nw - \tau \textstyle\sum_{j=1..n}\big[(1-p_a)x_j^{*G} + p_a w\big] = (1-p_a)\tau \textstyle\sum_{j=1..n}(w - x_j^{*G}) \qquad \text{(A.3)}$$

and the payoff difference for any player *i* is positive,

$$\begin{aligned}
\pi_i(G|T^e) - \pi_i(G|T^*) &= \beta_i^G(T^e - T^*) + p_a f\big(w - x_i^{*G}\big) - (1-p_a)\tau\big(w - x_i^{*G}\big) \\
&= \beta_i^G(1-p_a)\tau \sum_{j=1..n}\big(w - x_j^{*G}\big) + p_a f\big(w - x_i^{*G}\big) - (1-p_a)\tau\big(w - x_i^{*G}\big) \\
&\geq (1-p_a)\tau\left(\frac{1}{n-1}\sum_{j=1..n}\big(w - x_j^{*G}\big) - \big(w - x_i^{*G}\big)\right) + p_a f\big(w - x_i^{*G}\big) \\
&\geq \frac{(1-p_a)\tau}{n-1}\big(w - x_{-i}^{*G}\big) + p_a f\big(w - x_i^{*G}\big)
\end{aligned}$$

where the second equality follows from (A.3), the first weak inequality is implied by statement (*) whereas the second inequality follows from (A.2), the symmetry of citizen's optimal choices and $x_i^* \leq w$.

Part 2. Consider the following profile of strategies: the official funds the C-good (p^G=0) in the first r* rounds and the G-good in the remaining R-r* rounds. If no defection occurs then each player *i* claims w in rounds 1 to r* and x_i^G in the remaining R-r* rounds. Any defection at any round before r*+1 triggers funding of the G-good as of that round and claims of zero income as of the following round until the end of the game. No deviation can be profitable after round r* as all players are playing Nash. The most tempting deviating strategy for the official is to defect by funding the G-good (p^G=1) and declaring his G-optimal level of income instead of w as of round r* (instead of r*+1): The official's round payoff increases by

$$\Delta\pi_o^{NR} = \pi_o(x_o^G, w, 1) - \pi_o(w, w, 0)$$

$$= \begin{cases} (\beta_o^G - \beta_o^C)T^w + \tau(1-p_a)(1-\beta_o^G)(w - x_o^G) - p_a f(w - x_o^G), & if\ \beta_o^G < 1, \\ (\beta_o^G - \beta_o^C)n\tau w, & if\ \beta_o^G \geq 1. \end{cases}$$

The total payoff in the remaining R-r* rounds decreases by

$$\Delta\pi_o^{R-r^*} = (R - r^*)[\pi_o(x_o^G, x_c^G, 1) - \pi_o(x_o^G, 0, 1)]$$

$$= (R - r^*)\beta_o^G(1-p_a)\tau(n-1)x_c^G$$

Thus the official is better off not deviating at r*, i.e. $\Delta\pi_o^{R-r^*} > \Delta\pi_o^{NR}$ if R-r* is the smallest integer larger than the ratio of round r* gains and average future round losses; let δ^{NR} denote this ratio,

$$\delta^{NR} = \frac{\Delta\pi_o^{NR}}{\beta_o^G(1-p_a)\tau(n-1)x_c^G}$$

If m (instead of n-1) citizens use the punishing strategy (of claiming income 0 after a defection) then $\Delta\pi_o^{R-r^*} = (R-r^*)\beta_o^G(1-p_a)\tau m x_c^G$ whereas $\Delta\pi_o^{NR}$ is not affected. Hence, the number of rounds of the C-good being funded (no corruption), r*, increases with the number of citizens engaging in retaliation.

