

PRE-ANALYSIS PLAN

Elections and Embezzlement

Author information and acknowledgements omitted
(to facilitate anonymous peer review)

July 20, 2015

Abstract

This research design document describes an experimental study to investigate mechanisms through which elections can influence rent extraction by public decision-makers. The experiment will be carried out with 2360 randomly sampled adult citizens in 118 municipalities in Burkina Faso. In the experiment, groups of citizens from a municipality are provided with a budget, and a decision-maker in their midst can decide to embezzle or misappropriate any fraction of it for personal gain. The experiment varies independently (1) whether the decision-maker is elected or appointed at random, and, after the decision-maker has been selected, (2) whether the total amount of the budget (and hence the extent of embezzlement by the decision-maker) is revealed to the group (transparency), or remain private information of the decision-maker (no transparency).

Contents

1	About This Pre-Analysis Plan	4
2	Research Objectives and Relevance	5
3	Experimental Design and Procedures	9
4	Treatments and Treatment Assignment	11
5	Outcomes of Interest and Theoretical Expectations	12
	Embezzlement Decisions	13
	Trust in Decision-Makers	18
	Perceptions of Procedural Fairness	19
6	Outcome Measurement and Estimation of Treatment Effects	21
	Embezzlement of group resources	21
	Trust in the decision-maker	22
	Perceptions of procedural fairness	22
7	Interpretation and Generalizability of the Results	23
8	Supplementary Analyses	28
	Question 1.1: Do elections with minimal prior communication enable strangers to identify public-spirited leaders within their municipality?	28
	Question 1.2: To what extent can embezzlement behavior in the experiment be explained by baseline preferences?	30
	Question 1.3: How accurate are citizens' expectations regarding the extent of embezzlement?	32
	Question 2.1: Are citizens willing to incur personal costs to punish embezzlement?	33
	Question 2.2: How does a lack of transparency influence citizens' willingness to engage in costly sanctioning behavior?	36

Question 2.3: Does citizens' sanctioning behavior depend on whether decision-makers are elected or not?	38
9 Validation Strategy	40
Conceptual Validity of Outcome Measures	40
Validity of the Elections Treatment	41
10 Possible Extensions/Ancillary Experiments	43
Individual-level effects of being selected as a decision-maker	43
Group-level effects of the decision-maker's social identity	43
Female vs. male decision-makers	44
Co-ethnic vs. Non-co-ethnic decision-makers	45
Seniority of the decision-maker	46
Does selection into a position of responsibility increase perceived social status?	48
Ballot-order effects	48
11 Implementation and Timeline	49
12 Sampling and Power Calculations	51
Sampling and Study Population	51
Power Calculations	52
Appendix	57
Screenshots of the touch screen interface	57
R code for all figures in this pre-analysis plan	59
R code for power calculations	70
Protocol for on-site replacements of no-shows	72

1 About This Pre-Analysis Plan

This pre-analysis plan outlines the primary experimental comparisons that will be made in this study, as well as six supplementary research questions that will help with the theoretical interpretation of the experimental results. These supplementary research questions are divided into two lines of inquiry, corresponding to distinct causal mechanisms that could be responsible for the primary experimental results; (1) electoral selection effects and (2) citizens' willingness to sanction embezzlement of public resources by public decision-makers. Additionally, the pre-analysis plan outlines several ancillary experiments which arise as by-products of the original experiment and can potentially contribute to a better understanding of decision-, voting- and sanctioning behavior in the experiment.

The pre-analysis plan has been prepared and submitted after the research design has been finalized and after the procedures have been field tested, but before any of the experimental data has been received by the author. The experiment is being carried out by an independent implementing agency (Innovations for Poverty Action, IPA). The implementing agency provides the raw data from the experiment to the principal investigator, without engaging in any data cleaning or other treatment of the experimental data. Furthermore, IPA has no stake in the study's results and no access to this pre-analysis plan until the data collection is close to being complete. To preserve the integrity of the pre-analysis plan, IPA has been instructed not to release any data to the author until this pre-analysis plan has been submitted (and has not done so as of the submission date).

The modular structure of this pre-analysis plan (incorporating a hierarchy of supplementary questions and ancillary experiments) was adopted in order to illuminate the research design logic, while retaining some flexibility with respect to the eventual publication strategy. The relevance of the two lines of supplementary inquiry for the interpretation of the primary experimental comparisons is contingent on intermediate results. At the same time, the supplementary research questions and the ancillary experiments are interesting in their own right, above and beyond the value they add to the interpretation of the primary experimental results. Therefore, the supplementary analyses may be reported in separate research papers, especially since it would exceed the scope of a nor-

mal, article-length treatment to discuss all pre-specified questions and relationships among them comprehensively. For that reason, an additional objective of the pre-analysis plan is to help readers understand the original context and motivation for specific research design choices (especially if alternative design choices would have appeared superior ex-post) and to document the trajectory of learning and theory-development that took place over the course of the data analysis. The hierarchy among the different proposed analyses should become evident from the discussion.

All figures in this pre-analysis plan are based on computer-generated, simulated data. These simulations contain no actual empirical information. Their purpose is to illustrate the proposed initial analyses in the most detailed way possible and to test the code that will be used for the initial analyses of the experimental data.

2 Research Objectives and Relevance

Whenever public resources are managed by governments or other complex organizations, agency problems are almost inevitable. Public organizations are made up of individuals, and some of these individuals may be willing to pursue their personal advantage at the expense of the public interest, by shirking, accepting bribes or kickbacks, or directly misappropriating public funds. Unfortunately, the embezzlement of public funds is a very widespread phenomenon, as evidenced by countless audits of public agencies around the world. The prevalence of embezzlement by public decision-makers suggests that top-down bureaucratic oversight and the threat of judicial enforcement are often insufficient to deter it. It also raises the question what additional remedies exist to overcome the agency problems associated with the delegation of public responsibility to individual decision-makers. Two such potential remedies are democratic elections of public decision-makers (Ferejohn, 1986; Fearon, 1999; Besley, 2005) and greater transparency of public transactions (Reinikka and Svensson, 2005; Olken, 2007; Litschig and Zamboni, 2012).

Social scientists have only recently begun to systematically evaluate the impact of democratic elections on the misappropriation of public resources by local public officials in low- and middle-

income countries (Ferraz and Finan, 2011; Alatas et al., 2013; Beath et al., 2014; Lierl, 2014). Thus far, the results have been mixed and inconclusive, suggesting that both elected and non-elected leaders embezzle public resources to a considerable extent (Alatas et al., 2013; Beath et al., 2014; Lierl, 2014). Yet, from a theoretical perspective, there are plausible mechanisms through which democratic elections could reduce the embezzlement of public funds. First, elections enable citizens to select public-spirited decision-makers who are intrinsically motivated to refrain from embezzling public funds. Second, elections could increase accountability pressures on those decision-makers who are nevertheless tempted to embezzle public funds. However, it is challenging to assess to what extent each mechanism matters in practice and how they interact with transparency.

As far as electoral selection effects are concerned, existing research suggests that elections do indeed favor public-spirited leaders, who (at least to some extent) voluntarily refrain from embezzling public resources (Lierl, 2014; Beath et al., 2014). This evidence, however, is derived from studies of village-level governance, where voters usually know the candidates in person, even prior to incumbency. It therefore remains to be tested whether evidence for similar selection effects can be found in lower-information contexts, such as municipalities.

As far as accountability pressures are concerned, a widespread hypothesis is that elections and transparency interact to *enable* citizens to sanction corrupt or non-performing decision-makers, who would otherwise not have the ability to do so (Besley and Burgess, 2002; Ferraz and Finan, 2011; Humphreys et al., 2012; Malesky et al., 2012). According to this view, elections should only result in greater accountability and lower embezzlement if the decision-makers are office-motivated and eligible for re-election, and if there is sufficient transparency, so that citizens are made aware if public funds are embezzled. Ferraz and Finan (2008) and Ferraz and Finan (2011) appear to confirm this prediction in the case of Brazilian mayors. Other studies, however, find no evidence that better availability of information causes voters to vote against corrupt incumbents (Banerjee et al., 2010; Chong et al., 2013). On the other hand, civics training that provides citizens with normative benchmarks appears to increase voting against incumbents that do not serve their constituents well (Gottlieb, 2015). Further studies emphasize that elections are not the only way in which citizens can sustain accountability pressures on public decision-makers (Putnam, 1993; Tsai, 2007;

Lierl, 2014, 2015). However, even if informal accountability pressures are taken into account, it is not evident that transparency would actually create incentives for elected leaders to refrain from misappropriating public resources (Lierl, 2014). This raises the question whether citizens' ability to sanction corrupt decision-makers is the limiting factor, or rather their *willingness* to do so, and how citizens' willingness to punish the embezzlement of public resources is influenced by elections and transparency.

This study therefore seeks to shed light on whether democratic elections and transparency can affect the embezzlement of public resources through mechanisms other than changing citizens' *ability* to sanction corrupt decision-makers. To this end, the study simultaneously examines the generalizability of electoral selection effects to low-information contexts, as well as the mechanisms through which democratic elections and transparency influence citizens' willingness to punish embezzlement (including their trust towards decision-makers, their perceptions of procedural fairness, and their response to information about how much embezzlement has taken place). This is accomplished by evaluating the effect of elections and transparency on the embezzlement of public resources in a behavioral experiment where sanctioning opportunities are held constant and citizens have only very limited prior information about candidates. In order to distinguish the causal mechanisms of interest from context-specific influences, the study is built around a very generic decision situation, designed to emulate a fundamental governance dilemma: A decision maker is tasked with managing a certain amount of money on behalf of a group and has the opportunity to embezzle money for personal gain, but may face sanctions by the public.

This basic governance dilemma is recreated in 472 groups, involving 2360 randomly sampled adult citizens from 118 rural municipalities in Burkina Faso. In every group, there is a real payoff at stake. The experiment varies whether the decision maker (who has the opportunity to enrich her-/himself by embezzling group money) is elected or selected at random, as well as whether there is symmetric or asymmetric information between the group members and the decision-maker about the amount of money available to the group (transparency vs. no transparency). The most immediate outcomes of interest are the extent of embezzlement of group funds by the decision-maker, group members' trust in or suspicion towards the decision-maker (measured by their expectations of how much

embezzlement has taken place), and their perception of procedural fairness. These experimental comparisons are complemented with a set of additional, pre-specified observational comparisons and validation opportunities, designed to aid with the theoretical interpretation of the findings.

Throughout the study, the reference point against which elected decision makers are compared are randomly appointed decision makers. Comparing elections to random selection of decision makers is interesting in several respects: First, randomly selected decision makers represent how the “average citizen” would act in a comparable situation. Second, random selection of citizens constitutes an alternative “democratic” form of representative authority and various historical examples exist where this form of representative authority has been put into practice (Corazzini et al., 2014). Lastly, random selection is a common reference point in the literature on the behavioral effects of elections (Baldassarri and Grossman, 2011; Grossman and Baldassarri, 2012; Corazzini et al., 2014).

The remainder of this research design document is organized as follows. Section 3 outlines the experimental design and procedures, Section 4 describes the experimental treatments and treatment assignment, Section 5 describes the outcomes of interest and theoretical expectations, Section 6 deals with the measurement of outcomes and estimation of treatment effects, Section 7 discusses the interpretation and generalizability of the experimental results, proposing a set of self-standing supplementary research questions (SRQ) that explore the causal chain leading to the overall experimental results, Section 8 describes the supplementary analyses in detail, Section 9 proposes opportunities to validate the outcome measures and experimental treatments, Section 10 discusses potential extensions and secondary uses of the experimental data, Section 11 summarizes the implementation modalities and timeline, Section 12 describes the sampling procedure and power calculations.

3 Experimental Design and Procedures

In the experiment, groups of five randomly sampled citizens of a municipality are confronted with a stylized decision situation, where an unknown amount of money has to be allocated within the group by a single individual. It is varied experimentally whether this individual is elected by the group, or appointed at random. Independently, and only after the selection has taken place, it is also randomized whether the decision maker has private information about the amount of money that is to be distributed or whether this amount is public knowledge within the group.

The decision situation consists of the following steps:

1. Baseline decisions: Prior to the experiment, all study participants submit a proposal on how to split 5000 Francs between themselves and the group. One proposal per group is selected at random and implemented. These baseline allocation proposals and payoffs are recorded, but the results are not announced until the end of the exercise. This baseline decision serves to record every group member's allocation preferences and to familiarize the study participants with the decision situation.
2. Cheap talk/communication: The groups are given ten minutes to freely discuss (without intervention by the facilitator) who would be the best person to make the allocation decision and how the money should be split between group and its decision-maker. This cheap talk phase gives study participants the opportunity to make inferences about other candidates preferences through verbal and nonverbal cues.
3. Announcing the selection method: It is announced to the group how the decision maker is going to be selected: either by election within the group, or by random selection. Furthermore, it is announced to the study participants that there is an equal chance that the amount of money that has to be allocated is either private information of the decision maker, or public information within the group.
4. Selection of the decision maker: The decision maker is elected or randomly selected, depending on the treatment condition.

5. Announcing whether the size of the pie (the amount of money to be shared within the group) will be private or public information.
6. Allocation decision: On a touchscreen and in privacy, the decision maker is asked to divide banknotes worth 10000 Francs into two stacks. One stack is to be split equally among all five members of the group, including the decision maker. The other stack is captured by the decision maker and not shared with anyone else. The total amount is still unknown to the other group members.
7. Announcing the allocation decision: It is announced how much money was allocated to the group by the decision maker. In the public information condition, it is additionally announced how much money the decision maker kept for her-/himself.
8. Costly rewards/sanctions: After learning how much money has been allocated to them by the decision maker (and, in the public information condition, how much money the decision maker has captured for her-/himself), the other four group members can choose to reward or punish the decision maker at a cost to themselves. They receive a budget of 1000 Francs for that purpose, which they do not have to fully expend and for which they are the residual claimants. To reward the decision maker, group members can secretly send money to her/him. To punish the decision maker, group members can secretly pay an amount of money to have three times the amount of money deducted from the decision maker's payoff.

Prior to the decision exercise, all study participants complete a baseline survey. The baseline survey serves to produce municipal-level data about local governance quality and citizens' experiences with municipal authorities, as well as data on individual-level characteristics, such as age, education, occupation, household size, leadership experience, electoral participation in the last elections, and trust in the municipal administration.

Following the baseline survey, the decision exercises (baseline and experiment) are carried out in groups of five study participants under the supervision of two facilitators. Study participants are not allowed to communicate during the exercise, except when they are asked to. Video recorded instructions are played to the study participants in their vernacular language. The instructions

are divided into multiple logical blocks. Each video block is followed by a set of comprehension checks. If a study participant in the group does not pass a comprehension check, the video block is re-played to the entire group until all participants have understood the instructions. Votes and allocation decisions are made on specifically designed tablet computer interfaces. These are intuitively designed, and pilot tests have confirmed that they are easily understood, even by illiterate study participants. The video instructions include demonstrations of the tablet interface. Additionally, a practice tablet is passed around so that study participants can familiarize themselves with the interface.

Outcome data is collected at various stages of the decision exercise and includes: (1) Study participants expectations about the amount embezzled by the decision maker, (2) the decision makers allocation decision, (3) the amounts spent on rewarding or punishing the decision maker.

4 Treatments and Treatment Assignment

There are two cross-cutting experimental treatments.

(1) Election vs. random selection.

- In the election condition, all group members use a secret ballot to cast votes for two different candidates out of the five group members. Runoff elections are held with the candidates ranked first and second.¹ In the runoff elections, each group member casts one vote. The candidate who wins an absolute majority is selected.
- In the random selection condition, one of the five group members is selected at random.

(2) Symmetric vs. asymmetric information about embezzlement.

- In the symmetric information (transparency) condition, the amount of money the decision maker has captured for her-/himself is revealed to the group ex post.

¹In case of ties, there may be more than two runoff candidates, in which case several successive runoff elections may be held until one candidate has an absolute majority.

- In the asymmetric information (no transparency) condition, the amount of money the decision maker has captured for her-/himself remains private information of the decision maker.

In both treatment conditions, the size of the pie that is to be allocated is unknown at the outset. Whether the size of the pie will become public knowledge or private knowledge of the decision maker is announced only after the decision maker has been selected, but before the allocation decision is made. This ensures that the information condition cannot influence the election results.

The treatment and control conditions are randomized at the group level, using a 2×2 factorial design that results in four distinct combinations. The treatment assignment is blocked by municipality. Four groups are sampled in every municipality. Each group is assigned to a different combination of treatment conditions, so that all four possible combinations are carried out in every municipality.

Further treatment variations were considered based on theoretical interest, but could not be realized due to practical constraints and the limited sample size. These include (i) a no-sanctioning condition, (ii) a condition where deliberation takes place after leader selection instead of before, and (iii) further variation in the methods of leader selection, for example selection by consensus. The no-sanctioning condition would have been difficult to realize in a field setting. Given the non-anonymity of the study participants within their group, it would have been difficult to fully prevent subjects from interacting at the study site after the experiment. Therefore, it was considered superior to incorporate a formal sanctioning procedure into the decision exercise and to channel all sanctioning behavior into this formal procedure, by emphasizing to the study participants that this is the only way in which they are allowed to express approval or disapproval towards the decision maker. The other variations were abandoned due to limitations in the sample size, but may be taken up in further extensions of the study.

5 Outcomes of Interest and Theoretical Expectations

For three self-standing outcomes of interest (embezzlement decisions, group members' trust in the decision-maker, and perceptions of procedural fairness), prior working hypotheses about the impact

of the two experimental manipulations (elections and transparency) are outlined below.

Embezzlement Decisions

By design, the experiment allows for three mechanisms through which the experimental treatments (elections and transparency) can impact embezzlement outcomes. First, by influencing the selection of decision-makers. Second, by influencing the preferences of selected decision-makers. Third, by influencing the anticipated sanctioning behavior of the other group members towards the decision-maker (without changing the sanctioning institutions, i.e. the *ability* of group members to sanction the decision-maker). These design features and the underlying rationale are discussed in Section 7.

For each of the aforementioned mechanisms, the hypothesized influences of the elections and transparency treatments on embezzlement outcomes are summarized in Table 1. Overall, the prior working hypothesis is that both elections and transparency will reduce the embezzlement of group money.

