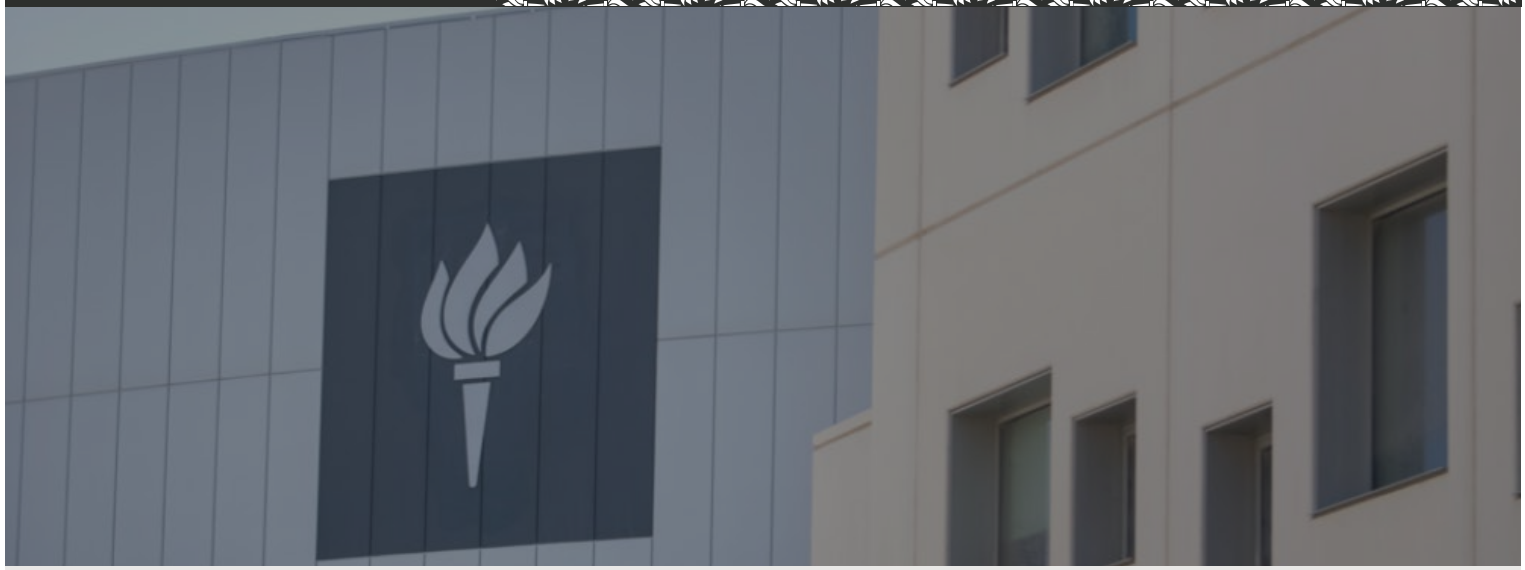
جامعة نيويورك أبوظبي

NYU | ABU DHABI



The ghost of institutions past: History as an obstacle to fighting tax evasion

Aaron Kamm, Christian Koch and Nikos Nikiforakis

Working Paper # 0008

October 2017

Division of Social Science Working Paper Series

New York University Abu Dhabi, Saadiyat Island P.O Box 129188, Abu Dhabi, UAE

https://nyuad.nyu.edu/en/academics/divisions/social-science.html

The ghost of institutions past: History as an obstacle to fighting tax evasion*

Aaron Kamm†, Christian Koch† and Nikos Nikiforakis†

This version: October 22, 2017

Abstract

If taxpayers believe past rates of compliance are indicative of the future, traditional measures for combating tax evasion can be compromised. We present evidence from a novel laboratory experiment with strategic complementarities showing that a history of low compliance can render a major institutional reform ineffective at reducing tax evasion. The experimental treatments manipulate the history of tax compliance by varying the percentage of tax revenue embezzled by a 'politician' – our measure of 'institutional quality'. We show that tax compliance is substantially higher in good-quality than bad-quality institutions when there is no history of tax evasion. When a bad-quality institution is replaced with a good-quality one, however, tax compliance remains low, as if the institutional change had not occurred. The reason is that the institutional change leaves expectations about future compliance largely unaffected. A history of high-quality institutions, on the other hand, shields tax compliance only partly from institutional deterioration. We discuss reasons for this, policy implications of our findings and evidence that a society-wide poll can assist in overcoming the 'ghost of institutions past'.

Keywords: tax evasion, interdependence, multiple equilibria, path dependence, experiment
JEL classification: C92, H26

*We would like to thank Dirk Engelmann as well as (seminar) participants in NYU Abu Dhabi, LMU Munich, HU Berlin, at the 2016 World Meeting of the Economic Science Association (Jerusalem), the 2016 workshop on "Enacting change in long-run behavior and attitudes" (Berlin), the 2017 International Symposium for Experimental Economics (Abu Dhabi), the 2017 European Public Choice Society Annual Meeting (Budapest), the 2017 International Meeting on Experimental and Behavioral Social Sciences (Barcelona), the 2017 World Meeting of the Economic Science Association (San Diego), the 2017 Verein für Socialpolitik (Vienna), the 2017 European Meeting of the Economic Science Association (Vienna).

†Division of Social Science, New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, United Arab Emirates. aaron.kamm@nyu.edu, christian.koch@nyu.edu, nikos.nikiforakis@nyu.edu.

You can run out your rules, but you know you can't outrun the history train.
– Paul Simon

1 Introduction

Tax compliance is essential for a well-functioning state. Understanding the determinants of compliance and finding remedies for tax evasion are therefore of obvious importance. The conventional approach is to model tax compliance as an individual decision-making problem in which each taxpayer weighs the monetary gains from evading against the probability of being apprehended and the resulting fines (Allingham and Sandmo 1972). This approach captures the intuition that the incentives provided by tax authorities are likely to matter, but the model has difficulty accounting for several empirical regularities such as the relatively high compliance rates despite low audit probabilities (Andreoni et al. 1998), the large within-country differences in spite of similar tax incentives (e.g., Vogel 1974, D'Attoma 2016), and the ineffectiveness of traditional policies at combating rampant evasion (e.g., Burgess and Stern 1993, Martinez-Vazquez and McNab 2000). A possible reason is that the standard model fails to account for the fact that "individuals' compliance decisions are interdependent" (IMF 2015, p. 21) which transforms tax compliance into a strategic decision.

Empirical evidence suggests that the cost of evading taxes for an individual are often positively correlated with the overall level of compliance, i.e., the higher the overall level of compliance in a society the more costly it is for an individual to evade in expected terms (Andreoni et al. 1998). Different factors can account for this relationship. At the country level, lower rates of compliance can reduce the resources available for identifying tax evaders in the future (Dubin and Wilde 1988, Burgess and Stern 1993);[1] even when the budget of the tax authority is unaffected, an increase in the number of evaders can create "enforcement congestion", reducing the probability of being apprehended and fined (Bassetto and Phelan 2008, Galbiatia and Zanella 2012). At the individual level, concerns for conformity, fairness and social image all increase the cost of deviating from what others do (Bordignon 1993, Myles and Naylor 1996, Kim 2003, Fortin et al. 2007, Traxler 2010, Dulleck et al. 2016). Irrespective of the reason, this strategic complementarity has important policy implications as it introduces the possibility of multiple equilibria, behavioral contagion and path dependence (Schelling 1978).

This paper presents evidence from a laboratory experiment exploring the extent to which past outcomes can shape tax compliance in the future, after a major institutional change occurs. We introduce a novel setting in which tax declarations are strategic

[1] Dubin and Wilde (1988) provide evidence that the relationship between past compliance rates and the future likelihood an individual is audited is causal. In line with this, Daley (2013) notes that "Greece [a country with relatively high tax evasion] has only 0.87 auditors for every 1,000 citizens, compared with Germany's rate of 1.36".

complements. While strategic complementarity implies past outcomes can influence future choices in a static environment, it is not obvious they should have an impact in the aftermath of a major institutional change, at least when the change is common knowledge. At the same time, one cannot rule out this possibility. Whether past outcomes can affect tax compliance in such instances is ultimately an empirical question, the answer to which will depend partly on the extent to which institutional change alters individual expectations about others' willingness to comply.

Understanding the extent to which past outcomes can shape tax compliance in the long run is clearly an important question. If taxpayers believe past rates of compliance are somehow indicative of future compliance rates, institutional changes (e.g., reducing corruption) or other policies aimed at combating evasion (e.g., increasing fines) can be less effective at promoting tax compliance in societies with a history of high evasion. In other words, just as past successes can have long-run benefits, past failures can be difficult to correct. The effect can be dramatic. In the recent debate on the chronic problem of tax evasion in Greece, for instance, it has been claimed that tax evasion emerged as a patriotic act and a form of resistance during the 400-year Ottoman rule which persisted after Greece gained independence in 1830 (Graham 2015, Karnitschig and Stamouli 2015).[2] Of course, anecdotal evidence does not suffice to answer our research question. Field data is also problematic given that institutional changes are endogenous. For this reason, we decided to conduct a laboratory experiment which allows us not only to vary institutions exogenously, but also to generate different histories prior to this change in a way that would be extremely difficult, if not impossible, otherwise.

Our experimental setting is as follows. In each period, players simultaneously decide how much of their income to declare. The income is taxed at a flat rate and the revenue collected is used to provide a public good. The benefits from the public good are equally divided among all group members, but part of the tax revenue is 'embezzled' by a politician. In some treatments, embezzlement is high (45% of the tax revenue) while in others it is low (5%). This variation is used to exogenously affect compliance rates.[3] The level of embezzlement is our measure of *institutional quality*. The treatments vary the quality of past and present institutions in a 2×2 design. We consider institutional improvement (from bad to good quality), institutional deterioration (from good to bad quality), and two control treatments without institutional change (always good quality, always bad quality). This design allows us to investigate whether a history of bad-quality institutions can undermine the efficacy of institutional reforms at promoting tax compliance, and,

[2]The persistence of tax evasion despite the major change in institutions may have been facilitated by the enormous debt accumulated during Greece's War of Independence which necessitated the imposition of heavy taxes making evasion an attractive choice. Path dependence can also help explain within-country differences in compliance, if different regions have had different histories (e.g., North- and South-Italy), and the relatively high compliance rates observed despite low audit probabilities.

[3]Field evidence indicates that tax compliance is negatively associated with embezzlement and corruption (Burgess and Stern 1993, Andreoni et al. 1998, Torgler 2005, Alm et al. 2016).

conversely, whether a history of good-quality institutions can prevent a rise in tax evasion when the quality of institutions deteriorates.

Strategic complementarity is introduced by making the audit probability endogenous. In every period, an individual is audited with a given probability. If s/he is caught evading, s/he must pay a fine proportional to the amount of the undeclared income. The probability of being audited increases with the overall level of tax compliance. This mechanism can be thought of either as capturing the idea that lower rates of compliance reduce the resources available for combating tax evasion or reflecting the lower probability of an individual being detected and reported when the number of honest taxpayers falls. Regardless of the interpretation, what is important is (*i*) that the individual cost for evading taxes is increasing in the overall level of compliance, and (*ii*) that it ensures the existence of multiple equilibria without having to rely on subjects' 'tax morale' which can be difficult to induce or control in a laboratory environment. In our setting, full tax compliance and full tax evasion are always equilibria of the game, irrespective of the level of embezzlement. This feature distinguishes our study from previous tax experiments which have either considered solitary decision-maker settings in the spirit of Allingham and Sandmo (1972), or games with a unique equilibrium (see next section).

Our findings indicate that past outcomes can cast a long shadow and influence future tax compliance, even after a major institutional change. While good institutional quality was successful in inducing high compliance rates in the absence of a past (about 50% higher than in bad-quality institutions), the connection between concurrent institutional quality and compliance breaks down when there is a past: when a bad-quality institution is replaced with a good-quality one, tax compliance remains low, as if the change in institutions had not occurred. This is remarkable given that the change involves a nearly 90% reduction in embezzlement levels (from 45% to 5%). It suggests that past outcomes can sometimes impede the effectiveness of measures for combating tax evasion greatly. This, to paraphrase Charles Dickens, is the 'ghost of institutions past'. The opposite, however, does not hold in our experiment; past successes appear to have limited ability to shield individuals from future failures. When a good-quality institution is replaced by a bad-quality one, tax compliance unravels.

Our analysis illustrates that a combination of factors determines whether the past will influence contemporaneous tax compliance. Importantly, we find no evidence to suggest that the institutional change affects individuals' expectations about future compliance, either in the case of institutional improvement or institutional deterioration. The impact of the change depends largely on the post-reform incentives and the pre-reform levels of compliance (as they shape expectations). With regards to the latter, an important factor is the existence of individuals who are not solely driven by pecuniary motives. A mixture model classifies a large fraction of our subjects as either 'evasion-averse' (who suffer a non-pecuniary cost from tax evading) or as 'inequality-averse' (who experience a non-pecuniary

cost when the inequality in earnings rises). These types can help explain why groups never converge to the full-evasion equilibrium even in bad-quality institutions, while selfish individuals can help explain why groups almost never converge to the full-compliance equilibrium. With regards to the post-reform incentives, we show that best-response dynamics can explain why a history of high compliance cannot prevent the increase of tax evasion when institutional quality deteriorates. A simulation in which individuals are assumed to be one of the three aforementioned behavioral types ('selfish', 'evasion-averse' and 'inequality-averse') and believe compliance will be the same as in past periods tracks well the patterns observed in our experiment.

Finally, we present evidence that the 'ghost of institutions past' can be overcome through a simple 'society-wide poll'. In a last part of the experiment, participants are asked to state their agreement or disagreement with the following statement: "Higher tax payments increase group earnings. Thus, let's FULLY pay our taxes and do not evade them!" When the number of participants agreeing/disagreeing with the message is revealed, groups in good-quality institutions that were experiencing low compliance due to a bad history are able to match the level of compliance in groups that never experienced bad-quality institutions. This finding further illustrates that the past matters because it allows individuals to establish mutually consistent expectations about the future: the poll improves expectations about future compliance, breaking the link with past outcomes.

The paper proceeds as follows. In Section 2, we offer an overview of the related literature. We present our experimental design in Section 3, and our main results in Section 4. In Section 5, we present evidence from a mixture model and a simulation analysis to better understand behavior in the experiment. We conclude in Section 6 with a discussion of the broader implications of our findings.

2 Related literature

Our study contributes to a long experimental literature on tax compliance. Although laboratory environments obviously differ in numerous ways from those in which taxpayers operate, the endogeneity problems arising from the fact that policies are not introduced randomly, the paucity of reliable data on actual tax declarations, and the need to test the micro-foundations of tax models have meant that laboratory experiments have played an important role in advancing our understanding of the determinants of tax compliance.

Early experimental studies focused on testing the predictions of the Allingham and Sandmo (1972) model and its more straightforward extensions, e.g., the effect of changing fines (Friedland et al. 1978), audit probabilities (Alm et al. 1992a), institutional uncertainty (Alm et al. 1992b), and strategic auditing (Alm et al. 1993). More recently, experimenters have allowed for individual actions to impose externalities on other players, e.g., by introducing a public good (Alm et al. 1992), a public audit (Coricelli et al. 2014), voting

(Feld and Tyran 2002), and social information (Fortin et al. 2007, Lefebvre et al. 2015). Even in these studies, however, there is always a unique equilibrium under the standard assumptions.[4] To the best of our knowledge, our experiment is the first on tax compliance to explore behavior in a game in which full-evasion and full-compliance are both equilibria.

Multiplicity of equilibria raises the issue of equilibrium selection. There is a rich experimental literature on coordination games (Devetag and Ortmann 2007). A robust finding is that payoff dominance often fails to select the equilibrium played by individuals when pitted against other equilibrium refinements such as risk dominance or max-min. This may be important in our experiment as the payoff-dominant equilibrium (full compliance) involves more risk than the full-evasion equilibrium. More importantly for our purposes, the existence of multiple equilibria introduces the possibility of behavioral contagion and path dependence. Path dependence is the idea that the set of decisions one faces at any given point in time is affected by past events, even if past circumstances may no longer be relevant. Path dependence can arise through different channels and for different reasons such as large fixed costs, falling variable costs, coordination effects or adaptive expectations (e.g., North 1990). Here we focus only on those of relevance to our study, namely, coordination effects when elements of the environment change.

We are aware of only few studies investigating how past outcomes affect contemporaneous behavior when elements of the environment change. The findings are as diverse as the environments studied. Brandts et al. (2016) and Hamman et al. (2007) show that changes in financial incentives which do not affect the set of equilibria allow groups to overcome a history of coordination failure in minimum-effort games. Similarly, Hossain and Morgan (2009) find no evidence of path dependence in a lab experiment on platform competition when participants start off on a particular platform; when a new, more efficient platform is introduced, participants have no difficulty to switch to the payoff-dominant equilibrium. Andreoni et al. (2017) and Smerdon et al. (2016) investigate experimentally conditions under which inefficient social norms can persist, while Wilkening (2016) considers the persistence of contractual-enforcement institutions. These studies find evidence that groups can fail to adapt optimally in changing circumstances. The main reason in Smerdon et al. (2016) and Wilkening (2016) is the problem of incomplete information. In contrast, in Andreoni et al. (2017) the problem appears to arise from preferences changing gradually and separately for each individual. Our findings contribute in this body of evidence by showing that, even under complete information, a major, simultaneous and commonly-known institutional change may fail to affect individuals' expectations and behavior.

[4] Although models of other-regarding or social preferences can admit multiple equilibria in social dilemmas, the requirements for cooperative outcomes to be supported in equilibrium in the absence of enforcement mechanisms tend to be strong (e.g., Fehr and Schmidt 1999).

Interestingly, although path dependence has a storied history in economics, there are hardly any references to it in the tax-compliance literature.[5] Different factors may account for this. The first is the traditional model by Allingham and Sandmo (1972) in which only present incentives matter, and the past has no role. The second is the limits of field data on tax compliance; most notably the endogeneity issues which hinder any attempts to safely identify the effect of past outcomes on present ones. The third is that, in countries with chronic evasion problems, major institutional reforms are easier said than done; even if agreed upon, their implementation can often be inadequate (e.g., Donadio 2012, Arapis 2015, IMF 2017). This makes it difficult to distinguish with field data, to what extent low compliance is due to path dependence or ineffective measures for fighting evasion.