About citizens, it can be verified that a citizen's defection at round r^* by claiming some other amount x instead of w changes the round payoff by

$$\Delta\pi_c^{NR} = \pi_c(w,w,0) - [p_a\pi_c(x,w,1) + (1-p_a)\pi_c 0]$$

$$= p_a f(w-x) + p_a(\beta_c^C - \beta_c^G)nw\tau + (1-p_a)(\beta_c^C - 1)(w-x)\tau$$

$$> p_a w\tau[(\beta_c^C - \beta_c^G)n - (\frac{1}{p_a} - 1)(1-\beta_c^C)\left(1 - \frac{x}{w}\right)]$$

$$> p_a w\tau[(\beta_c^C - \beta_c^G)n - (1-\beta_c^C)]$$

where the first inequality follows from the penalty function being positive whereas the second one follows form $(1-p_a)(1-x_c^G/w)<1$. Hence, for n large enough (**) one has $\Delta\pi_c^{NR} > 0$, so the citizen's round payoff decreases if he does not claim w. In addition the remaining rounds payoffs cannot increase either as with probability p_a defection is detected and claims of all players (but our citizen's claim) become 0 in response to defection, i.e., the change in future payoffs is

$$\Delta\pi_c^{R-r^*} = -(R-r^*)p_a\beta_c^G(1-p_a)\tau\left((n-2)x_c^G + x_o^G\right) < 0.$$

Q.E.D.

Proposition 2: Recall Game

1. The outcomes of the SPE are: under provision of the G-good, the only public good being funded and smaller payoff inequality than in the NoR game.
2. There exist Nash equilibria that are Pareto improvement of the SPE. The outcomes of such equilibria are of the following two types:
   a. Official is always recalled: C-good is funded during the first ra* rounds and G-good is funded during the remaining rounds, R-ra*, for some ra* *not larger than* r*.
   b. Official is not recalled if he funds the C-good: C-good is funded during the first rb* rounds and G-good is funded during the remaining rounds, R-rb*, for some rb* *larger than* both ra* and r*.

PROOF. Note that adding "always recall the official" to the profile of strategies of the NR-game SPE strategies remains SPE which concludes the proof of part 1. As the official is changing across rounds, players are taking rounds in enjoying the high payoff from the G-good, hence the payoff inequality is smaller.

About part 2a, consider the following extended profile of strategies reported in part 2 of Proposition 1: the official funds the C-good in the first ra* rounds and the G-good in the remaining R-ra* rounds. If no defection occurs then each player *i* claims *w* and votes against a recall in rounds 1 to ra* whereas in the remaining R-ra* rounds the declared income is x_i^G and the vote is in favor of a recall. Any defection at any round before ra*+1 triggers claiming earned income is 0, funding of the G-good and voting in favor of a recall until the end of the game. No deviation pays off after ra* as all players are playing Nash. As in the proof of part 2 of the NR game, a citizen's deviation at round ra* reduces the round payoff as well as future payoffs.

Suppose that the official defects by funding the G-good and declaring x_i^G as of round ra* (instead of ra*+1). The official's round gain is the same as in the NR game,

$$\Delta\pi_o^R = \pi_o(x_o^G, w, 1) - \pi_o(w, w, 0) = \Delta\pi_o^{NR}$$

Letting γ denote the probability of serving as an official in the remaining rounds, the total payoff in the remaining R-ra* rounds decreases by

$$\Delta\pi_o^{R-ra^*} = (R - ra^*)[\gamma(\pi_o(x_o^G, x_c^G, 1) - \pi_o(x_o^G, 0,1)) + (1-\gamma)(\pi_c(x_c^G, x_o^G, 1) - \pi_c 1]$$

The expression within the square brackets is smaller than the corresponding one in NR game if when the G-good is funded, others claiming 0 instead of their G-optimal level of income results in a citizen's loss smaller than the official's loss; formally is

$$\pi_c(x_c^G, x_o^G, 1) - \pi_c(x_c^G, 0,1) < \pi_o(x_o^G, x_c^G, 1) - \pi_o(x_o^G, 0,1)$$

which is equivalent with

$$\beta_c^G\left((n-2)x_c^G + x_o^G\right) < \beta_o^G(n-1)x_c^G$$

The last inequality for *n* big enough as the following inequality holds[30]

$$\frac{n-2}{n-1} + \frac{x_o^G}{(n-1)x_c^G} < \frac{\beta_o^G}{\beta_c^G}$$

Hence $\delta^{NR} > \delta^{Ra}$ from which it follows that ra* cannot be larger than r*. Therefore, just as in the case of SPE, the recall option cannot hinder corruption in this equilibrium either.