- HYPOTHESIS 1A: Elections reduce embezzlement of group resources both in the presence and in the absence of transparency.
- HYPOTHESIS 1B: If group members' ability to reward or punish the decision-maker is held constant, transparency reduces embezzlement of group resources by both elected and randomly appointed decision-makers.

Both hypotheses rest on the assumption that there is a social norm favoring equitable allocations of the group money and that embezzlement will always conflict with this norm. Under this assumption, elections can contribute to norm-conforming behavior, because they enable voters to select those candidates whom they consider most likely to act in their interest. If voters are able to correctly anticipate the candidates' behavior (see Supplementary Question 1.3), elections should favor public-spirited decision-makers (see Supplementary Question 1.1). Additionally, both the experience of having been elected (Corazzini et al., 2014), as well as the fact that their behavior in the transparency is socially observable (Haley and Fessler, 2005; Charness and Gneezy,

2008; Rigdon et al., 2009) could induce decision-makers to act more in line with social norms. Finally, transparency could influence citizens' willingness to punish behavior that deviates from social norms or to reward norm-conforming behavior, either because individuals are more hesitant to sanction decision-makers based on mere suspicion rather than factual knowledge (Croson et al., 2003, 144f.), or because they are systematically biased towards trusting decision-makers too much (Gottlieb, 2015). It is also possible that the elections influence citizens' willingness to sanction decision-makers, either because they are less tolerant of embezzlement by elected decision-makers, or because elections increase bias in citizens' expectations (i.e. cause them to be too trusting of decision-makers). Table 1 summarizes these expectations.

Mechanism	Hypothesized influence of	
	Elections	Transparency
Selection of decision-maker	Decrease embezzlement, because elections favor candidates who are intrinsically motivated to refrain from embezzlement (Lierl, 2014) → <i>SRQ 1.1, SRQ 1.3</i>	<i>No effect by design</i> , because the treatment condition is not revealed until after the selection stage.
Direct effects on selected decision-makers	Decrease embezzlement, because the experience of having been elected induces pro-social behavior (Corazzini et al., 2014) → <i>SRQ 1.2</i>	Decrease embezzlement, because social observability induces pro-social behavior (Haley and Fessler, 2005; Charney and Gneezy, 2008; Rigdon et al., 2009)
Willingness to punish embezzlement	Unknown how elections influence citizens' willingness to sanction decision-makers based on factual information about embezzlement, as well as their trust in decision-makers and their willingness to sanction suspected embezzlement. → <i>SRQ 2.1, SRQ 2.3</i>	Decrease embezzlement, because individuals might be more willing to punish others based on factual knowledge rather than mere suspicion (Croson et al., 2003, 144f.), → <i>SRQ 2.1, SRQ 2.2</i> or because citizens are biased towards being too trusting (Gottlieb, 2015). → <i>SRQ 1.3</i>
Ability to sanction the decision-maker	<i>No effect by design</i> , because sanctioning opportunities are held constant.	<i>No effect by design</i> , because sanctioning opportunities are held constant.
Overall effect	Elections decrease embezzlement (HYPOTHESIS 1A)	Transparency decreases embezzlement (HYPOTHESIS 1B)

Table 1 Hypothesized influences of the experimental treatments on the primary outcome of interest (embezzlement of group money), by causal mechanism. Supplementary research questions (*SRQ*) which examine potential counterarguments against these hypothesized causal relationships are indicated in italics.

There are several non-trivial reasons why Hypothesis 1a might be rejected, i.e. why elections might *increase* embezzlement in the experiment.

1. One possibility is that elections favor selfish candidates (adverse selection).
2. Another possibility is that the experience of having been elected makes people more selfish (or, loosely speaking, that “power corrupts”).
3. A third possibility is that citizens are more willing to tolerate embezzlement by elected leaders, either because they believe that elected leaders are entitled to a greater share of public resources, or because of selection effects, e.g. if electoral success is correlated with social status or other variables that influence how much tolerance for embezzlement an individual can expect.

The supplementary analyses outlined in Section 8 examine several of these potential counter-arguments to Hypothesis 1.

Similarly, there are several plausible reasons why Hypothesis 1b could be rejected, i.e. why transparency might increase embezzlement in the experiment.

1. If transparency causes decision-makers to conform to how they think the other study participants would act in the same situation and elections produce decision-makers who would otherwise be less willing to embezzle group resources than the average group member, then transparency might increase embezzlement in the elections condition, but not in the random appointment condition.
2. If knowledge of being observed causes a large proportion of decision-makers to act spitefully, or if they derive pleasure or pride from being perceived as ruthlessly pursuing their material self-interest, then transparency might contribute to increased embezzlement through these mechanisms.
3. If citizens engage in anti-social punishment (i.e. punish selfless individuals) or have a preference for non-zero embezzlement, then the anticipation of greater punishments for lower embezzlement might cause decision-makers to embezzle more under transparency than they

would otherwise be willing to.

4. If decision-makers anticipate systematic bias in the other group members' expectations, e.g. if citizens are overly suspicious towards elected and/or randomly selected decision-makers, then transparency might have the perverse effect of facilitating embezzlement by correcting these biases.

Hence, neither Hypothesis 1a nor Hypothesis 1b are self-evident.

With respect to the interaction effect of the elections and transparency treatments on embezzlement, holding group members' ability to sanction the decision-maker constant, the prior expectation is that elections and transparency are substitutes in reducing embezzlement, if citizens' ability to sanction decision-makers is held constant. One reason why elections and transparency are expected to be substitutes is that transparency can affect embezzlement through a sanctioning mechanism, but elections additionally through a selection mechanism (Fearon, 1999), or preference transformation mechanism (Corazzini et al., 2014), which can substitute for sanctioning. Hence, if transparency mainly strengthens sanctioning, but the importance of sanctioning is reduced in the elections condition, then elections and transparency should be substitutes, as long as citizens' ability to sanction decision-makers is held constant. Alternatively, if social norms provide a fixed reference point and, at the margin, individuals are less willing to sanction smaller deviations from the norm than larger deviations, then the effects of elections and transparency should be subadditive.

- HYPOTHESIS 2: If group members' ability to reward or punish the decision-maker is held constant, elections and transparency are *substitutes* in reducing the embezzlement of group resources.

Under Hypothesis 2, we would expect an interaction effect that is positive (i.e. leading to greater embezzlement), but not large enough to completely offset either of the main effects.

Hypothesis 2 would be rejected if elections and transparency are *complements* in reducing embezzlement, even if citizens' ability to sanction decision-makers is held constant. This would be plausible if citizens apply different social norms to elected and non-elected decision-makers. For

example, if transparency universally enhances the enforcement of norms and there is a norm that elected leaders are supposed to embezzle less than non-elected leaders, then transparency should have a greater effect under the elections treatment. Alternatively, elections and transparency could be complements if elections systematically biased expectation formation, i.e. if elections caused citizens to be too trusting towards decision-makers. Transparency would then remove this bias and induce appropriate responses. If elections and transparency are *complements*, we would expect a negative interaction effect of elections and transparency (i.e. leading to lower embezzlement).

Intermediate results, especially from the proposed lower-level analyses in Section 8 regarding the accuracy of citizens' expectations and the effect of transparency on norm enforcement, may have further implications as to whether elections and transparency should be considered complements or substitutes. If logically implied by lower-level results, such predictions could then be evaluated in lieu of Hypothesis 2.

Trust in Decision-Makers

Citizens' trust in or suspicion towards the decision-maker is an important intermediate outcome in understanding how elections alter the consequences public decision-makers face if they embezzle public resources: Do elected leaders enjoy greater trust from citizens? Is this trust warranted? Trust in the decision-maker is measured by their expectations about how much the decision-maker has embezzled. The initial working hypothesis is that elections and transparency influence trust towards decision-makers in ways that are consistent with their effects on actual embezzlement outcomes. Thus, if elections reduce embezzlement, it is expected that elected decision-makers enjoy greater trust. Analogously, if transparency reduces embezzlement, it is expected that decision-makers enjoy greater trust if their decisions are going to be revealed publicly.

- **HYPOTHESIS 3:** Study participants' expectations are consistent with the effects of elections and transparency on actual embezzlement outcomes.

This working hypothesis is based on the following reasoning:

1. If elections reduce embezzlement through a selection mechanism, it is plausible that the same process would also increase citizens' trust in the decision-maker: Citizens presumably vote for candidates whom they consider most trustworthy; therefore the winning candidates should on average be those to whom citizens attribute the greatest trustworthiness. Individual trust in the winning candidate might additionally be reinforced by the observation that a majority of the other group members favored that candidate.
2. If elections and/or transparency transform the preferences of selected decision-makers, it is plausible that group members are able to anticipate this through social intelligence.
3. If elections and/or transparency reduce embezzlement through a sanctioning mechanism, it appears plausible that group members would anticipate this, because otherwise they would have no actual reason to be more willing to punish embezzlement.

Therefore, the initial expectation is that the effects of elections and transparency on group members' trust have the same sign as their effects on actual embezzlement outcomes.

The experiment might fail to confirm Hypothesis 3, if elections and/or transparency have no detectable effect on citizens' trust, even though they have substantive effects on actual embezzlement outcomes. Per se, such a null finding would be uninformative, but it could be investigated further how citizens form expectations about the decision-maker's behavior (see the proposed analyses in Section 10).

However, Hypothesis 3 could also be contradicted by the experiment. This would be the case if expectation formation is biased, i.e. if citizens systematically over- or underestimate how much money is being embezzled, in ways that depend on the treatment condition. If the experimental treatments bias group members' expectations, then the effects of elections and/or transparency on citizens' expectations might be inconsistent with their effects on actual embezzlement. It is therefore important to also test how accurate citizens' expectations are and whether either elections or transparency introduce bias into citizens' expectations, rather than reducing existing bias (Supplementary Question 1.3).

Perceptions of Procedural Fairness

In prior research, it has been argued that democratic elections increase the perceived *legitimacy* of leadership, based on experimental findings that elected leaders are able to elicit greater compliance from others than randomly selected leaders (Baldassarri and Grossman, 2011; Grossman and Baldassarri, 2012). However, the reason why elections would increase the perceived legitimacy of leadership remains unclear. One possibility is that elections and random appointments differ in their perceived procedural fairness. The procedural fairness of an institution is an important prerequisite for its legitimacy (Levi et al., 2009). Although an institution may be more legitimate than another for reasons other than procedural fairness,² lower procedural fairness should be associated with lower legitimacy, because it is difficult to conceive of a reason why individuals should or would consent to procedure of assigning authority that they perceive as unfair. Therefore, understanding how perceptions of procedural fairness differ between elections and random appointments is an important step towards understanding whether and why elections actually enjoy greater legitimacy than random appointments.

The experiment therefore seeks to test whether elections differ from random appointments of decision-makers with respect to their perceived procedural fairness.³

- HYPOTHESIS 4: Elections and random appointments of decision-makers do not differ with respect to their perceived procedural fairness.

Whether elections or random appointments will be perceived as fairer is a non-trivial question, because the perceived procedural fairness of a leader selection method may depend both on equality of access and descriptive representation, but also on the fairness of the outcomes. Random

²For example, in comparison to random appointments, elections involve the opportunity to directly consent to a person's authority through the act of voting. The opportunity to consent could enhance legitimacy, because legitimacy is generally understood to be derived from explicit or implicit consent to an institution or procedure by those who are subjected to it.

³The reason for focusing on perceptions of procedural fairness as a first step, rather than on the more abstract notion of legitimacy, is that perceptions of procedural fairness can be measured much more directly and reliably in a study population with limited access to formal education. If study participants were asked how "legitimate" they perceive the decision-maker to be, it is unclear what sentiments and considerations their answers would reflect. Therefore, study participants are asked whether they think the selection procedure was "fair" (Fr.: *juste*). Fairness is a much more easily communicable and generally understood concept than legitimacy, whose practical definition remains the subject of much debate among social scientists and philosophers.

appointments entail more equal chances of being selected into a leadership position, and hence better descriptive representation. Elections, on the other hand, allow for consideration of individuals' preferences over different candidates. On the other hand, elections ultimately privilege the preferences of a majority over those of a minority, which may be considered unfair. The relative weight of these arguments to citizens may ultimately be shaped by the fairness of actual outcomes.

6 Outcome Measurement and Estimation of Treatment Effects

The primary outcome of interest is the fraction of group resources embezzled. Secondary outcomes of interest are group members' trust in the decision-maker (measured by their self-reported expectations about how much embezzlement has taken place), as well as their perception of the fairness of the procedure of selecting the decision-maker.

Embezzlement of group resources

The average effects of the elections and transparency treatments on these outcomes, as well as their interaction effect will be estimated via OLS regression of the form

$$y1_g = \beta_0 + \beta_1 e_g + \beta_2 t_g + \beta_3 e_g t_g + x'_g \gamma + \epsilon_g$$

where $y1_g$ is the fraction of group resources embezzled, e_g indicates whether the decision-maker was elected, t_g indicates whether the decision is transparent/public information, and x'_g is a vector of group-level covariates: The average baseline embezzlement decision in the group, number of women in the group, ethno-linguistic fractionalization of the group (Herfindahl concentration measure), geographic fractionalization of the group with respect to study participants' village of origin (Herfindahl concentration measure), number of pre-existing social ties within the group, highest level of education represented in the group (years of schooling completed).

With regard to Hypotheses 1a and 1b, p-values for $\beta_1 \geq 0$, $\beta_1 + \beta_3 \geq 0$, $\beta_2 \geq 0$ and $\beta_2 + \beta_3 \geq 0$

will be reported. With regard to Hypothesis 2, p-values for $\beta_3 \leq 0$ will be reported. Coefficients will be estimated with and without adjustment for group-level covariates.

Trust in the decision-maker

The effects of elections and transparency on average trust in the decision-maker will be estimated via OLS regression, adjusting for clustering of errors by group.

$$y_{2ig} = \beta_0 + \beta_1 e_g + \beta_2 t_g + \beta_3 e_g t_g + x_{2i}' \gamma + \epsilon_{ig}$$

where y_{2i} is an individual's expectation about the fraction of group resources embezzled, e_g indicates whether the decision-maker was elected, t_g indicates whether the decision is transparent/public information, and x_{2i}' is a vector of individual-level covariates: Individuals' baseline embezzlement decision, age, gender, and years of schooling completed.

To evaluate Hypothesis 3, the appropriate one- or two-sided tests will be defined after evaluation of Hypotheses 1a and 1b.

Perceptions of procedural fairness

Perceptions of procedural fairness are measured via the survey question: “*In your opinion, was the procedure for selecting the decision-maker fair?*”. Marginal effects from Logit regressions will be reported, with errors adjusted for clustering by group. As a group-level covariate, the transparency treatment indicator will be included, as individual-level covariates the baseline embezzlement decision, age, gender, and years of schooling completed. Specifications with and without covariates will be reported. For an evaluation of Hypothesis 4, the marginal effect of the elections treatment is of interest.

7 Interpretation and Generalizability of the Results

Of vital importance for the interpretation of the experimental findings is the extent to which they allow us to formulate expectations about how elections and transparency would affect the embezzlement of public resources in natural settings. For example, an analogous real-world problem would be the adoption of elections as a method of selecting municipal administrators.⁴ To facilitate the generalization of the experimental results to natural settings, the experimental setting has been designed to minimize both context-dependence and confirmatory bias in the results obtained with respect to Hypotheses 1a, 1b, and 3. This is achieved by imposing particularly high hurdles for elections to have *any* effect on embezzlement in the experimental setting, and by artificially limiting the set of plausible causal mechanisms through which elections could have such effects.

- In the experiment, there is no difference in sanctioning institutions between the election and random appointment conditions. By contrast, in natural settings elections often create additional opportunities for citizens to sanction incumbents, for example if incumbents are up for re-election, or if citizens can petition for a recall election. In the experiment, these possibilities have been eliminated by design. Therefore, the effects of elections in the experiment cannot be due to differences in citizens' *ability* to sanction decision-makers (which is highly context-specific), but might due to differences in their *willingness* to do so (which is the product of more context-independent behavioral mechanisms that may generalize more easily beyond the experimental setting). Citizens' willingness to sanction decision-makers may in turn depend on whether the decision-maker is elected or not and how much information is available to them.
- Since sanctions are costly in the experiment and there is no repeated interaction with the decision-maker, it is not ex-post rational for group members to sanction the decision-maker. Furthermore, voluntary sanctioning behavior may be reduced by the fact that sanctioning the decision-maker involves a collective action dilemma, because individuals can free ride on

⁴In Burkina Faso, the current, externally appointed municipal governments (*délégations spéciales*) are scheduled to be replaced by elected municipal governments in 2016.

other group members' sanctioning efforts. Hence, neither the election treatment, nor the transparency treatment introduce an *instrumental* reason for sanctioning the decision maker.

- The framing of the decision situation remains identical between the election and random appointment conditions. The differences in the instruction scripts are minimal and essentially limited to replacing “randomly selected” with “elected” and explaining the procedure of electing a decision maker in basic, neutral terms. The instructions are video recorded and the procedures are pre-programmed on a tablet computer application to ensure that the influence of interviewer/facilitator effects is minimized. Hence, the effects of elections and transparency in the experiment are unlikely to be due to framing, priming or surveyor effects.
- Lastly, there is no difference in prior access to information between the election and random appointment conditions in the experiment. This contrasts with natural settings, where elections usually involve campaigns, media coverage and public scrutiny of the electoral candidates that enable citizens to obtain information about them. In the experiment, the members of a group are generally strangers who are from the same municipality, but not from the same villages. The electoral selection of decision makers should therefore not be influenced by access to prior information about individuals' social reputation or the existence of social ties (except in the rare coincidence that two group members knew each other before).

As a consequence, the experimental design severely restricts the set of causal mechanisms through which elections could influence the extent of embezzlement of public resources, relative to natural settings. If elections have a detectable impact on embezzlement in the generic and restrictive setting of the experiment, only three relatively generic and context-independent causal mechanisms remain could plausibly explain such a result.

