How Do Electoral and Voice Accountability Affect Corruption? Experimental Evidence from Egypt

Sarah Mansour[1], Sally Wallace[2], Vjollca Sadiraj[3], Mazen Hassan[4]

Abstract

How far does democracy decrease corruption? And which specific aspects of democracy help generate such effects? Corruption is famously one of the strongest obstacles to social and economic development. Whereas there has been extensive research identifying the causes of corruption, there is little experimental research on the impact of political institutions on corruption using designs that control for significant confounders. This paper uses a series of laboratory experiments conducted in 2013 Egypt in which a government official decides whether to spend tax revenues paid by subjects on a self-serving good or a good that benefits everyone equally. We have two experimental manipulations (a) whether the official is electorally accountable to subjects or not; (b) whether subjects could send messages of protests to the official (and one another). We find evidence that *electoral accountability* does decrease the probability of the official choosing the self-serving good by 17% whereas *voice accountability* generates such outcome only in the authoritarian treatment (a reduction of corruption by 29%). We also find suggestive evidence that, in the authoritarian treatment, the likelihood of funding the self-serving good decreases by 27% when taxes paid by citizens fall short of the official's threshold. Our contribution to the literature is two-fold: (a) we are able to single out the effect of specific democratic mechanisms on government corruption; (b) we test outcomes of democratic mechanisms on a traditionally understudied subject pool.

Keywords: *Accountability, Corruption, Transition economies, Experiment, Public Goods.*

JEL classification: C920, K420, O170, E620.

[1] Economics Department, Faculty of Economics and Political Science, Cairo University, Egypt.
[2] Andrew Young School of Policy Studies, Georgia State University, USA, and African Tax Institute, University of Pretoria, South Africa.
[3] Economics Department and ExCEN, Andrew Young School of Policy Studies, Georgia State University, USA.
[4] Political Science Department, Faculty of Economics and Political Science, Cairo University, Egypt.

I. Introduction

The World Bank (2004) considers corruption 'the single greatest obstacle to economic and social development.' There is by now an extensive literature that establishes a strong association between corruption, growth, per capita income, child mortality, and literacy (Bai and Wei 2000; Burki and Perry 1998; Glynn et al. 1997; Kaufman et al. 1999) making it both theoretically and practically important to understand the underlying determinants of corruption. Indeed, there is substantial theoretical literature linking corruption to specific types of political institutions (Kunicová and Rose-Ackerman 2005). Democracy, in particular, has been an institution strongly linked with lower levels of corruption (e.g. Chowdury 2004, Sandholtz and Koetzle 2000). Disagreements exist however regarding which type of democratic mechanisms play the effective role in reducing corruption (Gigliolo 1996; Adsera et al. 2003; Van Rijckeghem and Weder 2001; Treisman 2000).

In this paper, we examine three possible mechanisms that could be used by citizens to reduce corruption in government: electoral accountability, voicing dissatisfaction and tax non-compliance. Whereas voice accountability and tax non-compliance have hardly been focused on before in this context, we study them along with electoral accountability in a controlled experimental setting and therefore are able to disentangle the effects of many other potential mechanisms which cannot be effectively controlled for when using observational data. We conduct an incentivized laboratory experiment to evaluate our arguments using a naturally occurring one-time rare opportunity for our study: Egypt during its brief democratic opening between 2011 and 2013 (our experiment was fielded in March 2013 while Mohamed Morsi was still in power).

Our experimental design adopts a public good approach to corruption which sees corrupt behaviour taking place whenever officials managing public goods transform those goods into

private ones (Booth and Cammack 2013; Rothstein and Torsello 2014; Rothstein and Varraich 2017) but further add to it by making public good contributions compulsory in the form of tax payments. At the beginning of the experiment, subjects earn money by performing a labour task and pay taxes according to their claimed income. They 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 fund a public good that benefits everyone equally and is more efficient (in equilibrium). Our first experimental manipulation is whether the official is subject to regular electoral accountability or not (*Electoral Accountability* versus *No-Electoral Accountability Treatment*). Our data suggest that electorally accountable officials are significantly less likely to choose the self-serving good (by 17%).

We then introduce a second accountability tool – sending messages to one another and to the official, or what we call *voice accountability* – to see how far subjects make use of such tool to voice dissatisfaction towards 'corrupt' officials and whether officials respond by increasing their choice of the 'fairer' good. We find evidence that subjects in the *No-Electoral Accountability* treatment do make effective use of such tool and that officials in this treatment do respond to such verbal accountability as the funding of self-serving good decreases (by 29%).

In all rounds, subjects can also respond to the official's behaviour by under-reporting their taxable incomes; this is the only economic channel available to discipline officials in the setting with no-electoral accountability. We, however, find no statistically significant effect of this potential mechanism – largely because our design makes it less economically rational for subjects to under-report their incomes.

From this point, this paper proceeds as follows. In the next section, we present how our experiment seeks to contribute to the literature. Section three outlines our theoretical construct and derivation of hypotheses. Sections four and five explain the experimental design and findings respectively. The sixth section concludes.

II. Literature and Contribution

On what factors could affect corruption, the literature has explored societal-historical factors (e.g. economic and political history, demographic and geographic factors), specific public policies (e.g. tax and trade policies) and institutional arrangements (for a review see Gerring and Thacker 2004). The effect of regime time on corruption has particularly been addressed by many theoretical and empirical, large-and-small-N, in addition to single case studies. The majority of such studies subscribe to a predominant view that democracies are less corrupt than autocracies (Chowdury 2004, Sandholtz and Koetzle 2000; Adsera et al. 2003; Besley and Case 2003; Ferraz and Finan 2011).[5] Deacon (2009) finds that autocracies provide significantly lower levels of public goods than democracies. Alt et al. (2009) show that economic growth is higher and taxes, spending, and borrowing costs are lower under reelection-eligible incumbents than under term-limited incumbents. Even within democracies, electoral rules that inject more accountability were shown to produce better outcomes (Ferraz and Finan 2011; see also Persson and Tabellini 1999).

The mechanisms focused on by such studies however have been different and certainly interacting with one another. The traditional and straightforward mechanism is that of elections. Regular electoral accountability is assumed to increase control by voters over officials to either

[5] A minority of studies however fail to find similar results (Mulligan et al. 2004; Lott 1999).

be less corrupt themselves or to more effectively fight corruption to extend their tenure in office (Ferejohn 1986; Persson and Tabellini 2000, Treisman 2000). Other mechanisms emphasize democracies' guarantee of free press (Giglioli 1996), well-informed electorates (Adsera et al. 2003), and effective judicial scrutiny (Rose-Ackerman 1999). The indirect effect of increased economic development in democracies which then decreases the incentives for corruption has also been explored (Van Rijckeghem and Weder 2001). On the other hand, some studies failed to find such a strong link between democracy and corruption (Ades and DiTella 1999; Treisman 2000) whereas research on East Asia – and the developmental state in particular – indicate that authoritarian regimes could perhaps be more effective in avoiding rent-seeking behaviour and corrupt responses to special interests (Haggard 1990; Evans 1995).

Such findings however are mostly based on large-N studies where the level of analysis is countries. Disentangling the effects of potential confounders however is usually best achieved in experimental designs where more control is possible and differentiated treatments could single out specific mechanisms (Plott 2001). Moreover, by adopting the country as the level of analysis, many of these studies do not involve the individual where the calculations of recalling/retaining an incumbent, and engaging in corrupt behaviour or not actually take place.

When it comes to relevant findings of experimental studies, there is evidence that elections have effects on decisions of punishment and rewards in public good games (Tyran and Feld 2006; Ertan et al. 2009; Walker et al. 2000; Kroll, Cherry, and Shogren 2007).[8] Abbink et al. (2002) introduce reciprocity games that mimic situations where corruption arises but make a third agent – a sudden death treatment – as the punishment of corrupt behaviour by others. Corazzini et al (2014) found that if elections are preceded by campaign promises, they induce officials to keep promises to avert the psychological burden of lying whereas Hamman et al.

(2011) found that elections make voters choose more pro-social officials. Bó et al. (2010) show that the impact and legitimacy of policies also increased when they were chosen democratically. Drazen and Ozbay (2014) present experimental evidence that elected leaders are significantly more likely to choose a policy not equal to their "type" than leaders who are appointed. Much of this literature however, by relying on simple public good games with voluntary contributions, fails to make the behaviour of the elected/appointed officials – managing those contributions – resemble as much as possible 'corrupt' real life behaviour. Indeed, behavioural drivers like fairness, reciprocity, and pro-sociality could be paying a role in standard public good games (Fehr and Schmidt 1999; Gintis 2003).

We seek to contribute to the literature on multiple grounds. *Firstly,* we bridge the literature on corruption that relies heavily on observational data with the control of the experimental methodology thereby controlling for democracy's other potential channels (e.g. educated electorate, free press, economic performance, etc.). *Secondly,* we aim to make the government official not responsible for distributing voluntary contributions in a public good game, but instead vested with the authority of distributing tax revenues paid by subjects who have earned their incomes after making real effort tasks. *Thirdly,* we adopt an experimental design that opens the black box of motives and emotions guiding the decisions of voters and those in command by allowing subjects to communicate (via text messaging) and then analyzing these chats. Finally, we conduct our experiment in a real-life transitional context.

III. Theory and Hypotheses[6]

Institutions affect behaviour because they incentivize certain actions and disincentivize others (North 1990). An institution/rule that allows for a direct or indirect punishment

[6] For a detailed theoretical construct and derivation of hypotheses, please check appendix I.

mechanism against politicians thus, such as the possibility of being forced out of office (direct) or tax evasion (indirect), can induce politicians to change their behaviour by aligning their actions with those of their electorates (Rose-Ackerman 1999). Authoritarian rulers, on the other hand, are likely to 'prey' upon societies if not restricted (North 1990).