Part 2b. Consider the profile of strategies as in part 2a with only one difference: in the first rb* rounds "vote in favor of recall only if the official funds the G-good," in rounds rb*+1 to R defection "vote against recall." No citizen is better off by deviating in rounds earlier than rb*.

If a citizen deviates and "votes in favor of a recall" after round rb* then his vote has no affect as the official leaves the office only if the majority (or the supermajority) votes for it. On the other hand, official's defection increases the round payoff by the same amount as in the NR game. That triggers claims of zero income, the official is recalled and remains out of the office until the end of the game. The ratio between the round gain and the average future rounds loss is smaller than in the NR game as

$$\delta^{Rb} = \frac{\pi_o(x_o^G, w, 1) - \pi_o(w, w, 0)}{\pi_o(x_o^G, x_c^G, 1) - \pi_c(x_c^G, 0,1)} < \frac{\pi_o(x_o^G, w, 1) - \pi_o(w, w, 0)}{\pi_o(x_o^G, x_c^G, 1) - \pi_o(x_o^G, 0,1)} = \delta^{NR}$$

where the inequality follows from $\pi_c(x_c^G, 0,1) < \pi_o(x_o^G, 0,1)$. Thus, rb* cannot be smaller than r*.

Q.E.D.

Appendix 2

Instructions in English

SUBJECT INSTRUCTIONS FOR THE NO RECALL TREATMENT

Welcome and thank you for participating in today's experiment.

This is an experiment in the economics of group decision making. Your earnings will be determined by your own decisions and the decisions of others as described in the following instructions. **SO, IT IS IMPORTANT THAT YOU READ THESE INTRUCTIONS CAREFULLY.**

This experiment is structured so that only you know your earnings. All of the money that you earn will be paid to you privately in cash immediately at the end of today's experiment. Various research agencies have provided the funds for the conduct of this research study.

If you have any questions, RAISE YOUR HAND and an experimenter will come up to you to answer questions in private. Please feel free to ask as many questions as you like.

Time

This experiment will last approximately two hours.

Scenario

In this experiment, you will be a member of a group of five individuals. You will be randomly assigned to a group and will remain in the same group for the entire experimental session.

Every group has an official who is selected randomly from among your group members by the computer at the beginning of the experiment (before period 1) and in the middle (before period 8) of the experiment. There are 14 decision periods in this experiment.

Anonymity

You will not know the rest of your group members, neither will they know you.

Monetary payoff

You earn money in Experimental Pounds (EP) in each decision period. This amount will be displayed on your computer screen at the completion of the decision period. At the end of today's experiment, your total accumulated earnings in experimental pounds divided by the number of periods will be converted into Egyptian pounds at the below mentioned conversion rate. The more experimental pounds you earn, the more Egyptian pounds you will be paid.

1 Experimental Pound = 10 Egyptian Pounds

The following section explains how to earn money in each decision period.

Task and Decision Making Process

In this experiment, you will go through the below mentioned sequence of events in each of 14 decision periods.

Event I. All subjects are given a simple task to find the spelling mistakes in a piece of text on the computer. You will be given two minutes to conduct the task. You can make corrections to the text by using your mouse to place your cursor in the correct area and make the correction. Use the mouse to move you to other parts of the text. You will earn two Experimental Pounds for each mistake that you correct accurately. There are 10 errors. This income will be displayed on your screen at the completion of the task.

Event II. Your earned income is what you earn in Event I. You will make the choice of how much of this earned income to report using the sliding scale on your screen. There is an income tax at 25 percent that you need to pay on the income you report. This tax rate is the same for all individuals belonging to the same group. As you move the slide to determine how much income you will report, you can see the consequences of your choice in terms of your net income if you are audited or not.

You can choose to report none of it, part of it or all of it. Consequently your reported tax liability is equal to: 25 percent * Reported Income.

Event III. Once you choose the level of income you will report, a random audit will be performed. One subject out of five in the group will be chosen for audit so the likelihood of a subject being audited is 20 percent. If you are chosen for the random audit, your earned income will be disclosed to the official. If the audited individual's reported income in Event II is less than the earned income in Event I, then the individual pays, in addition to the tax of 25 percent of the *earned* income, a tax penalty that increases in the difference between the earned income and reported income as in the table that has been handed out.