- Selection effects, i.e. the possibility that elections favor candidates who are intrinsically more motivated to refrain from embezzlement than the average group member. Evidence for positive selection effects in the experiment would imply that elections enable *even strangers with minimal prior communication* to identify public-spirited leaders from within their municipality. This would be a very strong result, suggesting that only minimal social interaction is

needed to . It would then be of interest to further explore the heuristics voters used to make their choices.

- Preference changes: The experience of having been elected could also make decision-makers more pro-social, as suggested by Corazzini et al. (2014).
- Differences in sanctioning behavior: Citizens might be less tolerant of embezzlement by elected leaders than by randomly appointed leaders, either because they hold them to higher standards of public-spiritedness, or because they are more willing to incur personal costs to enforce norms. In either case, the consequence would be that elected decision makers are confronted with a more severe threat of sanctions, which could cause them to refrain from embezzlement more than non-elected decision makers.

These mechanisms are of particular interest, because they reflect relatively generic and situation-independent aspects of human social cognition and behavior. If they can be shown to operate in the experiment, they are likely to also matter in natural settings, especially since the experiment has been designed to eliminate confirmatory bias in favor of either of these three mechanisms. For example, with regard to the sanctioning mechanism, the experimental setting removes any instrumental rationality for engaging in costly sanctioning behavior. If sanctions are nevertheless observed in the experiment, they must be due to reciprocity, a desire to enforce social norms, or spite. Furthermore, the framing of the decision situation is deliberately neutral, and it has been carefully avoided to include any cues in the instructions that could prime group members to hold elected decision-makers to higher moral standards than non-elected decision-makers or prime elected decision-makers to think that they have a particular moral obligation towards the other group members. Therefore, potential differences between the election and random appointment conditions with respect to study participants' sanctioning behavior or decision-makers' allocative preferences should not be due to framing or priming effects. Finally, with regard to the selection mechanism, the research design minimizes prior information about the other group members. The members of a group are generally strangers and communication between them is limited to a ten-minute group conversation immediately prior to the selection of the decision-maker. Hence, any electoral

selection effects that can be observed in the experiment must either be due to discrimination on persistent and easily observable attributes, such as seniority, gender, or ethnicity, or due to the social cues individuals can observe within this limited communication opportunity. If electoral selection can be shown to be effective even in this very limited-information setting, it seems likely that it would also be effective in more information-rich settings, such as day-to-day interaction in a village, candidate selection procedures within political parties, or electoral campaigns.

To shed light on each of the aforementioned mechanisms, the data from the experiment will be leveraged to gain insights into several additional research questions.

- Question 1.1: Do elections with minimal prior communication enable strangers to identify public-spirited leaders within their municipality?
- Question 1.2: To what extent can embezzlement decisions in the experiment be explained by baseline preferences?
- Question 1.3: How accurate are citizens' expectations regarding the extent of embezzlement?
- Question 2.1: Are citizens willing to incur personal costs to punish embezzlement?
- Question 2.2: How does a lack of transparency influence citizens' willingness to engage in costly sanctioning behavior?
- Question 2.3: Does citizens' sanctioning behavior depend on whether decision-makers are elected or not?

A limitation in analyzing the observed sanctioning behavior in the experiment is that the decision-makers will possibly anticipate the reactions of their constituents and choose endogenously how much to embezzle, depending on how much embezzlement they expect will be tolerated. Tolerance for embezzlement may in turn be influenced by observable or unobservable social characteristics of the decision-maker (such as gender, ethnicity, social status, economic need, etc.). The observed average punishment, conditional on the amount embezzled, should therefore be lower than the counterfactual punishment behavior that would have been observed if decision-makers' embezzlement choices were exogenous: Assuming that decision-makers do not systematically seek to *attract*

punishments or *avoid* rewards, those decision-makers who expect to get away with it will embezzle more than those who expect to be punished more severely. Therefore, the observable sanctioning behavior is likely to understate citizens' true willingness to punish embezzlement and to potentially overstate citizens' willingness to reward the decision-maker.⁵

Different techniques were considered to elicit information about counterfactual (off-the-equilibrium path) sanctioning behavior: First, the use of the strategy method to collect data on citizens' willingness to sanction counterfactual embezzlement decisions. Second, collecting data on decision makers' expectations about the sanctions they would receive, for a range of potential decisions. However, after careful consideration, neither of the aforementioned options were implemented, out of concern that the added measurement step would reduce the overall quality of the experimental data in the given field setting. One reason is that the added data collection stages would have been so disruptive to the decision exercise that study participants' overall understanding of the decision situation could have been negatively affected and the decision situation would have become unnecessarily staged and artificial. Another reason is that it would have been very difficult to explain the counterfactual situations to the study participants, given low levels of literacy and formal education. A third reason, concerning the possibility of eliciting the decision-maker's expectations, is that ex-post measurements (after the decision has been made) could be conditioned by the previous decision (which may have been affected by considerations other than expected sanctions), whereas ex-ante measurement could have primed decision-makers about sanctioning stage, attenuating the direct effects of the experimental treatments treatments. Therefore, it seemed preferable in this experiment to not even attempt to elicit counterfactual sanctioning behavior.

In the absence of information on counterfactual sanctioning behavior, a lack of observable sanctioning behavior in the experiment would not necessarily imply that individuals are unwilling to engage in costly punishment, because the threat of sanctions could have prompted the decision-maker to conform to the group members' normative expectations. However, wherever costly punishment behavior can be observed in the experiment, it would be reasonable to conclude that citizens' will-

⁵In the given setup, a rational, self-interested decision-maker with personal characteristics u who has accurate expectations about the sanctioning function $s(x, u)$ could maximize her payoff by embezzling a fraction x^* of the group money such that $\partial s(x^*, u)/\partial x = 4/5$.

ingness to punish counterfactual embezzlement decisions should be greater than or equal to their willingness to punish the decision-maker for their actual, observed embezzlement decision, because those decision-makers who expect to be punished most severely will have the greatest incentive to take evasive action. Hence, under the assumption that decision-makers do not systematically seek to attract punishments, the observable spending on punishments in the experiment should be considered a lower bound on study participants' actual willingness to punish embezzlement.

8 Supplementary Analyses

For each of the supplementary research questions, an initial hierarchy of tests and analyses is outlined below. This list of proposed analyses is not exhaustive. If it becomes apparent that the data allow for even greater scrutiny of the conclusions, additional tests may be reported.

Question 1.1: Do elections with minimal prior communication enable strangers to identify public-spirited leaders within their municipality?

To test whether even in a low-information context voters are able to identify public-spirited candidates, the distribution of the baseline embezzlement decisions among those who have been elected will be compared to the distribution of baseline embezzlement decisions in the remaining population. Since the baseline behavior is recorded before elections have taken place and before study participants even knew how the decision maker in the next round would be selected, it cannot have been influenced by the experience of having been elected or by the anticipation of becoming a candidate.

To reject the null hypothesis that winning candidates are just as willing to embezzle group resources at baseline as non-winning candidates, the p-value from a Mann-Whitney U test will be reported. The distributions will also be compared graphically:

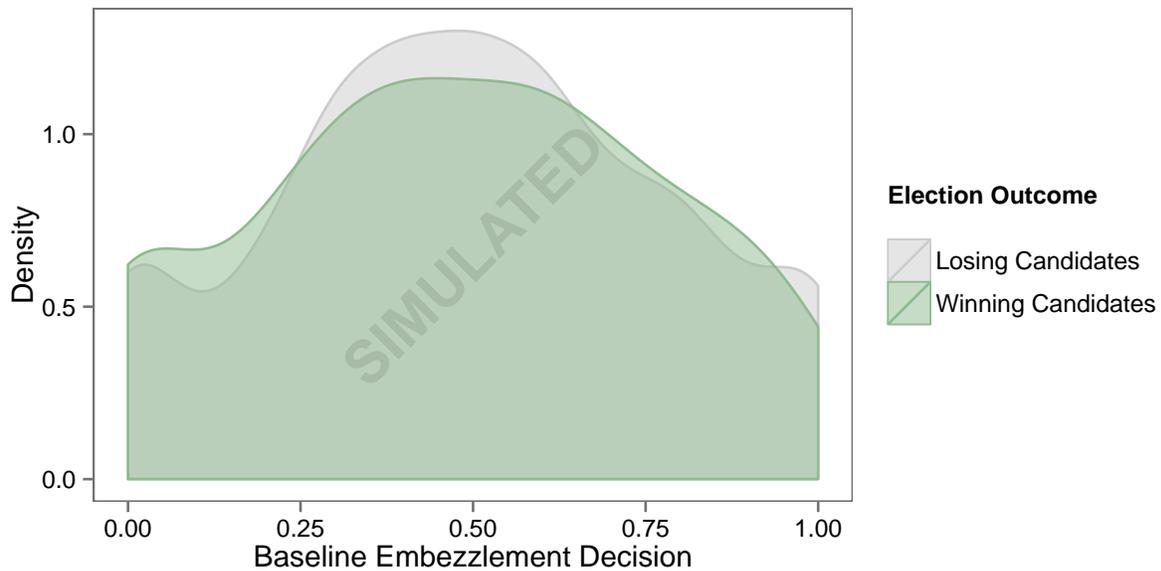


Figure 1.1.1: Comparison of baseline preferences of winning and losing candidates.

To further test whether there is a systematic relationship between public-spirited preferences and electoral success, the analysis will report an OLS regression of vote share in the first round of the election and of eventual electoral success on individuals' rank within the group in terms of baseline preferences (from 1=embezzled least to 5=embezzled most).

If there is evidence of selection effects, the next step will be to understand the heuristics voters used to identify candidates. This analysis will focus on the questions:

- What observable characteristics predict candidates' vote shares and their probability of winning?
- Do these characteristics also predict how public-spirited an individual is?

This analysis is exploratory. The three dependent variables of interest are (1) the number of votes received, (2) the probability of winning the election, (3) individuals' baseline embezzlement decisions. Regularized regression will be used to identify baseline covariates that best predict these outcomes. The initial set of observable features includes: gender, age, years of education, age rank within group, number of coethnics in the group, number of people of the same gender in the group,

prior leadership experience, self-rated wealth quartile of an individual's household within that individual's village, number of people in the group who knew the individual beforehand, number of co-villagers in the group, ballot order, color of badge.

Also if electoral selection does not favor public-spirited candidates, this analysis will be important, because it will help to understand whether voting is erratic, or whether voters focus on features which are poor predictors of public-spiritedness. If any meaningful observable predictors of electoral success can be identified, exogenous variation in the composition and in the identity of decision-makers in the random-appointment condition can be leveraged to further investigate whether features that predict electoral success are also associated with the actual embezzlement behavior of decision-makers, with group members' expectations about how much the decision-maker has embezzled, and with group members' willingness to sanction the decision-maker. Several of these potential extensions are outlined in Section 10.

Question 1.2: To what extent can embezzlement behavior in the experiment be explained by baseline preferences?

For a more complete understanding of the implications of electoral selection effects, which are measured via the preferences study participants revealed in the baseline decision, it is important to examine the relationship between baseline preferences and actual embezzlement decisions in the experiment. This relationship will be presented graphically (Figure 1.2.1). It will additionally be quantified by the cross-validated root mean squared error of a suitable parametric or nonparametric regression model that best fits the data. A strong dependence of decision-makers' choices on their baseline preferences would imply that electoral selection effects are consequential for embezzlement, at least in the experimental setting.

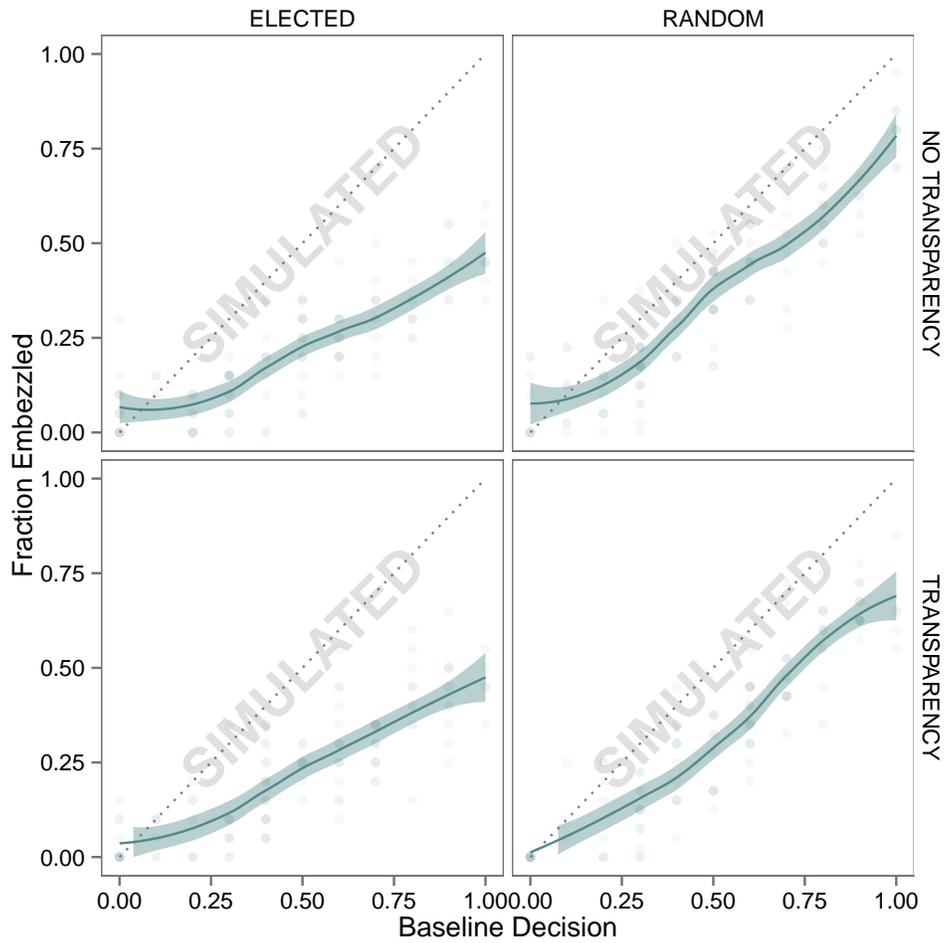


Figure 1.2.1 Local regression (LOESS) estimates of the mean fraction embezzled by baseline embezzlement decision (95% confidence intervals).

If baseline preferences are consequential for embezzlement outcomes, the next step would be to condition on selection effects and test whether there are any remaining, unexplained differences in the embezzlement behavior of elected and non-elected decision makers. To do this, the difference in means in the fraction of group resources embezzled will be reported (comparing elected and non-elected decision-makers), reweighting individual observations by inverse propensity score of being elected. Propensity scores will be estimated using baseline embezzlement decisions, as well as observable features that have been found to be associated with vote shares in Supplementary Question 1.1.

If there are differences between elected and random leaders in these weighted comparisons, those differences would be unlikely to be due to selection effects. They could be caused (1) by a possible change in decision-makers' preferences that was evoked by the experience of having been elected, or (2) by differences in the anticipated reactions of the other group members. Supplementary Question 2.3 will test whether the second mechanisms plays a role.

Question 1.3: How accurate are citizens' expectations regarding the extent of embezzlement?

To test whether citizens' expectations regarding the extent of embezzlement are consistent with actual outcomes at the level of individual decision makers, the relationship between expected and actual embezzlement will be examined in each of the four experimental conditions (Figure 1.3.1). There are two motivations for examining the accuracy of citizens' expectations at the level of individual decision-makers.

- Citizens' ability to predict the behavior of individual decision makers could influence whether pro-social candidates have an electoral advantage. This would be the case if citizens base their vote predominantly on the perceived trustworthiness of the candidates.⁶
- If transparency is lacking, citizens' expectations might be an important driver of sanctioning behavior. If citizens' expectation formation is biased in the direction of being too trusting, this could be exploited by decision-makers who have private information.⁷

⁶On the other hand, if citizens' expectations are inaccurate or uncorrelated with the actual behavior of decision-makers, this would not imply the absence of selection effects, because voters' choices might be based on heuristics other than perceived trustworthiness.

⁷Therefore, if citizens' expectations are inaccurate, we should expect that transparency has an indirect effect on sanctioning behavior, by correcting expectations. However, transparency may also have direct effects on sanctioning behavior, as discussed in Question 2.2.

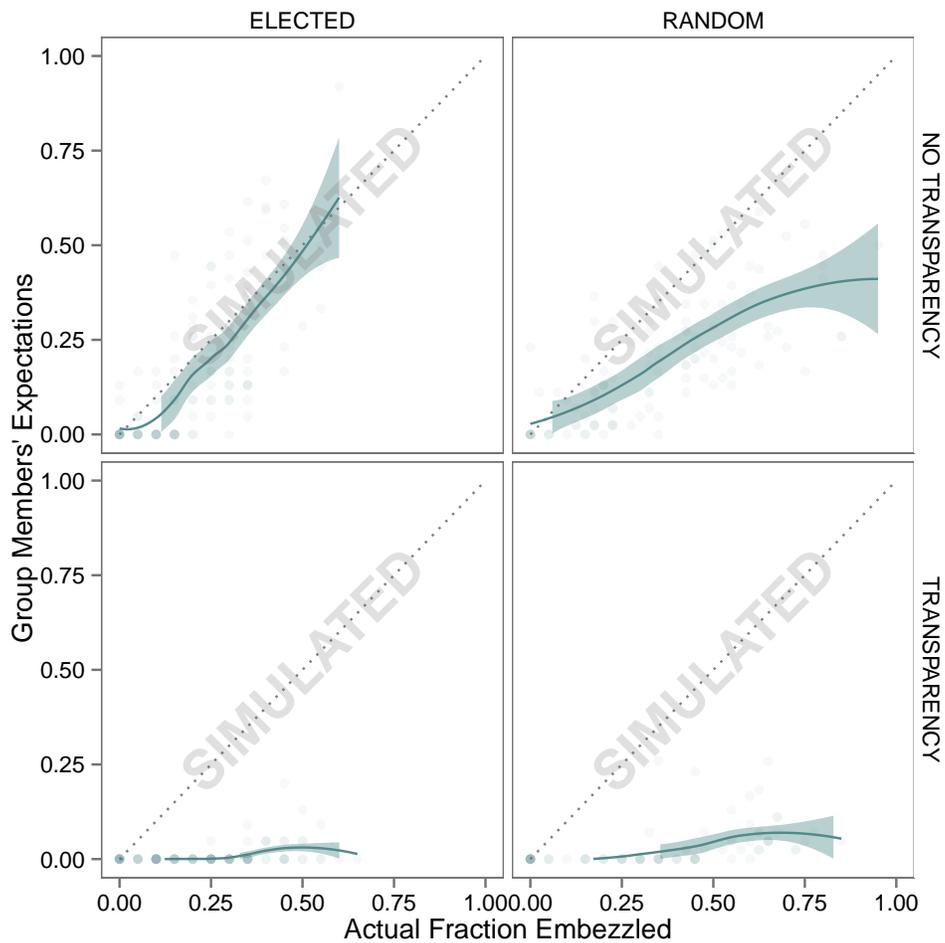


Figure 1.3.1 Local regression (LOESS) estimates of group members' expectations on the actual fraction of group resources embezzled (95% confidence intervals).