The only other study we are aware of to have explored the possible path dependence of tax compliance is Bruttel and Friehe (2014). Using an experiment in the tradition of the Allingham-Sandmo model, the authors show that past tax incentives, in some instances, continue to have an impact on future income declarations, after these incentives (i.e., the audit probability and the monetary fine for being caught evading) have changed. Although their and our study deal broadly with the same topic, there are important differences, both in the settings examined and the findings themselves. Unlike in our experiment, Bruttel and Friehe (2014) explore the choices of individual decision-makers whose actions have no payoff consequences for others. As tax revenue is used neither for the provision of public goods nor income redistribution, tax compliance is an individual choice and not a strategic decision as it is in our study. Also, with regards to behavior, unlike in our study where we observe that institutional improvement has no long-run effect whatsoever on compliance, Bruttel and Friehe (2014) observe that compliance always responds to changes in incentives in the expected direction. The authors attribute their findings to reference-dependent preferences. Their explanation is complementary to ours. We show that in a setting with strategic complementarities, a major change in institutions can fail to affect individual expectations about others' willingness to comply and subsequently tax compliance.

3 The experiment

3.1 Overview

Individuals earn an income by performing a real-effort task. Subsequently, they must decide how much of it to report. Taxes are paid on reported income only, but unreported

[5] Burgess and Stern (1993, p. 796) write: "the differences in levels and patterns of taxation across countries reflect a diversity of economic structures and administrative capabilities as well as widely varying political, cultural, and institutional histories and environments." Apart from this, the only other reference we could find is in a recent report by the IMF (2015, p.21) stating that "Individuals' compliance decisions are interdependent, implying the possibility of multiple equilibria: escaping a bad equilibrium of low compliance can then require a major push rather than marginal reform."

income may be discovered in which case the individual must pay a fine based on the unpaid taxes. The likelihood of an audit is a positive function of the overall level of compliance in the group. The tax revenue is used for the provision of a public good, but part of it is embezzled by a 'politician'. The treatments vary the degree and the history of this embezzlement. We discuss the experiment in detail below.

3.2 Earning an income

Asking individuals to declare their income in an environment in which all individuals earn the same seems artificial. For this reason, we decided to allow for two different levels of income among participants. To increase ecological validity, incomes are allocated based on one's relative performance in a real-effort task taken from Erkal et al. (2011). Participants are given a table assigning a number to each letter of the alphabet and have to use it to 'encrypt' a sequence of different words. They are given seven minutes to encrypt as many words as possible. All participants face the same words in the same order. After a subject has successfully encrypted a word s/he is presented with the next word. A subject cannot proceed to the next word until s/he encrypts the word correctly.

The translation of individual performance to incomes works as follows. At the start of the experiment, individuals are randomly divided in groups of $9+1$ participants. The "+1" refers to a randomly-selected participant who is allocated the role of the 'politician' and receives a high income without having to encrypt any words, I_{high}.[6] Among the remaining nine group members, the four best performing receive a high income, I_{high}. The other five receive I_{low}, with $I_{low} < I_{high}$. In order to minimize selection effects, the instructions do not explain the type of game that will follow afterwards (Erkal et al. 2011). After the seven minutes pass, individuals are informed whether they will have a high or a low income. The exact number of words that other players encrypted is only announced at the end of the experiment. The data from the Encryption Task is presented in the Online Supplementary Material (*OSM*).

3.3 The basic game

As mentioned above, we implement experimental societies of ten members ($N = 10$): one 'politician' and nine 'citizens' (called 'Player A' and 'Player B' in the instructions, respectively). While five citizens are 'poor' and receive a low income I_{low} at the beginning of each period of the tax game, four citizens (and the politician) are 'rich' and receive a high income I_{high}. In every period of the game, each participant – both citizens and the politician – must choose how much of this income to declare. This is the only decision each participant has to make in a given period. The individual pays taxes at a flat rate t

[6] As the experiment varies the percentage of tax revenue that is kept by the 'politician', we did not want him or her to compete with the other group members, as this could legitimize embezzlement. For the same reason, we decided that the politician should not have a low income in the experiment.

on every experimental dollar of income that is declared $d_i \in [0, I_i]$, while no taxes are paid on undeclared income. That is, $t \cdot d_i$ denotes the amount of taxes paid.

A fraction of the tax revenue $(1 - \delta)$ is used to fund a public good, while a fraction δ of it is (automatically) kept by the politician. We will call δ the level of embezzlement and use it as our measure of institutional quality. The public good provides an overall benefit of α which is equally distributed among all players (including the politician), such that the marginal return per capita is $m = \alpha/N$. Since $I_{low} < I_{high}$, poor players benefit more from redistribution through the public good, all else equal. Letting T denote the total tax revenue collected in the group (i.e., $T = \sum_{j=1}^{N} td_j$), the payoff of citizen i from declaring income d_i is:

$$\pi_i = \begin{cases} I_i - td_i + \frac{\alpha}{N}(1-\delta)T, & \text{if not caught,} \\ I_i - td_i + \frac{\alpha}{N}(1-\delta)T - \theta t(I_i - d_i), & \text{if caught.} \end{cases}$$

where $\theta t(I_i - d_i)$ is the penalty paid on the undeclared income.[7]

Participants are audited with a probability p. This probability is endogenously determined. In particular, the audit probability depends on the total tax revenue, i.e., the declared income of all members of the group, and is given by $p = T/c$ where the numerator denotes the total tax revenue and c is a scaling parameter ensuring that p is a well-behaved probability, $p \in [0, 1]$.[8] This modeling approach implies that compliance decisions are interdependent as the (expected) cost of tax evasion increases with the overall compliance rate and can potentially lead to multiple equilibria. Given that the probability of being audited depends on overall tax payments, denoting whether a subject is the politician by the indicator variable P_i, we can express subjects expected payoffs by:

$$E\pi_i = [I_i - td_i] + \left[\frac{\alpha}{N}(1-\delta)T\right] + [\delta T]P_i - \left[\frac{T}{c}\theta t(I_i - d_i)\right]. \tag{1}$$

The term in the first bracket refers to i's 'after-tax income', the expression in the second bracket describes the citizens' returns from the public good, the term in the third bracket denotes the fraction of tax revenues the politician embezzles, and the final bracket captures the punishment for tax evasion.

3.4 Equilibrium analysis and experimental parameters

What is the optimal behavior in the basic tax game? The fact that the audit probability increases with the level of compliance implies that multiple equilibria can exist in the basic game. The expression in equation (1) is convex implying that, if players care solely for their

[7] The politician faces a similar problem, with the difference that his/her payoff is higher by δT, i.e., the amount of tax revenue s/he embezzles.

[8] Notice that the level of embezzlement does not affect the probability of an audit. We made this decision to allow for a cleaner comparison across treatments by separating the impact of institutional quality from enforcement.

individual payoff, the only possible symmetric equilibria are full compliance, i.e., $d_i = I_i$ for all i, and full evasion i.e., $d_i = 0$ for all i (see *OSM*). From equation (1), it is easy to see that no compliance is always an equilibrium in our game as, if no one pays taxes, i.e., $d_i = 0$ for all i, the audit probability is zero, since $T = 0$.[9] Whether full compliance can be supported in equilibrium depends on whether deterrence is strong enough when everyone else is fully complying such that it outweighs the gains from evading. As we show in the *OSM*, full compliance is also an equilibrium, if and only if

$$1 - \frac{\alpha}{N}(1-\delta) - P_i\delta \leq \frac{\sum_{j\neq i} tI_j}{c}\theta \qquad \forall i \tag{2}$$

The intuition behind condition (2) is straightforward. The left-hand side reflects the benefit from evading, while the right-hand side reflects the expected fine from evading when others comply fully. As one would expect, full compliance is easier to support in equilibrium as the audit probability or the fine increase, as the level of embezzlement decreases and the benefit from the public good increases.[10]

We chose parameters such that condition (2) is always satisfied and that both full compliance and full evasion are equilibria, irrespective of the level of embezzlement. To create a considerable amount of inequality and, thus, justify a redistributive tax, we chose the income of the rich citizens (and the politician), I_{high} = E\$60, to be 50% higher than the poor, I_{low} = E\$40. We implemented a tax rate that is substantial but still takes less than half of a subject's income away, $t = 40\%$. In addition, the return of the public good is $\alpha = 2.5$, leading to a marginal per capita return of 0.25 (= 2.5/10). In line with previous experiments, we choose a fine rate $\theta = 2$. By implementing an audit probability ranging from 0% to 56% ($c = 360$), we ensure that deterrence is sufficiently high in case everyone pays their taxes. This is true for the two different levels of embezzlement that we implement in our experiment: *low embezzlement* $\delta_L = 0.05$ and *high embezzlement* $\delta_H = 0.45$.[11]

3.5 Experimental treatments

Our goal is to explore the extent to which past outcomes can shape tax compliance in the future, even after institutions have changed. For this, we need to exogenously

[9] One way for full evasion not to be supported in equilibrium is to have a positive audit probability even when $T = 0$. This probability however would have to be rather high: 38.1% when embezzlement is low and 44.4% when embezzlement is high.

[10] Condition (2) illustrates that incentives in our set up are similar to those in models in which individuals have an intrinsic motivation to comply, as the expected cost from evading increases with the overall level of compliance (e.g., Bordignon 1993, Myles and Naylor 1996, Kim 2003, Fortin et al. 2007, Traxler 2010).

[11] Two considerations drove our choice of embezzlement levels. First, we wanted to implement a fairly large variation in embezzlement to ensure that the manipulation has an effect and at the same time the politician should embezzle less than 50% of the tax revenue. Second, the low embezzlement level was selected to be different from zero for reasons of practicality. It ensures that low and high embezzlement groups face structurally identical instructions, facilitating the experimental manipulation of embezzlement levels over time.

Table 1: Experimental treatments

Treatment	Sample size	Part 1 (Periods 1-15)	Part 2 (Periods 16-30)	Part 3 (Periods 31-45)
BG	$N = 8; S = 80$	Bad quality	Good quality	Good quality
GB	$N = 8; S = 80$	Good quality	Bad quality	Bad quality
BB	$N = 8; S = 80$	Bad quality	Bad quality	Bad quality
GG	$N = 8; S = 80$	Good quality	Good quality	Good quality

Notes: The first letter of the treatment refers to the institutional quality in Part 1, whereas the second letter refers to that in Part 2 and beyond. N denotes the number of independent groups in a given treatment and S the total number of participants in a treatment.

manipulate both past outcomes and present institutions. Our treatments vary the (i) degree of embezzlement, δ, and (ii) the history of embezzlement in order to explore the tension between past and present institutions. As mentioned, we consider two levels of embezzlement; $\delta_L = 0.05$, and $\delta_H = 0.45$. For simplicity in exposition, we will refer to institutions with low embezzlement as 'good-quality institutions' or 'good institutions', and institutions with high embezzlement as 'bad-quality institutions' or 'bad institutions'.

Participants are informed at the start of the experiment that the session will be divided into parts. The first part is aimed to create different past experiences for participants. For reasons we explain in the next subsection, we anticipate higher compliance rates in good- than in bad-quality institutions. The second part is aimed to explore whether tax compliance is path-dependent by changing the quality of the institutions. Participants are aware some of the experimental parameters may change in the course of the experiment, but are not informed which parameter will change or when this will occur. Each part lasts 15 periods.[12]

The experiment consists of four treatments as shown in Table 1. Our two main treatments explore the impact of a change in the quality of the institutions. Treatment BG investigates tax compliance when there is an improvement in quality; from bad quality to good quality. Treatment GB considers a deterioration in institutional quality; from good quality to bad quality. In addition, for comparison, we study behavior in two treatments where the quality of the institutions remains the same throughout the experiment. In treatment BB, the institutional quality is bad in both parts 1 and 2 of the experiment,

[12]The instructions do not distinguish between parts in the way we do in the paper. Here, we separate parts 1 and 2 for expositional ease. In the experiment, the first 30 periods were presented as one part in the instructions and subjects were only told that in one of these 30 periods, one or more parameters might change in the following manner: "in one period, one or more of the following parameters will change: the punishment multiplier, the share of the tax revenues that player A automatically keeps for himself, the multiplier of the contributions to the common project". In all periods, the fine rate, the degree of embezzlement and the public good return, are shown to subjects on their computer screen. In period 16, the new degree of embezzlement is shown to subject in red. In addition, a summary sheet is distributed to subjects, outlining how the change affects payoffs (see *OSM*).

while in treatment GG, the institutional quality is always good. Comparing behavior between the two main and the two control treatments reveals to what extent past outcomes can influence future tax compliance, as the institutional environment is the same in Part 2 of BG and GG (as well as in GB and BB) but the history is different.

The third part was designed to study a possible remedy if the exogenous institutional reform in BG turns out not to be completely successful in increasing compliance. At the start of Part 3, participants are provided with additional instructions informing them that there will be another 15 periods with the same parameters as in the previous 15 periods, but that before Part 3 starts they would participate in a poll. They have to state whether they agree or disagree with the following statement: "Higher tax payments increase group earnings. Thus, let's FULLY pay our taxes and do not evade them!" The number of subjects (dis)agreeing with the message in their experimental society is communicated to the subjects before they play another 15 periods of the tax game. The institutional quality does not change between parts 2 and 3. The only difference is that subjects participate in the poll and learn its outcome, after which they play the basic game for 15 periods with the same parameters as in Part 2.

In order to understand the origins of path dependence, it is useful to measure individuals' expectations about overall compliance. This however poses certain difficulties. Apart from the fact that eliciting beliefs could affect the way in which participants make decisions (Gächter and Renner 2010), incentivized belief elicitation introduces a hedging problem (Blanco et al. 2010). Given that our first priority is to explore whether tax compliance is path dependent in our setting, we decided to measure beliefs indirectly. We provided participants with a payoff calculator, which we termed 'tax advisor'. In each period, before making a decision, participants could calculate the audit probability and the payoffs in case they were audited or not, by specifying how much of their income others would declare and how much they were thinking of declaring. Given the complexity of the payoff function, participants had an incentive to use the 'tax advisor' to make better decisions. Although individuals could of course sometimes enter values which may have not reflected their expectations, on average, the entries should help us identify how subjects' expectations are affected by the quality of the institutions and institutional change. As we will see in Result 3 and Result 4, this is indeed the case.

3.6 Behavioral hypotheses

The fact that both full evasion and full compliance are equilibria of the stage game raises the question of equilibrium selection. We chose parameters such that full compliance is always payoff-dominant in all treatments. Average payoffs are 78.5 (66.5) in Good (Bad) institutions in case of full compliance, and 50 in case of full evasion. The rich, the poor and the politician all earn more when everyone fully complies than when everyone evades, irrespective of the institutional quality (see Table A.2 in *OSM*). On the other hand, as

we show in the *OSM*, full evasion has a substantially larger basin of attraction than the full-compliance equilibrium, irrespective of institutional quality. These measures therefore do not allow us to make sharp predictions about how institutional quality and its history may impact on tax compliance.

To obtain testable hypotheses, we calculate 'compliance thresholds'. These indicate the fraction of individuals that must fully comply such that a risk-neutral individual is indifferent between fully complying and fully evading. If the individual expects the overall compliance rate (i.e., the amount of taxes paid by other subjects divided by the amount of taxes owed) to be above this threshold their best response is to fully comply, and to fully evade if it is below it. The thresholds differ depending on the individual's role in the experiment (i.e., rich, poor or politician). In good-quality (bad-quality) institutions, the threshold is 75% (84%) for the poor, 78% (88%) for the rich citizens, and 73% (42%) for the politician. So, for instance, a rich citizen will be better off fully complying in a good-quality institution if s/he expects the overall compliance rate to be above 78%.[13]

The compliance thresholds reveal a few interesting comparative statics. First, the politician is the one most likely to full comply since her effective tax rate is lowest; the difference in thresholds between rich and poor citizens is quite small. Second and most important for our purposes, a worse institutional quality increases the overall compliance rate needed for a citizen to have an incentive to fully comply. The reason is that embezzlement reduces the social return from paying taxes which makes compliance less attractive. Rich citizens in particular have little (monetary) incentive to fully comply when facing bad-quality institutions as they obtain similar earnings when everyone complies and when everyone fully evades (E$63.5 vs. E$60). The opposite holds for the politician: the threshold is lower in bad than in good institutions as her effective tax rate decreases given that s/he keeps more of the tax revenue for herself. Based on these insights, we derive our first hypothesis.[14]

Hypothesis 1 [The impact of institutional quality in the absence of a past]: *In Part 1, compliance rates will be higher in treatments with good-quality institutions (i.e., GG and GB) than in treatments with bad-quality institutions (i.e., BB and BG).*

In case Hypothesis 1 is supported and tax compliance is responsive to differences in institutional quality, the question arises about how changes in institutional quality will affect compliance. The analysis above illustrates that the answer will largely depend

[13] As one might expect from the fact that evading when others are complying caries a substantial risk of being punished, risk aversion lowers the threshold for which full compliance is the best response for citizens. For illustration, if we assume $U(\pi_i) = \sqrt{\pi_i}$, then in good (bad) institutions, the thresholds are 63% (71%) for the poor, 66% (74%) for the rich, and 62% (35%) for the politician.