You pay a tax penalty only if you are audited and if your reported income is less than the earned income.

Event IV. Income taxes in this experiment will go into your group fund; they will be used to fund a public project that is valuable (in terms of experimental pounds) to you and your group members.

Each experimental pound (EP) that goes in the public fund is tripled. Therefore,

Public fund = 3 * Income taxes collected from all the members in your group

(Note: Tax penalty is not added to the public fund)

There are two types of public projects available in this experiment, Type C and Type G. The choice of which project is made available to you and your group is made by the official who is a member of your group.

The type of good that is chosen will be highlighted in GREEN on your screen.

The benefits of Type C good are shared equally among all five members of the group.

The benefits of Type G good accrue 50 percent to the official with the remainder split among the other four group members.

Earnings if public project of Type C is funded

Public project earnings of:

- the official = Public fund / 5
- of each other member = Public fund / 5

Earnings if public project of Type G is funded

Public project earnings of:

- the official = Public fund / 2
- of each other member = Public fund / 8

For example, if

Income taxes collected by the government in your group = 20 EP

Public fund = 3*20 = 60 EP

Earnings from public project of Type C:

When this project is chosen, then all the group members earn equal amount and the money in public fund is equally divided between all the group members.

Public project earnings = 60 /5 = 12 EP

Earnings from public project of Type G:

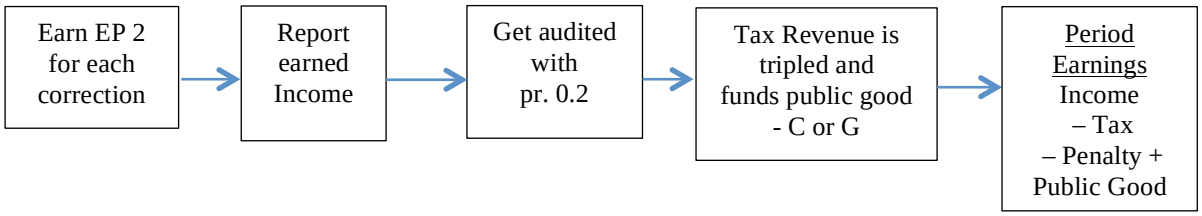
When this project is chosen, then the official will earn more than the rest of the group members: Half of the total amount of money in public fund is given to the official; the remaining half of the public fund is equally divided among all four remaining group members.

Public project earning of the official = 60/2 = 30 EP

Public project earnings of each other group members = 60/8 = 7.5 EP

The information below shows your total earnings or payoff in each decision period resulting from Events I to IV explained above.

The following diagram illustrates the sequence of events in every period



Earnings in each decision period

Scenario I: If you are not audited

Total earnings = Earned Income – tax liability + public project earnings

(Note: As explained above, public project earnings depend on the type of public project provided to the group by the official)

Scenario II: If you are audited

Total earnings = Earned Income – tax liability – tax penalty + public project earnings

(Note: Tax penalty is equal to zero if your reported income is equal to your earned income)

(Also as explained above, public project earnings depend on the type of project provided to the group by the official)

Final earnings at the end of the experiment = (Total earnings in 14 rounds)/14

Questionnaire and payment

At the end of today's experiment, you will complete a brief online questionnaire, receive payment of your earnings, and then the experiment is over. Information about your decisions will be kept without identifying information so no one can link you as an individual to the decisions that you make.

SUBJECT INSTRUCTIONS FOR THE RECALL TREATMENT

Welcome and thank you for participating in today's experiment.

This is an experiment in the economics of group decision making. Your earnings will be determined by your own decisions and the decisions of others as described in the following instructions. **SO, IT IS IMPORTANT THAT YOU READ THESE INTRUCTIONS CAREFULLY.**

This experiment is structured so that only you know your earnings. All of the money that you earn will be paid to you privately in cash immediately at the end of today's experiment. Various research agencies have provided the funds for the conduct of this research study.

If you have any questions, RAISE YOUR HAND and an experimenter will come up to you to answer questions in private. Please feel free to ask as many questions as you like.

Time

This experiment will last around two hours.