Question 2.1: Are citizens willing to incur personal costs to punish embezzlement?

In the experiment, the payoff-maximizing choice for group members is not to engage in any costly sanctioning activity. However, in line with behavior that is commonly observed in ultimatum games and other allocative dilemmas, individuals are willing to forgo own payoff in order to prevent others from receiving a greater, unequal payoff (Henrich et al., 2004). To verify whether this is also the case in the given experiment, it will be reported what fraction of study participants in the public

information condition spend any money to sanction the decision maker (i.e. to reward or punish her or him). Additionally, the average amount spent on rewards and costly punishments will be estimated as a function of embezzlement. Also the fraction of individuals who spend money on both rewards and punishments will be reported.⁸

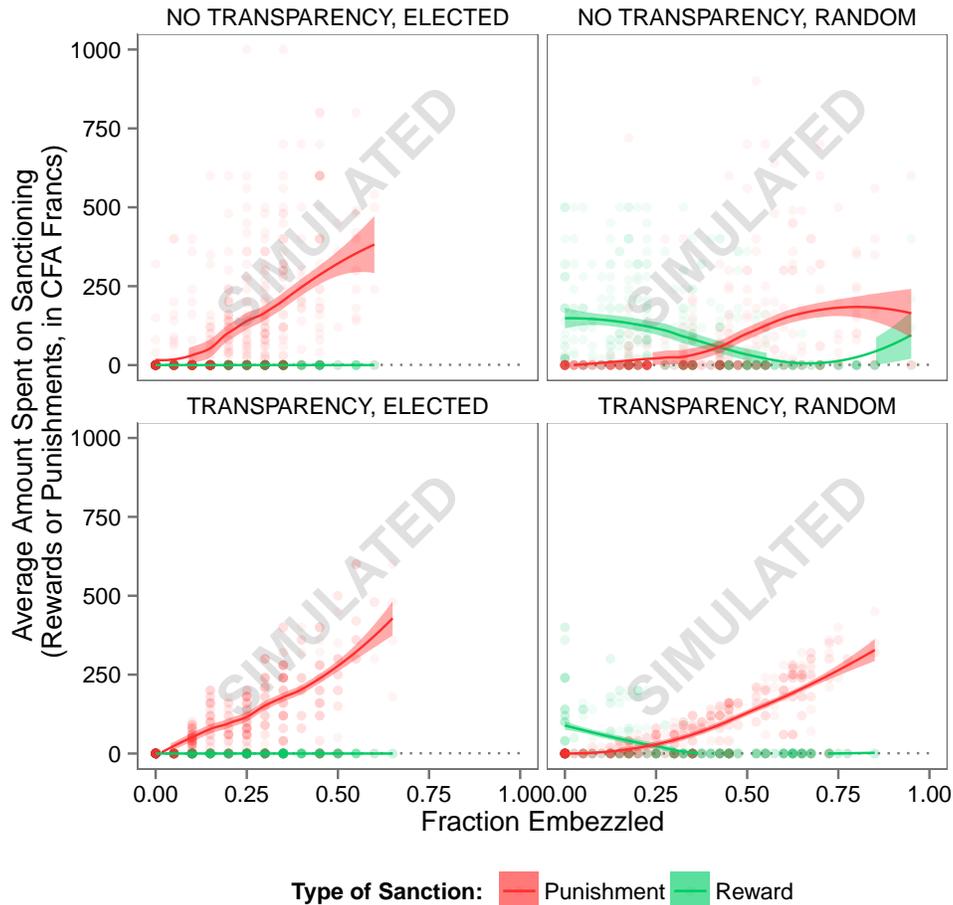


Figure 2.1.1 Local regression (LOESS) estimates of the average spending on punishments and rewards, as a function of the fraction of group resources the decision-maker has actually embezzled.

⁸During pilot tests, it was observed that some individuals chose to decrease the decision-maker's payoff by a very specific amount, e.g. by exactly 500 FCFA. A preference to decrease the decision-maker's payoff by a very specific amount, but not by any more or any less than that amount, could prompt study participants to spend, for example, two 100 FCFA coins on punishments (decreasing the decision-maker's payoff by 600 FCFA) and one 100 FCFA coin on rewards (lowering the total punishment to 500 FCFA, at an extra cost of 100 FCFA to the sender). However, if individuals allocate three or more coins to rewards, while also allocating coins to punishments, it is likely that they have not understood the logic of the sanctioning stage or made an irrational choice.

The prior expectation is that there will be sanctioning activity and that citizens' sanctioning behavior will loosely resemble the behavior of responders in an ultimatum game or in a dictator game with a second-party punishment option (Fehr and Fischbacher, 2004; Leibbrandt and López-Pérez, 2012). However, given the possibility of free riding on other group members' sanctioning efforts, sanctioning behavior in this specific experiment is expected to be lower than in a two-person dictator game with second-party punishment option.

Individual motives for sanctioning the decision-maker may vary. Inequity aversion is a likely motivation for engaging in costly sanctioning behavior (Fehr and Schmidt, 1999). Furthermore, citizens might reward the decision-maker out of positive reciprocity (for receiving any or a higher than desired share of the group money) or generosity, or punish the decision-maker out of negative reciprocity (for receiving no or a lower than desired share of the group money) or spite. Aside from personal motives, individuals might also derive an expressive benefit from enforcing a social norm of what is considered appropriate behavior in the given decision situation, especially if they are directly affected by it (Carpenter, 2007).

This study will specifically test the possibility is that individuals take their own behavior as a reference point and punish decision-makers for embezzling more than they would personally have been willing to, while rewarding them for embezzling less than they would personally have embezzled in the absence of any sanctioning capacity. To investigate this hypothesis, the average amount spent on rewards and punishments will be estimated as a function of the linear difference between the decision-maker's embezzlement decision (as a fraction of the group money) and study participants' own baseline embezzlement decision (also as a fraction of the group money). Elections and/or transparency might shift individuals' reference points/sanctioning thresholds away from their own baseline behavior. Figure 2.1.2 serves to visualize potential shifts in reference points across the four experimental conditions. If individuals sanctioning thresholds shift between the different experimental conditions, it would be consistent with the idea that elections and/or transparency alter the social norms individuals apply to the decision situation.

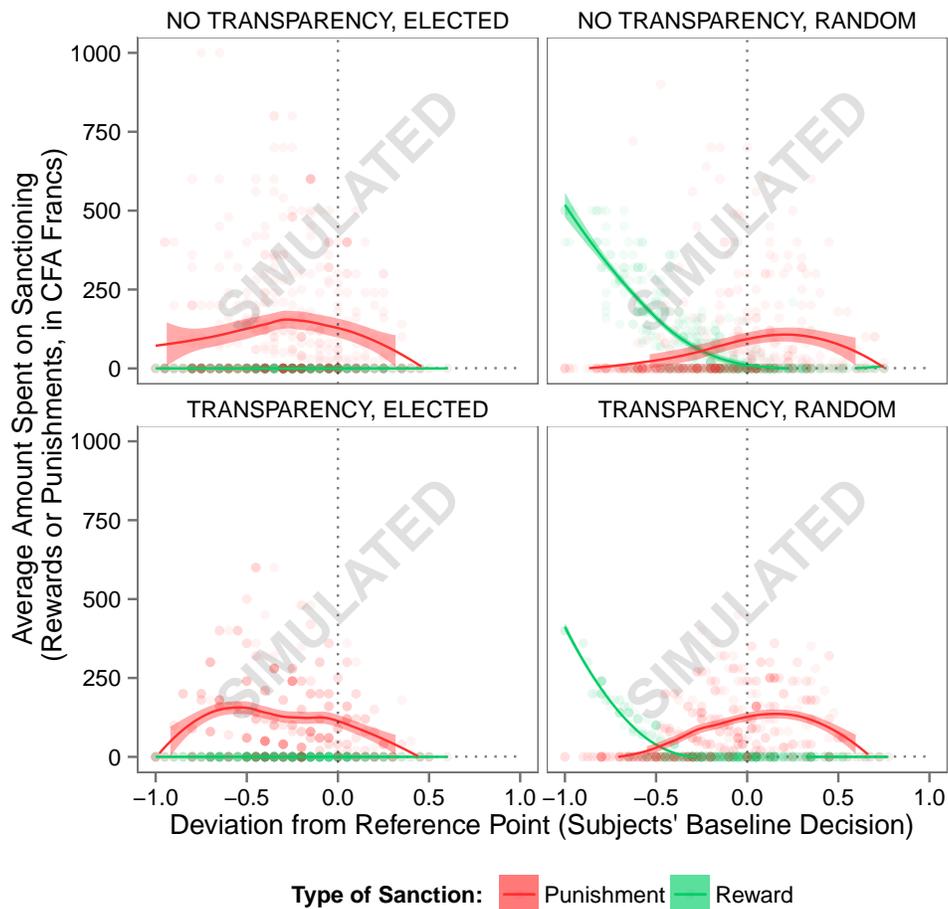


Figure 2.1.2 Do elections and/or transparency affect individual reference points/sanctioning thresholds?

Question 2.2: How does a lack of transparency influence citizens' willingness to engage in costly sanctioning behavior?

Based on the behavior observed in pilot tests of the experiment, the initial expectation is that individuals will be willing to engage in costly sanctioning activity, even if they are uncertain about how much embezzlement has actually taken place. This behavior is consistent with evidence from ultimatum games with one-sided imperfect information about the size of the pie, in which responders continue to reject offers at a cost to themselves. However, if responders do not know the size of the

pie, offers are lower on average and responders are willing to accept relatively lower offers (?, 145).

To test if uncertainty also reduces individuals' willingness to punish embezzlement by the decision-maker, the amount spent on rewards and costly punishments, as a function of perceived embezzlement, will be compared between the private information (no transparency) and public information (transparency) conditions. In the private information condition, perceived embezzlement is measured by study participants' stated expectations, in the public information condition by the actual, publicly revealed embezzlement. This analysis will be carried out for both the election and random selection conditions.

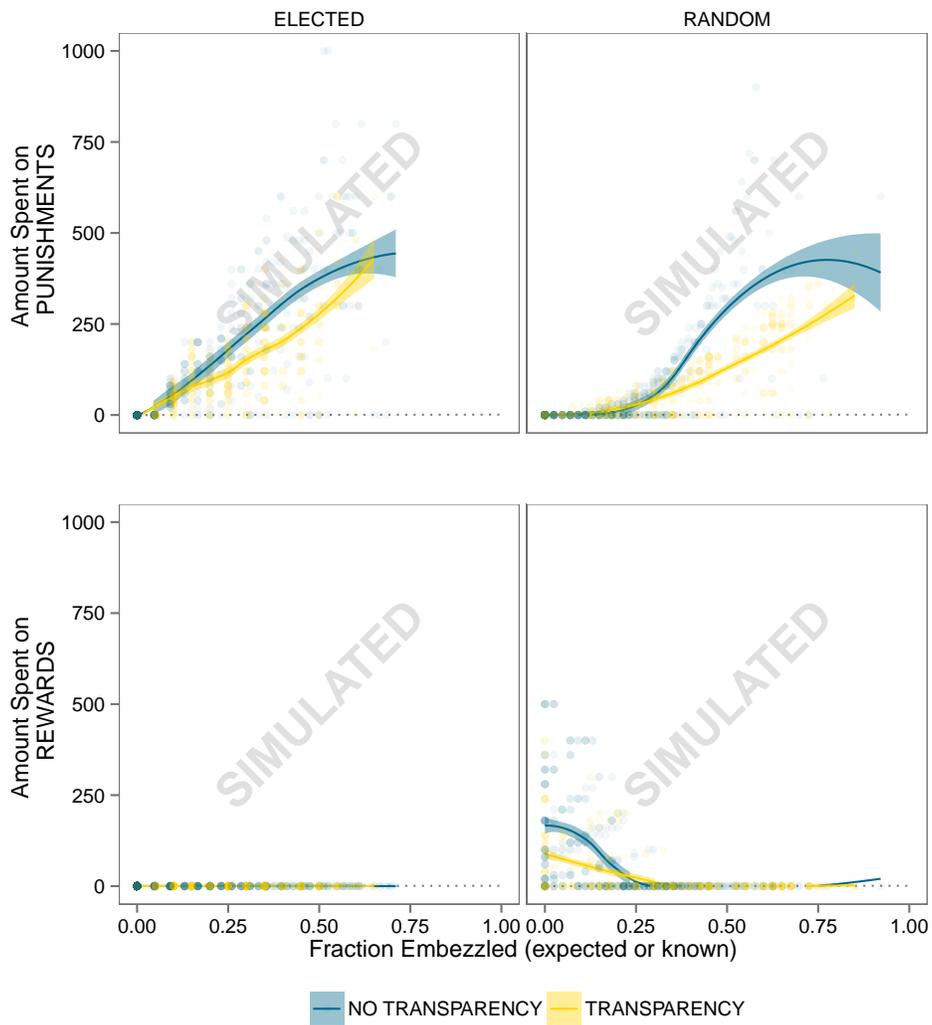


Figure 2.2.2 How does transparency affect individuals' willingness to reward or punish the decision-

maker?

Question 2.3: Does citizens' sanctioning behavior depend on whether decision-makers are elected or not?

Most importantly, citizens' willingness to sanction embezzlement of public resources might depend on whether the decision-maker is elected or not. For two reasons, citizens might be less tolerant of embezzlement by elected leaders than by randomly appointed leaders: Either because they hold them to higher standards of public-spiritedness, or because they are more willing to incur personal costs to enforce norms. In either case, the consequence would be that elected decision makers might be confronted with a more severe threat of sanctions, which could cause them to refrain from embezzlement more than non-elected decision makers.

To shed light on whether this might be the case, the study tests whether identical observed decisions by elected and non-elected leaders elicit differential responses from citizens. This is done by estimating the average sanction (punishment or reward) as a function of the fraction of group money embezzled by the decision maker (Figure 2.3.1). It should be emphasized that this test relates to the question whether elected and non-elected leaders *whose decisions are identical* elicit different reactions from citizens. This is distinct from the question in what ways citizens would sanction an *identical set of decision-makers* in the counterfactual situation that those decision-makers had been elected instead of randomly selected, or vice versa. In the experiment, as in reality, elected and non-elected leaders can be expected to differ in their observable and unobservable characteristics, including their intrinsic motivation to refrain from embezzlement and the social incentives they are confronted with (Lierl, 2014).

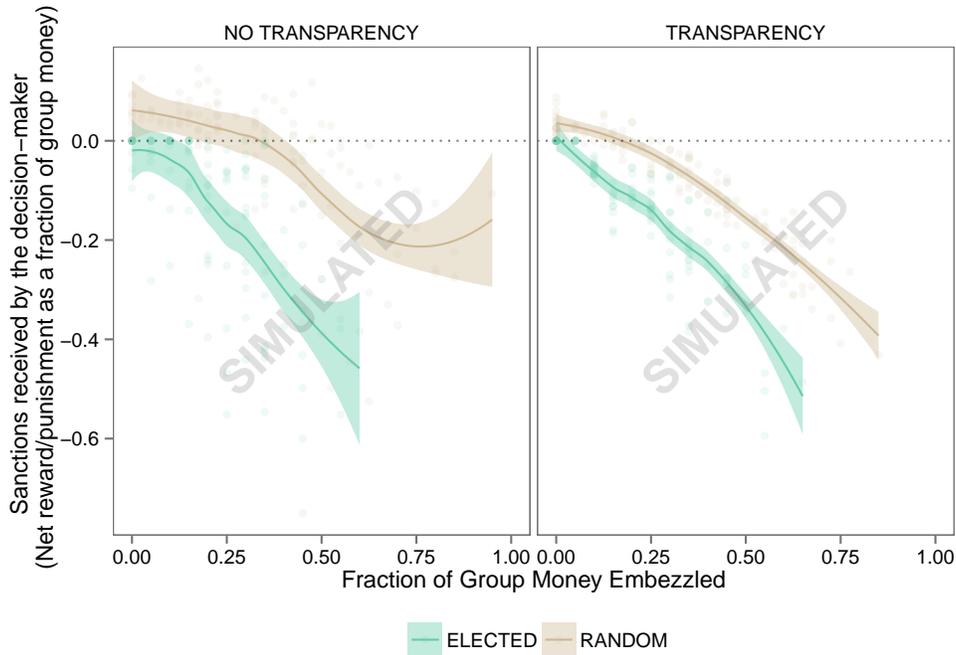


Figure 2.3.1 Observable sanctioning behavior in the election and random appointment conditions. The graphs show LOESS estimates of the average sanctions received by the decision-maker as a function of the fraction of group resources embezzled by the decision-maker.

As argued previously in Section 7, observed punishment behavior for a given level of embezzlement might be a biased representation of citizens' actual willingness to punish embezzlement, because decision-makers may be able to anticipate the group members' reactions and adjust their embezzlement decisions to avoid punishments. Specifically, if the decision-makers are able to correctly anticipate group members' reactions and to take evasive action, then the differences in group members' observable sanctioning behavior between the election and random appointment conditions could be attenuated by that, because embezzlement decisions would adjust endogenously to reduce differences in expected punishments between the two experimental conditions. This means that if elections have an impact on the average observed sanctioning decisions for a given level of embezzlement, the differences in counterfactual sanctioning functions (i.e. the difference we would expect to see if embezzlement decisions were held fixed) should be even more pronounced.