[14] Concerns for payoff inequalities would reinforce Hypothesis 1. In our game, the relative payoff equality varies with the quality of the institutions. In good quality institutions, the Gini-coefficient is smaller for full compliance than full evasion (0.05 vs. 0.10), whereas the opposite holds for bad quality institutions (0.17 vs. 0.10). As a reminder, the Gini-coefficient equals 0 when all earnings are equal and a value of 1 when one individual controls all the wealth.

on two factors: (*i*) how institutional change affects individual expectations about future compliance, and (*ii*) whether the resulting expectations about others' compliance will be above or below the compliance threshold. Broadly speaking, there are two possibilities concerning how the institutional change will affect expectations. The first is that, given the change in incentives, individuals will regard past outcomes as uninformative about compliance in the new institution. The second possibility is that past outcomes create mutually consistent expectations about what others will do, even after institutions change (David 1985). The latter may be exacerbated if groups have converged to one of the equilibria. We offer the following hypothesis:

Hypothesis 2 [The impact of a change in institutional quality]: *If Hypothesis 1 is supported, then if individual expectations about others' compliance are unaffected by the change in institutions and groups have converged to either full compliance or full evasion, past outcomes will affect future compliance such that compliance rates in the second part will be higher in GG than in BG, and higher in GB than in BB.*

Finally, we offer a hypothesis for the impact of the poll. Just like past outcomes, a 'society-wide poll' may help create mutually consistent expectations about what others are likely to do. Therefore, although the poll is a weak intervention due to its cheap talk nature, it may allow subjects to overcome the 'ghost of institutions past'. This may occur, if the high-quality institution in BG results in the majority of subjects favoring an increase in tax compliance. The publication of such outcome could realign individual expectations in favor of higher compliance. Therefore, the poll should have a stronger effect when the current institutional quality is good. Our final hypothesis is as follows.

Hypothesis 3 [The impact of a poll on overcoming the ghost of institutions past]: *If Hypothesis 2 is supported, then a poll will help increase compliance rates so long as the quality of institutions is good.*

3.7 Procedures

The experiment was conducted at the Experimental Economics Laboratory at the Technical University of Berlin in June 2016 using z-Tree (Fischbacher 2007). The 320 participants were students from universities in Berlin, randomly selected from a database of more than 3000 people using ORSEE (Greiner 2015). Earnings in the experiment are in experimental dollars (E$), which are converted to Euros at the end of the experiment at an exchange rate of E$100 = 5 EUR. Each individual participated in one of 16 sessions lasting approximately 105 minutes and earned 29 EUR on average, including a 5 EUR show-up fee.

At the start of the experiment subjects are provided with a set of general instructions that inform them that there are different parts to this experiment and that they will be informed about the content of the later parts only after the first had been completed (see *OSM*). Subjects play the game for a total of 45 periods. At the end of each period, they

receive information on whether they have been audited in this period, as well as their payoff in the preceding period, the average rate of compliance in their group by income level, and the average earnings by player type. The delayed feedback is implemented to capture the fact that it takes time to learn how others in society are behaving.

4 Results

The analysis of the data is divided into three subsections; one for each of our hypotheses. For statistical support, we use two-tailed tests that are based on statistical independent group averages, unless otherwise noted. To control for learning effects, the tests compare behavior in the last five periods of each part, unless otherwise noted. For simplicity, we will henceforth use boldface to highlight the present quality of institutions. For example, **G**B will indicate that we are investigating compliance in a good-quality institution in Part 1.

4.1 The impact of institutional quality in the absence of a past

Result 1: *In line with Hypothesis 1, average compliance rates in Part 1 of the experiment are about 50% higher with good (**G**G and **G**B) than bad institutional quality (**B**B and **B**G).*

SUPPORT: Figure 1 presents the average compliance rate in periods 1 to 15.[15] Compliance is substantially higher in the first part of the experiment when institutional quality is good (**G**G and **G**B: 75%) compared to when it is bad (**B**B and **B**G: 48%). The difference is statistically significant according to a Wilcoxon ranksum test ($p < .01, N = 32$).[16] ■

4.2 The impact of a change in institutional quality and the ghost of institutions past

Having established that our manipulation of the institutional environment leads to differences in compliance rates, we ask how an exogenous change in institutional quality impacts on compliance rates.

Result 2: *Past outcomes affect compliance even after a major change in the quality of institutions. The effect is especially pronounced in treatment BG, where an improvement in institutional quality has no impact on compliance rates. A weaker effect is observed in treatment GB, where deterioration in the quality of institutions leads to a reduction in compliance.*

[15]The average compliance rate is defined as the average percentage of income participants declared and does not take into consideration the fact that compliance has different overall implications for rich and poor citizens. We use this measure for simplicity, but note that we obtain very similar results if we instead use for our analysis the average income declared.

[16]Since there is no a priori reason why behavior would differ *in the first part* of treatment **G**G and **G**B (same for **B**B and **B**G) we pool the data. However, our sample allows us to detect significant differences between good and bad institutions even if we do not pool the data.

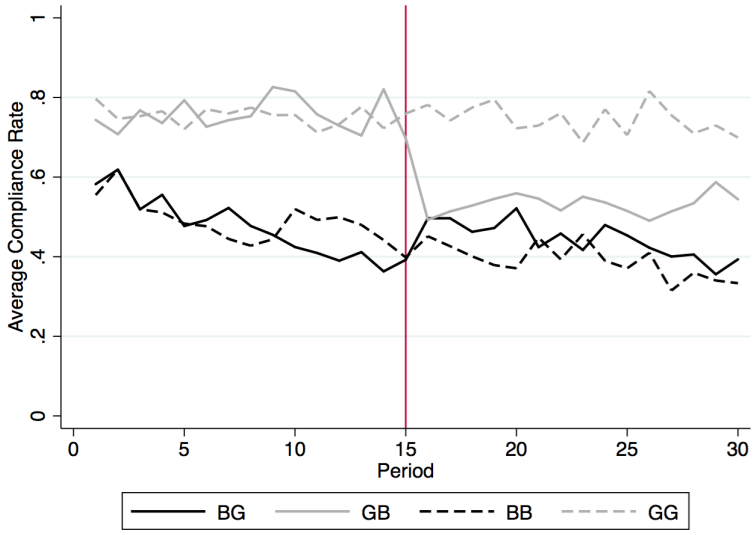


Figure 1: Average compliance rate in Part 1 and Part 2

Notes: This figure shows the average compliance rate over time for the four treatments. The vertical line marks period 15 after which in BG and GB institutional quality changed.

SUPPORT: As can be seen in Figure 1, when institutional quality improves in treatment BG, there is no noticeable increase in tax compliance, although embezzlement is reduced by nearly 90%. Comparing compliance rates in periods 11-15 and 26-30 in BG, a Wilcoxon signed-rank test fails to reject the hypothesis of no differences in compliance pre- and post-change ($p = 0.89, N = 8$). Similarly, we fail to reject the hypothesis of no differences in compliance rates in *Part 2* in treatments B**G** and B**B** (Wilcoxon ranksum: $p = 0.75, N = 16$), whereas we can reject the hypothesis that compliance rates are the same in treatments B**G** and G**G** (p=0.02, N =16). It is as if the institutional change did not occur. When institutional quality deteriorates in treatment GB, compliance rates decrease by about 20 percentage points. Comparing compliance rates in periods 11-15 and 26-30, a Wilcoxon signed-rank test finds this decrease to be significant ($p = 0.03, N = 8$). While compliance is weakly lower in G**B** than in G**G** (Wilcoxon ranksum: $p = 0.09, N = 16$) it is also weakly higher in G**B** than in B**B** (Wilcoxon ranksum: $p = 0.09, N = 16$), suggesting that a history of good institutions has a persistent – albeit modest – positive effect on compliance. ■

Result 2 is in line with Hypothesis 2. It suggests that an institutional reform, even a major one, may not be sufficient to increase compliance. Hypothesis 2 was based on the assumption that expectations about others' compliance are unaffected by the institutional change. Is this assumption supported by our data? To answer this question, we use the information entered in the 'tax advisor' calculator. Before we present our findings, however, two comments regarding the use of 'tax advisor' data appear to be pertinent.

As mentioned earlier, given the complexity of the payoff function, participants had an incentive to use the 'tax advisor' to make better decisions. Of course, there could

be no guarantee participants would not enter numbers randomly, but the evidence does not support this notion. There is a very clear pattern in our data suggesting individuals are not using the calculator randomly, but indeed to make better decisions. Individual inputs, in particular, are strongly correlated with the feedback participants received about behavior in previous periods. The overall correlation in between the average expectation an individual submitted to the 'tax advisor' and the average compliance rate observed in the previous period is 0.77 ($p < 0.01$) in periods 11-15, 0.49 ($p < 0.01$) across the first two parts, and excluding the first five periods of the experiment in which participants often use the calculator to explore the payoff function raises the strength of the correlation to 0.76 ($p<0.01$).

Another concern with the use of 'tax advisor' data is that, since subjects use the calculator voluntary, selection effects could affect our estimates. This can be a problem if those using the 'tax advisor' have systematically different expectations than those who do not use it. As it turns out, both in treatment GB and treatment BG, we fail to find significant differences in choices between subjects using and not using the calculator in both periods 15 and period 16 – the critical periods as they are just before and just after the institutional change. We proceed to explore how institutional change affects individual expectations about others' compliance.

Result 3: *We find no evidence to suggest that institutional change affects expectations about future compliance, even though institutional quality positively affects expectations in the absence of a history.*

SUPPORT: Table 2 shows the expectations about the average compliance rate in the two treatments in which there is a change in institutional quality: treatments BG and GB. We focus on three periods, which are of special interest. As can be seen, in both treatments, expectations seem not to differ much immediately before and after the change in institutions (i.e., periods 15 and 16). The differences are not significant (Wilcoxon signed-rank, BG: $p = .67, N = 8$; GB: $p = .61, N = 7$).[17] While a change in institutional quality does not appear to affect subjects' expectations, in period 1 (i.e., in absence of a past), expectations differ across institutions, with expectations in good-quality institutions being significantly higher than in bad-quality institutions (Wilcoxon ranksum: $p = .04, N = 16$). ■

4.3 The impact of a poll on overcoming the ghost of institutions past

So far we have seen that past outcomes in our experiment create mutually consistent expectations about what others intend to do, which a major institutional change cannot influence. The poll was designed as a cheap policy intervention which could potentially assist groups in overcoming past negative experiences by helping reboot individual expectations

[17] We do not find a significant change in the fraction of people using the calculator immediately before and after the change. Across conditions, 29.4% (47/160) use the calculator in period 15 and 34.4% (55/160) in period 16. The small increase is not statistically significant (Fisher's exact, $p = 0.40, N = 320$).

Table 2: Expectations about average compliance rate

Treatment	Period 1	Period 15	Period 16
BG	54.1 (49)	42.7 (26)	45.9 (28)
GB	65.2 (46)	63.2 (21)	66.3 (27)

Notes: Entries denote average expectation concerning others' compliance rate. In brackets is the number of subjects using the calculator in a given period.

and align them with current institutions. Result 4 reports on polling behavior and the effectiveness of the poll.

Result 4: *Past outcomes do not affect answers in the poll. In line with Hypothesis 3, the poll leads to a persistent increase in tax compliance when current institutional quality is good.*

SUPPORT: Table 3 shows the average number of votes in favor of full compliance across the four treatments (the maximum possible is 10). The majority of subjects appears to support full compliance, irrespective of the current institutional quality. Responses in the poll are very similar in treatments with the same current institutions, irrespective of their histories. A Wilcoxon ranksum test does not detect a significant difference between treatments with different histories (**B**B vs **G**B: $p = 0.54$, **G**G vs **B**G: $p = 0.79, N = 16$) indicating that past institutions have no effect on the polling behavior. In contrast, current institutions have a substantial effect on polling behavior. In good institutions on average 1.2 more people favor full compliance (Wilcoxon ranksum: $p < 0.01, N = 32$).

Table 3: Polling behavior

Treatment	Votes
BG	7.3
GB	6.8
BB	8.5
GG	8.1

Notes: Votes denotes the average number of votes in favor of the message (i.e., in favor of compliance).

Figure 2 reveals that there is a substantial increase in compliance rates following the poll in all treatments. The increase is statistically significant in all treatments if we compare the final five periods of Part 2 to the first five periods of Part 3 (Wilcoxon signed-rank: $p < 0.05, N = 8$, for all treatments).[18] The increase is especially pronounced in treatment

[18] As one might have expected, subjects' expectations about others' compliance as measured by the tax

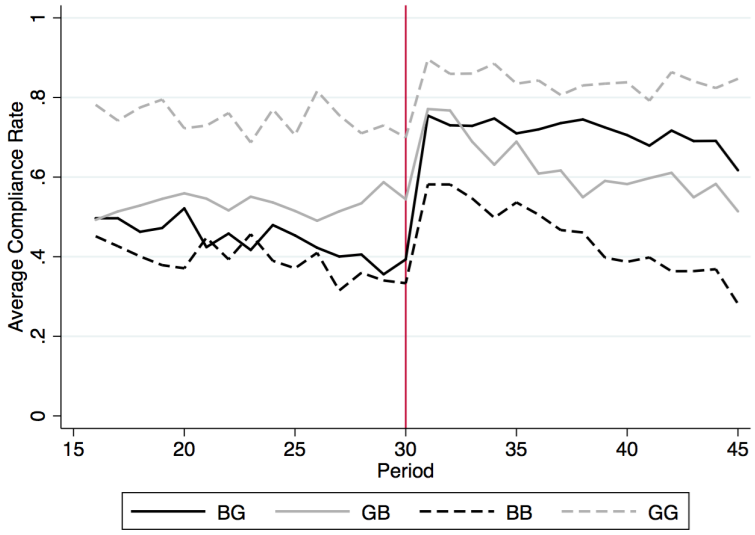


Figure 2: Average compliance rate in Part 2 and Part 3

Notes: This figure shows the average compliance rate over time for the four treatments. The vertical line marks period 30 after which the poll took place.

B**G** where compliance increases by 27 percentage points, but even in G**G** where compliance was already high in Part 2 the poll increases compliance by 10 percentage points. This increase however is persistent only in treatments with currently high institutional quality (i.e., G**G** and B**G**). Comparing compliance rates in the last five periods of Part 2 to the last five periods of Part 3, the difference is significant in GG and BG (Wilcoxon signed-rank: $p = 0.04$ in both cases, $N = 8$), but not in treatments GB and BB ($p = 0.48$ and $p = 0.89$, respectively). The increase in BG is large enough to overcome the negative effect of bad institutional history. We find no statistically significant differences in compliance rates in the last five periods of treatments BG and GG (Wilcoxon ranksum: $p = 0.46, N = 16$). ■

5 Discussion

Our results illustrate that past outcomes can impede the effectiveness of measures for combating tax evasion. Although the data supports our hypotheses, there are two observations which seem puzzling at first and warrant further investigation. The first is that behavior does not convergence to either full compliance or full evasion. The second regards the asymmetric reaction to the institutional change in treatments BG and GB. We explore these observations in turn.

advisor calculator increase significantly in all four treatments: from 41.7 (GB), 36.0 (BG), 65.0 (GG), and 41.4 (BB) in the period before the poll (period 30), to 72.1 (GB), 76.4 (BG), 88.4 (GG), and 74.5 (BB) following the poll, in period 31. This suggests that the poll is able to reset expectations and break the link with past institutions.

5.1 Understanding the non-convergence to full compliance and full evasion

One possible explanation for the patterns in Figure 1 could be that the average compliance rate is simply masking the fact that some groups converge to full compliance while other to full evasion. This is however not the case. Using a generous criterion for classifying a group as fully complying (evading), i.e., overall compliance rate above 75% (below 25%) in the last five periods of Part 1, we find that 10 out of 32 (3 out of 32) groups can be classified as fully complying (evading) across conditions (see *OSM*). The fact that hardly any group converges to full evasion, even in bad-quality institutions, is somewhat surprising. Even though the rate of compliance decreases over time in bad-quality institutions (Wilcoxon signed-rank test, comparing behavior in periods 1-5 to that in periods 11-15: $p = 0.01, N = 16$), overall compliance remains substantial as seen in Figure 1.

Another explanation is that, unlike what we assumed in our theoretical analysis, not all individuals care solely about their pecuniary payoff. Concerns for equality and honesty have been found to play a role in previous studies. Compliance will not fully break down if some individuals are evasion-averse, i.e. suffer a non-pecuniary cost from tax evading. Similarly, individuals may not comply fully if they are inequality-averse and dislike the increasing difference between their and the politician's earnings. The experimental data and the responses to our post-experiment survey suggest that these motives also play a part in our experiment. For example, 72 of the 320 subjects (22.5%) declared on average more than 90% of their income, while 31% of the subjects stated that they paid their taxes "to be honest" in the post-experiment survey. Similarly, 38% of participants reported that an important reason for their decision to evade taxes was to prevent high earnings for the politician.

To explore the relative importance of these non-pecuniary motivations, we first estimated a mixture model allowing for three behavioral types: 'selfish', 'inequality-averse' and 'evasion-averse'. All types are assumed to rationally best-respond to other players' past compliance and care about their earnings from the experiment. In addition, inequality-averse types care about their earnings but suffer a disutility from earning inequalities, while evasion-averse individuals experience a disutility when they evade taxes which is proportional to the extent of their evasion.[19] Assuming one population across all treatments and parts, 37.8% of subjects are classified as 'selfish', 36.7% as 'evasion-averse' and the remaining 25.5% as 'inequality-averse'.

[19]For evasion-averse individuals, we assume they have a utility function $U(d) = E\pi(d) - \beta(I-d)^2$, where d is the player's income declaration, I their income, $E\pi(d)$ the expected monetary payoff given d, and β is a parameter capturing their sensitivity to evading, similar to models of lying aversion (Kartik 2009, Dufwenberg and Dufwenberg 2016). For inequality-averse types, we assume they have a utility function $U(d) = E\pi(d) - \eta(\frac{E\pi(d)}{10\hat{\pi}} - \frac{1}{10})^2$, where d is the player's income declaration, $E\pi(d)$ is the expected monetary payoff given this declaration, η is the inequality-aversion parameter and $\hat{\pi}$ is the expected average monetary payoff in the group. As long as $\eta > 0$, the utility of such individuals will decrease as their relative payoff ($\frac{E\pi(d)}{10\hat{\pi}}$) deviates from the equal share (1/10), similar to the model by Bolton and Ockenfels (2000). For details of the estimation see *OSM*.