Scenario

In this experiment, you will be a member of a group of five individuals. You will be randomly assigned to a group and will remain in the same group for the entire experimental session.

Every group has an official who is selected randomly from among your group members by the computer at the beginning (before period 1) and in the middle (before period 8) of the experiment in the absence of a recall "election." Thus, a selected official remains the official of the group for seven periods unless the majority of members vote for a recall election. In case of a recall election, another official is selected randomly from among the eligible members of the group. A member of the group is eligible if he/she has not been a subject of a recall election during the last three elections. There are 14 decision periods in this experiment.

Anonymity

You will not know the rest of your group members, neither will they know you.

Monetary Payoff

You earn money in Experimental Pounds (EP) in each decision period. This amount will be displayed on your computer screen at the completion of the decision period. At the end of today's experiment, your total accumulated earnings in experimental pounds divided by the number of periods will be converted into Egyptian pounds at the below mentioned conversion rate. The more experimental pounds you earn, the more Egyptian pounds you will be paid.

1 Experimental Pound = 10 Egyptian Pounds

The following section explains how to earn money in each decision period.

Task and Decision Making Process

In this experiment, you will go through the below mentioned sequence of events in each of 14 decision periods.

Event I. All subjects are given a simple task to find the spelling mistakes in a piece of text on the computer. You will be given two minutes to conduct the task. You can make corrections to the text by using your mouse to place your cursor in the correct area and make the correction. Use the mouse to move you to other parts of the text. You will earn two Experimental Pounds for each mistake that you correct accurately. There are a total of 10 errors. This income will be displayed on your screen at the completion of the task.

Event II. Your earned income is what you earn in Event I. You will make the choice of how much of this earned income to report using the sliding scale on your screen. There is an income tax at 25 percent that you need to pay on the income you report. This tax rate is the same for all individuals belonging to the same group. As you move the slide to determine how much income you will report, you can see the consequences of your choice in terms of your net income if you are audited or not.

You can choose to report none of it, part of it or all of it. Consequently your reported tax liability is equal to: 25 percent * Reported Income

Event III. Once you choose the level of income you will report, a random audit will be performed. One subject out of five in the group will be chosen for audit so the likelihood of a subject being audited is 20 percent. If you are chosen for the random audit, your earned income will be disclosed to the official. If the audited individual's reported income in Event II is less than the earned income in Event I, then the individual pays, in addition to the tax of 25 percent of the *earned* income, a tax penalty that increases in the difference between the earned income and reported income as in the table that was handed out to you.

You pay a tax penalty only if you are audited and if your reported income is less than the earned income.

Event IV. Income taxes in this experiment will go into your group fund; they will be used to fund a public project that is valuable (in terms of experimental pounds) to you and your group members. Each experimental pound (EP) that goes in the public fund is tripled. Therefore,

Public fund = 3 * Income taxes collected from all the members in your group

(Note: Tax penalty is not added to the public fund)

There are two types of public projects available in this experiment, Type C and Type G. The choice of which project is made available to you and your group is made by the official who is a member of your group.

The benefits of Type C good are shared equally among all five members of the group.

The benefits of Type G good accrue 50 percent to the official with the remainder split among the other four group members.

Earnings if public project of Type C is funded

Public project earnings of:

- the official = Public fund / 5
- of each other member = Public fund / 5

Earnings if public project of Type G is funded

Public project earnings of:

- the official = Public fund / 2
- of each other member = Public fund / 8

For example, if

Income taxes collected by the government in your group = 20 EP

Public fund = 3*20 = 60 EP

Earnings from public project of Type C:

When this project is chosen, then all the group members earn equal amount and the money in public fund is equally divided between all the group members.