9 Validation Strategy

The interpretation of the experimental results can be strengthened by addressing two potential concerns regarding their conceptual validity. The first concern is whether the outcome measurements are valid: Do embezzlement decisions in the experiment reflect embezzlement behavior of public decision-makers in the real world? The second potential relates to the validity of the experimental setting as a whole: Are the causal effects observed in the experiment relevant in natural settings, i.e. in analogous real-world situations outside the experiment?

Conceptual Validity of Outcome Measures

Since study participants' willingness to embezzle public resources cannot be observed in natural settings, a direct validation of the primary outcome measure (embezzlement decisions) is difficult. The reason is that study participants are usually not in a position to embezzle public resources outside the experiment, because they are not public decision-makers (and if they were, it would be very difficult to measure the extent to which they embezzle public resources, let alone in a way that is comparable across subjects).

One way of circumventing this challenge is to look at aggregate, municipality-level relationships instead. Local norms and shared experiences with respect to the embezzlement of public resources may vary across municipalities. It could therefore be tested whether embezzlement behavior in the experiment (measured by the average baseline embezzlement decision in a municipality) is correlated with measures of municipal governance quality, which in turn should depend on corruption and embezzlement in the municipal administration.

As a source of data on the quality of public services in a municipality, municipal performance scorecards will be used, which are maintained by the *Programme d'Appui aux Collectivités Territoriales*. Specifically, the dependent variables will be the 2013/14 overall point rating for public service delivery (under the previous elected municipal governments), which is a weighted sum of performance scores for individual municipal services.

Alternatively, at the individual level, it may be possible to validate the primary outcome measure indirectly, via citizens' expectations (the secondary outcome measure). If individuals' expectations in the experiment are correlated with their expectations about embezzlement in the real world, this would be consistent with the idea that there are similarities in how they perceive both situations. To test this, three baseline measures of individuals' expectations about their municipal leaders have been incorporated into the study:

- Willingness to donate anonymously to the municipal treasury: Study participants are given the opportunity to anonymously donate any fraction of their show-up compensation of 2000 CFA Francs to the municipal treasury. Individuals' donations should reflect both their willingness to contribute to local public goods in their municipality, as well as their generalized trust that municipal resources are managed well. If, conditional on their baseline embezzlement decision (which is a proxy measure for how public-spirited they are), individuals' expectations of how much embezzlement has taken place in the experiment are correlated with their trust in municipal decision-makers in the real world, then this would be consistent with the claim that individuals think about embezzlement in the experiment in similar ways as they think about embezzlement by their municipal administration.
- Expectations about self-seeking behavior by elected municipal councilors: *“In your commune, did most of the elected municipal council members use their position to enrich themselves?”*
- Expectations about self-seeking behavior by the mayor: *“In your opinion, do you think that the mayor of your commune is more willing to enrich himself than the mayors of most other communes?”*

Validity of the Elections Treatment

Heterogeneity in the effects of the elections treatment across municipalities could be leveraged to validate that the causal effects of elections in the experiment are relevant in the real world. If variation in real-world measures of governance quality by elected municipal governments can be explained by heterogeneity in the effects of elections on embezzlement in the experiment (conditional

on how much embezzlement is observed in the absence of elections), then this could suggest that the causal effects observed in the experiment capture a phenomenon that also shapes governance outcomes in the real world. For example, across municipalities, the effects of the election treatment may vary due to differences in voters’ ability to identify public-spirited candidates, due to differences in the social norms associated with elected leadership, or due to differences in the effect of elections on citizens willingness to punish embezzlement. These sources of heterogeneity might also influence governance outcomes in the real world.

To test this hypothesis, the study takes advantage of the fact that treatment assignment is blocked by municipality, such that there is one group in each of the four treatment conditions in every municipality. This makes it possible to obtain an unbiased (but presumably highly unreliable, given the small number of groups in a municipality) estimate of the average treatment effect of elections on embezzlement in every municipality. Regression models of the following form will be estimated to examine the relationship between the effect of elections in the experiment and real-world governance outcomes:

$$p_m = \beta_0 + \beta_1\tau_m + \beta_2z_m + \beta_3x_m + \epsilon_m$$

where p_m is a measure of governance quality in the municipality m , τ_m is the average treatment effect of elections on embezzlement in the municipality (pooling across the transparency and no-transparency conditions), z_m is the average level of embezzlement in the random appointment condition, and x_m is a vector of municipal-level covariates: region, total population and population density. Of interest is the coefficient β_1 . Under the null hypothesis that the effect of the elections treatment on embezzlement in the experiment is unrelated to the effects of elections on governance quality in the real world, β_1 should be zero.

As measures of municipal governance quality, the following data will be used: (1) the aforementioned scorecard rating of the quality of public services, (2) the fraction of respondents in a municipality affirming that “most of the elected municipal council members used their position to enrich them-

selves”, (3) the fraction of respondents in a municipality affirming that the mayor of their commune “is more willing to enrich himself than the mayors of most other communes”.

Since the proposed real-world outcome measures do not directly quantify embezzlement, the assumed causal relationships that link the effects in the experiment with the real-world outcomes are very indirect, and any relationship between them should be very weak to begin with. Additionally, the fact that only four distinct data points are observed per municipality weakens the reliability of the municipality-level treatment effect estimates, which further decreases the statistical power to detect an existing correlation between the effect of elections in the experiment and real-world governance outcomes. Null findings in these proposed validation checks should therefore not be taken to imply that the elections treatment lacks conceptual validity; they would be uninformative. Instead, the proposed validation opportunity should be seen as one of several potential ways of justifying the conceptual validity of the elections treatment in the experiment, which is worth trying, even if it is most likely to remain uninformative.

10 Possible Extensions/Ancillary Experiments

Group-level effects of the decision-maker’s social identity

Exogenous selection of the decision-maker in the random appointment condition makes it possible to estimate the impact of the decision-maker’s identity (in terms of salient social characteristics) on group members’ expectations towards them, as well as on the relationship between decisions and observed sanctioning behavior. In the local context, three dimensions of social discrimination are potentially salient: gender, co-ethnicity, and seniority. If any of these social characteristics are correlated with electoral success (as explored in Supplementary Question 1.1), then it becomes important to understand whether they simultaneously influence group members’ expectations and their willingness to sanction the decision-maker. This can be tested by estimating the effects of being assigned to a male or female, co-ethnic or non-co-ethnic, junior or senior decision-maker in the random appointment condition (adjusting for the demographic proportions that determine the

respective treatment assignment probabilities).

Female vs. male decision-makers

To test whether female decision-makers face different expectations and reactions than male decision-makers, coefficients from weighted least squares regressions will be reported. The dependent variable is a study participant's expectation about how much the decision-maker has embezzled, the independent variable is an indicator of whether the decision-maker is female. The results will be presented separately by gender of the group member. If ballot order and/or badge color qualify as instruments for electoral success, this analysis can potentially be extended to the election condition as well.

Additionally, group members' sanctioning behavior (as a function of actual embezzlement) will be compared across male and female decision-makers (Figure 3.1.1). In the transparency condition, the comparison of sanctioning behavior towards male and female decision-makers for identical embezzlement decisions should shed light on whether there are gender differences in how much embezzlement is socially accepted. In the no-transparency condition, differences in the sanctioning behavior might additionally reflect gender-based prejudice in group members' expectations of how much money the decision-maker has embezzled.

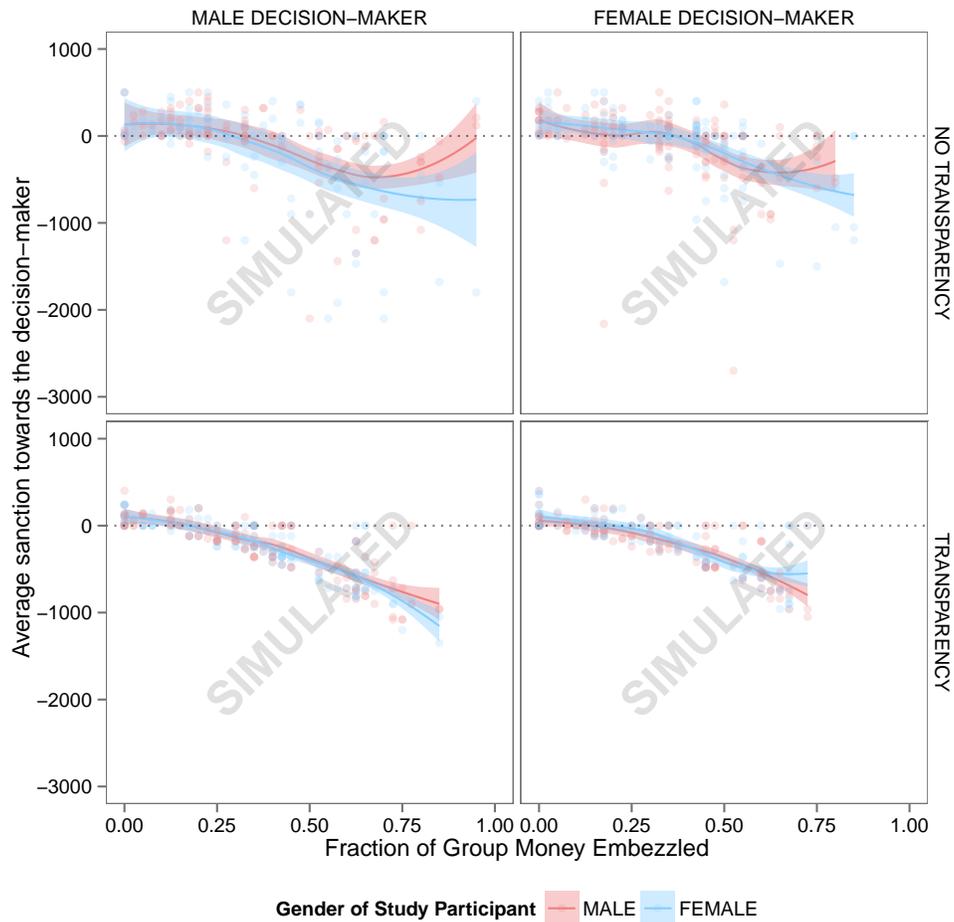


Figure 3.1.1 Sanctioning behavior towards male and female decision-makers in the random appointment condition, by gender of the study participant.

Co-ethnic vs. Non-co-ethnic decision-makers

To test whether citizens are more tolerant of embezzlement by decision-makers of their own ethnicity, individuals' sanctioning behavior will be compared across the cases of decision-makers of the same ethnicity and decision-makers of a different ethnicity. Estimated mean sanctioning decisions conditional on how much embezzlement has taken place will be presented graphically (Figure 3.2.1) for both cases. Depending on the functional form and outcome distribution, an appropriate parametric or semi-parametric regression model will be used to estimate a p-value for the null hy-

pothesis that the conditional mean function is identical in both cases. In this analysis, observations will be weighted by the inverse probability of being subjected to a co-ethnic decision-maker.

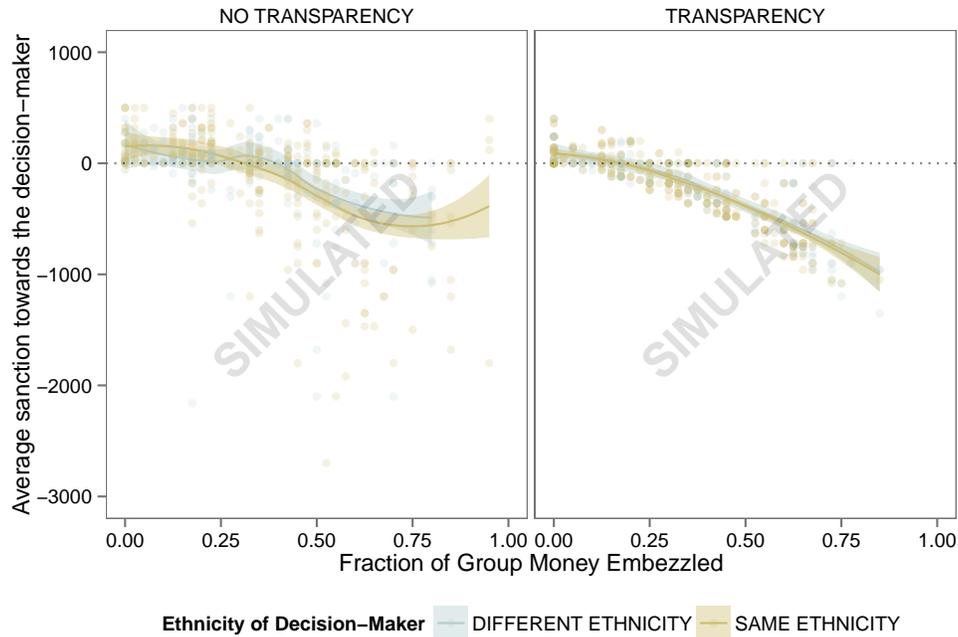


Figure 3.2.1 Sanctioning behavior towards co-ethnic and non-co-ethnic decision-makers in the random appointment condition.

Furthermore, citizens' expectations towards decision-makers of their own ethnicity and decision-makers of a different ethnicity will be compared. The average treatment effect of being assigned to a co-ethnic decision-maker on group members' expectations will be estimated via weighted least squares regressions, adjusting for the respective treatment assignment probabilities.

Seniority of the decision-maker

To test whether relatively more senior decision-makers face different expectations and reactions than relatively more junior decision-makers, the data analysis will compare group members' expectations by age rank of the decision-maker within the group and juxtapose it to actual decision outcomes (Figure 3.3.1). Mean age by age rank will be reported.

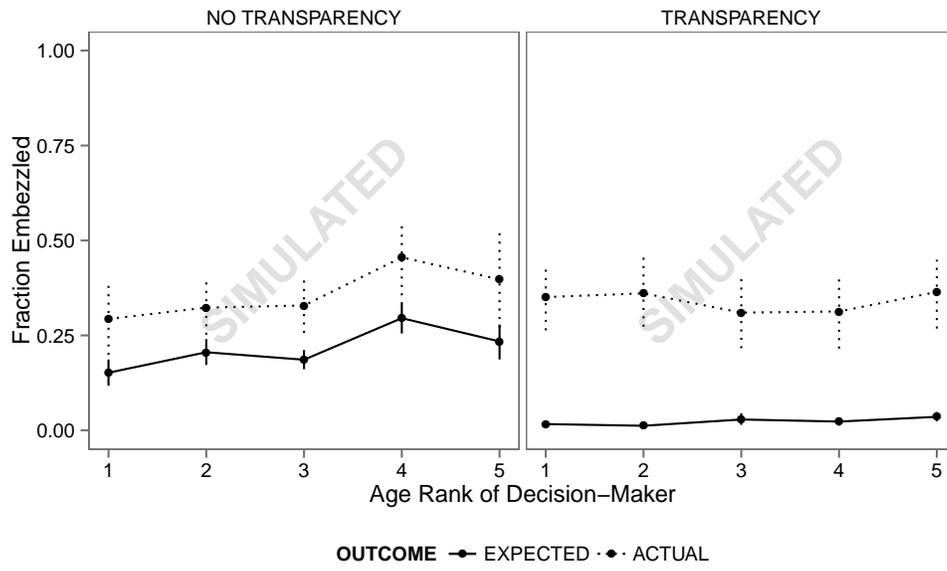


Figure 3.3.1: Expectations and actual decisions, by seniority of the decision-maker within the group.

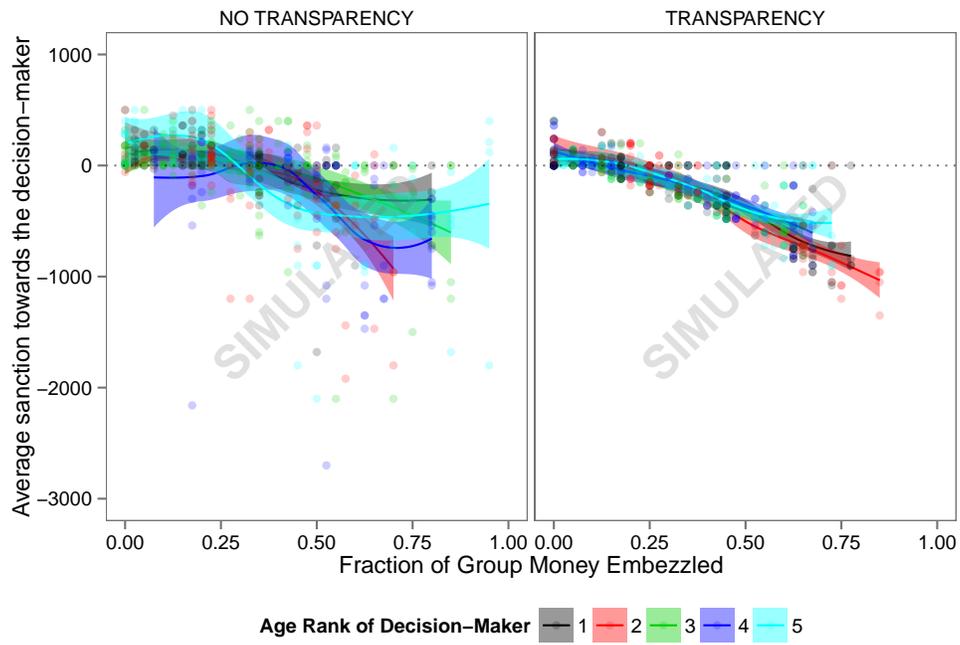


Figure 3.3.2: Observed sanctions by embezzlement decision.

Individual-level effects of being selected as a decision-maker

The random selection of the decision maker is by itself an interesting ancillary experiment, through which both the individual-level effects of being put in a position of responsibility, as well as the group-level effects of the decision-maker's social identity (especially with respect to gender, coethnicity, and seniority) can be evaluated.

Does selection into a position of responsibility increase perceived social status?

This experiment examines the impact of being randomly selected into a position of responsibility (deciding over the allocation of the group money) on the perceived relative social status of an individual within the group. Two outcomes are of interest: (1) the number of other group members who consider that a person has the highest social status within the group and (2) whether an individual considers that she/he has the the highest social status within the group. This is measured through a survey question after the experiment, asking every group member to indicate who within the group they believe has the highest social status.