Principal-agent approach, in particular, has been central in explaining the actions of officials (for a review see Besley and Case 2003). Although originally designed to examine relations within firms, the principal–agent model became the dominant framework in analyzing political accountability (Kiewiet and McCubbins 1991). According to this approach, elected politicians are the agents in an asymmetrical relationship whose other end – the principal – are voters empowered to discipline the agent on a regular basis for misbehaviour (Downs and Rockle 1994). Politicians calculate the costs and benefits of engaging in corrupt behaviour accordingly – whether material or non-material (e.g. societal retribution). As such mechanism operates more effectively in a democracy rather than an autocracy – because the threat of removing an official is more credible in the former – the costs of engaging in a corrupt behaviour are much higher in a democracy, making corruption less likely.[7] Our first hypothesis therefore is as follows:

> ***Hypothesis One****: Officials who are electorally accountable are less likely to engage in corrupt behaviour, compared to officials who face no such office threat.*

Electoral accountability however is only one source of accountability. Democracies usually have other non-electoral accountability mechanisms than elections. Moreover, elections have proved to have multiple problems in practice. On the one hand, they could happen with long intervals raising concerns about what forms of accountability could be carried out between

[7] Surely, legal prosecution for corrupt practices is also more effective in a democracy where judiciary is likely to be more independent. However, this is one other potential confounder associated with democracies that we do not focus on and thus whose effect we seek to disentangle in our theory and design.

elections. On the other hand, they are used to judge multiple issues and therefore reduce several dimensions into a single ballot making them sometimes potentially ineffective in enforcing accountability (Przeworski, Stokes and Manin 1999). Moreover, voting is often times influenced by intervening motives that might obscure the desire to punish or reward an official. Ideological, ethnic, or religious voting for example do not have to correspond to – and sometimes even contradict with – objective evaluation of incumbents. An ideological voter might continue to vote for an ideologically close official even if that official is corrupt just to block an ideologically-distant candidate out of power (Evans 2003). Furthermore, pork-barrel behaviour – which could take place in democracies as well as autocracies – has been repeatedly shown to weaken the accountability function of elections (Ferejohn 1974; Stratmann and Baur 2002).

We therefore wanted to examine how *voice* could also serve as a check on government corruption, especially in today's digital world of social platforms. The fact that our experimental context is Egypt – considered at the time of our fieldwork a major story of the Arab Spring – where the 2011 protests were called for, organized and largely managed on social media platforms (Clarke 2018) makes the testing of such accountability tool even more relevant. These are non-electoral but still vertical tools of accountability that aim at exposing governmental wrongdoing (Smulovitz and Peruzzotti 2000). Fung and Wright (2001) label them 'empowered participatory governance' that ensure accountability over public goods as schooling, policing and environmental protection. Sharma (2008) argues that citizens' capacity to express and exercise their views has the potential to influence government priorities and governance processes. Even if voice is not directly associated with material measures, the rhetoric it employs is seen as an effective mechanism in destroying flawed legitimacy claims (Rodgers 1987). We label this intervention 'voice accountability' (see Kaufmann 2004). In real life, it includes signing

petitions, writing complaints, lobbying or protesting. Our second hypothesis therefore is as follows:

> ***Hypothesis Two****: Officials who are exposed to criticism by subjects because of their perceived misuse of public money, are less likely to engage in corrupt behaviour.*

Non-compliance with the laws, especially laws generating the very revenue that citizens believe is being mismanaged by the government, is another mechanism of expressing dissatisfaction and checking the government's financial misconduct. In this paper, we focus particularly on tax non-compliance. In general terms, there are multiple factors that shape attitudes towards tax compliance. These include stigma of non-compliance (Cowell 1990), the effect of guilt and shame on compliance (Erard and Feinstein 1994), perception of others (Frey and Torgler 2007) and concern about others' welfare (Bosco and Mittone 1997). Politically motivated tax non-compliance has also been argued for as early as Karl Marx (Ireland 2019). More recently, tax non-compliance has been called for by protestors opposing diverse set of government policies, like nuclear armament, the Vietnam War and most recently anti-Trump protests[8]. Besley et al. (1997) have shown empirical evidence of politically motivated tax non-compliance in their analysis of how the response to the poll tax in the United Kingdom under Thatcher was driven by attitudes towards the Conservative party and where non-payment was an act of defiance or civil disobedience.

In an extensive review of the literature, examining religious, economic and philosophical views, McGee (2006) points to three views regarding the ethics of tax evasion, ranging from

[8] 'We will not pay: the Americans withholding their taxes to fight Trump', *The Guardian*, February 15th, 2017. Link: https://www.theguardian.com/us-news/2017/feb/15/tax-refusing-pay-protest-trump.

being never justified (because it is a duty to God, the government or fellow citizens) to always justified (mainly view of anarchists), with the middle and more widespread view making it evasion conditional on government actions. Case-study research finds that such middle view is the one held by tax payers in many countries (e.g. McGee and Galina 2006). More relevant to this paper, the Muslim view towards tax evasion seems to belong to the view seeing justifiability of tax evasion as dependent on circumstances (McGee, 1997; Ahmad 1995; Yusuf 1971). In the 2012 Egyptian wave of the World Values Survey, 60.7% of Egyptians said that cheating on taxes when one has the chance is never justified whereas around 39.3% indicated different degrees of justification.[9] Given the social – and even legal – desirability bias usually associated with answering such a question, the nearly 40% justifying tax evasion to some degree is likely to be reflecting a much higher actual rate.

The only way to discipline the official in the "No-electoral accountability" system is through low compliance. This is so as the payoff from the public good depends on the total tax revenue and the marginal return. While the marginal return of self-serving public good is larger for the official, she is worse off funding it if a consequence of doing so is substantially less tax revenue. Citizens can accomplish this by tax evasion. We therefore expect subjects in our experiment to use tax non-compliance as a tool in response to officials' corrupt behaviour and that officials would therefore respond accordingly so that:

> ***Hypothesis Three****: Officials who face higher rates of tax evasion as a protest against their rule by subjects, are less likely to engage in corrupt behaviour.*

[9] The complete data on the 2012 Egyptian wave is on the following link: http://www.worldvaluessurvey.org/wvs.jsp.

IV. Experimental Design

Our experiment had two experimental manipulations. The first is a*cross subjects,* measuring the effect of electoral accountability via two treatment arms (*Electoral Accountability* treatment and *No-Electoral Accountability* treatment). The second manipulation is *within subjects,* measuring the effect of voice accountability within each treatment arm. Each of the *Electoral Accountability* and the *No-Electoral Accountability* treatments had 14 rounds. In both settings, subjects are randomly matched into groups of five at the beginning of the experiment; groups remain fixed during the entire experiment (see Table 1).

Table 1: Summary of Treatments

Treatment Name	Decision Period	Electoral Accountability	Voice Accountability	Sessions (Groups)	Total Subjects
No-Electoral Accountability	1-7	No	No	2 (12)	60
	8-14	No	Yes, at midterm (after round 10)		
Electoral Accountability	1-7	Yes	No	2 (12)	60
	8-14	Yes	Yes, at midterm (after round 10)		
Total				4 (24)	120

At the beginning of round 1 for each group, an initial official is randomly selected by the computer from among the five group members. At the beginning of each round, subjects perform a task to earn experimental pounds (up to EGP 20). The task is as follows. For each subject, a paragraph written in Arabic language appears on their computer screen.[10] There are 10 mistakes

[10] We designed this real-effort task to make detecting mistakes fairly easy. The purpose is not to discriminate between subjects based on their Arabic language skills.

per paragraph and subjects are asked to identify and correct these spelling mistakes on the screen. They are paid 2 experimental pounds per accurate correction and the maximum earnings is 20 experimental pounds. At the end of the task, they are told how much income they have earned from their editorial task.[11] All subjects (citizens and officials) then decide how much income to report given an announced tax rate of 25%. 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.[12] Subjects are told that one out of the five members of each group will be randomly selected to be audited.[13]

Total taxes paid by group members are used to fund one of two feasible public goods (C-good and G-good) to be chosen by the official of the group.[14] If the group official decides to fund the C-good, then tripled tax revenues are equally distributed among group members (i.e., mpcr=3/5).[15] However, if the group official decides to fund the G-good, then half of the amount goes to the official (mpcr=3/2) whereas the remaining half is distributed equally among other group members, i.e., the other four citizens (mpcr=3/8). Thus, while the benefits from the C-good are the same across group members, the G-good provides more benefit (four times as much as the citizen) to the official at a cost of reduced marginal benefit (from 3/5 to 3/8) to the citizens. Funding a G-good (i.e. using tax revenues to disproportionally enrich the group official at the expense of others) is therefore our measurement of corrupt behaviour. The frequency of

---

[11] Please check appendix II for more details on the task as per the instructions given to subjects.
[12] Please check appendix III for the penalty structure.
[13] The purpose of the labor task, auditing and penalties was to have subjects feel entitled to their earnings and experience trade-offs similar to the ones faced in the real life where consequences of non-compliance are stochastic.
[14] 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, public good benefits and other details. Instructions (in English) are included in Appendix II.
[15] 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 in our experiment) of the total individual investments in the public good (i.e., total tax revenue in our setting) among group members.

choosing self-serving public goods usually is a stylized measure of corruption (Booth and Cammack 2013; Rothstein and Torsello 2014). According to Rothstein and Varraich (2017), when a public good is distributed according to the private wishes of those managing it, a conversion of the public good into a private one happens and corruption takes place.

After the choice between C and G goods is made by the official, the net of the round's earnings for each subject is calculated (earned income minus taxes less penalties, if audited, plus the payoff from the public good chosen by the official). The game is played for 14 rounds. Net earnings from all rounds are averaged for final payment at the end of the experiment as explained to the subjects before the experiment is run.

The *No-Electoral Accountability* and *Electoral Accountability* settings differ as follows. In the *No-Electoral Accountability* treatment, the official remains in office for seven rounds after which a new official is randomly selected and the experiment continues for seven more rounds (i.e., until the 14th round). In the *Electoral Accountability* treatment, the official could be recalled after each of the 14 rounds if the majority of group members (including the official)[16] agree to recall him/her, in which case a new official is randomly selected by the computer from eligible members.[17] A random selection of an official takes place only in *Electoral Accountability* treatment if the initial official was never recalled for seven consecutive rounds. Whereas elections in reality combine the decision to hold the incumbent to account while at the same time also selecting potential successor (Powell 2000; Corazzini et al. 2014), untying the two mechanisms – accountability and selection – is what we aim to achieve from such a design. By

[16] This is our implementation of supermajority as the majority here is the same as three out of four citizens voting to vote out the official. As an official would not vote to recall himself (confirmed in our data by 98.21% of our "officials" votes), 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.

[17] A group member is eligible if he/she has not been recalled during the last three elections.

allowing subjects to recall officials but not elect them, we are able to single out – and hence validly test – the effect of accountability on incumbent's behaviour.