Public project earnings = 60 /5 = 12 EP

Earnings from public project of Type G:

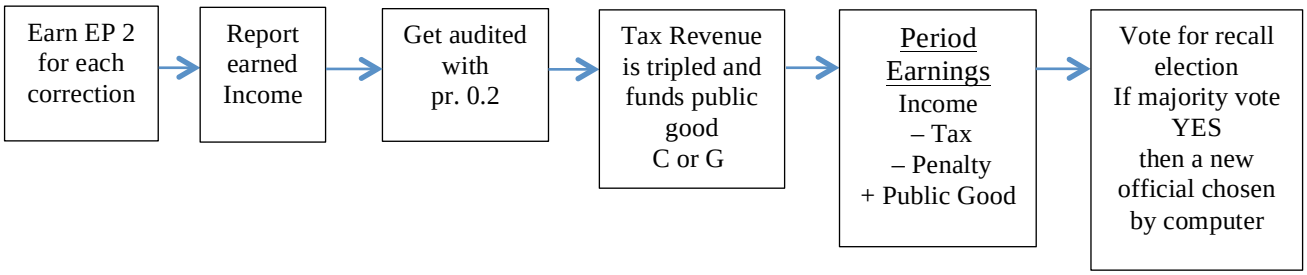
When this project is chosen, then the official will earn more than the rest of the group members: Half of the total amount of money in public fund is given to the official; the remaining half of the public fund is equally divided among all four remaining group members.

Public project earning of the official = 60/2 = 30 EP

Public project earnings of each other group members = 60/8 = 7.5 EP

Event V. Once the public good decision is made, you will see a screen that asks whether you would like a recall election or not. If the majority of the group chooses yes, then the computer will choose a new official.

The following diagram illustrates the sequence of events in every period



Section IV below shows your total earnings or payoff in each decision period resulting from Events I to IV explained above.

Earnings in each decision period

Scenario I: If you are not audited

Total earnings = Earned Income – tax liability + public project earnings

(Note: As explained above, public project earnings depend on the type of public project provided to the group by the official)

Scenario II: If you are audited

Total earnings = Earned Income – tax liability – tax penalty + public project earnings

(Note: Tax penalty is equal to zero if your reported income is equal to your earned income)

(Also as explained above, public project earnings depend on the type of project provided to the group by the official)

Final earnings at the end of the experiment = (Total earnings in 14 rounds)/14

Questionnaire and payment

At the end of today's experiment, you will complete a brief online questionnaire, receive payment of your earnings, and then the experiment is over. Information about your decisions will be kept without identifying information so no one can link you as an individual to the decisions that you make.

Appendix 3

Penalty Structure

Unreported Income	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Tax Penalty if Audited	0.10	0.28	0.52	0.80	1.12	1.47	1.85	2.26	2.70	3.16	3.65	4.16	4.69	5.24	5.81	6.40	7.01	7.64	8.28	8.94

Appendix 4

Questionnaire

INSTRUCTIONS

Below are several questions relating to your demographic information, your views concerning some economic and political issues, and experience with tax reporting. These questions may be of a sensitive nature. Although your name will not be matched with your responses in any way and all information provided will be kept strictly confidential, you may be uncomfortable or unable to answer all questions. Please indicate if you prefer not to answer a particular question or if you would like to leave the study at any time. If you choose to answer the questions, please answer them honestly and to the best of your ability.

1. In what year were you born?

Year:_____

2. Are you?

- ☐ Male
- ☐ Female

3. What is your current grade point average?_____

4. What is your field of study?

_____

5. What is your religious affiliation?

- ☐ Muslim
- ☐ Copt
- ☐ Catholic
- ☐ Protestant
- ☐ Other
- ☐ No Religion
- ☐ Prefer Not to Answer

5. Are you currently working?

- ☐ Yes, I have a full-time job
- ☐ Yes, I have a part-time job
- ☐ Yes, I am self-employed
- ☐ No, I am still studying
- ☐ No
- ☐ Prefer Not to Answer

6. Have you ever had a paid job?

- ☐ Yes
- ☐ No
- ☐ Do not know
- ☐ Prefer not to answer

7. What is your year in university now?

- ☐ Freshman
- ☐ Sophomore
- ☐ Junior
- ☐ Senior
- ☐ Graduate Student
- ☐ I am not currently enrolled in university
- ☐ Prefer Not to Answer

8. What is <u>your</u> current marital status?

- ☐ Single
- ☐ Engaged
- ☐ Married
- ☐ Separated
- ☐ Divorced
- ☐ Widowed
- ☐ Prefer Not to Answer