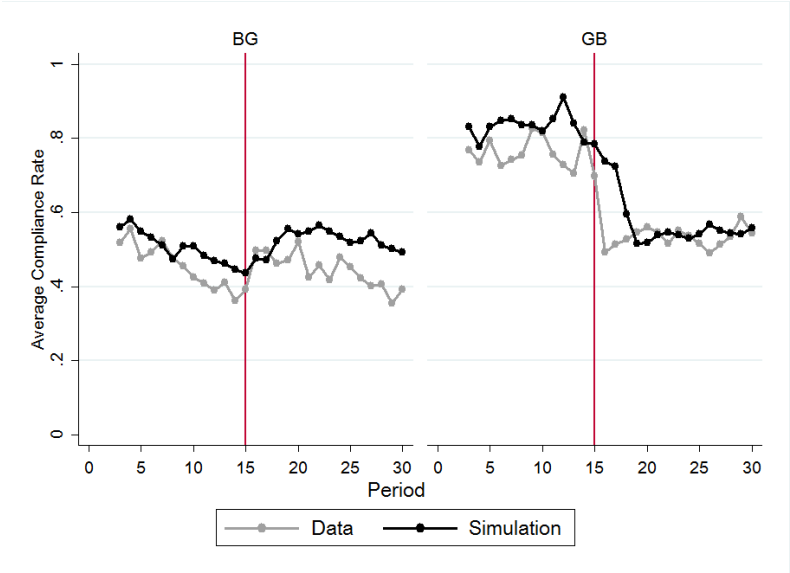


Figure 3: Simulated and actual behavior

Notes: This figure shows the observed as well as the simulated average compliance rate over time for treatments BG and GB. The vertical line marks period 15 after which in BG and GB institutional quality changed.

The existence of individuals who care about equality and honesty can in principle help account for the lack of convergence to full compliance and full evasion. But whether they explain the patterns in our experiment is a different matter. For this reason, we carried out a simulation analysis the details of which can be found in the *OSM*. Here, we present the main findings from this exercise for the two treatments in which there was a change in institutions. In the simulation, individuals are assumed to be either 'selfish' with probability 37.8%, 'evasion-averse' with probability 36.7%, or 'inequality-averse' with probability 25.5%, i.e., the probabilities estimated in the mixture model. All types are assumed to best respond to the levels of compliance they observed in the previous two periods.[20]

Figure 3 presents the results from the simulation and compares them to the actual data. The simulated behavior appears to track the observed levels of tax compliance closely. When compliance levels are low, selfish players generally best respond by not paying taxes. Evasion- and inequality-averse types are the ones who prevent compliance from fully deteriorating (see *OSM*). As it turns out, it is not the inequality-averse types who prevent convergence to full compliance in the first part of GB, but selfish players. Whereas some of them fully comply as compliance levels are high in their group, some

[20]In the *OSM*, we also present separately the behavior of each type. Note that, in the simulation, the experimental data is continuously fed in over time, such that in period 4, for example, types best-respond to the actual compliance rate in periods 1 and 2. (Subjects cannot best-respond to behavior in period 3 as they receive feedback with one period delay). Alternatively, one could only use the data from periods 1 and 2 and simulate behavior without any actual data updating. This kind of simulation however does not capture the noise that comes along with the real data. Since this noise matters for equilibrium convergence, this alternative approach leads to a worse fit.

find it optimal to fully evade as compliance is lower. In other words, many groups in GB have levels of (initial) compliance that are high but just not high enough to make it a best response for selfish players to pay their taxes. Compliance in these groups does not unravel but remains at relatively high levels since both inequality- and evasion-averse players help to sustain it. Finally the impression that the simulation fits the experimental data well is further supported by a regression where observed group behavior is regressed on predicted group behavior resulting in an R^2 of 0.76.

5.2 Why is tax compliance asymmetrically path-dependent?

In section 3.6, we hypothesized past outcomes will affect future compliance such that compliance rates will be higher in the second part of the experiment in GG than in BG, and higher in GB than in BB (Hypothesis 2). Although the hypothesis was confirmed (see Result 2), the data suggests an asymmetry in the reactions to institutional change. Whereas the improvement in institutions in treatment BG left compliance rates unaffected, as if the change never occurred, the deterioration in treatment GB had a clear negative impact on compliance, even if compliance was weakly higher in GB than in BB. What could account for this given that expectations appeared to be equally unaffected by the change (see Result 3)?

To answer this question, it is helpful to consider not only the long-term effect of the institutional change (as we did in Result 2), but also the immediate reaction to the change in institutions, i.e., the change in behavior between periods 15 and 16. Figure 1 indicates that the institutional reform in treatment BG led to a small short-term increase in compliance rates of about 10 percentage points (Wilcoxon signed-rank: $p = 0.07, N = 8$). Although this increase is only half as large as that in treatment GB where compliance rates dropped by 20 percentage points, it is evidence of an immediate reaction to the institutional change in both treatments. Therefore, a satisfactory answer must be able to account for both the immediate change in compliance in both BG and GB, and the fact that the change is persistent only in GB. As it turns out, the behavioral heterogeneity captured by the mixture model can account for both findings.

The simulation in Figure 3 captures the temporary increase in compliance in treatment BG. When looking deeper into the results of the simulation, we observe that inequality- and evasion-averse individuals increase compliance after the change, but not the selfish ones (see *OSM*). The reason is that the improvement in institutions makes taxation more effective in redistributing wealth, while simultaneously reducing the effective tax rate. However, the increase is small and neither of the two types ever reaches the compliance levels observed in Part 1 of GB, leading to a progressive decline over time. As a result, compliance does not reach levels that would induce selfish players to pay their taxes. In contrast, in treatment GB, we observe a permanent drop in compliance after the deterioration of institutions.[21]

[21]The drop is less immediate and drastic in the simulation. One reason is that although on average

This change is driven by selfish players some of which find themselves in groups in which suddenly full compliance is no longer a best response. Inequality- and evasion-averse types do not initially adjust their compliance, but the increasing rate of tax evasion due to the selfish types leads them to eventually declare smaller fractions of their income.

To summarize, the asymmetric reaction to institutional change in treatments GB and BG can be explained by the existence of different behavioral types and the associated best-response dynamics.

6 Conclusion

Spurred by recent theoretical models and empirical evidence suggesting that the individual cost of evading taxes can be positively associated with the overall level of compliance, we developed a new experimental game in which tax declarations are strategic complements. We used this game to investigate whether past outcomes can affect tax compliance in the future, even after institutions change. Our findings show that a history of low compliance can render even a major institutional reform ineffective at reducing tax evasion in our setting. The reason is that the institutional change appears to leave expectations about future compliance unaffected. The latter holds both in the case of institutional improvement and deterioration.

By showing that history can be an obstacle to fighting tax evasion, our findings have potentially important policy implications. Measures for combating tax evasion are introduced neither randomly nor in a vacuum; they are introduced in countries where tax evasion is a problem, in the place of previously inadequate policies or institutions. Our results suggest that it is precisely in these situations, i.e., where levels of compliance are typically low, where measures for combating tax evasion may be least effective. A history of tax evasion can create mutually consistent expectations between citizens that institutional reforms can fail to change. This may help explain the persistence of tax evasion in many developing countries (Burgess and Stern 1993), but also in developed countries with chronic tax problems such as Greece.

As an illustration of the potentially important policy implications of our findings, during the recent discussions on Greece's debt crisis, German finance minister Wolfgang Schäuble suggested sending 500 German tax specialists to Greece to increase tax compliance. With this he noted that "it is almost undisputed among economists that institutions matter", and hinted that the tax problem in Greece is due to problems with current institutions and the ineffective implementation of tax policies (Kirschbaum 2012). Our findings suggest

expectations do not change (see Result 3), some subjects appear to anticipate the treatment effect leading to bigger initial reaction than implied by the simulation. Another reason is the out-of-trend spike in compliance in period 14. Given the simple belief formation assumed in the simulation, this sudden upturn leads to optimistic beliefs about compliance in period 16. Of course, in the actual experiment, subjects may place less weight in this out-of-trend behavior. Since the simple simulation captures the main patterns in our data, however, we saw no reason to provide a more complex model of belief formation.

that such intervention in itself may be insufficient to have the desired effect. Experiences with past institutions can in some instances continue to 'haunt' economies long after the institutions have been replaced. Policies which help reset taxpayers' expectations may have beneficial effects, so long as the other causes of tax evasion such as incentives, have been effectively dealt with (see Result 4).

The discussion on policies raises the inevitable question of generalizing experimental results. At the most fundamental level, our data provides evidence that even major institutional reforms can fail to affect expectations and improve tax compliance. That is, our study provides a 'proof of existence' about the path dependency of tax compliance. We stress that, in our experiment, the reforms are common knowledge and their effectiveness beyond dispute. Moreover, the scale of the reforms – an instantaneous 90% reduction in the level of embezzlement – is arguably difficult to implement in reality. Given that these factors provide ideal conditions for institutional change to affect behavior, all else equal, one would think expectations would not be more responsive to institutional change in daily life. Of course, all things will not be equal. Can our theoretical analysis provide us with some guidance concerning the limits of our results?

An important factor in our design is the existence of strategic complementarities. As overall compliance decreases the cost for an individual to evade paying taxes decreases too. When compliance is low, the expected cost of evading is therefore low, despite the reform. An obvious remedy to consider therefore is to increase the expected cost for low levels of compliance. In the *OSM*, we show that, if we were to double the penalty for evading (from $\theta = 2$ to $\theta = 4$) or introduce a 15% probability of being audited when tax revenue is zero, a risk-neutral, selfish individual in a good-quality institution would have an incentive to fully comply only if s/he expected a 40% level of compliance in her group. While this is close to the compliance rate observed pre-reform in bad-quality institutions, these numbers are several times higher than what is observed in reality, e.g., Andreoni et al. (1998) report that the overall audit probability in the USA in 1995 was 1.7% and θ was close to 1.75.[22] In other words, strategic complementarities may lead to low costs of evading in bad quality institutions, but these costs do not appear to be lower than those in the field, suggesting that these complementarities do not drive our main result.

The main effect of the strategic complementarity is that it allows full compliance to be supported in equilibrium. In the absence of a history of tax evasion, compliance was maintained at high levels in good-quality institutions. Our results showed however that, whereas breaking out of the 'bad equilibrium' can be difficult, falling into it may be easier, even if expectations are slow to adjust to the institutional deterioration. The reason is that full compliance is a best-response only when the overall level of compliance is

[22]The probability of being audited in the US has been decreasing in recent years due to budget and staff cuts in the aftermath of the 2008 financial crisis. According to Ohlemacher (2017), in 2016, only 0.7% of all individuals were audited, and even individuals with incomes above $200,000 had only a 1.7% chance of being audited.

high. Since many groups failed to converge to full compliance prior to the institutional deterioration, the best-response dynamics led compliance to unravel even if the set of equilibria was unaffected by the change. Our theoretical analysis illustrates that the 'compliance threshold' for full compliance to be a best-response is high for a very large range of parameters (see *OSM*). This suggests that the fragility of high compliance is not an artefact of the parameters used in the experiment.

To summarize, we have presented evidence illustrating tax evasion can be path-dependent and that even major institutional reforms may fail to increase tax compliance sufficiently in societies with a history of high evasion. While our findings do not allow us to claim that the persistence of rampant tax evasion in developing and certain developed countries is due to path dependence, they do suggest that overcoming a history of tax evasion is likely to be more difficult than simply improving problematic institutions. Escaping the 'ghost of institutions past' may require non-traditional policies which will assure citizens of the willingness of others to break with the past. Such policies should be carefully designed as, unlike in our experiment, participation in society-wide polls will be costly and voluntary which could reduce their efficacy.

References

Allingham, M. G. and A. Sandmo (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics 1*(3/4), 323–338.

Alm, J., M. B. Cronshaw, and M. McKee (1993). Tax compliance with endogenous audit selection rules. *Kyklos 46*, 27–45.

Alm, J., B. R. Jackson, and M. McKee (1992a). Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal 45*(1), 107–115.

Alm, J., B. R. Jackson, and M. McKee (1992b). Institutional uncertainty and taxpayer compliance. *American Economic Review 82*(4), 1018–1026.

Alm, J., J. Martinez-Vazquez, and C. McClellan (2016). Corruption and firm tax evasion. *Journal of Economic Behavior & Organization 124*, 146–163.

Alm, J., G. H. McClelland, and W. D. Schulze (1992). Why do people pay taxes? *Journal of Public Economics 48*(1), 21–38.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax compliance. *Journal of Economic Literature 36*(2), 818–860.

Andreoni, J., N. Nikiforakis, and S. Siegenthaler (2017). Social change and the conformity trap. *working paper*.

Arapis, T. (2015). Why reform will fail in Greece. *Huffington Post* (http://www.huffingtonpost.com/author/theodoros-arapis).

Bassetto, M. and C. Phelan (2008). Tax riots. *Review of Economic Studies 75*(3), 649–669.

Blanco, M., D. Engelmann, A. Koch, and H.-T. Normann (2010). Belief elicitation in experiments: Is there a hedging problem? *Experimental Economics 13*(4), 412–438.

Bolton, G. and A. Ockenfels (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review 90*(1), 166–193.

Bordignon, M. (1993). A fairness approach to income tax evasion. *Journal of Public Economics 52*(3), 345–362.

Brandts, J., D. J. Cooper, E. Fatas, and S. Qi (2016). Stand by me - Experiments on help and commitment in coordination games. *Management Science 62*(10), 2916–2936.

Bruttel, L. and T. Friehe (2014). On the path dependence of tax compliance. *European Economic Review 65*, 90–117.

Burgess, R. and N. Stern (1993). Taxation and development. *Journal of Economic Literature 31*(2), 762–830.

Coricelli, G., E. Rusconi, and M. C. Villeval (2014). Tax evasion and emotions: An empirical test of the re-integrative shaming theory. *Journal of Economic Psychology 40*(1), 49–61.

Daley, S. (2013). Greece's aggressive pursuit of tax evaders appears to collect more anger than money. *The New York Times, October 28* (http://www.nytimes.com/2013/10/29/world/europe/greeces-aggressive-pursuit-of-tax-evaders-appears-to-collect-more-anger-than-money.html).

D'Attoma, J. (2016). Divided nation: North-South cleavages in Italian tax compliance. *Polity 42*(1), 66–99.

David, P. A. (1985). Clio and the economics of QWERTY. *American Economic Review 75*(2), 333–337.

Devetag, G. and A. Ortmann (2007). When and why? A critical survey on coordination failure in the laboratory. *Experimental Economics 10*(3), 331–344.

Donadio, R. (2012). For Greek tax reformers, good ideas aren't enough. *The New York Times, February 2* (http://www.nytimes.com/2012/02/03/world/europe/the-powerful-resist-change-to-greek-tax-system.html?mcubz=0).

Dubin, J. A. and L. L. Wilde (1988). An empirical analysis of federal income tax auditing and compliance. *National Tax Journal 41*(1), 61–74.

Dufwenberg, M. and M. Dufwenberg (2016). Lies in disguise – A theoretical analysis of cheating. *CESifo Working Paper Series No. 6208*.

Dulleck, U., J. Fooken, C. Newton, A. Ristl, M. Schaffner, and B. Torgler (2016). Tax compliance and psychic costs: Behavioral experimental evidence using a physiological marker. *Journal of Public Economics 134*, 9–18.

Erkal, N., L. Gangadharan, and N. Nikiforakis (2011). Relative earnings and giving in a real effort experiment. *American Economic Review 101*(7), 3330–3348.

Fehr, E. and K. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics 114*(3), 817–868.

Feld, L. P. and J.-R. Tyran (2002). Tax evasion and voting: An experimental analysis. *Kyklos 55*(2), 197–221.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fortin, B., G. Lacroix, and M.-C. Villeval (2007). Tax evasion and social interactions. *Journal of Public Economics 91*(11-12), 2089–2112.

Friedland, N., S. Maital, and A. Rutenberg (1978). A simulation study of income evasion. *Journal of Public Economics 10*(1), 107–116.

Gächter, S. and E. Renner (2010). The effects of (incentivized) belief elicitation in public good experiments. *Experimental Economics 13*(3), 364–377.

Galbiatia, R. and G. Zanella (2012). The tax evasion social multiplier: Evidence from Italy. *Journal of Public Economic 96*(5-6), 485–494.

Graham, D. A. (2015). How Greece became European. *The Atlantic, July 7* (https://www.theatlantic.com/international/archive/2015/07/is-greece-a-part-of-europe/397796/).

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association 1*(114-125).

Hamman, J., S. Rick, and R. A. Weber (2007). Solving coordination failure with "all-or-none" group-level incentives. *Experimental Economics 10*(3), 285–303.

Hossain, T. and J. Morgan (2009). The quest for QWERTY. *American Economic Review 99*(2), 435–440.

IMF (2015). Current challenges in revenue mobilization: Improving tax compliance. *IMF Staff Report*.

IMF (2017). Insolvency and enforcement issues in Greece. *IMF Country Report No. 17/41*.

Karnitschig, M. and N. Stamouli (2015). Greece struggles to get citizens to pay their taxes. *The Wall Street Journal, February 25* (https://www.wsj.com/articles/greece-struggles-to-get-citizens-to-pay-their-taxes-1424867495).

Kartik, N. (2009). Strategic communication with lying costs. *Review of Economic Studies 76*(4), 1359–1395.

Kim, Y. (2003). Income distribution and equilibrium multiplicity in a stigma-based model of tax evasion. *Journal of Public Economics 87*(7-8), 1591–1616.

Kirschbaum, E. (2012). German tax collectors volunteer for duty in Greece. *Reuters, February 25* (http://www.reuters.com/article/us-eurozone-germany-greece/german-tax-collectors-volunteer-for-duty-in-greece-idUSTRE81O0F120120225).

Lefebvre, M., P. Pestieau, A. Riedl, and M. C. Villeval (2015). Tax evasion and social information: An experiment in Belgium, France, and the Netherlands. *International Tax and Public Finance 22*(3), 401–425.

Martinez-Vazquez, J. and R. McNab (2000). The tax reform experiment in transitional countries. *National Tax Journal 53*(2), 273–298.

Myles, G. D. and R. A. Naylor (1996). A model of tax evasion with group conformity and social customs. *European Journal of Political Economy 12*(1), 49–66.

North, D. C. (1990). *Institutions, institutional change, economic performance*. Cambridge University Press.

Ohlemacher, S. (2017). Chances for a tax audit have rarely been this low. *Associated Press, March 6* (https://www.cnbc.com/2017/03/06/chances-for-a-tax-audit-have-rarely-been-this-low.html).

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. New York: Norton.