To imbed *voice accountability* in the design, we allow group members to communicate via sending text messages after the tenth round in both treatments. Sent messages are shown to all group members and are saved by the experiment software (we show the analysis of these messages below). Subjects were not allowed to communicate with one another during the experiment other than the messaging allowed after round 10.[18] Such an intervention allows for within-subjects testing of how potentially voicing dissatisfaction towards the official could affect his/her behaviour within each treatment.

All 120 subjects (60 subjects in each treatment; two sessions per treatment where each session was run with 30 subjects) who participated in the experiment were volunteers from undergraduate classes at an Egyptian public university. Each subject participated only once in the experiment. At the end of the experiment, subjects were paid the average of all 14 rounds. Average payoffs were US$ 25 per subject.[19] The experiment lasted approximately two hours and was conducted in Arabic.

V. Experimental Results

5.1 Predictions and Incentives

Before we report subjects' behaviour, it is important to look at subjects' incentives across the *No-Electoral Accountability* and *Electoral Accountability* treatments, given the parameters used in the experiment. We begin by noting that the maximum feasible group payoff is reached

[18] After completion of the decision tasks, subjects completed an online questionnaire that included questions designed to get information about idiosyncratic individual characteristics.

[19] At the time of the experiment, the exchange rate was: 1 USD = 6.78 EGP. An average hourly rate in 2013 is 33 EGP (CAPMAS 2013). Thus, each subject earned at least twice what he could have earned outside the lab per hour.

when there is full compliance and it is EGP 150 (see Table 2, last row). This is so regardless of which good is funded. The type of the good being funded, however, affects the distribution of earnings.[20] If the C-good (equal distribution of revenues) is funded, it is optimal to report 65% of income for citizens as well as the official. If the G-good is funded however, it is optimal for the official to fully report earned income whereas for a citizen to report 15% of earned income. For optimal claims of income, the expected (round) payoff is EGP 27 for everyone if the C-good is funded; the round payoffs are EGP 32 and EGP 21 for the official and the citizen, respectively, if the G-good is funded. Thus, funding C-good is both more just and more efficient as optimal declared income is larger. Economic efficiency is 90% (=135/150) for the C-good and down to 77% (=116/150) for the G-good. Yet, the official can increase his/her round payoff by 18.5% if he or she funds the G-good.

The subgame perfect equilibria (SPE - see appendix I) predicts that only the G-good will be funded in either treatment/game. Given the parameters used in our experiment though, playing the game for seven rounds is not long enough to support funding of the C-good in equilibrium. To see why, note that if the official defects by funding the G-good, then the instantaneous gain is 22.5 whereas any future round comes with a loss of 1.5. As there are at most six future rounds, it is thus optimal for the official to always fund the G-good.[21] Thus, it is predicted that in the *No-Electoral Accountability* treatment, the official gets corrupt and uses tax

[20] With full compliance, penalty is 0: If the C-good is funded, each group member earns EP30; if the G-good is funded then official's and citizen's earnings are EP52.5 and EP24.38 respectively.

[21] Under full cooperation, each subject's income tax is 5(=0.25*20), so the total tax revenue is 25. If the official funds the C-good then his own payoff from the public good is (3/5)*25=15. If the official funds the G-good then his payoff from the public good is (3/2)*25=37.5. So, the instantaneous gain is 22.5. If the G-good funding triggers full retaliation (citizens claiming 0 income) then in the remaining rounds the expected total tax revenue is 5 (paid by official) plus an expected 4 paid by citizens (0 if no citizen is audited an event with likelihood 1/5 in our experiment and 5 if a citizen is audited, an event with likelihood 4/5). Therefore, the expected tax revenue for each round after r^* is 9 and official's return from the G-good is 13.5 (=(3/2)*9). Summarizing, the sequence of public good benefits from funding good C is (15, 15, 15 …) whereas from funding good G the expected stream of public good benefits is (37.5, 13.5, 13.5, …). For the earlier to be preferred by a selfish official the game needs to continue for at least 12 rounds after r^*.

proceeds to fund the public good that benefits own self at the expense of citizens. A detrimental consequence of this is low tax compliance by citizens.[22] In the *Electoral Accountability* treatment however, if C-good is always funded, with full cooperation, the sequence of payoffs is (30, 30, 30, 30, ….)[23] whereas with a defection to funding the G-good, with full retaliation, the sequence of payoff is (52.5, 19.5, 19.5, 19.5…..) as the official gets recalled and becomes not eligible for three rounds. Therefore, the defection is profitable in the last two rounds but not before. Thus, given the parameters used in our experiment, the alternative hypotheses (derived in appendix I) are one sided:

H1b*: The G-good (self-serving public spending) is likely to be more often chosen in the No-Electoral Accountability than in the Electoral Accountability treatment.*

H2b*: Tax Compliance is likely to be higher in the Electoral Accountability than in the No-Electoral Accountability treatment.*

5.2 Findings

Table 2 reports summary statistics for both, the *Electoral Accountability* and the *No-Electoral Accountability* treatments, along with the embedded *no-voice accountability* (rounds 4 to 7) and *voice accountability* interventions (rounds 11 to 14).[24] The G-good was chosen much less frequently (44% of the time) in the *Electoral Accountability* treatment compared to the *No-Electoral accountability* treatment (56% of the time) which supports our theoretical expectation. Also confirming our expectation is the fact that the highest percentage of G-good choice happened when *No-Electoral Accountability* was combined with *No-Voice Accountability* (65 %

[22] Without retaliation, the optimal compliance rates of a payoff maximizing citizen are 15% if G-good is funded and 65% if C-good is funded (see Table 2). With full retaliation compliance rate is obviously 0.
[23] 15 from the public good and 15 from the after-tax income.
[24] The first three rounds in each block are not included to control for learning and order effects.

of the time). The introduction of *Voice Accountability* within each treatment seems to have decreasing effects on choosing G-good in the *No-Electoral Accountability* treatment but a positive effect in the *Electoral Accountability* treatment. As could also be seen from table 2, citizens' tax compliance rate (reported income/actual income) was higher in the *No-Electoral Accountability* treatment (77.07.%) compared to the *Electoral Accountability* treatment (69.56%).

Table (2): Summary statistics of subject's decisions and SPE predictions

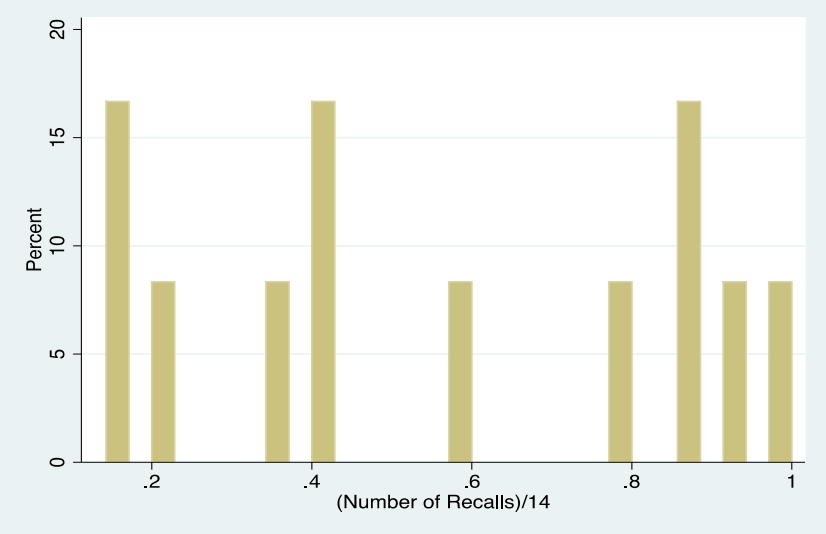
Treatment/Institution	Voice accountability (rounds)	Frequency of G-good	Tax Compliance (% of Income)		Earnings	
			Citizen	Official	Citizen	Official
No-Electoral Accountability	No (4 to 7)	0.65 (0.483)	77.07 (0.316)	92.53 (0.178)	22.09 (3.655)	34.97 (9.083)
Electoral Accountability		0.44 (0.501)	69.56 (0.325)	73.24 (0.355)	22.06 (4.696)	31.02 (9.288)
No-Electoral Accountability	Yes (11 to 14)	0.42 (0.498)	75.59 (0.358)	81.11 (0.326)	23.37 (4.856)	32.51 (9.952)
Electoral Accountability		0.56 (0.501)	68.92 (0.350)	81.52 (0.270)	22.48 (4.421)	33.53 (8.947)
Predictions (payoff maximizing)		1.00	15 if G 65 if C	100 if C 65 if C	21 if G 27 if C	32 if G 27 if C
Maximum Feasible			100	100	24.38 if G 30 if C	52.5 if G 30 if C

Notes: Standard deviations in parentheses.

Role Assignment. Of the 60 subjects who participated in the *No-Electoral Accountability* treatment, 40 subjects (67%) never served as a group official, 16 subjects (27%) served as group officials for 7 rounds, and 4 subjects (7%) served for 14 rounds. In the *Electoral Accountability* treatment, on the other hand, of the 60 subjects, there were 6 subjects (10%) who never served as officials, 5 (8%) subjects served for 7 or 8 rounds, and no subject served for more than 8 rounds; half of the subjects (30) served as group officials for 2 or 3 rounds.

Figure 1 shows the empirical distribution of recalls across the 12 groups. The mean likelihood of a recall is 56 percentage points (st.dev. =0.498). Clearly, our subjects weren't shy of exercising the option to recall the group official. However, the 56% is a far cry from 100% rate of recall predicted by the SPE (hypothesis H3o in appendix I). So, what determines the likelihood of recalling the government official?[25]

Figure 1: Histogram of Recalls



Recalls. If an intrinsic need for power is the main driver for a recall then the likelihood of recalling the group official should not depend on his/her choice of the public good, consistent with null hypothesis *H3o*. On the other hand, an official can reduce the likelihood of being recalled in the *Electoral Accountability* treatment by choosing to fund the C-good (alternative hypothesis *H3a* in appendix I). It should be noted that an individual voting for a recall does not necessarily mean he/she is retaliating (or negatively reciprocating) as such a vote is self-serving for it improves the odds of the individual to serve as the group official. Examining the data at the aggregated level, we find that the likelihood of a recall is 16.28 percentage points following a C-

[25] Voting an official out of office requires at least three votes. We can safely rule out that the high rate of vote outs is a result of trembles/ noise (such as subjects submitting 'vote out' when they meant to submit 'do not vote out').

good and almost six times as high, 97.56 percentage points, following a G-good funding. In addition, we looked at two categories of groups: committed (6 groups that recalled fewer than half of their officials) and volatile (6 groups that recalled more than half of their officials). The likelihood of recalling well-behaved officials is a low 6.45% for committed groups but is rather high, 41.67% for volatile groups. On the other hand, the bums are thrown out almost always: 90.91% (committed groups) and 100% (volatile groups) of officials are recalled following a G-good funding.