9. I seek opportunities for doing things that I never did before.

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

10. I don't worry about the consequences of what I do.

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

11. I never get lucky breaks.

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

12. I frequently get jittery and worry about things.

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

13. I proceed with care in most endeavors.

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

14. I tend to do dangerous things without adequate precautions.

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

15. While at university, did you take part in social activities?

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

16. If yes in answer 15, in which social activities did you take part?

17. Do you have friends?

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

18. Do you share your secrets with some of them?

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

19. Would you say that most people can be trusted?

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

20. Do you think democracy, with multiple political parties and free elections, is the best system for governing Egypt?

- ☐ Agree
- ☐ Disagree
- ☐ Don't know
- ☐ Prefer not to answer

21. Do you think the following institutions are trustworthy?

	Agree	Disagree	Don't know	Prefer not to answer
Judiciary				
Parliament				
Government				
Religious leaders				
State media				
Private media				

22. Thinking now of the country as a whole, do you think compared with five years ago, standards of living have?

- ☐ Fallen a great deal
- ☐ Fallen a little
- ☐ Stayed the same
- ☐ Risen a little
- ☐ Risen a lot
- ☐ Don't know
- ☐ Prefer not to answer

23. Here is a list of existing problems in Egypt today. Tick the biggest problem and the second biggest problem:

	a. Biggest problem	b. Second biggest problem
Poor public goods and services		
Unemployment		
Poverty		
Corruption		
Security/crime		
Protests		
Wages and salaries		

24. What do you think about the following statement?

	Agree	Disagree	Don't know	Prefer not to answer
Free elections are the means to solving the above mentioned problems.				

25. Are you generally satisfied with the quality of public goods and services provided by the government?

- ☐ Yes
- ☐ No
- ☐ Don't know
- ☐ Prefer not to answer

26. What do you think about the following statements?

	Agree	Disagree	Don't know	Prefer not to answer
It is okay not to declare everything one earns to the tax authorities				
Most people try to avoid paying their fair share of tax				

27. Have you participated in an economics experiment previously?

- ☐ Yes
- ☐ No
- ☐ Don’t know
- ☐ Prefer Not to Answer

28. Have you filed tax return before?

- ☐ Yes
- ☐ No
- ☐ Don’t know
- ☐ Prefer Not to Answer

NOTES

1. There is an argument that corruption may reduce other transactions costs associated with investment and economic development but there is little empirical support for this "corruption greasing the wheels" hypothesis (see, for example, Fuest et al. 2013).
2. "…public goods often face a double jeopardy: market failure compounded by government failure…" (Kaul et al. 1999).
3. The 'right to recall' exists in parliamentary systems under the name 'no confidence vote' where the parliament can initiate a motion to recall the prime minister. In presidential systems, however, there is no such right in the constitution, with the exception of Venezuela. In the US, for example, there are 'right to recall' governors but not presidents. We do not consider impeachment as a 'right to recall' institution.
4. As per the new constitution of Egypt (January 2014), the right to recall the president has been enshrined as a constitutional right—probably for the first time in a semi-presidential system. According to article 161, a two-thirds majority of parliament can initiate a motion to withdraw confidence from the president. Such a motion, however, has to be approved by the electorate in a public referendum. If rejected, the president remains in office and parliament is automatically dissolved. At the time of our experiments, the right to recall was not institutionalized in the political system.
5. During the last three decades, various organizations have collected and published data on corruption. However, most corruption indicators are about perceived and not actual levels of corruption.
6. In practice, these other costs may include a loss of institutional knowledge due to high leader turnover and pecuniary costs and social costs associated with frequent recalls.