The average effects of having been selected as decision-maker will be estimated via an OLS regressions of the form

$$s_i = \beta_0 + \beta_1 d_i + x_i' \beta_2 + \epsilon_i$$

where, s_i is the outcome of interest, and d_i indicates whether an individual i has been selected as the decision-maker. x_i is a vector of individual-level covariates believed to be associated with social status: Prior leadership experience, self-rated wealth quartile of the individual's household within their village, gender, age, and years of schooling completed.

Ballot-order effects

In the elections condition, the order of candidates on the ballot is randomized at the group level, as is the color of the badges (black, blue, green, red, yellow) by which they are identified. To investigate whether attributes which are independent of a candidate's qualification or character

can influence electoral outcomes, the average number of votes received in the first round will be compared by ballot order position and badge color. A Kruskal-Wallis test will be used to test whether any ballot order position or any badge color stochastically dominates another in terms of electoral outcomes.

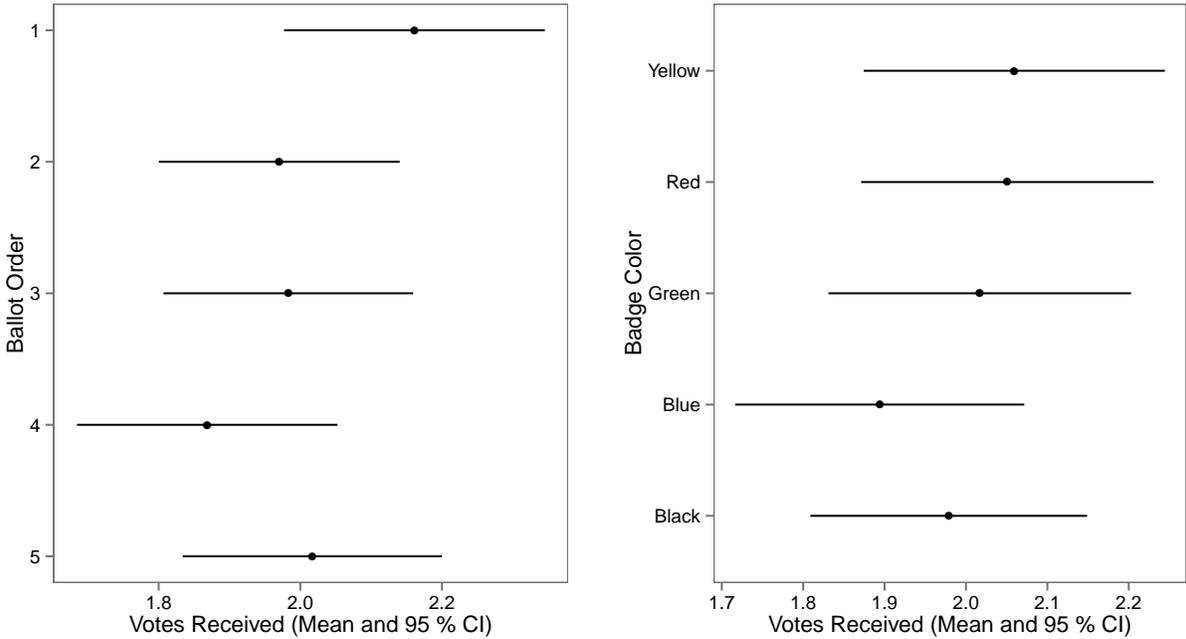


Figure 4.1.1 Average first-round election results by ballot order and badge color.

The purpose of testing whether these randomly assigned characteristics (ballot order and/or badge color) have a substantial influence on voting outcomes (which is not implausible in a low-information environment) is to explore whether they qualify as instrumental variables for electoral success.

11 Implementation and Timeline

The experiment is being carried out in 118 rural municipalities in Burkina Faso in June and July 2015. 30 randomly sampled adults from different villages of a municipality are being invited to a central location (typically the meeting room of the municipal administration). Of these 30 potential

study participants, 20 are part of the original sample, and 10 are backup participants.⁹ Backup participants are used to replace those individuals in the original sample who do not show up at the study site or choose not to participate.

All invited study participants (including the backup participants and those who come to the study site but refuse to participate), receive an unconditional show-up compensation of 2500 FCFA, which is enough to cover their transportation costs and time. Upon arrival at the study site, all study participants complete the informed consent procedure, during which they have the opportunity to ask questions and decide whether they participate. Subsequently, sampled study participants or their replacements receive a randomly assigned badge with a barcode by which study participants are identified during the data collection session as well as a pictogram by which surveyors identify which group a study participant has been assigned to. At the study site, study participants are required to wear around their neck at all times during their participation in the study. The same set of badges is re-used at each study site and the different badges are assigned at random among the 20 study participants at each site. Within a session, the Badge ID serves as a unique identifier, making it possible to link study participants responses in the baseline survey to their decisions and responses in the experiment, which are recorded through different software systems. The Badge ID is also used to process payments of the payoffs study participants earned during the experiment, without the need to record study participants' names or any other identifying information.

After completing the baseline survey in a one-on-one setting with a surveyor, study participants are assembled by group in separate locations. The survey teams are instructed to ensure that no communication takes place between members of different groups, nor among study participants of the same group, except as required by the experimental protocol. Every group of five study participants is supervised by two surveyors, one of whom is responsible for recording data and facilitating the decision exercise, while the other is responsible for administering the instructions and comprehension tests to the study participants. While one facilitator fulfills his or her responsibilities, the other facilitator simultaneously watches that no verbal or non-verbal communication takes place

⁹In consultation with the PI, the number of backup participants was decreased in several municipalities where sufficiently high turnout and participation rates were anticipated, in order to avoid unnecessary costs and logistical efforts.

between study participants and the decision exercise is shielded from any external interruptions or bystanders. If a facilitator becomes unavailable during a decision exercise for whatever reason, he or she is replaced by the team leader so as to not interrupt the dual supervision.

In order to minimize interviewer effects, to prevent errors and deviations from the protocol, and to ensure exact replicability, the entire experimental and data collection procedures have been standardized and pre-programmed into tablet computer applications. For the baseline survey, SurveyCTO is used. For the decision-exercise, a custom-designed Android application has been developed. This application is run from two tablets. One tablet is used to administer pre-recorded video instructions to the study participants in the correct sequence. These video instructions have been synchronized into nine vernacular languages. The video instructions are divided into small blocks, which are automatically followed by comprehension checks. The comprehension checks as well as other very simple instructions and directions are administered orally by the facilitator, following a strictly specified protocol. The other tablet is used to record data and to provide intuitive touch screen interfaces for the allocation decisions and for the voting procedure (see screenshots in Appendix A1). The tablet application records study participants' decisions and background data in real time and in an anonymized form.

12 Sampling and Power Calculations

Sampling and Study Population

The experiment is carried out in all 118 rural municipalities that are located within six out of 13 administrative regions of Burkina Faso: Sahel, Centre-Nord, Centre-Sud, Centre-Est, Plateau Central and Cascades. In every municipality, ten villages were sampled at random with equal probability, except in municipalities with fewer than ten villages. In municipalities with fewer than ten villages, all villages were included in the sample. As a result, the sample includes 1125 villages.

In each of the villages sampled, a comprehensive census of the population aged 18-65 was carried out in 2013. This census serves as a sampling frame for multiple studies and interventions and was

financed and implemented by Burkina Faso's *Programme d'Appui aux Collectivités Territoriales* (PACT) with technical support from the World Bank and from the author of this study. For the census, the eligible population of every *quartier* (hamlet) of the sampled villages was listed in a standardized, numbered register, consisting of several volumes if necessary. Within every *quartier*, individuals were listed by *concession* (cluster of households) and household. This census served as the sampling frame for this study. In every village, ten individuals were sampled by randomly drawing index numbers from the census booklets of that village, in proportion to the total number of individuals listed in that booklet. Since booklets never included more than one quartier, this procedure ensured that within villages, the sample is self-weighted and stratified by quartier.

The ten individuals sampled within a village were placed into a random priority order. Individuals were invited to participate in the study according to this priority ordering. If an individual declined to participate, had moved, or was unable to participate without elevated risk to her-/himself or others, the next individual in the list would be approached. In municipalities with ten or more villages, a total three individuals were invited to participate in the study, starting from the top of the priority ordering. In municipalities with fewer than ten villages, the number of invited study participants per village was adjusted upwards, so that per municipality 30 individuals were invited to participate in the study. Of the invited study participants, two thirds (i.e. 20 individuals per municipality) were randomly designated as original study participants, the remaining third as backup study participants. The backup study participants were used to replace original study participants who did not show up at the study site or declined to participate. The procedure for replacements followed a strict protocol, summarized in the flowchart in Appendix A4.

Power Calculations

Statistical power will be most constrained for the evaluation of group-level outcomes (Hypotheses 1a, 1b and 2). Power calculations for the group-level outcomes are reported below for a range of effect sizes, assuming a sample size of 472 groups and Bonferroni correction for three comparisons (since there are a total of three outcome variables in the experiment). These power calculations

have been obtained via simulation (see R code in Appendix A3).

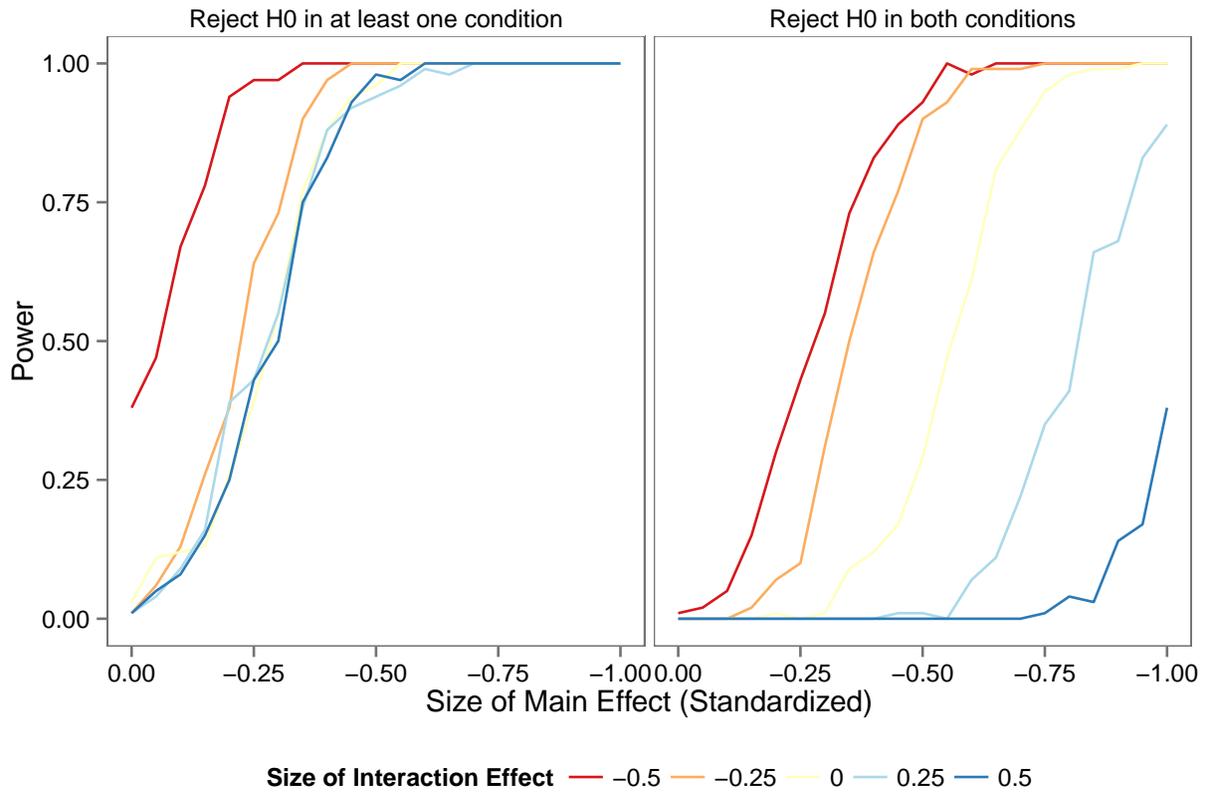


Figure 5.1.1: Statistical power by effect size, assuming no covariate adjustment and a significance level $\alpha = 0.05/3$.

References

- Alatas, V., Olken, B. A., and Wai-poi, M. (2013). Does Elite Capture Matter? Local Elites and Targeted Welfare.
- Baldassarri, D. and Grossman, G. (2011). Centralized sanctioning and legitimate authority promote cooperation in humans. *Proceedings of the National Academy of Sciences*, 108(27):11023–11027.
- Banerjee, A. V., Kumar, S., Pande, R., and Su, F. (2010). Do informed voters make better choices? Experimental evidence from urban India.

- Beath, A., Christia, F., and Enikolopov, R. (2014). Do Elected Councils Improve Governance? Experimental Evidence on Local Institutions in Afghanistan.
- Besley, T. (2005). Political Selection. *Journal of Economic Perspectives*, 19(3):43–60.
- Besley, T. and Burgess, R. (2002). The Political Economy of Government Responsiveness: Theory and Evidence from India. *The Quarterly Journal of Economics*, 117(4):1415–1451.
- Carpenter, J. P. (2007). Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. *Games and Economic Behavior*, 60(1):31–51.
- Charness, G. and Gneezy, U. (2008). What’s in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior and Organization*, 68(1):29–35.
- Chong, A., de la O, A. L., Karlan, D. S., and Wantchekon, L. (2013). Looking Beyond the Incumbent: Exposing Corruption and the Effect on Electoral Outcomes.
- Corazzini, L., Kube, S., Maréchal, M. A., and Nicolò, A. (2014). Elections and Deceptions: An Experimental Study on the Behavioral Effects of Democracy. *American Journal of Political Science*, 58(3):579–592.
- Croson, R., Boles, T., and Murnighan, J. K. (2003). Cheap talk in bargaining experiments: Lying and threats in ultimatum games. *Journal of Economic Behavior and Organization*, 51(2):143–159.
- Fearon, J. D. (1999). Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance. In Przeworski, A., Stokes, S. C., and Manin, B., editors, *Democracy, Accountability, and Representation*, chapter 2, pages 55–97. Cambridge University Press, Cambridge.
- Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2):63–87.
- Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

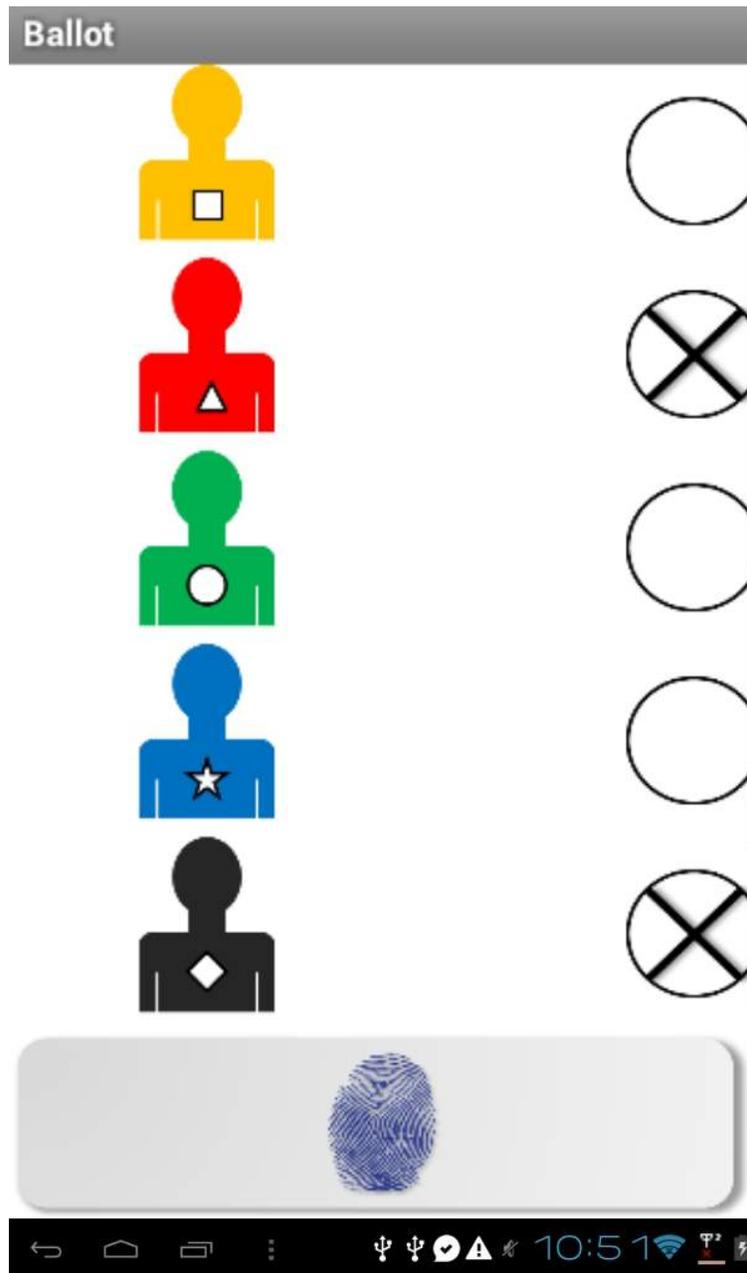
- Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice*, 50(1-3):5–25.
- Ferraz, C. and Finan, F. (2008). Exposing Corrupt Politicians: The Effects of Brazil’s Publicly Released Audits on Electoral Outcomes. *Quarterly Journal of Economics*, 123(2):703–745.
- Ferraz, C. and Finan, F. (2011). Electoral Accountability and Corruption: Evidence from the Audits of Local Governments. *American Economic Review*, 101(4):1274–1311.
- Gottlieb, J. (2015). Greater Expectations: A Field Experiment to Improve Accountability in Mali. *forthcoming, American Journal of Political Science*.
- Grossman, G. and Baldassarri, D. (2012). The Impact of Elections on Cooperation: Evidence from a Lab-in-the-Field Experiment in Uganda. *American Journal of Political Science*, 56(4):964–985.
- Haley, K. J. and Fessler, D. M. T. (2005). Nobody’s watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26(3):245–256.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., and Gintis, H. (2004). *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford University Press, Oxford.
- Humphreys, M., Sanchez de la Serra, R., and van der Windt, P. (2012). Social and Economic Impacts of Tuungane: Final Report on the Effects of a Community Driven Reconstruction Program in the Eastern Democratic Republic of Congo. Technical Report June, Columbia University, New York.
- Leibbrandt, A. and López-Pérez, R. (2012). An exploration of third and second party punishment in ten simple games. *Journal of Economic Behavior and Organization*, 84(3):753–766.
- Levi, M., Sacks, a., and Tyler, T. (2009). Conceptualizing Legitimacy, Measuring Legitimizing Beliefs. *American Behavioral Scientist*, 53:354–375.
- Lierl, M. (2014). Preferences or Incentives: What Motivates Village Leaders to Refrain From Misappropriating Public Resources?