Smerdon, D., T. Offerman, and U. Gneezy (2016). Everbody's doing it: On the emergence and persistence of bad social norms. *Tinbergen Institute Discussion Paper No. 16-023/I*.

Torgler, B. (2005). Tax morale in Latin America. *Public Choice 122*(1-2), 133–157.

Traxler, C. (2010). Social norms and conditional cooperative taxpayers'. *European Journal of Political Economy 26*(1), 89–103.

Vogel, J. (1974). Taxation and public opinion in Sweden: An interpretation of recent survey data. *National Tax Journal 27*(4), 499–513.

Wilkening, T. (2016). Information and the persistence of private-order contractual enforcement institutions. *European Economic Review 27*(4), 194–215.

Online Supplementary Material

The ghost of institutions past:
History as an obstacle to fighting tax evasion

Aaron Kamm, Christian Koch and Nikos Nikiforakis

Contents

Appendix A Theoretical Analysis **3**
A.1 General Analysis 3
A.2 Equilibrium analysis 4
A.3 Equilibrium properties 6
A.4 Robustness 7

Appendix B Performance in the encryption task **11**

Appendix C Mixture model and simulation **12**
C.1 Model estimation 12
C.2 Simulation 14

Appendix D Group-level heterogeneity **17**

Appendix E Instructions and screenshots of the experiment **20**
E.1 Instructions 20
E.2 Summary sheet for low embezzlement 30
E.3 Summary sheet for high embezzlement 31

Appendix F Bibliography **32**

Appendix A Theoretical Analysis

A.1 General Analysis

Player i's expected payoff is given by:

$$E\pi_i = [I_i - td_i] + \left[\frac{\alpha}{N}(1-\delta)T\right] + [\delta T]P_i - c(d_i). \tag{A.1}$$

The definition of the variables is as follows:

- I_i is player i's income,
- t is the tax rate,
- d_i is income declared by player i,
- α is the return of the public good,
- N is the size of the group,
- T is the total amount of taxes collected (or the tax revenue), $T = \sum_{j=1}^{N} td_j$,
- δ is the fraction of the tax revenue that is embezzled,
- P_i is an indicator that only takes value one if player i is the politician,
- $c(\cdot)$ is the expected cost of tax evasion.

The cost of tax evasion is given by the product of detection probability and fine:

$$c(d_i) = \frac{b+T}{c} \cdot \theta t(I_i - d_i) \tag{A.2}$$

The definition of the variables is as follows:

- b/c is the baseline probability of tax evasion being discovered.[1]
- $c > b + \sum_{j=1}^{N} tI_j$ is a factor scaling the detection probability such that it is a well-behaved probability.
- θ is the punishment factor, i.e. what multiple of the evaded amount has to be paid as a fine.

The expected payoff is given by the following function:

$$E\pi_i = [I_i - td_i] + \left[\frac{\alpha}{N}(1-\delta)T\right] + [\delta T]P_i - \left[\frac{b+T}{c} \cdot \theta t(I_i - d_i)\right] \tag{A.3}$$

[1]Deviating from the model of the main text, we consider a baseline audit probability that is in effect even if everyone pays zero taxes (see footnote 9 – main text). This is done to be more general. In our experiment, $b = 0$.

We can compute the payoff maximizing income declaration, assuming that agents are risk neutral. The relevant first order condition is:

$$\frac{\partial E\pi_i}{\partial d_i} = -t + \left[\frac{1-\delta}{N}\alpha + P_i\delta\right] t + \frac{b+T}{c}\theta t - \frac{t}{c}\theta t(I_i - d_i) \equiv 0 \tag{A.4}$$

The second order condition is:

$$\frac{\partial^2 \pi_i}{\partial d_i^2} = \frac{2}{c}\theta t^2 > 0 \tag{A.5}$$

This implies that the objective function is convex and therefore the payoff maximizing income declaration, d^*, is given by a corner solution, $d^* = 0$ or $d^* = I_i$. Consequently, a player pays her taxes if the amount of taxes paid by others $\hat{T}$ is higher than the level at which she is indifferent between paying and evading taxes. The underlying idea is that the audit probability has to be sufficiently high to ensure that a player's best response is to pay taxes.

A.2 Equilibrium analysis

Specifically, $\hat{T}$ is defined by the following expression:

$$\begin{aligned} & \pi_i(d=0) = p_i(d=I_i) \\ \Leftrightarrow\ & I_i + \left[\frac{1-\delta}{N}\alpha + P_i\delta\right]\hat{T} - \frac{b+\hat{T}}{c}[\theta\ t\ I_i] = (1-t)I_i + \left[\frac{1-\delta}{N}\alpha + P_i\delta\right](\hat{T} + tI_i) \\ & \Leftrightarrow \hat{T} = \frac{c}{\theta}\left(1 - \left[\frac{1-\delta}{N}\alpha + P_i\delta\right)\right] - b \end{aligned} \tag{A.6}$$

Given that subjects either fully comply or fully evade, an interesting question is under what conditions the extreme outcomes of collective full compliance and collective full evasion are an equilibria.

Full compliance is an equilibrium as long as $\hat{T}$ is smaller than the maximum possible tax revenue, i.e. the amount of taxes collected if everyone else is declaring all their income. Thus, exchanging the amount other players pay in taxes, $\hat{T}$, with the maximum amount those players could pay, $\sum_{j\neq i} tI_j$, provides a condition when (fully) paying taxes can be a best response to others paying their taxes:

$$\frac{c}{\theta}\left(1 - \left[\frac{1-\delta}{N}\alpha + P_i\delta\right)\right] - b < \sum_{j\neq i} tI_j \qquad \forall i. \tag{A.7}$$

Full evasion is an equilibrium as long as $\hat{T}$ is positive for all players, i.e.:

$$\frac{c}{\theta}\left(1 - \left[\frac{1-\delta}{N}\alpha + P_i\delta\right]\right) - b > 0 \qquad \forall i. \tag{A.8}$$

We choose the parameters of the experiment such that conditions (A.7) and (A.8) are satisfied. With these parameters, one can calculate compliance thresholds such that an individual fully complies if she expects the overall compliance to be above this threshold and fully evades if it is below. These thresholds are determined by the amount of taxes others have to pay, $\hat{T}$, such that complying becomes a best response. To aid the interpretation of these thresholds, we normalize them by the maximum amount of taxes other could pay. Thereby, we reframe the thresholds in terms of the overall compliance level by other group members. These values are shown in Table A.1.

Table A.1: Compliance thresholds

	Politician	Rich citizen	Poor citizen
Low embezzlement	73%	78%	75%
High embezzlement	42%	88%	84%

Notes: Entries show for each player type and embezzlement level the resulting compliance thresholds in terms of the overall compliance level by others.

One important question is whether given our parameter choice other equilibria alongside the full-compliance and full-evasion ones exist. We restrict our attention to pure-strategy equilibria. Notice that all equilibria are type-symmetric (i.e. citizens with the same income choose the same action) because all players of the same type face symmetric incentives. We consider 8 equilibrium candidates:

1. Full compliance:

   This is an equilibrium since for all types the thresholds are below 100%.

2. Full evasion:

   This is an equilibrium since for all types the thresholds are above 0%.

3. Only the poor comply:

   In this case the total collected taxes are 80 ($= 5 \times 16$) which is not sufficiently high for the poor to comply since for them condition A.7 is not satisfied.

4. Only the rich comply:

   In this case the total collected taxes are 96 ($= 4 \times 24$) which is not sufficiently high for the rich to comply since for them condition A.7 is not satisfied.

5. Only the politician complies:

   In this case the total collected taxes are 24. This is not sufficiently high for the politician to comply since for them condition A.7 is not satisfied.

6. The poor and the rich comply:

In this case the collected taxes are 176 ($= 5 \times 16 + 4 \times 24$) but then also the politician would comply since condition A.7 is satisfied for the politician.

7. The poor and the politician comply:

   In this case the collected taxes are 104 ($= 5 \times 16 + 24$) but this is not sufficiently high for the poor to comply since for them condition A.7 is not satisfied.

8. The rich and the politician comply:

   In this case the collected taxes are 120 ($= 5 \times 24$) but this is not sufficiently high for the rich to comply since for them condition A.7 is not satisfied.

Therefore, the only pure-strategy equilibria are full compliance and full evasion.

A.3 Equilibrium properties

This section discusses various properties of the two equilibria and how they relate to equilibrium-selection criteria.

First, we consider the equilibrium payoffs associated with the two equilibria which are shown in Table A.2. As is easy to see, the full-compliance equilibrium is payoff-dominant for both levels of embezzlement and, furthermore, each type prefers the full-compliance to the full-evasion equilibrium.

Table A.2: Equilibrium payoffs

	Politician	Rich citizen	Poor citizen	Weighted Average
full evasion	60	60	40	50
full compliance (low embezzlement)	93.5	83.5	71.5	78.5
full compliance (high embezzlement)	153.5	63.5	51.5	66.5

Notes: Entries show for each player type the payoffs resulting from the two equilibria.

Next to payoff dominance, risk dominance (Harsanyi and Selten 1988) is a prominent tool for equilibrium selection. Unfortunately, there does not exist a consensus on how to generalize this concept to games with more than two players.[2] Therefore, we consider the relative stability of the two equilibria as measured by the size of equilibria's basin of attraction.[3]

To compute the basin of attraction we randomly generate initial behavior by assigning each player a compliance level uniformly distributed between 0 and 100%. In the next

[2] See, for instance, Kim (1996) and Peski (2010) for papers proposing such generalizations.

[3] An equilibrium's basin of attraction consists of all the combinations of initial behavior that over time converge to this equilibrium.

Table A.3: Basin of attraction

	Full evasion	Full compliance
Low embezzlement	9,995,219	4,781
High embezzlement	9,999,995	5

Notes: Entries show the number of simulations converging to each of the two equilibria.

step, we compute each player's optimal behavior given the other players' behavior. This best response is then taken to again compute optimal behavior. We repeat this process ten times such that behavior has converged to an equilibrium. Overall, we generate 10,000,000 draws of initial group behavior and then count how many of these draws converge to the full-evasion and full-compliance equilibrium, respectively.[4] The relative frequency of the full-compliance equilibrium then describes the size of the basin of attraction. Table A.3 shows the convergence pattern for the two embezzlement levels.

As becomes apparent from Table A.3, the basin of attraction of the full-evasion equilibrium is much larger than the basin of attraction of the full-compliance equilibrium. The reasons for this are the high equilibrium thresholds which exert strong pressure towards the full-evasion equilibrium unless a substantial majority of the players start with very high initial compliance levels. Furthermore, we see that the likelihood of converging to full compliance starting from random initial play is substantially higher when embezzlement is low but it is very small, nevertheless.

A.4 Robustness

This section discusses how the compliance thresholds react to changes in the baseline audit probability, the fine multiplier, the marginal-per-capita-return of the public good and the embezzlement level.

Figure A.1 shows that as the baseline detection probability increases the compliance levels necessary to make full compliance a best response decrease linearly. For very high baseline probabilities the thresholds even become negative indicating that full compliance is a dominant strategy.

Figure A.2 shows that as the fine multiplier increases the compliance levels necessary to make full compliance a best response decrease but the speed of this decline is diminishing. This is caused by the absence of a baseline probability which implies that even for very high fine rates full evasion is an equilibrium unless overall compliance is sufficiently high.

Figure A.3 shows that as the public good multiplier increases the compliance levels necessary to make full compliance a best response decrease linearly. Nevertheless, even for very high values of the public good multiplier substantial compliance by others is necessary

[4]The underlying code is available upon request.

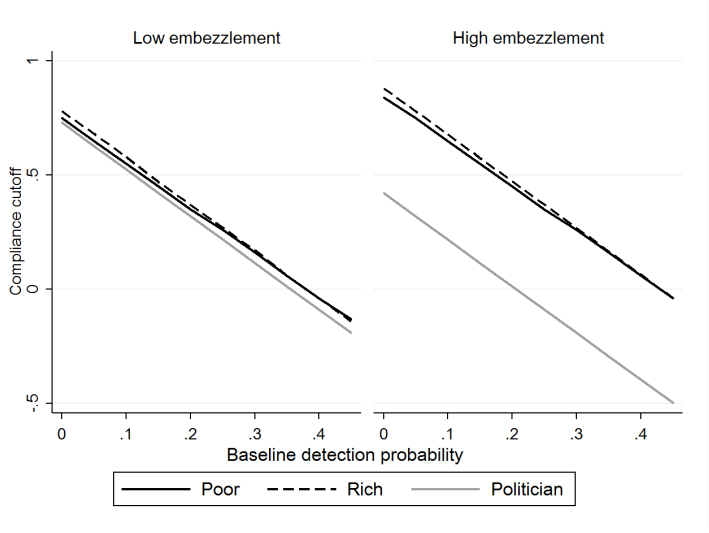


Figure A.1: Baseline detection probability

Notes: This figure shows compliance thresholds by type for varying levels of the baseline detection probability. All other variables are fixed at the value used in the experiment.

to make full compliance a best-response. Figure A.3 furthermore shows that the exogenous manipulation of compliance in Part 1 could have potentially been achieved by varying the public good multiplier. Introducing embezzlement by the politician has two implications: First, it makes the provision of the public good less effective. For instance, with high embezzlement, only 55% of the tax revenue is actually spent on the public good, implying an effective public good multiplier of only 1.4 ($= 0.55 * 2.5$). Second, in the presence of embezzlement, full compliance does not only reduce inequality among the citizens but also increases the inequality in earnings between citizens and the politician. This implies that while the compliance thresholds resulting from embezzlement could also have been implemented by varying the public good multiplier, the two are not equivalent.

We decided to manipulate the level of embezzlement for two reasons. First, as outlined in the main text, there is clear empirical evidence that tax compliance is negatively associated with embezzlement and corruption. Second, even though a direct manipulation of the multiplier could create similar differences in the compliance threshold, it is not clear whether it would create a sufficient difference between different treatments in Part 1. Introducing the politician reinforces the difference in thresholds across treatments because it creates substantial inequality when the politician is able to embezzle a lot of the tax revenue thereby making cooperation more difficult.

Figure A.4 shows that as the embezzlement level increases the compliance thresholds of citizens and politician diverge. While the politician becomes more willing to pay her taxes (since her effective tax rates decreases) the citizens become less likely to comply (since they get less of their taxes refunded through the public good).

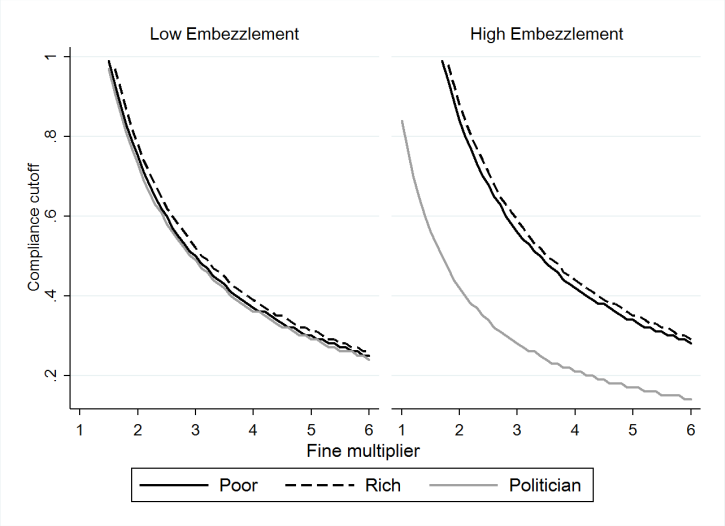


Figure A.2: Fine Multiplier

Notes: This figure shows compliance thresholds by type for varying levels of the fine multiplier. All other variables are fixed at the value used in the experiment.

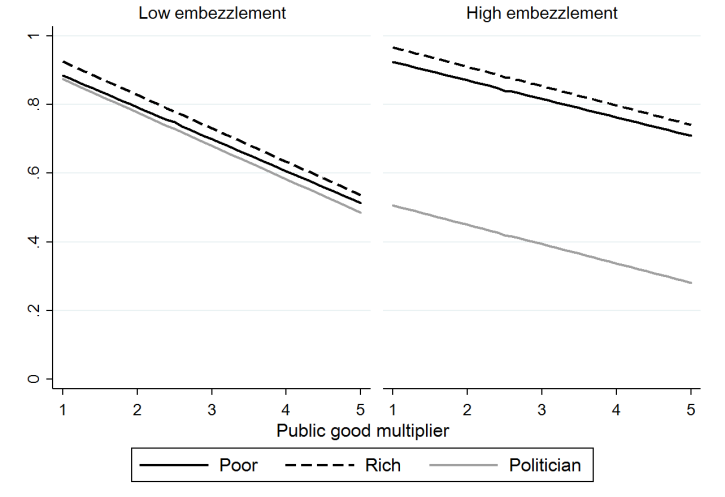


Figure A.3: Public-good multiplier

Notes: This figure shows compliance thresholds by type for varying levels of the public good multiplier. All other variables are fixed at the value used in the experiment.

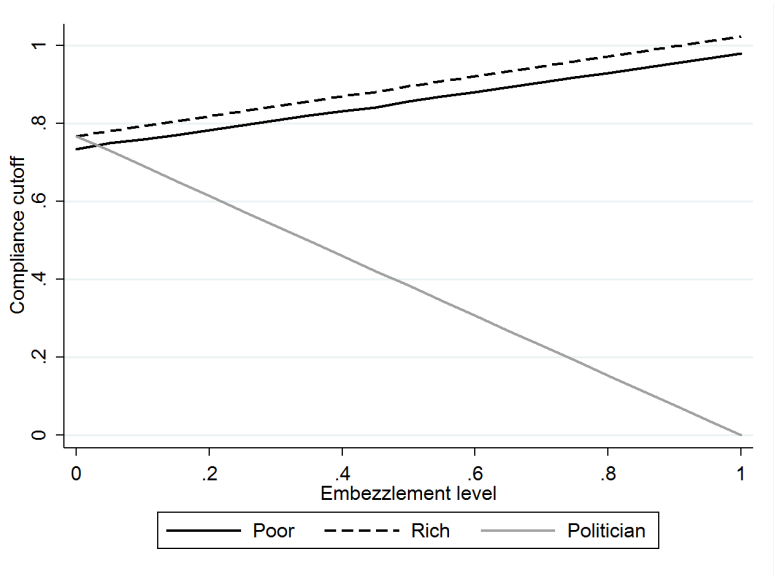


Figure A.4: Embezzlement level

Notes: This figure shows compliance thresholds by type for varying values of the embezzlement level. All other variables are fixed at the value used in the experiment.