Data from the volatile groups thus suggest that there is some evidence for intrinsic need for power behind recalls. However, the effect of official's choice to fund the G-good seems to be the main cause of recalls. A probit regression (with standard errors clustered at group level) supports this hypothesis (see table 3). 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 percentage points. We conclude that our data reject the null hypothesis (*H3o*) of intrinsic need for power in favour of the alternative hypothesis (*H3a*) that corruption triggers recalls.

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

Table 3: Probit Regression of Recalls (Marginal Effects)

G-good Funding (D)	0.816***
	(0.059)
Voice	-0.094
	(0.216)
Female	0.072
	(0.135)
Muslim	-0.177
	(0.172)
Senior students	0.023
	(0.117)
Being Pro-Democracy	-0.057
	(0.116)

Period	0.029
	(0.024)
Observations	168
Number of Clusters	12
Pseudo R-squared	0.601
Wald chi2(7)	310.0
P > Chi2	0.000

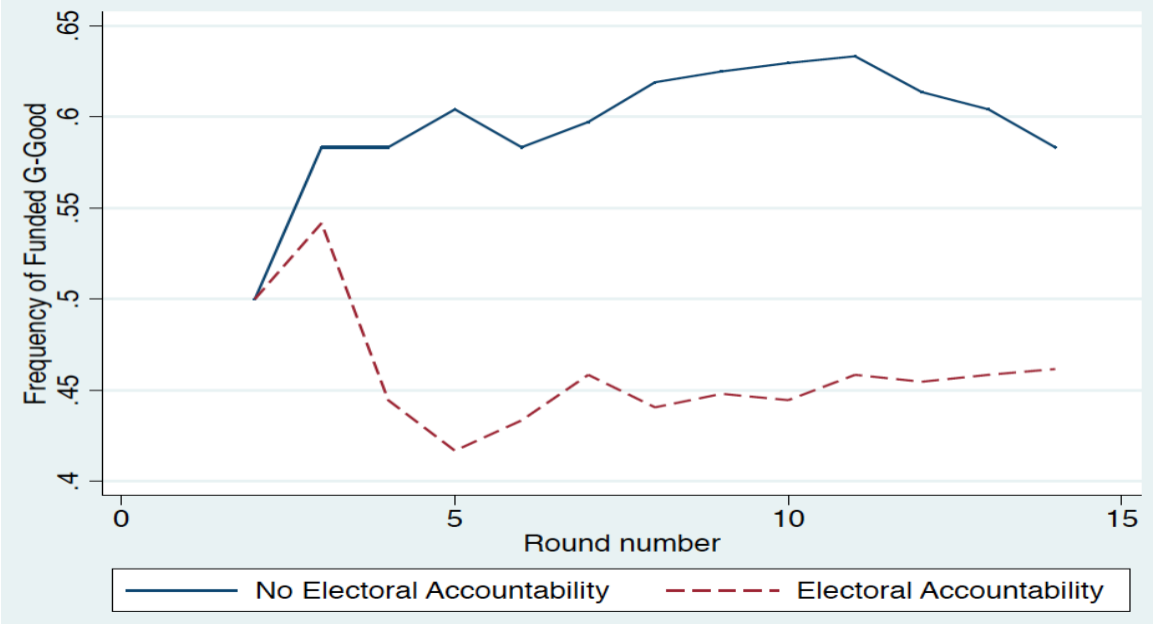
*Notes: Data used is for the Electoral Accountability treatment only. Number of clusters is the same as the number of groups, 12. Number of observations is 168 (=12 groups times 14 periods.) Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Official's Behaviour. The strong effect of officials' decisions on the likelihood of a recall brings to the forefront the question of the interaction between political institutions and governance failure. Data from the end game (round 14) supports the hypothesis that it is the fear of being recalled 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 *Electoral Accountability* treatment. In the last round of the experiment (when the recall comes with no consequences) we observe that 83% of the officials fund the G-good which is not statistically different (Pearson chi2(1) = 0.25, p-value = 0.615) from behaviour of officials in the *No-Electoral Accountability* treatment: 75% of officials choose to fund the G-good in round 7, the last round before a scheduled official replacement in the *No-Electoral Accountability* treatment takes place.[26] On the contrary, for round 7 (which is not the end of term in office for 83% (10 out of 12) of officials in the *Electoral Accountability* treatment) only 33.33% of the officials funded the G-good (Pearson chi2(1) = 4.20, p-value = 0.041).

[26] Data from round 14 in the *No-Electoral Accountability* treatment are less informative for comparison as we find a strong effect of communication (after round 10) in the *No-Electoral Accountability* treatment but not in the *Electoral Accountability* treatment.

Data from all rounds is consistent with hypothesis *H1b*, that *Electoral Accountability* institution is performing better than the *No-Electoral Accountability* one in restraining the official from taking advantage of being empowered with making decisions on the use of public funds. With groups as the unit of observation, we find that the mean of G-good funding rate, until the voice event, is 63% (95% confidence interval is (0.41, 0.86)) in the *No-Electoral Accountability* treatment and down to 46% (95% confidence interval is (0.27, 0.64)) in the *Electoral Accountability* treatment. To capture the dynamics at the 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 rate that the G-good is funded up to round *t*. Figure 2 shows evolution of TFG values across the two treatments.[27]

Figure 2: Time Frequencies (at group level) of G-good being funded



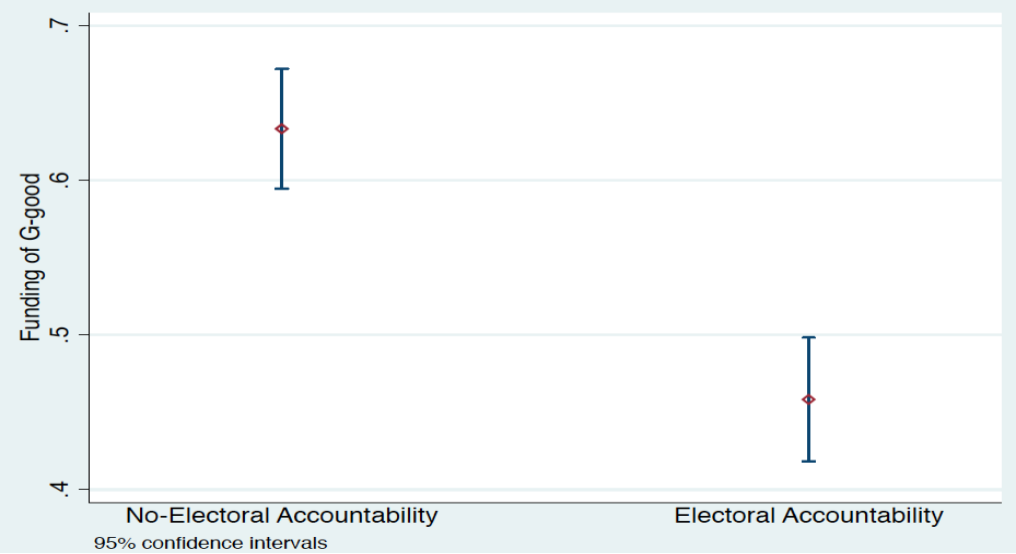
A visual inspection of Figure 2 suggests that: (i) the prevalence of corruption is negatively affected by the existence of electoral accountability; dotted line is everywhere (except

[27] Data points at each round correspond to the averages of the TFG across groups at a given treatment.

at the beginning rounds) below, by at least 10%, the solid line (*No-electoral Accountability*), (ii) officials in the *Electoral Accountability* treatment learn quickly (as early as round 3) to fund the G-good less often with behaviour seemingly stabilized around 45%, but not in the *No-Accountability* treatment where (iii) there is a persistent upward trend in the frequency of G-good funding until the voice event.

Testing Electoral Accountability. Figure 3 below shows means of corruption across our electoral and no-electoral accountability treatments, the number of times the official chose to fund the G-good in each group in each round. We label this dependent variable "*Level of Corruption"*. As shown in the figure, officials were significantly more *corrupt* in the *No-Electoral Accountability* treatment.

Figure 3: Mean Level of Corruption across treatments, before introduction of voice accountability[28]



[28] To get clean treatment effects of the *Electoral Accountability* intervention, we restrict our results of this treatment to the first 10 rounds before the introduction of the *voice accountability* mechanism which had its own effect on officials' corrupt behaviour as will be shown below.

To examine how far this effect survives while controlling for potential confounders, we ran a probit regression with the dependent variable, corruption rate, being a dummy that takes the value of 1 (0) if the group official decides to fund G-good (C-good). We control for factors like the historical rate of G-good being funded in the official's own group, whether in the preceding round the official funded the C-good and remained in office, individual idiosyncratic characteristics of the official such as gender, religion, attitudes towards democracy, and year of study. Table 4 reports the marginal effects on the probability of funding G-good. The results support the conclusion that in the absence of electoral accountability the likelihood of corruption goes up by 17 percentage points, rejecting the null hypothesis *H1o* in favour of the alternative hypothesis *H1b*. Estimates also reveal that previous frequency of corrupt behaviour (within the official's group) and the official being a risk-lover are positively associated with official's self-serving choice of funding G-good.[29]

Result 2: *Electoral accountability institution leads to less corruption in government.*

Table 4: Probit Regression of G-good Funding (Marginal Effects) absent Voice Opportunity

	(1) All Data
Provision of C-good and staying in office in previous round	-0.063
	(0.115)
Historical Frequency of G-good funding in official's group	0.578***
	(0.147)
No-Electoral Accountability	0.168**
	(0.083)
Official being a Female	-0.040

[29] Strangely, being pro-democracy makes the official also more corrupt. However, given that the democratic support question traditionally suffers from social desirability bias makes us cautious in making conclusions about this result.