7. Corruption may be carried out by others including bureaucrats but we do not specifically investigate those other channels in this paper.
8. For an earlier review of the literature, see (Abbink 2006).
9. In either game, we do not have elections per se as citizens do not have any control over who will come into office. This setting is close to Powell's (Powell 2000) classification with respect to voters' objectives at election time and which makes voters use elections to reward or punish incumbents, instead of using elections to choose between prospective teams of future policymakers.
10. Penalties do not go into the public pool of funds; they go to cover administrative costs of auditing and are considered a loss. The G and C goods are produced at the same constant marginal cost.
11. Another way to think of payoffs from the G-good is a transfer of $\left(1-\frac{\beta_c^G}{\beta_c^C}\right)T$ to the official's account (which captures rent extraction) and use the remaining of the tax proceeds, $\left(\frac{\beta_c^G}{\beta_c^C}\right)T$ to fund the C-good. In this interpretation, (which is payoff equivalent for citizens to the one above with two public goods) there is only one public good to be funded that is equally valuable to everyone (think of defense) but the official makes a decision on how much of the total tax revenue T goes to funding it (while the rest is appropriated by the official).
12. The supermajority rule is preferred to the simple majority in protecting non-corrupt officials. It is also superior to the unanimity rule if "vote buying" is added to the equation as a corrupt official would need to "buy" one vote to survive a recall.
13. For simplicity we assume homogenous income and that decision of how much to work are not part of the problem of our decision-maker. Since the optimal strategies have the dominance property these assumptions are innocuous.

14. If we let x^G denote the vector of optimal declared income when the G-good is funded (i.e, $p^G=1$) then in the Nash equilibrium of the stage game the payoff of individual *i* is $\pi_i\left(x_i^G, x_{-i}^G, 1\right)$ which is larger than the minmax payoff, $\pi_i\left(x_i^G, 0,1\right)$ in which the official funds the G-good and every player but i declares zero income; the expected difference of the two payoffs is $\beta_i^G(1-p_a)\tau \sum_{j\neq i} x_j^G \ (>0)$.
15. The instructions (in Arabic) were distributed in hardcopy to the subjects to ensure that subjects could refer to them at any time during the experiment for information on the audit rate, penalty structure, the value of the two public goods to officials and citizens and other details. Instructions are included in Appendix 2.
16. Accumulated payoffs in experimental pounds were converted at the end of the experiment into Egyptian pounds.
17. All subjects in our experiment knew that they faced the same tax rate as all other subjects.
18. Penalties are not added to the public fund and are therefore considered wasted resources.
19. This is our implementation of supermajority as the majority here is the same as three out of four citizens voting to recall the official. As an official would not vote to recall himself (confirmed in our data as 98.21 percent of our "officials" did so), in the instructions we elected to go for allowing the official to vote as well and implement the majority rule as this was easier to explain to subjects.
20. A group member is eligible if he has not been a subject of recall elections during the last three elections.
21. At the time the experiment was run, the exchange rate was: 1 USD = 6.78 EGP. The subjects' earnings were between 180 EGP and 406 EGP. An average hourly rate is 33 EGP

(CAPMAS, 2013). Thus each subject earned at least twice what he could have earned outside the lab per hour.

22. EP386 (=7*32+7(32/5+21*4/5)), EP378 (=14*27) and EP338 (=10*21+4*32).
23. To vote an official out of office requires at least three votes. We can safely rule out that the high rate of recall is a result of trembles/ noise (such as subjects submitting ‘recall’ when they meant to submit ‘do not recall’).
24. A linear regression (with clusters at the group level) with dependent variable the number of votes for recall tells a similar story. The estimate of the G-good being funded is 2.40 (robust standard error.=0.264, p=0.000, R2=0.663), that is, funding G-good increases the number of votes in favor of a recall by 2.4, which for the group size of five and the majority rule results in the official being voted out of office. There is no round effect, nor any chatting effect, on the number of votes in favor of a recall.
25. Data from round 14 in the no-recall treatment are less informative for comparison as we find a strong effect of communication (after round 10) in the No-Recall treatment but not in the Recall treatment. Further study is warranted to investigate the interaction between communication and officials’ behavior.
26. Data points at each round correspond to the averages of the TFG across groups at a given treatment.
27. The persistency of the effect of communication on corruption remains a question for another study; our design is not well-suited to address it as the experiment continued only for four rounds after the chatting.
28. Income is measured as the final earnings, i.e., income after tax and transfers.

29. Take for example using minmax strategies (that punish the official) in three sequential rounds: the payoff to a citizen in the No-Recall treatment is 5.6(=3*15/8) whereas in the Recall treatment is four times higher, 26.25 (=2*15/8 (out of the office) + 22.5 (in the office)).
30. Recall that optimal claims do not depend on n, so the left hand side converges to 1 as n goes to infinity whereas the right hand side is strictly larger than 1 as the G-good is more valuable to the official than the citizen.