Appendix B Performance in the encryption task

Figure B.1 shows a histogram of subjects' performance in the Encryption Task. As we can see there is substantial variation in the task performance with the number of encoded words ranging from 13 to 51 with 50% of subjects encoding between 27 and 35 words.[5] Furthermore, even though the subjects assigned to the role of politician knew before performing the task that it would not have any influence on their earnings, they nevertheless exerted substantial effort and encoded on average 28 words. This is only three words less than the 'citizen'-subjects encoded, for which the task performance had an influence on earnings (a Wilcoxon rank sum test with $N = 320$ finds the difference between citizens and politicians only significant at the 10%-level, $p = .08$).

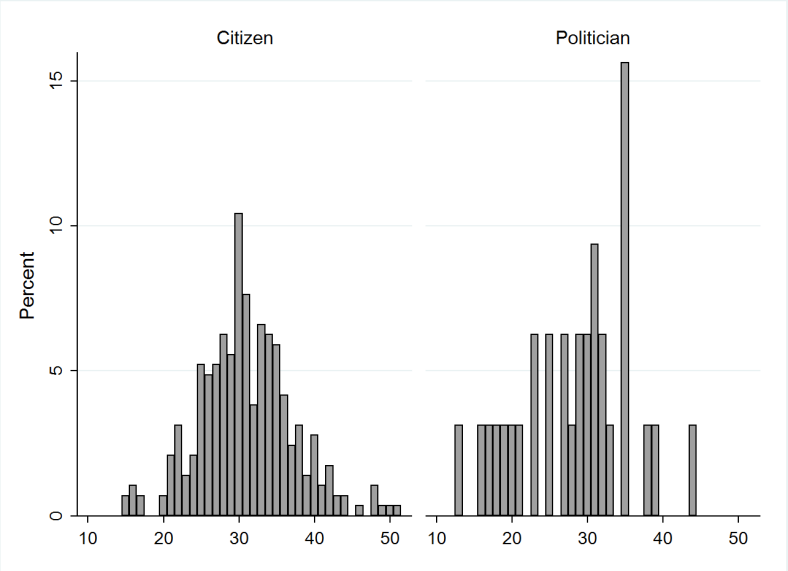


Figure B.1: Histogram of task performance

Notes: This figure shows the number of words encoded by 'citizen'- and 'politician'-players, respectively.

[5] A Wilcoxon ranksum test finds no significant difference in citizens' task performance across initial institutional quality ($p = 0.36, N = 288$).

Appendix C Mixture model and simulation

Below, we will outline our mixture-model estimation and the details of the simulation based on this model. A mixture model is a probabilistic model that takes into account the presence of different types within a population (see also, Cappelen et al. 2007, Bardsley and Moffat 2007, Moffat 2016). Notably, based on a player's behavior the model identifies which type she belongs to. At the same time, the mixture model cannot identify the underlying types by itself but relies on externally given types. Theoretical considerations as well as post-experimental questionnaire data in which subjects were asked about their motives to pay taxes guided the selection of types in our model.

C.1 Model estimation

In the baseline model, we assume that a player is either a selfish, evasion-averse or inequality-averse type. A selfish player's compliance rate is described by a random draw from $N(\mu_S, \sigma_S)$ truncated at 0 and 1, where μ_S is the compliance level d_i maximizing π_i (given by A.3). An evasion-averse player's compliance level is described by a random draw from $N(\mu_{EA}, \sigma_{EA})$ truncated at 0 and 1, where μ_{EA} is the compliance level d_i maximizing $\pi_{EA} = \pi - \beta(I_i - d_i)^2$.[6] An inequality-averse player's compliance is distributed according to $N(\mu_{IA}, \sigma_{IA})$, truncated at 0 and 1, where μ_{IA} is the compliance level d_i maximizing $\pi_{IA} = \pi_i - \eta(\frac{\pi_i}{10 \cdot \hat{\pi}} - \frac{1}{10})^2$.[7]

Given these behavioral assumptions, for each observed choice x_i, we can compute the likelihood of this choice being made by each of the three types. The contribution of this observation to the likelihood function to be optimized is then given by the likelihood for each type weighted by its frequency in the population (denoted by P_S, P_{EA}, P_{IA}). The objective function of the maximum likelihood estimation is therefore given by

$$LL(x, \sigma_S, \sigma_{EA}, \sigma_{IA}, P_S, P_{EA}, \delta, \beta)$$
$$= P_S \cdot LL_S(x, \sigma_S, \mu_S) + P_{EA} \cdot LL_{EA}(x, \sigma_{EA}, \mu_{EA}, \beta) + (1 - P_S - P_{EA}) \cdot LL_{IA}(x, \sigma_{IA}, \mu_{IA}, \eta).$$

In order to estimate the frequency of the three types and the amount of noise in their behavior we need to compute the players' choices which depend on the other players' behavior. We assume that players have a memory of two periods and best respond to each player's average behavior over the past two periods. We decided against purely myopic behavior where only the last period is considered when making choices to reduce the influence of potential outliers on behavior.

[6] We also considered a more extreme version of evasion-aversion where subjects are completely honest (i.e. $\beta = \infty$) but a likelihood ratio rejected $\beta = \infty$.

[7] Additionally, we also considered a special type of inequality aversion where players are not only averse to inequality in general but also to inequality relative to the politician (i.e. they are embezzlement-averse). In that case $\pi_{IA} = \pi_i - \eta(\frac{\pi_i}{10 \cdot \hat{\pi}} - \frac{1}{10})^2 - \gamma(\frac{\pi_{pol}}{10 \cdot \hat{\pi}} - \frac{1}{10})^2$. A likelihood ration test failed to reject $\gamma = 0$.

Table C.1: Basin of attraction

	Selfish	Inequality-averse	Evasion-averse
Fraction of population	37.8%	36.7%	25.5%
Standard deviation of choices	74	33	251

Notes: Entries show estimated frequency of the three types and the standard deviation of their choices.

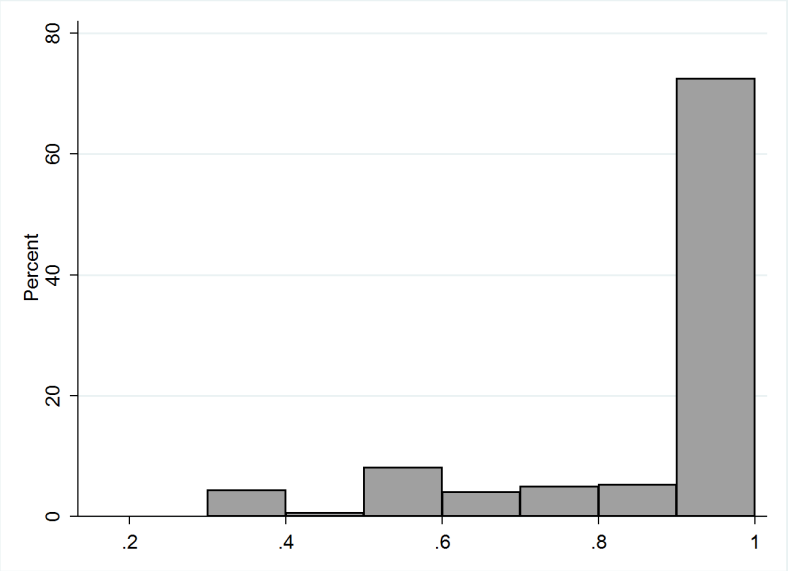


Figure C.1: Distribution of posteriors

Notes: This figure shows the distribution of posterior probabilities assigning a player to her most likely type.

When computing the frequencies of types, we assume that the distribution of types does not vary across treatments. We therefore consider types as innate characteristics of the players that are not influenced by the environment they are experiencing. Additionally, to avoid over-fitting, we only use the first 15 periods to estimate the frequency of types and the amount of noise in their behavior.[8] Table C.1 shows the resulting estimates. We find that selfish and inequality-averse (with $\eta = 1,000,000$)[9] account for very similar shares and are more common than the evasion-averse players (with $\beta = 0.0045$) who account for a quarter of the population. Figure C.1 shows the implied posterior probabilities, i.e. the certainty with which a player can be assigned to a certain type. As we can see the clear majority of subjects can be assigned to one of the three types with great certainty. For instance, for around 70% of subjects the model is more than 90% certain in its type assignment.

[8]Using the first 30 periods leads to qualitatively similar results.

[9]This extremely high inequality-aversion parameter is due to the small influence a player's choice has on inequality owing to the relatively large number of players in the group. Therefore, for inequality-aversion to have a noticeable impact on a player's utility a high degree of inequality-aversion is necessary.

C.2 Simulation

Based on the results of the mixture model, we can simulate the behavior implied by the type distribution. To do so, we assume (as in the estimation) that subjects have a limited memory and believe compliance in the next period to be the average of the previous two periods. Each type then best responds according to her type and this best response is taken as her predicted choice for the simulation. Figures C.2-C.4 show the average simulated behavior. The prediction for behavior on the group level is then given by the average of the type's behavior weighted by this type's estimated frequency in the population.

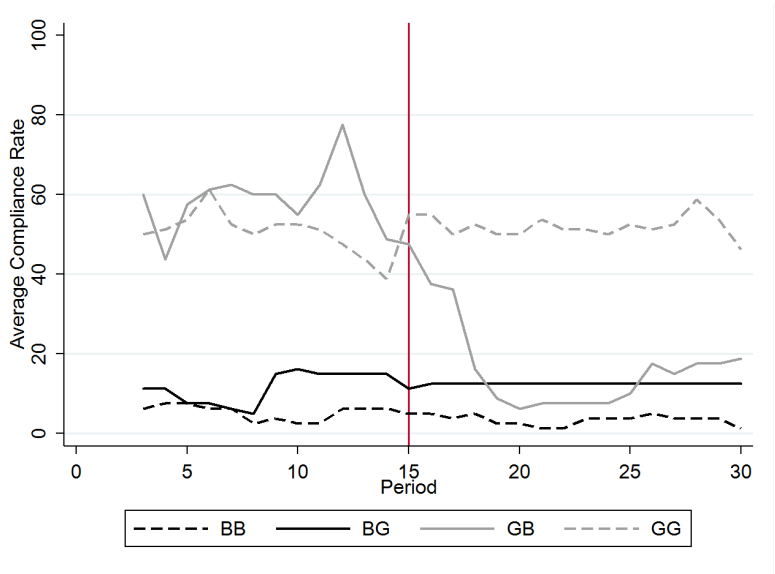


Figure C.2: Selfish players

Notes: This figure shows the average simulated compliance rate of selfish players over time for the four treatments. The vertical line marks period 15 after which in BG and GB the embezzlement level changed.

Figure C.2 shows the (simulated) behavior of the selfish players in Part 1 and 2. In Part 1, we observe a large difference in compliance – about 50 percentage points – between good and bad institutions. With bad institutions, in the large majority of groups compliance is not sufficiently high to induce selfish players to pay their taxes. With good institutions, compliance is high enough in slightly more than half the groups, although there is quite a bit of fluctuations. The latter finding reflects that many groups' compliance levels are close to the compliance threshold, sometimes above and sometimes below. In Part 2, selfish players do not increase their compliance in BG, despite the improvements in incentives. The reason is that the overall level of compliance remains insufficient to induce compliance. The worsening of incentives in GB, however, has a profound impact on compliance. Because many groups have been close to the compliance threshold in Part 1, there is an immediate drop in compliance as it is no longer a best response to comply in these groups. This drop becomes even more pronounced in later periods when subjects observe that others have

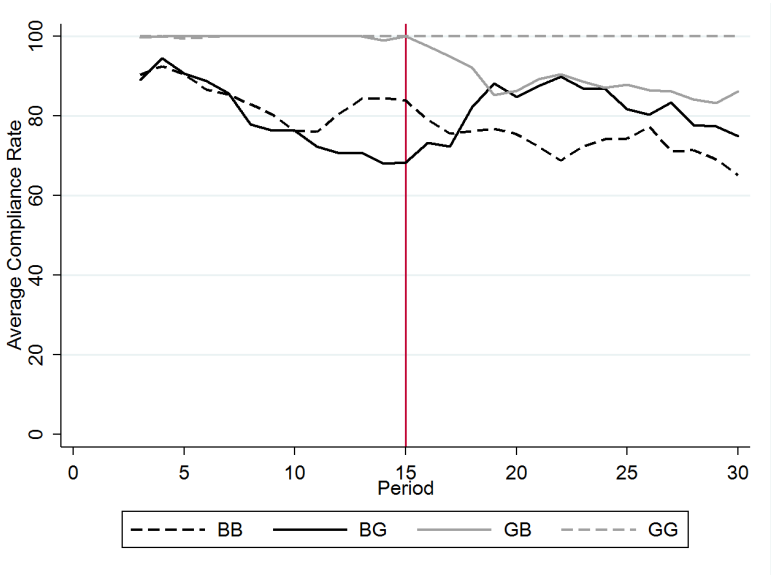


Figure C.3: Inequity-averse players

Notes: This figure shows the average simulated compliance rate of inequality-averse players over time for the four treatments. The vertical line marks period 15 after which in BG and GB the embezzlement level changed.

reduced their compliance and therefore even fewer groups have overall compliance rates sufficient to induce compliance. Overall, the behavior of the selfish players can help to explain why many groups do not converge completely to full compliance, why we observe an initial treatment difference across institutions and why compliance deteriorates after switching to bad institutions.

Figure C.3 and C.4 show the (simulated) behavior for the non-selfish players: the inequality-averse and the evasion-averse types. The qualitative behavior of the two player types is remarkably similar. While aversion to inequality can provide both an incentive to reduce inequality by paying redistributive taxes as well as an incentive to evade taxes to avoid enriching the politician, the former motive seems to be empirically predominant.[10] In Part 1, the compliance level is very high and the observed difference between good and bad institutions is relatively small (roughly 20 percentage points). This suggests that both types help to sustain compliance in both environments, although compliance incentives still make a small difference. In Part 2, compliance increases (decreases) slightly in BG (GB). These responses come with a delay, are fairly symmetric and not very large. Due to the latter fact, the increase in compliance in BG for the two non-selfish types is not sufficient to induce selfish players to pay their taxes. Moreover, as reflected in the delay, the decrease in compliance in GB seems more a response towards the steep decline in compliance of selfish players than towards the change in conditions. Overall, non-selfish

[10] As discussed in footnote 7, we also considered a player type that puts additional weight on inequality against the politician. As outlined above, we do, however, not find statistical support for this type of player.

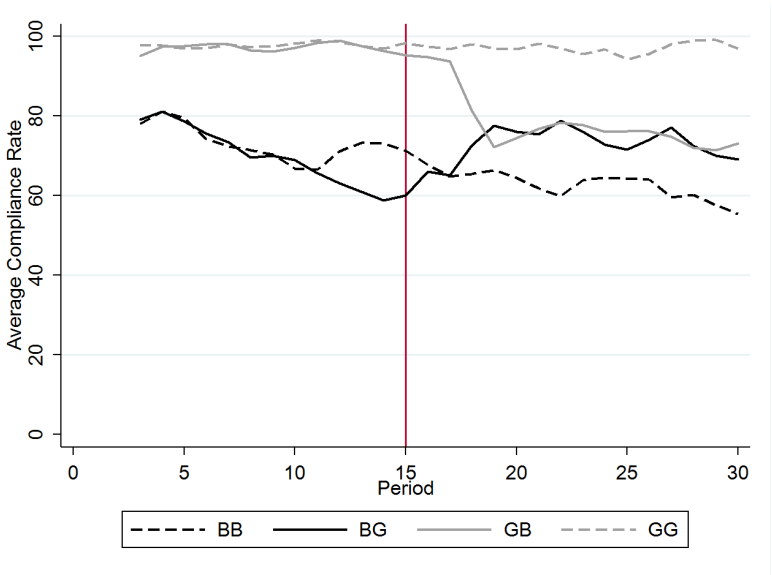


Figure C.4: Evasion-averse players

Notes: This figure shows the average simulated compliance rate of evasion-averse players over time for the four treatments. The vertical line marks period 15 after which in BG and GB the embezzlement level changed.

players help to sustain compliance even in cases in which selfish players have no incentive to comply. Their reaction to institutional change is fairly moderate.

Appendix D Group-level heterogeneity

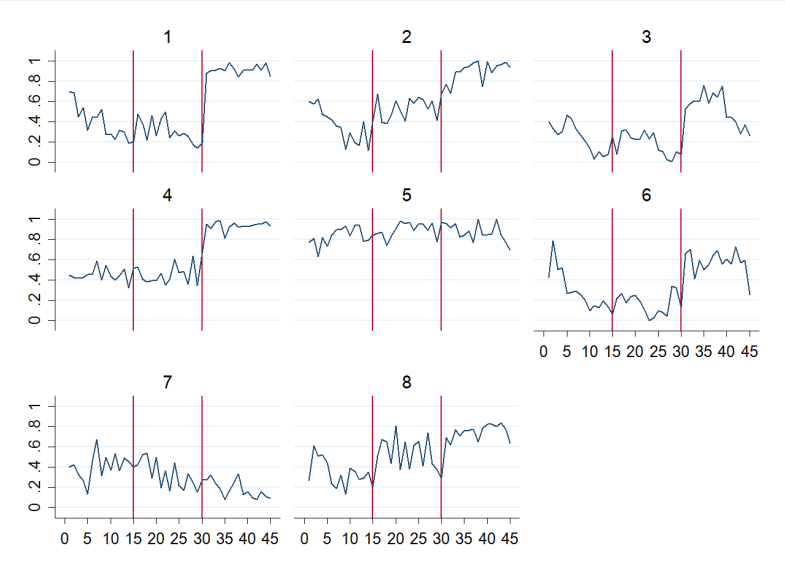


Figure D.1: Behavior on group-level in BG

Notes: This figure shows the average observed compliance rates for each group. The vertical lines mark period 15 (after which in BG and GB the embezzlement level changed) and period 30 after which the poll took place.