	(0.100)
Official being a Muslim	-0.029
	(0.121)
Official being Prodemocracy	0.206**
	(0.087)
Official being a senior student	-0.157
	(0.114)
Period	0.004
	(0.014)
# of Observations	216
# of Clusters	68
Wald chi2(8)	41.82
Prob>chi2	0.000
Pseudo R2	0.201
Obs. Pr (G-good)	0.551
Predicted Pr(G-good)	0.564

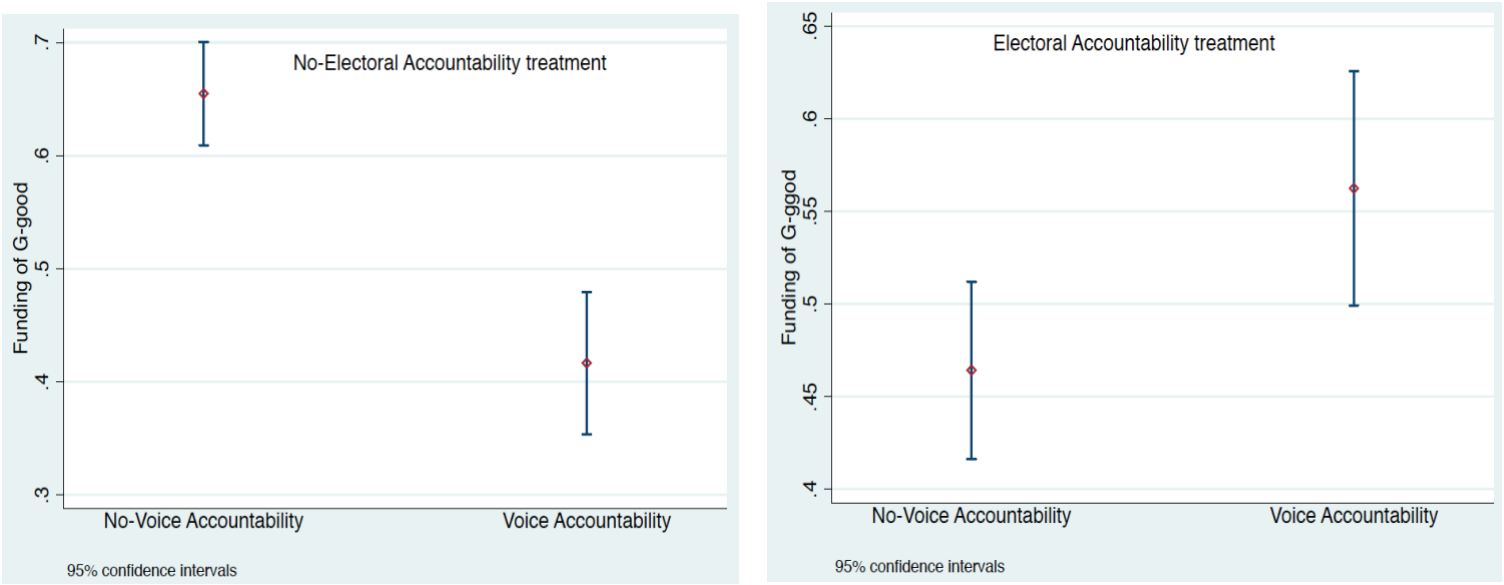
*Notes: Data are only from period 2 to 10. Period 1 data not included as the first regressor is defined for t >1. After chat data (period 11 to 14) not included as voice opportunity is available after round 10. Number of clusters is the same as the number of subjects who served as officials, 68 from period 2 to period 10. Number of observations, 216 is 24 groups times 9 periods. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.*

Testing Voice Accountability. We now move to test the second hypothesized variable; *voice accountability*. The design allows us to test the effect of such mechanism within subjects (before and after chatting in each treatment). As per figure 2 above, *voice* (after round 10 when it was introduced) seems to have a positive effect on restraining self-serving choices by officials in the *No-Electoral Accountability* treatment but not in the *Electoral Accountability* treatment. In Figure 4, we show the mean level of corruption before and after voice introduction. The probability of funding G-good decreased significantly from 63.33 to 41.67 percentage points after chatting was introduced in the *No-Electoral Accountability* treatment. Surprisingly, in the

Electoral Accountability treatment, voice produced significant results but in the opposite direction; increasing officials' corrupt behaviour.

Figure 4: Mean Level of Corruption in the *No-Electoral Accountability* and *Electoral Accountability* treatments, with and without Voice Accountability[30]

(a) **(b)**



Testing this effect via a probit regression (see table 5) that includes a number of controls, confirms the statistically significant decreasing effect of *voice accountability* on corruption in the absence of electoral accountability (model 1). The positive effect in the *Electoral Accountability* treatment does not survive the regression test (model 2). When we interacted *voice* with *No-Electoral Accountability*, the effect appears negative and statistically significant (model 3). We

[30] When testing the effect of voice accountability, we only compare rounds 11-14 (the post-chat rounds) to rounds 4-7 to ensure a balanced comparison.

therefore conclude that, in the absence of electoral accountability, voice accountability significantly lowers the likelihood of corruption by almost 28.6 percentage points.[31]

Result 3: Voice *accountability leads to less corruption in government, only when there is no electoral accountability.*

Table 5: Probit Regression of G-good Funding (Marginal Effects)

	Model (1) No-Electoral Accountability treatment	Model (2) Electoral Accountability treatment	Model (3) All Data
Provision of C-good and staying in office in previous round	-0.280*	-0.382***	-0.371***
	(0.163)	(0.104)	(0.090)
Voice Accountability	-0.286*	-0.016	0.092
	(0.170)	(0.161)	(0.118)
Official being a Female	0.047	-0.085	-0.047
	(0.117)	(0.096)	(0.077)
Official being a Muslim	0.003	-0.245*	-0.022
	(0.126)	(0.135)	(0.120)
Official being Prodemocracy	0.430***	0.043	0.203***
	(0.097)	(0.100)	(0.074)
Official being a senior student	-0.329**	-0.176	-0.182*
	(0.135)	(0.130)	(0.098)
Period	-0.007	0.021	0.005
	(0.016)	(0.020)	(0.013)
Voice with No-Electoral Accountability			-0.362***
			(0.119)
No-Electoral Accountability			0.200**
			(0.083)
# of Observations	156	156	312
# of Clusters	20	54	74
Wald chi2	-	18.05	35.00
Prob>chi2	-	0.012	0.000
Pseudo R2	0.254	0.143	0.172
Obs. Pr (G-good)	0.577	0.487	0.532
Predicted Pr(G-good)	0.585	0.485	0.534

Notes: Data from period 1 not included as the first regressor is defined for t >1. Number of clusters is the same as the number of subjects who served as officials: 20 and 54. Number of

[31] The persistency of the effect of communication on official's behaviour 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.

*observations, 156 is 12 groups times 13 periods. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

There are multiple ways to interpret the conditional effect of voice accountability on corruption. It could be that voice accountability is used more effectively in the *No-Electoral Accountability* treatment compared to the *Electoral Accountability* one given that it is the only accountability tool available for subjects. It is therefore not potentially crowded out by electoral accountability. From another perspective, there could have been a time effect. Subjects in the *No-Electoral Accountability* treatment had to wait 10 rounds being unable to express any protest against their corrupt officials – either in action by voting them out or in words – which then made voice accountability (when finally made available) amalgamate accumulated feelings of anger. Both explanations could have practical implications for institutional design. On the one hand, regular and evenly-balanced tools of oversight are likely to ensure that corrupt behaviour by officials could be tackled early on, avoiding huge societal losses. Secondly, longevity of corrupt officials in office does not mean that repressed subjects would get submissive over time, but in fact are likely to get 'angrier' in demanding accountability when the chance finally arises. In real life, this could mean mass protests, uprisings and revolutions similar to the bursts we have seen in the Arab World since 2011 and the subsequent waves seen in Sudan and Algeria in 2019.

As a robustness check, we examined whether voice accountability was actually used more prominently in the *No-Electoral Accountability* treatment by analyzing the content of the sent messages. The purpose is to investigate whether – and how far – the chat in the *No-Electoral Accountability* treatment included messages demanding the official to be less corrupt by choosing the C-good more than G-good. To that end, we hand-coded the content of the messages in both treatments along seven categories as shown in table 6. The total number of messages sent by subjects was 415. As could be seen from the table, messages of protest were much more

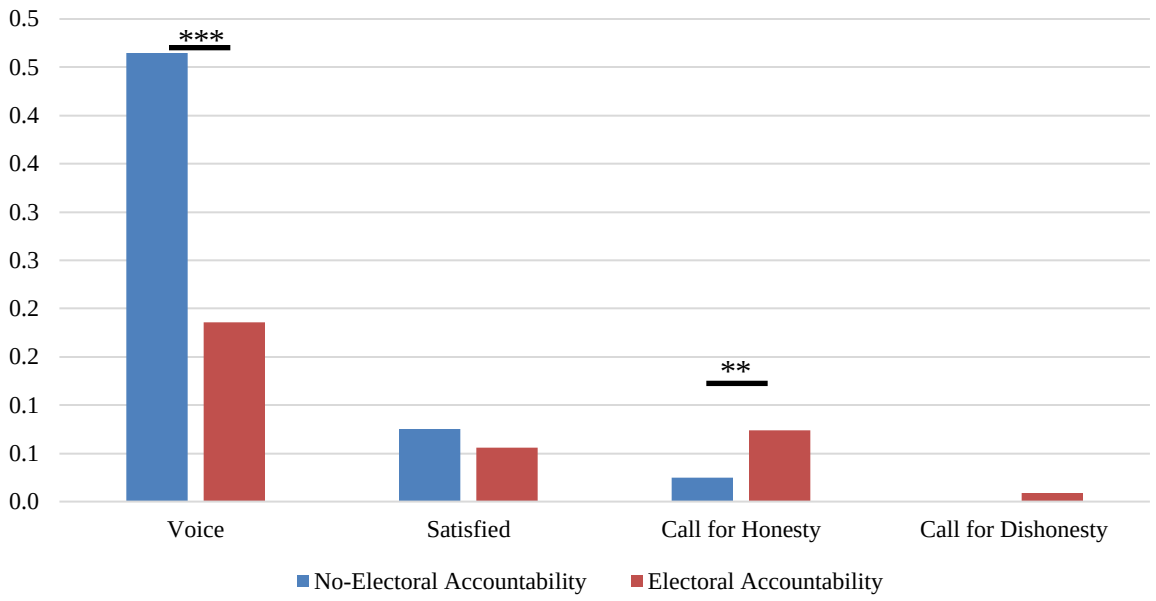
prevalent in the *No-Electoral Accountability* treatment (mentioned in 46.5% of all messages) compared to *Electoral Accountability* treatment (mentioned in only 19% of messages) – a difference that is statistically significant (see figure 5). This outcome indicates that subjects in the *No-Electoral Accountability* treatment used the chats as a replacement mechanism for the fact they could not recall corrupt officials and that the official's reduced frequency of choosing the G-good is likely to be a response to such demands.[32]