- Lierl, M. (2015). Social Sanctions and Informal Accountability: Evidence from a Laboratory Experiment. *forthcoming, Journal of Theoretical Politics*.
- Litschig, S. and Zamboni, Y. (2012). Audit risk and rent extraction: Evidence from a randomized evaluation in Brazil.
- Malesky, E., Schuler, P., and Tran, A. (2012). The Adverse Effects of Sunshine: A Field Experiment on Legislative Transparency in an Authoritarian Assembly. *American Political Science Review*, 106(4):1–25.
- Olken, B. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115(2):200–249.
- Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press, Princeton.
- Reinikka, R. and Svensson, J. (2005). Fighting Corruption to Improve Schooling: Evidence from a Newspaper Campaign in Uganda. *Journal of the European Economic Association*, 3(2-3):259–267.
- Rigdon, M., Ishii, K., Watabe, M., and Kitayama, S. (2009). Minimal social cues in the dictator game. *Journal of Economic Psychology*, 30(3):358–367.
- Tsai, L. (2007). Solidary Groups, Informal Accountability, and Local Public Goods Provision in Rural China. *American Political Science Review*, 101(2):355–372.

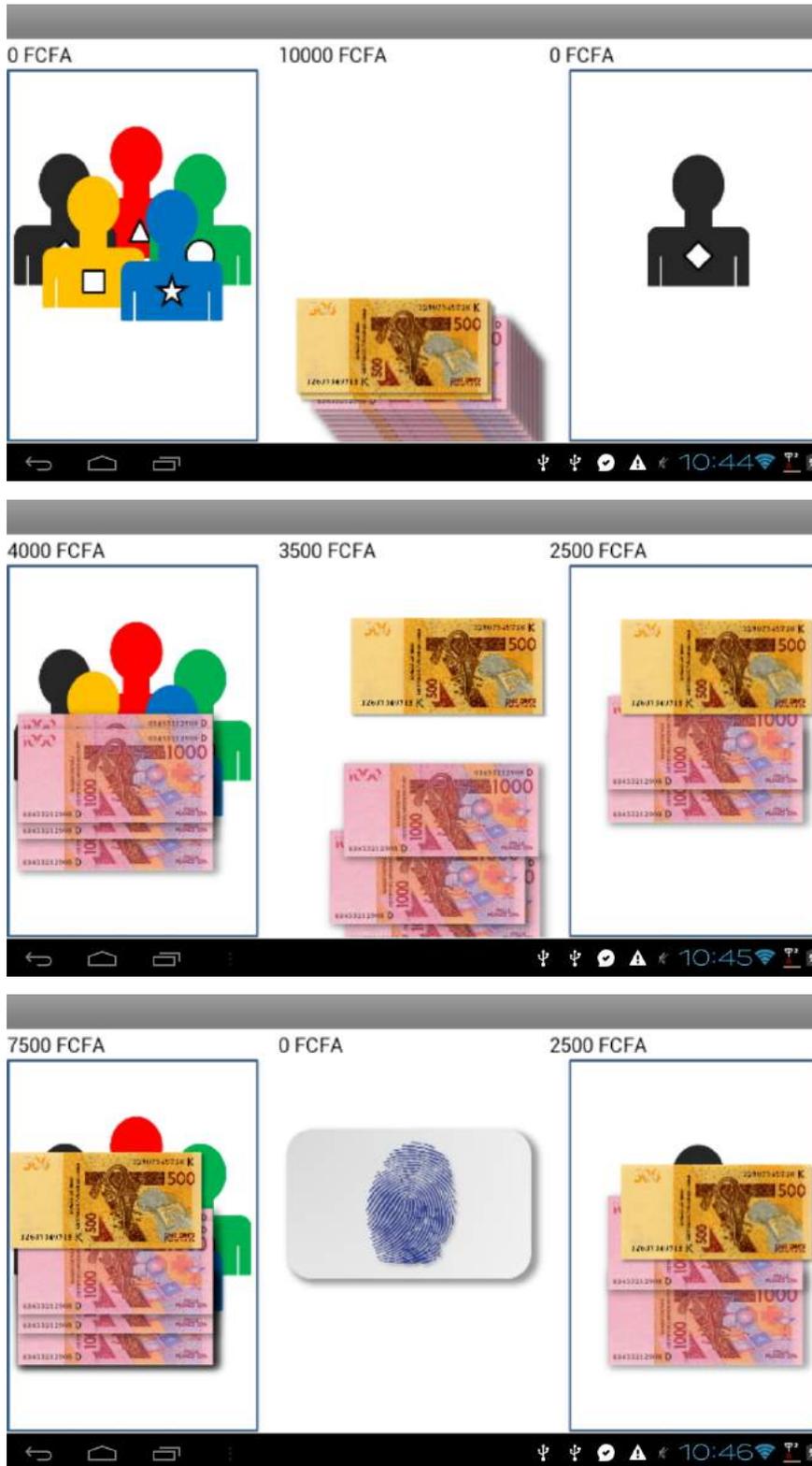
Appendix

A1. Screenshots of the touch screen interface

Voting Decision



Embezzlement Decision



A2. R code for all figures in this pre-analysis plan

```
#####  
# DATA ANALYSIS CODE FOR "ELECTIONS AND EMBEZZLEMENT" PRE-ANALYSIS PLAN #  
# *Author Omitted*, July 20, 2015 #  
# For further information see submission at EGAP social science registry #  
#####  
  
rm(list=ls())  
  
require(ggplot2)  
require(ggthemes)  
require(grid)  
require(gridExtra)  
require(Hmisc)  
require(rms)  
require(mfx)  
require(foreign)  
  
## SET WORKING DIRECTORY  
setwd("C:/Users/Admin/Dropbox/Research/Burkina Baseline Experiments/Pre-  
Analysis Plan")  
experiment<-read.dta("Simulated Data for PAP.dta") #load simulated data  
  
## DISCLAIMER WHILE TESTING CODE ON SIMULATED DATA  
disclaimer<-annotation_custom(grid.text(label="SIMULATED",rot=45, gp=gpar(  
cex=2,col="gray40",fontface = "bold", alpha = 0.2)),xmin=-Inf, ymin=-  
Inf, xmax=Inf, ymax=Inf)  
  
## RESCALE OUTCOME VARIABLES  
experiment$baseline_self<-experiment$baseline_self/5000  
experiment$decision_self<-experiment$decision_self/10000  
experiment$embezzlement<-experiment$embezzlement/10000  
experiment$expected_total<-experiment$expectation+10000-experiment$  
embezzlement  
experiment$decision_expected_capture<-experiment$expectation/experiment$  
expected_total  
  
## CONVERT TREATMENT INDICATORS TO FACTORS  
experiment$treatment_election<-factor(experiment$treatment_election)  
levels(experiment$treatment_election)<-c("RANDOM","ELECTED")  
experiment$treatment_election<-ordered(experiment$treatment_election,  
levels=c("ELECTED","RANDOM"))  
table(experiment$election,experiment$treatment_election)  
experiment$treatment_public<-factor(experiment$treatment_public)  
levels(experiment$treatment_public)<-c("NO TRANSPARENCY","TRANSPARENCY")  
experiment$treatment<- "NA "  
experiment$treatment[experiment$election==1&experiment$public==1]<- "  
ELECTIONS/PUBLIC "
```

```

experiment$treatment[experiment$election==1&experiment$public==0] <-"
  ELECTIONS/PRIVATE"
experiment$treatment[experiment$election==0&experiment$public==1] <-"RANDOM
  /PUBLIC"
experiment$treatment[experiment$election==0&experiment$public==0] <-"RANDOM
  /PRIVATE"

#### AVERAGE TREATMENT EFFECTS

### EMBEZZLEMENT

## Visualize embezzlement outcomes

barpl<-ggplot(data=subset(experiment, leader==1))+theme_few()+geom_boxplot(
  aes(treatment, decision_self))
barpl+disclaimer

plot.embezzlement<-ggplot(data=subset(experiment, leader==1))+theme_few()+
  facet_wrap(~treatment_public)+
  geom_density(aes(decision_self, color=treatment_election, fill=treatment_
    election), alpha=0.5)+
  xlab("Fraction of Group Money Embezzled")+ylab("Density")
plot.embezzlement+disclaimer

## Estimate treatment effects without covariates
reg.embezzlement<-lm(decision_self~election*public, data=subset(experiment,
  leader==1))
summary(reg.embezzlement)

#p-value for b1>=0 (one-sided)
t.e<-(reg.embezzlement$coef[2]/sqrt(vcov(reg.embezzlement)[2,2]))
p.e<-pt(t.e, df=summary(reg.embezzlement)$df[2])
#p-value for b1+b3>=0 (one-sided)
t.sum.e<-(reg.embezzlement$coef[2]+reg.embezzlement$coef[4])/sqrt(vcov(reg
  .embezzlement)[2,2]+vcov(reg.embezzlement)[4,4]-vcov(reg.embezzlement)
  [2,4])
p.sum.e<-pt(t.sum.e, df=summary(reg.embezzlement)$df[2])
#p-value for b2>=0 (one-sided)
t.t<-(reg.embezzlement$coef[3]/sqrt(vcov(reg.embezzlement)[3,3]))
p.t<-pt(t.t, df=summary(reg.embezzlement)$df[2])
#p-value for b2+b3>=0 (one-sided)
t.sum.t<-(reg.embezzlement$coef[3]+reg.embezzlement$coef[4])/sqrt(vcov(reg
  .embezzlement)[3,3]+vcov(reg.embezzlement)[4,4]-vcov(reg.embezzlement)
  [3,4])
p.sum.t<-pt(t.sum.t, df=summary(reg.embezzlement)$df[2])
#p-value for b3<=0 (one-sided)
t.int<-(reg.embezzlement$coef[4]/sqrt(vcov(reg.embezzlement)[4,4]))
p.int<-1-pt(t.int, df=summary(reg.embezzlement)$df[2])

#Hypothesis 1a
p.e

```

```

p.sum.e

#Hypothesis 1b
p.t
p.sum.t

#Hypothesis 2
p.int

## Also estimate treatment effects with covariate adjustment [to be added
  later]

### EXPECTATIONS

## Estimate treatment effects without covariates
require(rms)
reg.expectations<-robcov(ols(decision_expected_capture~election*public,x=
  TRUE,y=TRUE,data=subset(experiment,leader!=1)),cluster=subset(
  experiment,leader!=1)$group)
reg.expectations
reg.expectations.vcov<-vcov(reg.expectations)

#p-value for b1=0 (two-sided)
t.e<-reg.expectations$coef[2]/sqrt(reg.expectations.vcov[2,2])
p.e<-2*(1-pt(abs(t.e),df=reg.expectations$df))
#p-value for b1+b3=0 (two-sided)
t.sum.e<-(reg.expectations$coef[2]+reg.expectations$coef[4])/sqrt(reg.
  expectations.vcov[2,2]+reg.expectations.vcov[4,4]-reg.expectations.vcov
  [2,4])
p.sum.e<-2*(1-pt(abs(t.sum.e),df=reg.expectations$df))
#p-value for b2=0 (two-sided)
t.t<-reg.expectations$coef[3]/sqrt(reg.expectations.vcov[3,3])
p.t<-2*(1-pt(abs(t.t),df=reg.expectations$df))
#p-value for b2+b3=0 (two-sided)
t.sum.t<-(reg.expectations$coef[3]+reg.expectations$coef[4])/sqrt(reg.
  expectations.vcov[3,3]+reg.expectations.vcov[4,4]-reg.expectations.vcov
  [3,4])
p.sum.t<-2*(1-pt(abs(t.sum.t),df=reg.expectations$df))

p.e
p.sum.e
p.t
p.sum.t

## Also estimate treatment effects with covariate adjustment [to be added
  later]

### PERCEPTIONS OF PROCEDURAL FAIRNESS

## Estimate treatment effects without covariates

```

```

reg.fairness<-robcov(lrm(endline_fairprocedure ~ election, x=TRUE, y=TRUE,
  data=experiment), cluster=experiment$group)
reg.fairness
logitmfx(endline_fairprocedure ~ election,data=experiment,robust=TRUE,
  clustervar1="group")

## Also estimate treatment effects with covariate adjustment [to be added
  later]

#### SUPPLEMENTARY QUESTIONS

### QUESTION 1.1: Do elections with minimal prior communication enable
  citizens of a municipality to identify public-spirited leaders?

## Report p-value from Mann-Whitney U test
wilcox.test(decision_expected_capture~leader,data=subset(experiment,
  election==1))

## Figure 1.1.1

scale.election<-scale_color_manual("Election Outcomes\n",labels = c("
  Losing Candidates", "Winning Candidates"), values = c("blue", "red"))

plot.baseline<-ggplot(data=subset(experiment,election==1))+theme_few()+
  geom_density(aes(baseline_self,color=factor(leader),fill=factor(leader))
  ,alpha=0.5)+
  xlab("Baseline Embezzlement Decision")+ylab("Density")+
  scale_color_manual("Election Outcome\n",labels = c("Losing Candidates",
  "Winning Candidates"), values = c("gray80", "darkseagreen"))+
  scale_fill_manual("Election Outcome\n",labels = c("Losing Candidates", "
  Winning Candidates"), values = c("gray80", "darkseagreen"))
plot.baseline+disclaimer
ggsave("Figure1-1-1.pdf",height=3.5,width=7)

## Calculate individuals' rank within group, in terms of baseline
  embezzlement decision
#1= least, 5=most embezzlement
experiment$baseline_rank<-NULL
for (g in experiment$group) {
  experiment$baseline_rank[experiment$group==g]<-rank(experiment$baseline_
  self[experiment$group==g])
}

## Regression of votes received in first round on baseline rank/baseline
  embezzlement decision
reg.votes.received.1<-lm(votes_received~baseline_rank,data=subset(
  experiment,treatment_election==1))
summary(reg.votes.received.1)
reg.votes.received.2<-lm(votes_received~baseline_self,data=subset(
  experiment,treatment_election==1))

```

```

summary(reg.votes.received.2)

## Exploratory analysis of voters' heuristics [to be added later]

### QUESTION 1.2: To what extent can embezzlement behavior in the
  experiment be explained by baseline preferences?

## Figure 1.2.1

#Define labels for plot
labeli<-function(variable,value) {
  if (variable=="treatment_election") {
    lnames<-list("0"="Random","1"="Elected")
    return(lnames[value])
  }
  if (variable=="treatment_public") {
    lnames<-list("0"="No Transparency","1"="Transparency")
    return(lnames[value])
  }
  else{return(value)}
}

plot.decision.vs.baseline<-ggplot(aes(baseline_self,decision_self),data=
  experiment)+
  geom_line(aes(c(0,1),c(0,1)),lty=3,alpha=0.5) + stat_smooth(color="
    darkslategray4",fill="darkslategray4",alpha=0.4) +
  geom_point(color="cadetblue4",alpha=0.05)+
  theme_few() + xlim(c(0,1)) + ylim(c(0,1)) + xlab("Baseline Decision")+
  ylab("Fraction Embezzled")+
  facet_grid(treatment_public~treatment_election)
plot.decision.vs.baseline+disclaimer
ggsave("Figure1-2-1.pdf",height=6,width=6)

## Also report root MSPE for each cell [to be added later]

### QUESTION 1.3: How accurate are citizens' expectations regarding the
  extent of embezzlement?

## Figure 1.3.1

plot.expectation.vs.decision<-ggplot(aes(decision_self,decision_expected_
  capture),data=experiment)+
  geom_line(aes(c(0,1),c(0,1)),lty=3,alpha=0.5) + stat_smooth(color="
    darkslategray4",fill="darkslategray4",alpha=0.4) +
  geom_point(color="cadetblue4",alpha=0.05)+
  theme_few() + xlim(c(0,1)) + ylim(c(0,1)) + xlab("Actual Fraction
    Embezzled")+ylab("Group Members' Expectations")+
  facet_grid(treatment_public~treatment_election)
plot.expectation.vs.decision+disclaimer
ggsave("Figure1-3-1.pdf",height=6,width=6)

```

```
### QUESTION 2.1 Are citizens willing to incur costs to punish
embezzlement?
```

```
## Figure 2.1.1
```

```
cols <- c("Reward"="springgreen3","Punishment"="firebrick1")
plot.sanctioning.vs.decision<-ggplot(data=subset(experiment,leader==0))+
  geom_line(aes(c(0,1),c(0,1)),lty=3,alpha=0.5) +
  geom_point(aes(embezzlement,sanctions_reward,color="Reward"),alpha=0.05)
  +
  geom_point(aes(embezzlement,sanctions_punishment,color="Punishment"),
    alpha=0.05)+
  geom_smooth(aes(embezzlement,sanctions_reward,color="Reward",fill="
    Reward"),alpha=0.4,label="Reward")+
  geom_smooth(aes(embezzlement,sanctions_punishment,color="Punishment",
    fill="Punishment"),alpha=0.4,label="Punishment")+
  theme_few() + xlim(c(0,1)) + ylim(c(0,1000)) + xlab("Fraction Embezzled"
    )+ylab("Average Amount Spent on Sanctioning \n(Rewards or Punishments
    , in CFA Francs)")+
  facet_wrap(treatment_public~treatment_election)+
  scale_fill_manual("Type of Sanction: ",values=cols)+scale_color_manual("
    Type of Sanction: ",values=cols)+
  theme(legend.position="bottom")
plot.sanctioning.vs.decision+disclaimer
ggsave("Figure2-1-1.pdf",height=6,width=6)
```

```
## Distribution of sanctioning choices by experimental condition
hist.sanctions<-ggplot(data=subset(experiment,leader==0))+theme_few()+xlab(
  "Sanctioning Decision")+
  geom_histogram(aes(sanction),color="black",fill="white")+facet_grid(
    treatment_election~treatment_public)
hist.sanctions+disclaimer
```

```
## Figure 2.1.2
```

```
# Sanctions by deviation from reference point
```

```
experiment$deviation<-experiment$embezzlement-experiment$baseline_self
```

```
cols <- c("Reward"="springgreen3","Punishment"="firebrick1")
plot.sanctioning.vs.decision<-ggplot(data=subset(experiment,leader==0))+
  geom_line(aes(c(0,1),c(0,1)),lty=3,alpha=0.5) +
  geom_point(aes(deviation,sanctions_reward,color="Reward"),alpha=0.05)+
  geom_point(aes(deviation,sanctions_punishment,color="Punishment"),alpha
    =0.05)+
  geom_smooth(aes(deviation,sanctions_reward,color="Reward",fill="Reward")
    ,alpha=0.4,label="Reward")+
  geom_smooth(aes(deviation,sanctions_punishment,color="Punishment",fill="
    Punishment"),alpha=0.4,label="Punishment")+
  theme_few() + xlim(c(-1,1)) + ylim(c(0,1000)) + xlab("Deviation from
    Reference Point (Subjects' Baseline Decision)")+ylab("Average Amount
    Spent on Sanctioning \n(Rewards or Punishments, in CFA Francs)")+
  geom_vline(aes(x=0),lty=3,alpha=0.5)+
```

```

facet_wrap(treatment_public~treatment_election)+
scale_fill_manual("Type of Sanction: ",values=cols)+scale_color_manual("
  Type of Sanction: ",values=cols)+
theme(legend.position="bottom")
plot.sanctioning.vs.decision+disclaimer
ggsave("Figure2-1-2.pdf",height=6,width=6)

```

QUESTION 2.2: How does a lack of transparency influence citizens' willingness to engage in costly sanctioning behavior?