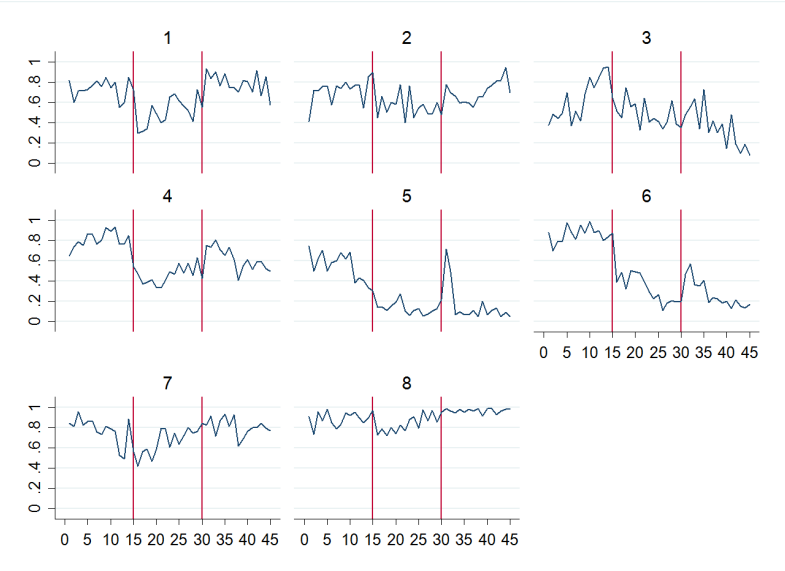


Figure D.2: Behavior on group-level in GB

Notes: This figure shows the average observed compliance rates for each group. The vertical lines mark period 15 (after which in BG and GB the embezzlement level changed) and period 30 after which the poll took place.

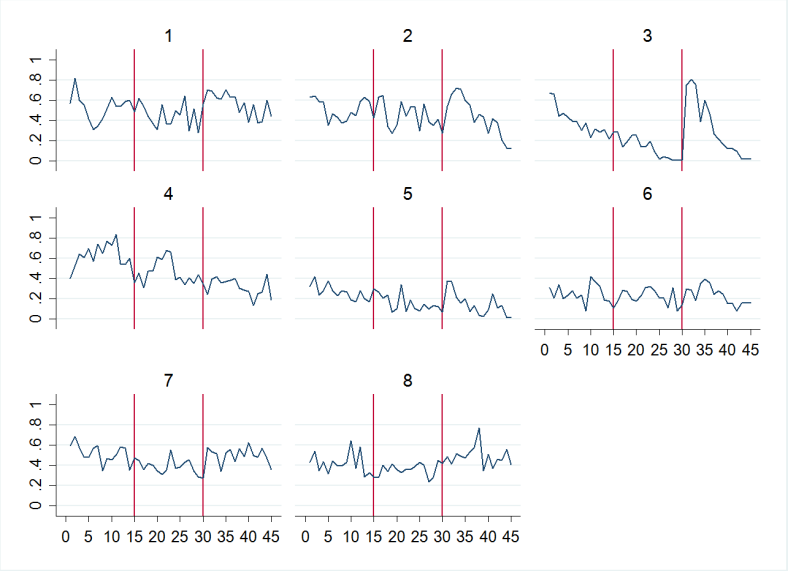


Figure D.3: Behavior on group-level in BB

Notes: This figure shows the average observed compliance rates for each group. The vertical lines mark period 15 (after which in BG and GB the embezzlement level changed) and period 30 after which the poll took place.

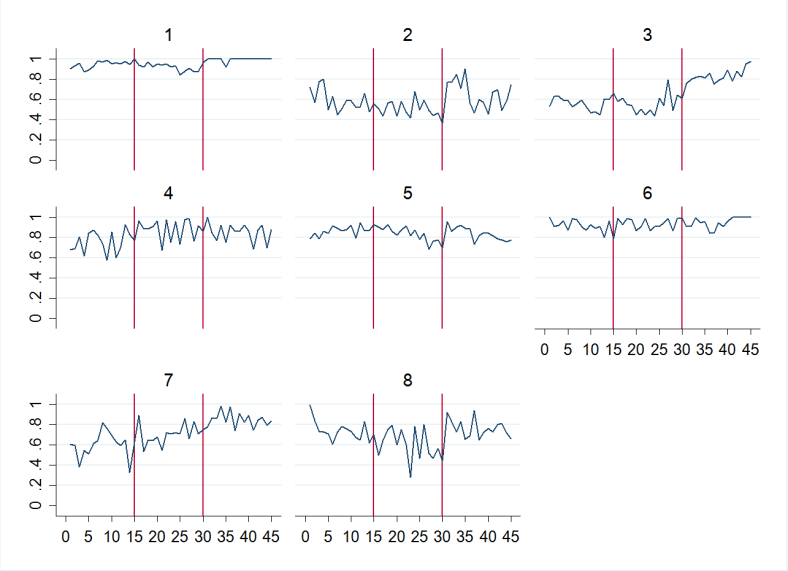


Figure D.4: Behavior on group-level in GG

Notes: This figure shows the average observed compliance rates for each group. The vertical lines mark period 15 (after which in BG and GB the embezzlement level changed) and period 30 after which the poll took place.

Table D.1: Group-level behavior

Group	Part 1	Trend 1	Part 2	Trend 2	Part 2 vs Part 1	Part 3	Trend 3	Part 3 vs Part 2	vote
BG-1	0.40	-	0.30	-	0	0.89	0	+	9
BG-2	0.34	-	0.53	0	0	0.91	+	+	8
BG-3	0.29	-	0.24	-	0	0.55	-	+	6
BG-4	0.54	0	0.52	0	0	0.94	0	+	9
BG-5	0.85	0	0.92	0	+	0.89	0	0	10
BG-6	0.35	-	0.18	0	-	0.58	0	+	9
BG-7	0.50	0	0.37	-	-	0.22	-	-	6
BG-8	0.42	0	0.52	0	0	0.76	0	+	8
BG	*0.46*	-	*0.45*	-	*0*	*0.72*	-	*+*	*8.1*
GB-1	0.75	0	0.54	+	-	0.80	0	+	7
GB-2	0.74	+	0.62	0	0	0.76	0	+	7
GB-3	0.67	+	0.54	0	0	0.44	-	0	7
GB-4	0.82	0	0.52	+	-	0.67	-	0	8
GB-5	0.53	-	0.19	0	-	0.20	0	0	7
GB-6	0.86	0	0.34	-	-	0.29	-	0	6
GB-7	0.75	-	0.65	+	-	0.79	0	+	8
GB-8	0.90	0	0.86	+	0	0.97	0	+	8
GB	*0.75*	*0*	*0.53*	*0*	-	*0.62*	-	*0*	*7.3*
BB-1	0.60	0	0.53	0	0	0.63	-	0	8
BB-2	0.46	0	0.40	0	0	0.42	-	0	9
BB-3	0.46	-	0.21	-	-	0.40	-	+	7
BB-4	0.67	0	0.53	0	-	0.39	0	-	4
BB-5	0.30	-	0.22	0	0	0.24	-	0	7
BB-6	0.35	0	0.32	0	0	0.35	-	0	7
BB-7	0.57	-	0.44	0	0	0.56	0	+	7
BB-8	0.47	0	0.44	0	0	0.55	0	0	5
BB	*0.49*	-	*0.39*	-	-	*0.44*	-	*0*	*6.8*
GG-1	0.94	+	0.92	0	0	0.99	0	0	10
GG-2	0.64	0	0.53	0	0	0.70	0	+	8
GG-3	0.63	0	0.62	0	0	0.85	+	+	7
GG-4	0.78	0	0.88	0	0	0.87	0	0	8
GG-5	0.87	+	0.83	-	0	0.84	-	0	8
GG-6	0.91	0	0.94	0	0	0.95	+	0	8
GG-7	0.57	0	0.70	0	0	0.83	0	+	9
GG-8	0.75	0	0.61	0	-	0.78	0	+	10
GG	*0.76*	*0*	*0.75*	*0*	*0*	*0.85*	*0*	*+*	*8.5*

Notes: This table shows group-level behavior across periods and its changes. The first column identifies the group with the letters denoting the treatment and the numbers corresponding to the groups in Figure D.1-D.4. The columns 'Part1' ('Part2'; 'Part3') present the average compliance level in periods 1-15 (16-30; 31-45). A '+'/ '-'/ '0' in column 'Trend1' ('Trend2'; 'Trend3') denotes that the Pearson correlation between period and compliance level is significantly positive/significantly negative/not significant at the 5%-level across periods 1-15 (16-30; 31-45). A '+'/ '-'/ '0' in column 'Part2 vs Part1' denotes that a signed-rank test ($N = 10$ on group-level and $N = 8$ on treatment level) finds compliance in periods 16-30 to be significantly higher/significantly lower/not significantly different at the 5%-level from compliance in periods 1-15. A '+'/ '-'/ '0' in column 'Part3 vs Part2' denotes that a signed-rank test ($N = 10$ on group-level and $N = 8$ on treatment level) finds compliance in periods 31-45 to be significantly higher/significantly lower/not significantly different at the 5%-level from compliance in periods 16-30. The column 'vote' denotes the number of votes agreeing with the statement that everyone should be paying their taxes.

Appendix E Instructions and screenshots of the experiment

In this appendix, we present an English translation of the German instructions we provided to the subjects. The instructions also include screenshots of the most important screens the subjects saw during the experiment.

E.1 Instructions

Welcome to our experiment![11]

During our experiment you are not allowed to use electronic devices or to communicate with other subjects. Please do not use other programs and functions than those of the experiment. Please do not communicate with other participants. In case, you have a question, please raise your hand. We will come to you and answer your question in private. Please do not ask questions publicly. In case the question is relevant for all participants, we will repeat your question and answer it publicly. In case you do not stick to this rule, we have to exclude you from the participation in the experiment and any experimental earnings.

If you read the following instructions carefully, you can, depending on your and the other subjects' decisions, earn a considerable amount of money. It is therefore important you take your time to understand the instructions.
During the experiment we shall not speak of euros, but of Experimental Dollars (E$). Your entire earnings will be calculated in E$. We will later inform you precisely how E$ will be converted to Euros. In addition to your experimental earnings, you will receive a show-up fee of **5 Euro**.

At the beginning of the experiment, participants will be randomly divided into **groups of 10**. You will therefore be in a group with 9 other participants. **The composition of each group will remain the same throughout the experiment.**

The experiment is divided into **three parts**. We will begin by explaining the first part of the experiment. Once the first part is finished, you will receive detailed information about the second part of the experiment. After the second part is finished, you will receive detailed information about the third part.

[11] These are the instructions for the BG treatment in which groups start with bad quality institutions. The instructions for GB, BB, and GG are analogous and available upon request.

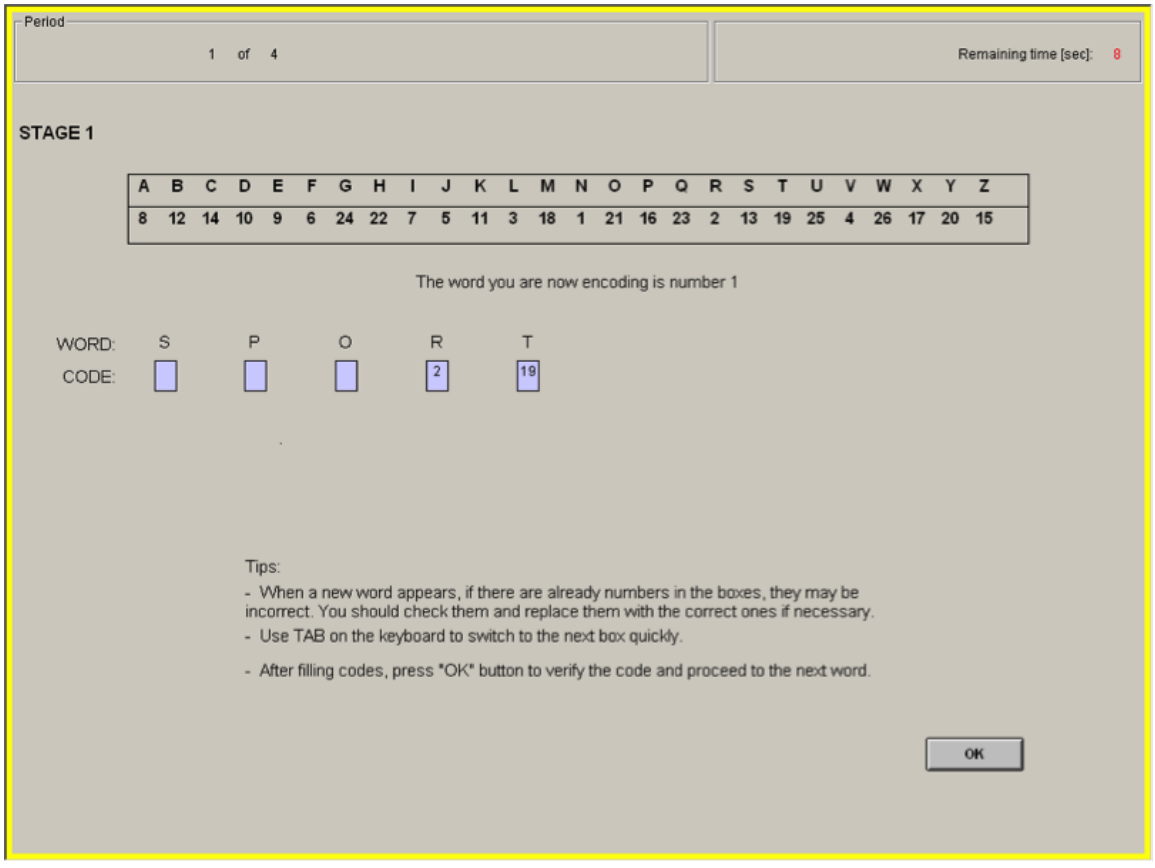
Instructions – Part 1

In the first part, all participants will perform the same task. **You will be presented with a number of words and your task is to encode these words by substituting the letters of the alphabet with numbers using a table on your computer screen.** The task decision screen is seen in Figure 1.

Example: You are given the word SPORT. The letters in Table 1 show that S=13, P=16, O=21, R = 2, and T=19. (Tip: This word is also the first one you have to encode.)

Once you encode a word correctly, the computer will prompt you with another word to encode. Once you encode that word, you will be given another word and so on. **This process will continue for 7 minutes** (420 seconds). Everyone will be given the same table and the same words to encode in the same sequence.

Figure E.1: Encryption Task



The purpose of the task

The task is not associated with a direct payoff, i.e. you do not receive any direct earnings from the first part of the experiment. **Nonetheless, it is still worthwhile to encode as many words as possible**: With a high likelihood, you earnings in the second and

the third part of the experiment will depend on how many words you encoded in the first part. In the following, we will provide the details.

First of all, there will be **two types of players** in the second and third part of the experiment: **One** player A and **nine** players B. Whether you are a player A or B will be **randomly** determined by the computer in this part of the experiment, and you will be informed about the computer's decision before you perform the task described above.

In case you are one of the nine **B players**, the relative performance of each individual in this part will influence your earnings in the next parts of the experiment. This will happen in the following way.

In the second and third part, each participant will receive a certain income at the start of each round: Four of the nine B players will receive an income of E$60 in each period whereas the other five will receive an income of E$40. At the end of the Part 1, the computer will rank all nine B players based on the number of words they encoded.

The four players that rank first, second, third, and fourth will receive an income of E$60. That is, the four B players with the highest number of encoded words will receive the higher income. **The five players that rank fifth, sixth, seventh, eighth, and ninth among the B players will receive an income E$40**. If two or more participants tie, the computer will determine their rank randomly.

In case you are the only **A player**, you will receive an income of E$60 in the next two parts of the experiment independent of your performance in the task. You still have the opportunity to take part in the task but your performance will not have any effect on your income.

Once Part 1 is over, B players will be informed as to whether they receive an income of E$40 or E$60 and a new set of instructions will be distributed. You will not be informed about the precise number of words encoded by each group member until the end of the experiment.

Instructions – Part 2

Summary

This part of the experiment consists of 30 rounds. At the beginning of each round, you will receive an income and you have to decide how much income you would like to declare to the (experimental) tax authority. While taxes have to be paid on declared income, undeclared income naturally remains free of tax. Every group member can, however, be audited with a certain probability. In case you are audited and have not fully declared your income, you have to pay a fine. Notably, the total amount of taxes collected from the

group will be redistributed to its members although not every player will receive the same amount.

Four random rounds will be chosen by the computer for your final payoff. For this purpose, the $E you have earned will be converted to Euros at the rate of **E$10 = 0.50 EUR**. Moreover, you will get an additional payment from part 3 of the experiment.

Detailed explanation

There will be 30 rounds in Part 2. As outlined in Part 1, there will be two types of players: one player A and nine players B. Player A as well as the four best-performing players B in Part 1 will receive an income of E$60 in each round. The remaining five players B will receive an income of E$40 in each round. Each of the 30 rounds consists of **two stages**.

Stage 1: Income declaration

In the first stage, everyone declares their income. This is **the only decision** you will have to make in each round. Your **declared income** may be smaller than your actual income but not larger. If your income is E$60, you can choose as a declared income any integer between 0 and 60. If your income is E$40, you can choose as a declared income any integer between 0 and 40.

Declared income is subject to an income tax. The **tax rate** is 40%, i.e. for every E$1 you declare in income you have to pay E$0.40 taxes. Undeclared income is not taxed. Thus, in case you declare an income of E$0, you will pay E$0 taxes. In case you declare an income of E$40, you will pay taxes of E$16 (=0.40*40). If you declare an income of E$60, you will pay taxes of E$24 (=0.40*60).

Below we present some numerical examples to help you understand better the experiment.

All examples in the text are used for illustration purposes only and should not be considered as guides for how to behave in the experiment.