Table 6: Percentage of mentions of each code per treatment

Coding Category	Content of Code	*No-Electoral Accountability*	*Electoral Accountability*
Message of Protest	Expression on anger/dissatisfaction towards the official. Demanding more choice of C-good over G-good.	46.5%	19%
Expressing Satisfaction	Satisfaction with pay-offs or behaviour of others, or experiment in general.	7.5%	5.6%
Honesty Call	Call upon others to report their true incomes.	2.5%	7.4%
Dishonesty Call	Call upon others to cheat when reporting their taxes, or bragging about one's cheating behaviour.	0%	1%
General	Getting to know one another, asking irrelevant questions, being funny, etc.	29.5%	56.3%
Official sending a message	The official making comment, while making it clear that he/she is/was the official.	7.5%	8.8%
Facts	Stating the difference between C-and-G-goods, how many times one got audited, how many rounds remain in the experiment, etc.	6.5%	0%
Revealing ID	Revelations of one's name, appearance, or clothes.	0%	2%

[32] Certainly, there is also the effect of having less corrupt officials in the Electoral Accountability treatment which would then make citizens less critical of their officials in the voice event.

Figure 5: Chi[2] test comparing the different messages' codes across treatments[33]



Testing the effect of tax non-compliance. Moving to our third and final accountability tool subjects could have exerted on a corrupt official – tax non-compliance – we ran probit regressions for each treatment and then for the whole data set where we included tax compliance as a predictor of officials' corrupt behaviour.[34] We measured tax compliance by two variables; the amount the official paid in taxes "*Tax Paid by Official*" as her benchmark, and "Average Tax Paid by Others < Tax paid by Official" which is a dummy variable that, in any given period, takes value 1 if the average tax paid by the other four group members is smaller than the tax paid by the official.[35] As per table 7, we find statistical significance of tax compliance on officials' behaviour. Specifically, "Average Tax paid by Others < Tax paid by Official" has a negative and significant effect on the dependent variable "G-good funding". This is consistent with H3

[33] The fact that calls for honesty in reporting income are significantly more in *the Electoral Accountability* treatment reflects – as per the calculations above (and in appendix I) – that it is better to honestly report income in this treatment.
[34] We acknowledge an anonomous reviewer for the the idea to add "Average Tax Paid by Others < Own Tax" and "Tax Paid Official" in the list of explanatory variables to act as a benchmark for the extent of relative tax evasion.
[35] The result of the audit in each round was never formally revealed to the official.

suggesting that behaviorally tax non-compliance can be an effective institution in the no-electoral accountability treatment; decreasing corruption by 27%.

Table 7: Probit Regression of Level of Corruption

	(1) No-Electoral Accountability treatment	(2) Electoral Accountability treatment	(3) All Data
Average Tax paid by Others < Tax paid by Official	-0.270**	-0.057	-0.168*
	(0.117)	(0.143)	(0.092)
Tax paid By Official	0.095	-0.011	0.063*
	(0.081)	(0.044)	(0.033)
Provision of C-good and remaining in office	-0.269*	-0.376***	-0.367***
	(0.162)	(0.104)	(0.091)
Voice Accountability	-0.308*	-0.027	0.071
	(0.176)	(0.161)	(0.119)
Official being a Female	0.026	-0.088	-0.039
	(0.138)	(0.097)	(0.081)
Official being a Muslim	0.047	-0.253*	-0.034
	(0.121)	(0.138)	-0.119
Official being pro-Democracy	0.421***	0.038	0.211***
	(0.102)	(0.102)	(0.073)
Official being a senior student	-0.322**	-0.164	-0.174*
	(0.141)	(0.130)	(0.098)
Period	-0.007	0.024	0.006
	(0.016)	(0.020)	(0.013)
Voice x No-Electoral Accountability			-0.359***
			(0.125)
No-Electoral Accountability			0.188**
			(0.084)
# of Observations	156	156	312
# of clusters	20	54	74
Wald chi2	_	18.48	42.99
Prob > chi2	_	0.030	0.000
Pseudo R2	0.268	0.146	0.179
Obs. Pr (G-good)	0.577	0.487	0.532
Predicted Pr(G-good)	0.587	0.485	0.535

Notes: Data from period 1 not included as the third regressor is defined for t >1. Number of clusters is the same as the number of subjects who served as officials: 20 and 54 for the No-Electoral Accountability and the Electoral Accountability treatments respectively. *Number of observations, 156 is 12 groups times 13 periods. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1*

To check whether the effect of voice depends on the nature of the messages, we created a new variable, "$\Delta Corruption$", a within-group difference between average G-good funding three periods after and before the chat; that is for each group *i*,

$$\Delta Corruption_i = \frac{1}{3}\sum_{t=11}^{13} G_i(t) - \frac{1}{3}\sum_{t=8}^{10} G_i(t)$$

where *G(t)* is a binary variable that takes value 1 if G-good is funded in period *t*. The list of regressors include percentages of group messages in categories Protests, Honesty Call, Facts, and Expressing Satisfaction. Table 8 shows the least square estimates. One percent increase in protest messages decreases the group's mean corruption level by 1.3 percentage points. The effect of "honesty call" messages is estimated to be twice as much.

Table 8: Least Square Estimates of Voice Effect on Corruption

Dep. Variable: $\Delta Corruption$	
Protests	-0.013***
	(0.004)
Honesty Call	-0.024***
	(0.008)
Facts	0.017
	(0.013)
Satisfaction	-0.009
	(0.007)
Electoral Accountability	0.068
	(0.185)
Constant	0.378
	(0.249)
# of Observations	24
F(5,18)	3.465
Prob>F	0.023
R-squared	0.490

Notes. Number of observations is the number of groups in the experiment. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

VI. Conclusion

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 self-serving officials and other pork barrel projects. Can institutions of electoral accountability restrain self-serving behaviour of representatives? Through a laboratory experiment, we differentiated officials' behaviour regarding use of public funds in situations that allow "citizens" to vote out the official. We find evidence that the culture of power abuse is sensitive to the political institution in place. We find that the prevalence of corrupt behaviour falls early and significantly in the electoral accountability treatment. We also observe positive effects on corruption of giving citizens voice accountability tools even when they are not regularly empowered to change their governments via elections.

We had a one-time rare opportunity for our study: Egypt during its brief democratic opening between 2011 and 2013 (our experiment was fielded in March 2013 while Mohamed Morsi was still in power). We took advantage of that period for three purposes. *Firstly,* having experienced a mass uprising that toppled the 30-year rule of former president Hosni Mubarak just 18 months earlier – in addition to several waves of mass protests in the following months that managed to remove two interim prime ministers – makes Egypt at the time a suitable backdrop of recent successful experience of enforcing mass accountability. One of the primary motivations for the uprising was indeed a feeling among the protestors that corruption was rampant in government (Lahlali 2014). *Secondly,* Egypt in March 2013 provided a real-life milieu for a country in transition. It held its first post-uprising democratic parliamentary and presidential elections in January and June 2012. But on the other hand, it was still writing its new constitution and still had not made all its institutions popularly accountable. This transitional democracy status makes the Egyptian context – at the time – theoretically interesting given

research findings that the reductionist effect of democracy on corruption does not manifest itself unless a democratic culture takes root over time (Mohtadi and Roe 2003; Rock 2008; Rose-Ackerman 1999). In fact, processes of democratization were shown to have unpredictable effects on corruption (Moran 2001) with the literature on Southeast Asia, Latin America, and former Soviet Union countries indicating that corrupt practices in fact increase with democratization processes (Cohen, 1995; Harris-White and White, 1996; Pellegata 2013). *Thirdly*, we are responding to a traditional criticism to the experimental literature that it draws most of its conclusions from lab experiments done among subjects in western, industrialized, rich, democracies. By conducting our experiment with Arab, predominantly Muslim subjects who have lived under authoritarianism for most of their lives, we have a unique opportunity to explore dynamics usually overlooked by previous experimental studies, except for few (Hassan and Shalaby 2019; Hassan et al. 2020; Haas et al. 2020).

Our results have significant implications. Firstly, subjecting officials even to the minimal checks of accountability matter as far as reducing corruption is concerned. In our experiment, officials did not have to face challenging candidates, effective and independent judiciary or internal party rivalry. Instead, just facing the threat of a vote of censure on a regular basis incentivised them to be less corrupt compared to officials who did not face such threats. This result indicates that although scholars and practitioners do – justifiably – consider and defend democracy as a whole system of institutions and checks and balances, maintaining the very basic instrument of democracy, even in its simplest forms (regular electoral accountability) does produce significant effects on governance. Secondly, even in settings where regular elections do not take place, the windows available to voice criticism (i.e. simple forms of freedom of

expression) would be duly capitalized on by subjects to express anger in which case there is evidence that governing officials could respond accordingly.

We would like to end however with an important caveat. Our design is certainly not complex enough to reflect real-world dynamics. For example, our subjects in the authoritarian setting could express anger towards officials without the fear of physical punishment in return (perhaps only economic punishment via the continuation of government corruption which eventually deceases their payoffs). Our officials' only payoffs were their own incomes plus tax revenues (if they choose to be corrupt) whereas in real life they could have access to multiple others resources and spoils that could affect their decisions. Finally, all our subjects are students who are likely to be more pro-social – and perhaps less corrupt – than many actual government officials. These three aspects – and many others – could be embedded in future designs that could help increase our understanding of how public accountability could affect government corruption.

References

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

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

Ades A, Di Tella R. 1997. The new economics of corruption: a survey and some new results. Polit. Stud. 45(3):496–515

Adsera A, Boix C, Payne M. 2003. Are you being served? Political accountability and the quality of government. J. Law Econ. Org. 19(2):445–90

Ahmad, M.: 1995, Business Ethics in Islam (The International Institute of Islamic Thought and the International Institute of Islamic Economics, Islamabad, Pakistan).