Figure 2.2.2

```

#Individual beliefs: prior expectations in no-transparency condition;
  actual outcomes in transparency condition
experiment$updated.beliefs<-experiment$decision_expected_capture
experiment$updated.beliefs[experiment$public==1]<-experiment$embezzlement[
  experiment$public==1]

```

```

cols.transparency <- c("TRANSPARENCY"="gold","NO TRANSPARENCY"="
  deepskyblue4")

```

```

plot.punishments.vs.expectation<-ggplot(aes(updated.beliefs,sanctions_
  punishment),data=subset(experiment,leader==0))+
geom_line(aes(c(0,1),c(0,1)),color="black",lty=3,alpha=0.5) +
geom_smooth(aes(updated.beliefs,sanctions_punishment,color=treatment_
  public,fill=treatment_public),alpha=0.4)+
geom_point(aes(updated.beliefs,sanctions_punishment,color=treatment_
  public,fill=treatment_public),alpha=0.05)+
theme_few() + xlim(c(0,1)) + ylim(c(0,1000))+ ylab("Amount Spent on \n
  PUNISHMENTS")+xlab(element_blank())+
facet_wrap(~treatment_election)+
scale_fill_manual("",values=cols.transparency)+scale_color_manual("",
  values=cols.transparency)+
theme(axis.text.x=element_blank(),axis.ticks.x=element_blank(),axis.
  title.x=element_blank(),
  legend.position="none",plot.margin=unit(c(0,0,0,0),"cm"))
plot.punishments.vs.expectation+disclaimer

```

```

plot.rewards.vs.expectation<-ggplot(aes(updated.beliefs,sanctions_reward),
  data=subset(experiment,leader==0))+
geom_line(aes(c(0,1),c(0,1)),lty=3,alpha=0.5) + geom_smooth(aes(updated.
  beliefs,sanctions_reward,color=treatment_public,fill=treatment_public
  ),alpha=0.4)+
geom_point(aes(updated.beliefs,sanctions_reward,color=treatment_public,
  fill=treatment_public),alpha=0.05)+
theme_few() + xlim(c(0,1)) + ylim(c(0,1000))+xlab("Fraction Embezzled (
  expected or known")+ylab("Amount Spent on \n REWARDS")+
facet_wrap(~treatment_election)+
scale_fill_manual("",values=cols.transparency)+scale_color_manual("",
  values=cols.transparency)+

```

```

  theme(legend.position="bottom",strip.text=element_blank())
plot.rewards.vs.expectation

pdf("Figure2-2-2.pdf",width=7,height=7.5)
gl = lapply(list(plot.punishments.vs.expectation+disclaimer+coord_fixed(
  ratio=1/1000),plot.rewards.vs.expectation+theme(legend.position="bottom
  ")+disclaimer+coord_fixed(ratio=1/1000)), ggplotGrob)
library(gtable)
g = do.call(rbind, c(gl, size="first"))
g$widths = do.call(unit.pmax, lapply(gl, "[[", "widths"))
grid.draw(g)
dev.off()

```

QUESTION 2.3: Does citizens' sanctioning behavior depend on whether decision-makers are elected or not?

Figure 2.3.1

```

cols.election <- c("ELECTED"="mediumaquamarine","RANDOM"="wheat3")

# average punishment received vs actual embezzlement, by treatment
condition
plot.punishment.vs.embezzlement<-ggplot(aes(decision_self,sanctions_
received/10000,color=treatment_election,fill=treatment_election),data=
subset(experiment,leader==1))+
stat_smooth(alpha=0.4) +
geom_point(alpha=0.1)+
geom_hline(aes(0),lty=3,alpha=0.5)+
theme_few() + xlim(c(0,1)) + xlab("Fraction of Group Money Embezzled")+
ylab("Sanctions received by the decision-maker \n (Net reward/
punishment as a fraction of group money)")+
scale_fill_manual("",values=cols.election)+scale_color_manual("",values=
cols.election)+
facet_wrap(~ treatment_public)+theme(legend.position="bottom")
plot.punishment.vs.embezzlement+disclaimer
ggsave("Figure2-3-1.pdf",height=5,width=7)

```

ANCILLARY EXPERIMENTS

GROUP-LEVEL EFFECTS OF THE DECISION-MAKER'S SOCIAL IDENTITY

FEMALE VS. MALE DECISION-MAKERS

Figure 3.1.1

```

experiment$gender[experiment$female==1]<-"female"
experiment$gender[experiment$female==0]<-"male"
experiment$leader.female<-NULL
for (g in unique(experiment$group)) {
  if (experiment$gender[experiment$group==g&experiment$leader==1]=="female
  ") {

```

```

    experiment$leader.female[experiment$group==g]<-1
  }
  else if (experiment$gender[experiment$group==g&experiment$leader==1]=="
    male") {
    experiment$leader.female[experiment$group==g]<-0
  }
}
experiment$leader.female<-as.factor(experiment$leader.female)
levels(experiment$leader.female)<-c("male","female")

cols.female <- c("male"="skyblue1","female"="lightcoral")
#average sanction vs embezzlement
levels(experiment$leader.female)<-c("MALE DECISION-MAKER","FEMALE DECISION
-MAKER")
plot.gender<-ggplot(aes(embezzlement,sanction,color=gender,fill=gender),
  data=subset(experiment,leader==0&election==0))+
  stat_smooth(alpha=0.4) +
  geom_point(alpha=0.2)+
  geom_hline(aes(0),lty=3,alpha=0.5)+
  theme_few() + xlim(c(0,1)) + ylim(c(-3000,1000)) + xlab("Fraction of
  Group Money Embezzled")+ylab("Average sanction towards the decision-
  maker")+
  scale_fill_manual("Gender of Study Participant",values=cols.female,
  labels=c("MALE","FEMALE"))+scale_color_manual("Gender of Study
  Participant",values=cols.female,labels=c("MALE","FEMALE"))+
  facet_grid(treatment_public ~ leader.female)+theme(legend.position="
  bottom")
plot.gender+disclaimer
ggsave("Figure3-1-1.pdf",height=7,width=7)

## CO-ETHNIC VS. NON-CO-ETHNIC DECISION-MAKERS

## Figure 3.2.1

experiment$leader.ethnicity<-NULL
for (g in unique(experiment$group)) {
  experiment$leader.ethnicity[experiment$group==g]<-rep(experiment$
  ethnicity[experiment$group==g&experiment$leader==1])
}
experiment$leader.ethnicity<-as.factor(experiment$leader.ethnicity)
levels(experiment$leader.ethnicity)<-levels(experiment$ethnicity)
table(experiment$leader.ethnicity,experiment$ethnicity)
experiment$leader.coethnic<-as.factor(experiment$leader.ethnicity==
  experiment$ethnicity)
levels(experiment$leader.coethnic)<-c("DIFFERENT ETHNICITY","SAME
  ETHNICITY")
table(experiment$leader.coethnic)

#average sanction by actual embezzlement decision
cols.ethnicity<-c("DIFFERENT ETHNICITY"="lightcyan3","SAME ETHNICITY"="
  lightgoldenrod3")

```

```

plot.ethnicity.1<-ggplot(aes(embezzlement, sanction, color=leader.coethnic,
  fill=leader.coethnic), data=subset(experiment, leader==0&election==0))+
  stat_smooth(alpha=0.4) +
  geom_point(alpha=0.2)+
  geom_hline(aes(0), lty=3, alpha=0.5)+
  theme_few() + xlim(c(0,1)) + ylim(c(-3000,1000)) + xlab("Fraction of
  Group Money Embezzled")+ylab("Average sanction towards the decision-
  maker")+
  scale_fill_manual("Ethnicity of Decision-Maker", values=cols.ethnicity)+
  scale_color_manual("Ethnicity of Decision-Maker", values=cols.
  ethnicity)+
  facet_wrap( ~ treatment_public)+theme(legend.position="bottom")
plot.ethnicity.1+disclaimer
ggsave("Figure3-2-1.pdf", height=5, width=7)

```

SENIORITY OF THE DECISION-MAKER

```

experiment$seniority<-NULL
for (g in unique(experiment$group)) {
  experiment$seniority[experiment$group==g]<-rank(experiment$age[
    experiment$group==g], ties.method="min")
  experiment$leader.seniority[experiment$group==g]<-rep(experiment$
    seniority[experiment$group==g&experiment$leader==1])
}

experiment$dec<-experiment$decision_expected_capture
experiment$dec[experiment$leader==1]<-experiment$decision_self[experiment$
  leader==1]

```

Figure 3.3.1

```

plot.seniority.1<-ggplot(aes(leader.seniority, dec, lty=factor(leader)), data
  =subset(experiment, election==0))+
  stat_summary(fun.data="mean_cl_boot", geom = "pointrange")+
  stat_summary(fun.data="mean_cl_boot", geom = "line")+
  scale_linetype_manual("OUTCOME", values=c(1,3), labels=c("EXPECTED", "
  ACTUAL"))+
  theme_few() + ylim(c(0,1)) + xlab("Age Rank of Decision-Maker")+ylab("
  Fraction Embezzled")+ theme(legend.position="bottom")+
  facet_wrap( ~ treatment_public)
plot.seniority.1+disclaimer
ggsave("Figure3-3-1.pdf", height=4.5, width=7)

```

Figure 3.3.2

```

plot.seniority.2<-ggplot(aes(embezzlement, sanction, color=factor(leader.
  seniority), fill=factor(leader.seniority)), data=subset(experiment, leader
  ==0&election==0))+
  stat_smooth(alpha=0.4) +
  geom_point(alpha=0.2)+
  geom_hline(aes(0), lty=3, alpha=0.5)+

```

```

theme_few() + xlim(c(0,1)) + ylim(c(-3000,1000)) + xlab("Fraction of
  Group Money Embezzled")+ylab("Average sanction towards the decision-
  maker")+
scale_fill_manual("Age Rank of Decision-Maker",values=c(1:5))+scale_
  color_manual("Age Rank of Decision-Maker",values=c(1:5))+
  facet_wrap( ~ treatment_public)+theme(legend.position="bottom")
plot.seniority.2+disclaimer
ggsave("Figure3-3-2.pdf",height=5,width=7)

### INDIVIDUAL-LEVEL EFFECTS OF BEING SELECTED AS A DECISION-MAKER

#[to be added later]

### BALLOT-ORDER AND BADGE-COLOR EFFECTS

## Kruskal-Wallis Tests

#by ballot order
kruskal.test(votes_received~ballot_order,data=experiment)

#by badge color
kruskal.test(votes_received~badge_color,data=experiment)

## Figure 4.1.1

plot.ballotorder<-ggplot(aes(ballot_order,votes_received),data=subset(
  experiment,election==1))+
  stat_summary(fun.data="mean_cl_normal", geom = "pointrange")+coord_flip
  ()+xlim(c(5,1))+
  ylab("Votes Received (Mean and 95 % CI)")+xlab("Ballot Order")+theme_few
  ()
plot.ballotorder
ggsave("Figure4-1-1a.pdf",height=5,width=4.5)

plot.badgecolor<-ggplot(aes(factor(badge_color),votes_received),data=
  subset(experiment,election==1))+
  stat_summary(fun.data="mean_cl_normal", geom = "pointrange")+coord_flip
  ()+
  ylab("Votes Received (Mean and 95 % CI)")+xlab("Badge Color")+theme_few
  ()
plot.badgecolor
ggsave("Figure4-1-1b.pdf",height=5,width=4.5)

```

A3. R code for power calculations

```
#####  
# POWER CALCULATIONS for "Elections and Embezzlement" #  
# *Author omitted* #  
# See submission at EGAP social science registry for further details #  
#####  
  
rm(list=ls())  
require(ggplot2)  
require(ggthemes)  
  
##ASSUMPTIONS (GROUP-LEVEL OUTCOMES)  
n<-118*4 #Assume 472 groups  
#No covariate adjustment in these power calculations  
#One-sided tests for Hypotheses 1a and 1b (group-level outcomes)  
alpha<-0.05/3 #Bonferroni correction for three comparisons  
  
#Function to simulate hypothesis tests  
simulate.test<-function(beta,n,alpha) {  
  elections<-c(rep(0,n/2),rep(1,n/2))  
  transparency<-c(rep(0,n/4),rep(1,n/4),rep(0,n/4),rep(1,n/4))  
  cons<-rep(1,n)  
  X<-cbind(cons,elections,transparency,elections*transparency)  
  e<-rnorm(n,mean=0,sd=1)  
  y<-X%%beta+e  
  sim<-data.frame(y,elections,transparency)  
  reg.embezzlement<-lm(y~elections*transparency,data=sim)  
  #p-value for  $b_1 \geq 0$  (one-sided)  
  t.e<-(reg.embezzlement$coef[2]/sqrt(vcov(reg.embezzlement)[2,2]))  
  p.e<-pt(t.e,df=summary(reg.embezzlement)$df[2])  
  #p-value for  $b_1 + b_3 \geq 0$  (one-sided)  
  t.sum.e<-(reg.embezzlement$coef[2]+reg.embezzlement$coef[4])/sqrt(vcov(  
    reg.embezzlement)[2,2]+vcov(reg.embezzlement)[4,4]-vcov(reg.  
    embezzlement)[2,4])  
  p.sum.e<-pt(t.sum.e,df=summary(reg.embezzlement)$df[2])  
  output<-cbind(p.e,p.sum.e)  
  #testing  $b_1 \geq 0$  or  $b_1 + b_3 \geq 0$  or both  
  at.least.one<-as.numeric(length(output[output<=alpha])>0)  
  #testing  $b_1 \geq 0$  and  $b_1 + b_3 \geq 0$   
  both<-as.numeric(length(output[output<=alpha])==length(output))  
  return(c(at.least.one,both))  
}  
  
#Function to replicate hypothesis tests with given parameters and  
  calculate power  
#Returns power for two scenarios:  
#(a)  $H_0$  rejected in at least one condition  
#(b)  $H_0$  rejected in both conditions  
sim.reps<-function(beta,n,reps,alpha) {  
  parameters<-matrix(data=beta,nrow=length(beta),ncol=reps)
```

```

sims<-apply(parameters,2,simulate.test,n,alpha)
sims<-t(sims)
power<-rbind(apply(sims,2,sum)/reps)
colnames(power)<-c("At least one","Both")
return(power)
}

#Run simulation for a range of constant effect sizes
set.seed(20150706)
#range of main effect
range.main<-seq(0,-1,-0.05)
#range of interaction effect
range.int<-seq(0.5,-0.5,-0.25)
#loop over parameter values
sim.data<-NULL
for (e in range.int) {
  eff.size<-cbind(range.main)
  len<-length(eff.size)
  eff.size.int<-rep(e,len)
  b.in<-cbind(rep(0,len),eff.size,eff.size,eff.size.int)
  b.in<-t(b.in)
  sim.all<-apply(b.in,2,sim.reps,n,100,alpha)
  sim.all<-t(sim.all)
  sim.data<-rbind(sim.data,data.frame(eff.size,eff.size.int,sim.all))
}
names(sim.data)<-c("eff.size","eff.size.int","at.least.one","both")

#Stack data
sim.data.a<-sim.data
sim.data.a$test<-rep("Reject H0 in at least one condition",nrow(sim.data.a
))
sim.data.a$power<-sim.data.a$at.least.one
sim.data.b<-sim.data
sim.data.b$test<-rep("Reject H0 in both conditions",nrow(sim.data.b))
sim.data.b$power<-sim.data.b$both
sim.data.2<-rbind(sim.data.a,sim.data.b)

#Plot power under different effect size assumptions
power.plot<-ggplot(aes(eff.size,power,color=factor(eff.size.int)),data=sim
.data.2)+
  geom_line()+
  theme_few()+xlab("Size of Main Effect (Standardized)")+ylab("Power")+
  scale_color_brewer(name="Size of Interaction Effect",palette="RdYlBu")+
  theme(legend.position="bottom")+
  facet_wrap(~test)
power.plot

#Export plot
setwd("C:/Users/Admin/Dropbox/Research/Burkina Baseline Experiments/Pre-
Analysis Plan")
ggsave("PowerPlot.pdf",width=7,height=5)

```


A4. Protocol for on-site replacements of no-shows

CONFIDENTIAL: DO NOT SHARE WITH ANYONE OUTSIDE THE SURVEY TEAM

PROTOCOL FOR THE REPLACEMENT OF NO-SHOWS AT THE DATA COLLECTION SESSION