Examples:

*(a) In case you have an income of $E40 and declare half your income, E$20, you will have to pay E$8 (=E$20*0.40) in taxes since the tax rate is 40%. In this case, your undeclared income is E$20 (=E$40 – E$20), the other half of your actual income. This implies that you evade taxes of E$8.*

*(b) Similarly, in case you have an income of E$60 and again declare half of your income, E$30, you will have to pay E$12 (=E$30*0.40) in taxes. In this case, your undeclared income is also E$30 (=$E60-E$30), implying that you also evade taxes of E$12.*

Stage 2: Tax audit

Probability of an audit

In every round, your tax declaration can be audited with a certain probability. Notably, a higher amount of overall collected taxes is associated with a higher **individual audit probability**. The precise relation is given by the following formula.

$$audit\ probability = \frac{taxes\ collected}{3.6}\%$$

where taxes collected refer to the sum of taxes that all group members have paid. **In case no one pays taxes, no taxes are collected** and, hence, **the audit probability is 0%**. Thus, in this case, your tax declaration will never be audited. In case, however, some (or all) members of the group pay their taxes, the audit probability will be larger than 0%. **The maximum probability with which you will be audited is 56%** (=(200/3.6)%). This probability, however, only occurs in case **everyone *fully* declares their income.** Then, E$200 (=5*16+5*24) are collected in overall taxes since the five participants with an income of E$40 pay E$16 in taxes and the five participants with an income of E$60 pay E$24 in taxes.

Example (continued):

*Assume all group members declare half of their income. The five participants with an income of E$40 pay E$8 in taxes individually, and all together E$40 (=5*8). The five participants with an income of E$60 pay E$12 in taxes individually, hence, E$60 (=5*12) all together. Thus, the overall amount of taxes collected is E$100 (=40+60), and the audit probability is 28% (=(100/3.6)%).*

Fine

What happens in case you are audited? If you did not declare all of your income in Stage 1, you will have to pay a fine. The fine will be equal to the amount of taxes you evaded multiplied by a factor of 2. We will refer to this factor as the fine multiplier. Of course, in case you fully paid your taxes in Stage 1, you will not have to pay any fine if you are audited.

Examples (continued):

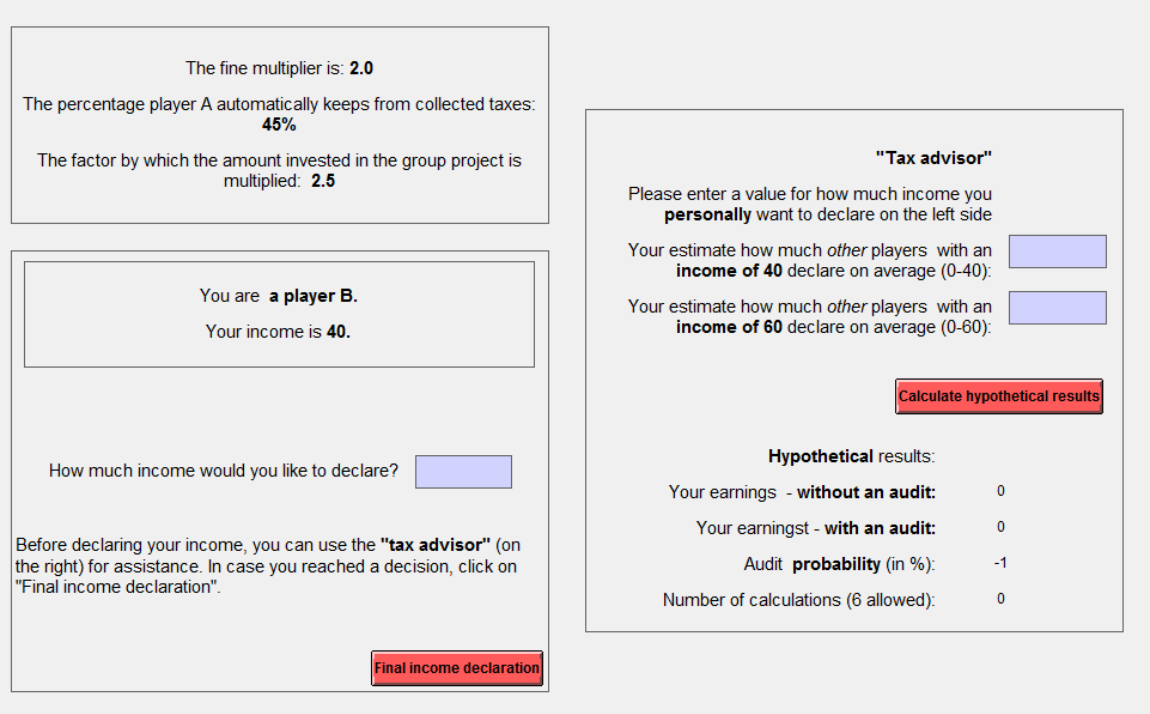
*(a) Assume that your undeclared income is E$20 and you, thus, evaded E$8. In case you are audited, you will have to pay a fine of E$16 (=2*E$8). If you are not audited, you do not have to pay a fine.*

*(b) Assume that your undeclared income is E$30 and you, thus evaded E$12. In case of an audit, you have to pay a fine of E$24 (=2*E$12). If you are not audited, you do not have to pay a fine.*

The calculation of the audit probability as well as the calculation of your earnings is a bit cumbersome. For this reason, you will be allowed to consult a "**tax advisor**" in every round before you declare your income. The tax advisor is a calculator that will inform

you about the potential audit probability and your earnings. After these instructions, we will more precisely instruct you how to use the calculator. The tax advisor can be seen in the right-hand side of Figure 2. On the left-hand side, you see information about you, and can use the box to declare your income.

Figure E.2: Tax declaration screen



Earnings at the end of each round

What are your earnings at the end of each round? This obviously first of all depends on whether you receive an *income* of E$40 or E$60. Moreover, you have to pay taxes on *declared income*. Undeclared income naturally remains free of tax. In case you are audited, you have to pay a *fine* in case you did not declare all of your income. Notably, the total amount of taxes collected is not wasted. At least a part of it will be invested in a group projected and redistributed to the group members although not every player will receive the same amount. We call this redistribution tax benefit and explain it in more detail in the next section.

Round Earnings = Income - (Declared Income * 40%) – Fine * S + Tax benefit (where S=1 if the individual is audited and evaded taxes, and 0 otherwise)

Tax benefit

What is the magnitude of the tax benefit? First of all, it is important that player A and the B players receive a different tax benefit. This is because *player A automatically keeps a certain fraction of the total amount of collected taxes*. This fraction is **45%** of the total

taxes collected. Only the remaining 55% of the total amount of taxes collected from all participants will be invested in a group project. This amount will be increased by a **factor of 2.5**. The resulting amount will be equally distributed to all group members (including player A). As noted earlier taxes collected refer to the total amount of taxes that are paid by all group members. Notably, *fines* paid by group members do not increase this variable.

B players receive the following tax benefit:

$$tax\ benefit\ (player\ B) = \frac{[55\% * taxes\ collected * 2.5]}{10\ group\ members} = 0.1375 * taxes\ collected$$

Participants assigned the role of player A get the same benefit plus 45% of overall collected taxes they automatically keep.

$$tax\ benefit\ (player\ A) = 0.1375 * taxes\ collected + 0.45 * taxes\ collected$$

In case no one pays taxes, the tax benefit will obviously be zero. Thus, the tax benefit will be zero both for player A and the B players. In this case, everyone (*also player A*) will just earn his initial income, either **E$40** or **E$60**. Moreover, since the probability of an audit is zero (no taxes are collected), no fines have to be paid.

In case everyone fully declares their income, the overall amount of taxes collected is 200 (=5*16+5*24) and the tax benefit will be E$27.5 (=0.1375*200) for B players and E$117.50 (=0.45*200+27.5) for player A. This implies that those with an income of E$40 pay E$16 in taxes and get a tax benefit of E$27.50, resulting in round earnings of **E$51.50** for that round. Those with an income of E$60 pay E$24 in taxes and get a tax benefit of E$27.50, resulting in round earnings of **E$63.50**. Only player A gets a higher benefit of E$117.50, resulting in round earnings of **$E153.50**. These results hold for both cases, with and without an audit since no fines have to be paid when everyone pays their taxes.

Example (continued):

*Let us assume again that each of the 10 players declares half of their income. As described in the example above: those with income E$40 pay E$8 in taxes and those with income E$60 pay E$12 in taxes. The overall amount of taxes collected then is 100 (=5*8+5*12) for all 10 players, leading to a tax benefit of E$13.75 (=0.1375*100) for players B and E$ 58.75 (=0.45*100+13.75) for player A.*

Hence, those with an income of E$40 pay E$8 in taxes and get a benefit of E$13.75, resulting in earnings of E$45.75. Those with an income of E$60 pay E$12 in taxes and get a tax benefit of E$13.75, resulting in earnings of $E61.75. Only player A gets a higher

*benefit of E$58.75, resulting in round earnings of E$106.75. Note that these round earnings are only valid when no audit occurs. With a 28% likelihood, an audit occurs and a fine of either E$16 (=8*2) for participants with an income of E$40 or E$24 (=12*2) for participants with an income of E$60 has to subtracted from these incomes.*

Note that, the more of their income participants declare, the higher the overall earnings of the whole group will be. *Crucially, different players benefit from high tax compliance in a different way.* B players with an income of E$60 benefit less strongly than those with an income of E$40. All B players get the same tax benefit but players with a higher income have to pay higher taxes (at least in case they declare their income). *Notably, player A has the biggest benefit from collected taxes.* A higher amount of taxes collected is associated with a rising inequality in the tax benefits of player A and the B players. You are given a sheet that summarizes the examples provided in this text.

As can be seen in Figure 3, at the end of the second stage, you will be provided with the following information as feedback for the current round: your *declared income*, your *undeclared income*, whether you were audited in this round or not, and whether you were fined. Some feedback however arrives with a one-period delay. In particular, you will only be informed about how much taxes were collected in total, your tax benefit, your earnings, the average earnings of B players, and player A's earnings for the *previous* round.

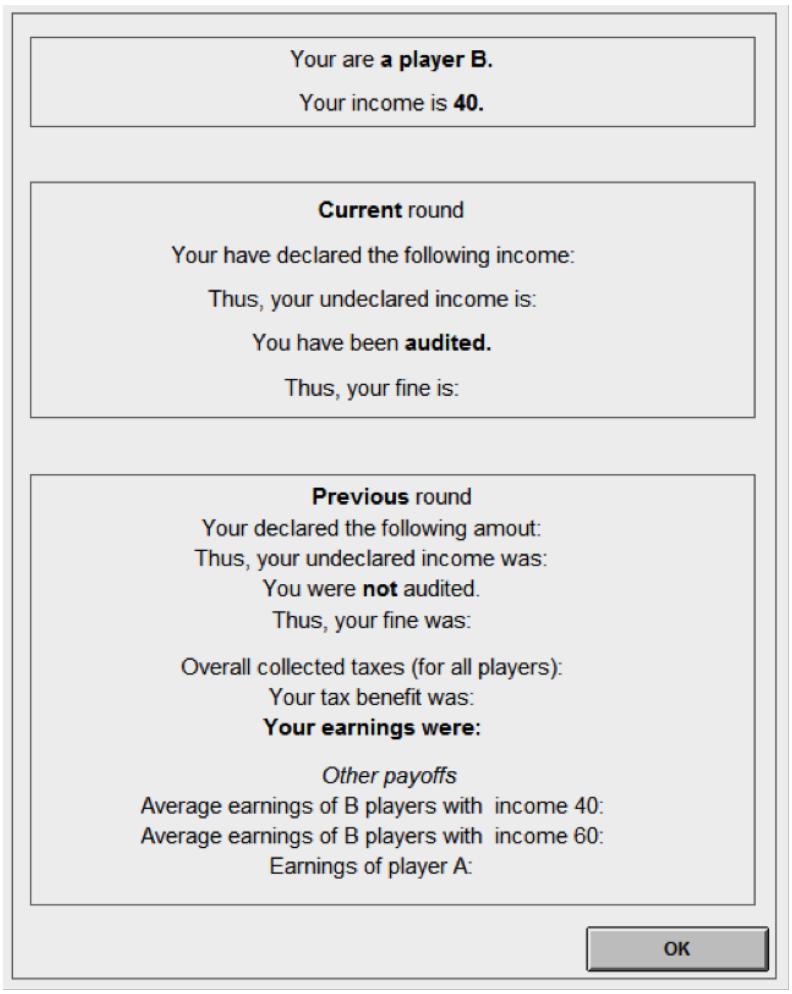
Final remarks

After one round is finished, we will directly go to the next round until all 30 rounds are finished. Importantly, in one of these rounds, one or more of the parameters outlined in these instructions might be changed. Notably, this does not need to happen. Figure 2 shows the three parameters that may potentially change (the fine multiplier, the percentage player A automatically keeps from the overall collected taxes and the factor by which the amount invested in the group project is multiplied). In case, any of the parameter changes, it will appear in red on the declaration screen shown in Figure 3. Moreover, you will be given a new sheet that summarizes the examples provided in these instructions under the changed parameter. Note that, after a potential change has occurred, no further changes will be made until the end of this part of the experiment.

Summary

- Depending on your performance in Part 1, you will receive an income of either **40** or **60** in every round of Part 2. You have to decide how much of this income you would like to **declare**. While declared income is taxed at a rate of **40%**, no taxes are paid on undeclared income.
- Crucially, however, your tax declaration will be **audited** with a certain probability. This **probability** increases with the amount of overall collected taxes. It ranges

Figure E.3: Tax declaration screen



between **0%** and **56%** (no one pays taxes vs. everyone fully pays their taxes). In case an audit occurs and you have not declared your full income, you will have to pay a **fine** that is equal to the amount of taxes you evaded **multiplied by 2**.

- Collected taxes will not be wasted and you will get a **tax benefit**. Part of the collected taxes will be invested in a group project such that the invested amount will be **multiplied by 2.5** and equally redistributed to all group members. Importantly, however, **player A automatically keeps 45%** of total collected taxes and only the remaining 55% will be invested in the group project.

- In case, **no one pays taxes**, the tax benefit is E$0 and all players just earn their initial incomes, either **E$40** or **E$60**. Crucially, in this case player A also earns **E$60**. In case, **everyone fully declares their income**, the tax benefit is E$27.5 for B players and B players with an initial income of E$40, hence, earn **E$51.5** while B players with an initial income of E$60 earn **E$63.5**. Crucially, in this case player A earns **E$153.5** since his/her tax benefit is E$117.50.

Instructions – Part 3

Part 3 of the experiment consists of 15 rounds. The group composition, individual roles (i.e., player A, B) and income remain the same as in Part 2. The setting is also the same as in Part 2. The only difference is that there will be a poll at the beginning of the third part. You have to decide whether you agree or disagree with the following statement:

"Higher tax payments increase group earnings. Thus, let's FULLY pay our taxes and do not evade them!"

After all group members have made a decision whether they agree or disagree with the statement, you will be informed about how many group members agreed with the statement. (This number includes yourself in case you agreed with the statement.)

Similar to Part 2, two rounds will be randomly chosen and the round earnings of these two rounds will be converted to Euros at the rate of **E$10 = 0.50 EUR**, as before.

E.2 Summary sheet for low embezzlement

Summary of the examples from the text

	Everyone declares …					
	… no income		*… half their income*		*… all their income*	
	Income **40**	Income **60**	Income **40**	Income **60**	Income **40**	Income **60**
1 - Income declaration						
Declared Income	0	0	20	30	40	60
Tax (tax rate: 40%)	0	0	8	12	16	24
Undeclared income	40	60	20	30	0	0
Evaded taxes	16	24	8	12	0	0
2 - Audit						
Tax collected	0		100 (=5*8+5*12)		200 (=5*16+5*24)	
Audit probability = (taxes collected/3.6)%	0%		28%		56%	
Fine (=2* evaded taxes)	/	/	16	24	0	0
3 - Earnings						
tax benefit (player B)	0		23.75		47.5	
tax benefit (player A)	0		28.75		57.5	
Earnings1 (without audit) = Income - taxes + tax benefit						
Player B Earnings1 (without audit)	**40**	**60**	55.75	71.75	**71.5**	**83.5**
Player A Earnings1 (with audit)		**60**		76.75		**93.5**
Earnings2 (with audit) = Earnings1 - Fine						
Player B Earnings2 (with audit)			39.75	47.75	**71.5**	**83.5**
Player A Earnings2 (with audit)				52.75		**93.5**

E.3 Summary sheet for high embezzlement

Summary of the examples from the text

	Everyone declares …					
	… no income		*… half their income*		*… all their income*	
	Income **40**	Income **60**	Income **40**	Income **60**	Income **40**	Income **60**
1 - Income declaration						
Declared Income	0	0	20	30	40	60
Tax (tax rate: 40%)	0	0	8	12	16	24
Undeclared income	40	60	20	30	0	0
Evaded taxes	16	24	8	12	0	0
2 - Audit						
Tax collected	0		100 (=5*8+5*12)		200 (=5*16+5*24)	
Audit probability = (taxes collected/3.6)%	0%		28%		56%	
Fine (=2* evaded taxes)	/	/	16	24	0	0
3 - Earnings						
tax benefit (player B)	0		13.75		27.5	
tax benefit (player A)	0		58.75		117.5	
Earnings1 (without audit) = Income - taxes + tax benefit						
Player B Earnings1 (without audit)	**40**	**60**	45.75	61.75	**51.5**	**63.5**
Player A Earnings1 (with audit)		**60**		106.75		**153.5**
Earnings2 (with audit) = Earnings1 - Fine						
Player B Earnings2 (with audit)			29.75	37.75	**51.5**	**63.5**
Player A Earnings2 (with audit)				82.75		**153.5**

Appendix F Bibliography

Bardsley, N. and P. G. Moffat (2007). The experimetrics of public goods: Inferring motivations from contributions. *Theory and Decision 62*(2), 161–193.

Cappelen, A. W., A. D. Hole, E. O. Sorensen, and B. Tungodden (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review 97*(3), 818–827.

Harsanyi, J. C. and R. Selten (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Kim, Y. (1996). Equilibrium selection in n-person coordination games. *Games and Economic Behavior 15*(2), 203–227.

Moffat, P. G. (2016). *Experimetrics: Econometrics for Experimental Economics*. Palgrave Macmillian.

Peski, M. (2010). Generalized risk-dominance and asymmetric dynamics. *Journal of Economic Theory 145*(1), 216–248.