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.

Besley, T., Preston, I., and Ridge, M. 1997, ‘Fiscal Anarchy in the UK: Modelling Poll Tax Noncompliance’, Journal of Public Economics, 64: 137–152.

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

Booth, D., & Cammack, D. 2013. Governance for Development in Africa: Solving Collective Action Problems. New York, NY: Zed Books

Bosco, L. and L. Mittone: 1997, ‘Tax Evasion and Moral Constraints: Some Experimental Evidence’, KYKLOS, 50, 297–324.

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.

Chowdury, S.K. 2004. The Effect of Democracy and Press Freedom on Corruption: An Empirical Test, Economics Letters. 85: 93-101.

Clarke, K., & Kocak, K. 2018. Launching Revolution: Social Media and the Egyptian Uprising’s First Movers. *British Journal of Political Science*, 1-21.

Cohen, A., 1995. Crime and Corruption in Eurasia: A Threat to Democracy and International Security. Washington, DC: The Heritage Foundation.

Corazzini, Luca, Sebastian Kube, Michel Andr´e Mar´echal, and Antonio Nicolo. 2014. “Elections and deceptions: an experimental study on the behavioral e↵ects of democracy.” American Journal of Political Science 58 (3): 579–592.

Cowell, F.: 1990, Cheating the Government: The Economics of Evasion (The MIT Press, Cambridge, MA).

Deacon, Robert T. 2009. “Public Good Provision Under Dictatorship and Democracy.” Public Choice 139:241-62.

Downs GW, Rocke DM. 1994. Conflict, agency, and gambling for resurrection: the principal-agent problem goes to war. Am. J. Polit. Sci. 38:362–80

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.

Erard, B. and J. S. Feinstein: 1994, 'The Role of Moral Sentiments and Audit Perceptions in Tax Compliance', Public Finance 49, 70–89.

Ertan, A., Page, T. and Putterman, L., 2009. Who to punish? Individual decisions and 780 majority rule in mitigating the free rider problem. Eur Econ Rev, 53, 495–511.

Evans, Jocelyn. 2003. *Voters and Voting: an introduction*. London: Sage.

Evans, Peter. 1995. Embedded Autonomy: States and Industrial Transformation. Princeton, NJ: Princeton University Press.

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

Fehr E., Schmidt KM. A theory of fairness, competition, and cooperation. Q J Econ. 1999;114(3):817–868.

Ferejohn, J. A.. 1974. *Pork barrel politics: Rivers and harbors legislation. 1947–1968*. Stanford, CA: Stanford University Press.

Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." Public Choice 50(1): 5–26.

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

Frey, B. S. and B. Torgler: 2007, 'Tax Morale and Conditional Cooperation', Journal of Comparative Economics 35, 136–159.

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.

Fung, Archon, and Erik Olin Wright. 2001. Deepening democracy: Innovations in empowered participatory governance. Politics & Society 29:5–41.

Gerring J, Thacker S. 2004. Political institutions and corruption: the role of unitarism and parliamentarism. Br. J. Polit. Sci. 34(2):295– 330.

Giglioli, P.P., "Political Corruption and the Media: The Tangentopoli Affair," International Social Science Journal 1996 (48), 381–394.

Gintis H, Bowles S, Boyd R, Fehr E. Explaining altruistic behavior in humans. Evol Hum Behav. 2003;24(3):153–172.

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.

Haas, Nicholas, Mazen Hassan, Sarah Mansour and Rebecca B. Morton. 2020. "Polarizing Information and Support for Reform", European Journal of Political Economy, forthcoming.

Haggard, Stephan. 1990. Pathways from the Periphery: the Politics of Growth in the Newly Industrializing Countries. Ithaca: Cornell University Press.

Hamman, John, Roberto Weber and Jonathan Woon (2011),"An Experimental Investigation of Electoral Delegation and the Provision of Public Goods" American Journal of Political Science, 55(4), 738-752.

Harris-White, B. and G. White (eds.), Liberalization and New Forms of Corruption (Brighton: Institute of Development Studies, 1996).

Hassan, Mazen and Marwa Shalaby. 2019. "Drivers of Tolerance in Post-Arab Spring Egypt: Religious, Economic, or Government Endorsements?", Political Research Quarterly, 72:2, pp. 293-308.

Hassan, Mazen, Sarah Mansour, Stefan Voigt and May Gadallah. 2020. "When Syria was in Egypt's Land - Egyptians Cooperate with Syrians, but less with each other", Public Choice, forthcoming.

Ireland, David. 2019. "What Marxist Tax Policies Actually Look Like." Historical Materialism, 27(2): 188-221.

Kaufmann, D. 2004. "Corruption, Governance and Security: Challenges for the Rich Countries and the World." The Global Competitiveness Report 2004–2005. World Economic Forum. New York: Oxford University Press.

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.

Kiewiet, D. Roderick and Mathew D. Mccubbins. *The Logic of Delegation: Congressional Parties and the Appropriations Process* (Chicago: The University of Chicago Press, 1991)

Kroll, S., Cherry, T. L., & Shogren, J. F. (2007). Voting, punishment, and public goods. *Economic Inquiry*, 45, 557–570.

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.

Lahlali, El Mustapha. 2014. "The Discourse of Egyptian Slogans: from 'Long Live Sir' to 'Down with the Dictator." Arab Media & Society, 19: 1-14.

McGee, R. W. and Galina G. Preobragenskaya, Accounting and Financial System Reform in Eastern Europe and Asia (pp. 299–334), New York: Springer, 2006.

McGee, R. W.: 1996, 'When Is Tax Evasion Unethical?', Working Paper, Policy Analysis No. 11, International Business & Technical Consultants, Inc.

McGee, R. W.: 1997, _The Ethics of Tax Evasion and Trade Protectionism from an Islamic Perspective, Commentaries on Law & Public Policy 1, 250–262.

McGee, R. W.: 2006, 'Three Views on the Ethics of Tax Evasion', Journal of Business Ethics 67, 15–35.

Mohtadi, H. and T. Roe (2003). Democracy, Rent Seeking, Public Spending and Growth, Journal of Public Economics. 87:.445-466.

Moran, J. (2001). Democratic Transition and Forms of Corruption, Crime, Law, and Social Change. 36: 379-393.

North, Douglass (1990) *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.

Pellegata, A. (2013) 'Constraining Political Corruption: An Empirical Analysis of the Impact of Democracy', Democratization, 20: 1195–1218.

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.

Plott, Add, 2001. Public Economics, Political Processes and Policy Applications. Collected Papers on the Experimental Foundations of Economics and Political Science, Volume One. Cheltenham, UK: Edward Elgar Publishing.

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

Przeworski, Adam, Susan C. Stokes and Bernard Manin (eds.) 1999. *Democracy, Accountability, and Representation*, Cambridge: Cambridge University Press

Rock, M. (2008). Corruption and Democracy, Journal of Development Studies. 45: 55-75

Rodgers, Daniel, 1987. Contested Truths: Keywords in American Politics since Independence, New York: Basic Books.

Rose-Ackerman, S. 1999. Corruption and Government: Causes, Consequences, and Reform. New York: Cambridge University Press

Rothstein, Bo and Aiysha Varraich, 2017, *Making Sense of Corruption*, Cambridge: Cambridge University Press.

Rothstein, Bo, and Davide Torsello. 2014. "Bribery in Pre-Industrial Societies: Understanding the Universalism-Particularism Puzzle" *Journal of Anthropological Research* 70(2): 263-282

Sandholtz W, Koetzle W. 2000. Accounting for corruption: economic structure, democracy, and trade. *International Studies Quarterly*, 44:31–50

Sharma, B. (2008). Voice, accountability and civic engagement. A conceptual overview. ODI, commissioned by Oslo Governance Centre, Bureau for Development Policy, UNDP.

Smulovitz, Catalina and Enrique Peruzzotti, "Social Accountability in Latin America, 2000. *Journal of Democracy,* 11: 147–58.

Stratmann, T., and Baur, M. 2002. "Plurality rule, proportional representation, and the German bundestag: How incentives to pork-barrel differ across electoral systems." *American Journal of Political Science*, 46: 506–514.

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

Treisman, Daniel. 2000. "The Causes of Corruption: A Cross-National Study," *Journal of Public Economics*, 76(3): 399-457.

Tyran, J.-R. and L. P. Feld. 2006. Achieving Compliance When Legal Sanctions Are Non-deterrent. *The Scandinavian Journal of Economics*, 108 (1): 135-156.

Van Rijckeghem C, Weder B. 2001. Bureaucratic corruption and the rate of temptation: Do wages in the civil service affect corruption and by how much? *Journal of Developmental Economics*, 65:307– 31

Yusuf, S. M.: 1971, *Economic Justice in Islam,* Lahore: Sh. Muhammad Ashraf.

Supporting Information

Appendix I

THEORETICAL ANALYSIS AND DERIVATION OF HYPOTHESES

In this section, we construct a model for our two treatments. We thus have two games; no-Recall game (noR-game) that corresponds to our no-Electoral accountability treatment, and Recall game (R-game) corresponding to our Electoral accountability treatment. We start by an illustration of the structure of the two games, then we move to stating the equilibrium analysis of the two games.

Earned income is private information making income reported for tax purposes a strategic variable; some incentives for truthful revelation are provided through an auditing mechanism. In the no-Recall game (noR-game) the official sits as the incumbent for one fixed term (with a known duration in terms of rounds of the experiment). In the Recall game (R-game) the official can be challenged while in office; if a recall is voted for, a new official is chosen among eligible group members.[36] In both games officials are exogenously selected and they can be thrown out of office in the R-game but not in the noR-game.[37]

A. The Stage Game

In both political settings, subjects earn income by performing an editing task and subsequently report their earned income, which is used to determine income tax liability given tax and audit rates and a penalty structure for less than full disclosure of income. It is a common knowledge that only a subset of players is being audited.[38] A player who is not audited pays according to declared income whereas an audited player pays tax on actual earned (not on declared) income plus a penalty (fine), which is a convex and increasing function of unreported income.[39]

[36] A group member is eligible if he has not been a subject of recall during the last three elections.

[37] Note that 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 (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.

[38] The assumption that only a subset of (randomly selected) players gets audited reflects the constraint on the resources available to conduct the audits.

[39] Please check appendix III for the penalty structure.

Tax proceeds[40] are used to finance one of two available linear public goods. It is common knowledge that the G-good favours the official at the expense of other citizens whereas the C-good benefits all players equally but the production technology is the same. If we let β_i^j denote the marginal per capita return (mpcr) of *j*-good to *i*-player, marginal valuations of public goods G and C across players satisfy the following set of inequalities,[41]

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

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). Because the citizen's mpcr of investing in the public good is smaller than 1, in the absence of audits, full evasion of taxes would be a dominant strategy for preferences defined over own income. Implementation of audits provides incentives for reducing tax evasion, so unlike in the common studies of linear public good games, positive investment (meaning less than full tax evasion in our game) can be optimal even for selfish players. On the other hand, full compliance is socially optimal by statement (*).

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. 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 a 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, Recall enabled replacement of the official (R-game) and Fixed scheduled replacement (noR-game).

[40] Penalties do not go into the public pool of funds; they go to cover administrative costs of auditing.

[41] 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 of the remaining of tax proceeds, $\left(\frac{\beta_c^G}{\beta_c^C}\right) T$ to fund the C-good. In this interpretation, which is closer to the conventional concept of corruption, there is only one public good to be funded that is equally valuable to everyone (think of defence) 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). The game is payoff equivalent for citizens to the one with two public goods. Behaviour, however, may be different between the two scenarios as a self-serving official may be more tolerated in the two public good scenario case than in the "tax appropriation'' one.

B. Finitely Repeated Game

Let the stage game be played for a known number, R of rounds. In the noR-game the official serves uncontested for R rounds whereas in the R-game the official can be challenged at any time and if recalled he loses the right to be a representative for three consequent elections. The main question of interest is whether recall-enabled rather than fixed-scheduled replacement of officials is a more effective institution in terms of public good provision. There is no a priori clear yes or no answer to this question as well-behaved officials can also be thrown out of office if craving for political power is widespread among group members.[42] Theoretically, the level of efficiency of public good provision is expected to be the same in both games if one appeals to subgame perfect equilibria (SPE). However, there are other Nash equilibria, with players using "minmax" strategies (to discipline officials) out of the equilibrium path, in which the efficiency of public good provision differs across the two games: higher efficiency is expected in the R-game if refraining from recalling the official during the end periods of his tenure is part of the strategy profile (see part 2 of the Main Result in this section). The intuition behind this result is that the official in either game is better off funding the C-good as long as the instantaneous benefits from funding of the G-good are smaller than future losses because of low compliance and, in the R-game, a recall vote. But while the instantaneous benefits are the same across the two games the future losses differ as (i) the likelihood of being in the office is lower in the R-game and (ii) for the G-good the low compliance is more costly for the official than for a citizen. So, unless citizens refrain from exercising the recall option during the late years of the official in the office, the low compliance is expected to be more effective in the noR-game than in the R-game. In the two following sections we state the equilibrium analysis of the two games (details below in part D of this Appendix).

C. Equilibrium Analysis

The following notation will be used: w is the individual's income, is the tax rate, p_a and $f(\cdot)$ 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 marginal return from the public good investment 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

[42] The supermajority rule is preferred to the simple majority in protecting well-behaved (benevolent) officials. It is also superior to the unanimity rule if "vote buying" is added to the equation as a corrupt official would then need to "buy" only one vote to survive a recall.

income by others, player i's expected payoff in the stage game from reporting x_i (when the real income is w) 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, $y_i = \tau(p_a w + (1 - p_a)x_i)$ is the expected payment by individual i as income tax.

It can be verified that in player i's optimal declared income: (i) is a dominant strategy but, unlike in much studied linear public good games, (ii) full free riding (declaring 0 income here) is not optimal for penalty functions that are sufficiently convex, (iii) increases in the public good return, which implies low tax proceeds from if G-good is funded.

The SPE outcome (see Proposition 1.1 and 2.1 below in section D) is the same level of inefficient public good provision across the two games. Note, however, that players' equilibrium payoffs in the stage game are larger than the minmax payoff.[43] Hence, there are Nash equilibria in which players' payoffs are sufficiently close to a desirable strictly enforceable payoff profile provided the game is played long enough (Benoit and Krishna 1987). In such an equilibrium, we see efficient public good provision (C-good and full compliance) followed by inefficient public good provision (G-good and low compliance) only during last R-r* rounds, for some r*. The efficacy of the institution depends on the length of end-game, R-r* between the two games. It turns out that the length of the end-game can be (weakly) shorter (and therefore more efficient public good provision) in the R-game than in the noR-game IFF the recall option is not exercised during the end rounds. So, although intuitively we might expect more efficient public good provision in the recall-enabled replacement institution than in the fixed-scheduled replacement of officials a formal reasoning (see part D in this Appendix) reveals that the performance depends on strategies used during the end of the game.

Theoretical predictions for the play between two games are summarized in the following proposition.

[43] Player i's stage game minmax payoff, $\pi_i(x_i^G, 0,1)$ is strictly smaller than the equilibrium payoff, $\pi_i(x_i^G, x_{-i}^G, 1)$ where x^G is the vector of optimal declared incomes when G-good is provided, $\pi_i(x_i^G, x_{-i}^G, 1)\pi_i(x_i^G, 0,1)$as the expected difference is $bt\mathring{a}_{j^1 i} I_{nA}(j)x_j^G > 0$ where $I_{nA}(j)$ is the indicator function for 'not being audited'.$\beta_i^G(1 - p_a)\tau\sum_{j\neq i} x_j^G (> 0)$

Proposition. Let players' preferences on the payoff space be represented by equation (1).

1. SPE predict the same level and type of public good funding in both games.
2. There are Nash equilibria[44] that predict efficient public good funding in all but the end rounds. The required number of end rounds with low efficiency of public good provision is:
    a. *lower* in the noR-game than in the R-game if the *recall* option is always exercised.
    b. *higher* in the noR-game than in the R-game if the *recall* option is exercised after defections during the non-end rounds but not during the end rounds.

Proof: See part D in this Appendix.

Our first hypotheses that follow from the SPE predictions are:

H1o*: Representatives fund the self-serving public good (G) in either game.*

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 result of proposition 2, part 1 in section D) so we have the third null hypothesis

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

The alternative hypotheses that follow from Part 2.b of the above Proposition are:

H1a*: Likelihood of the self-serving public good (G) is higher in the noR-game*

H2a*: Efficiency of public good provision is higher in the R-game.*

The one-sided alternative hypotheses for the Nash equilibria of Part 2.a of the above proposition are the opposite: (i) the self-serving G- good is more often funded in the R-game and (ii) public good provision is more efficient in the no-R game.

In the R-game, if a funding of the G-project by the official triggers recalls then the one-sided alternative hypothesis to H3o is

H3a*: Funding of the G-good has a positive effect on the likelihood of recall.*

D. PROOFS

[44] The official funds the C-good ($p^G=0$) in the first r* rounds and the G-good in the remaining R-r* rounds. Each player *i* claims *w* in round 1. If no defection occurs then player i claims *w* in rounds 2 to r* and x^G in the remaining R-r* rounds. Any defection at any round before r*+1 triggers strategies that minmax defector's payoff.

\We provide equilibrium analysis separately for the two games. 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

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

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

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

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 II

Instructions in English

SUBJECT INSTRUCTIONS FOR THE NO-ELECTORAL ACCOUNTABILITY 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 <u>less than</u> the earned income in Event I, then the individual pays, <u>in addition to the tax of</u> 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 <u>public project</u> 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:

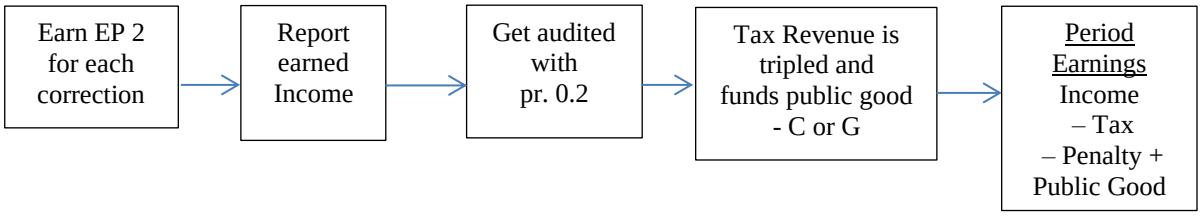
When this project is chosen, then the official will earn more than the rest of the group members: Half of the total amount of money in public fund is given to the official; the remaining half of the public fund is equally divided among all four remaining group members.

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 ELECTORAL ACCOUNTABILITY 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 <u>less than</u> the earned income in Event I, then the individual pays, <u>in addition to the tax of</u> 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 <u>public project</u> 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:

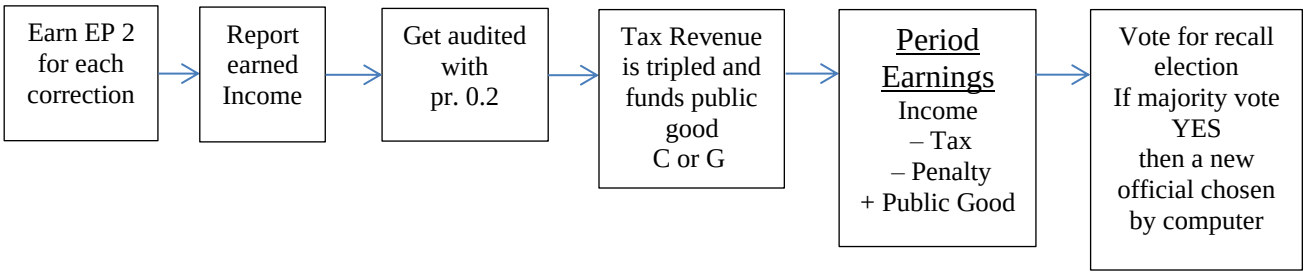
When this project is chosen, then the official will earn more than the rest of the group members: Half of the total amount of money in public fund is given to the official; the remaining half of the public fund is equally divided among all four remaining group members.

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 III

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 IV

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 <u>five years ago</u>, 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 $\pi_i(x_i^G, x_{-i}^G, 1)$ which is larger than the minmax payoff, $\pi_i(x_i^G, 0,1)$ 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' behaviour.
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.