Pre-Analysis Plan
Accountability and taxation

Ingrid Hoem Sjursen*[1]

[1]The Choice Lab & Department of Economics, Norwegian School of Economics

May 24, 2017

*Ingrid.Sjursen@nhh.no

1 Introduction

Corruption and mismanagement of public revenue is a widespread and serious obstacle to social and economic development in many poor countries (Lederman et al., 2005; UNDP, 2008; Ferraz and Finan, 2011). The so called *Rentier State Hypothesis* states that when governments are financed through taxation, as opposed to through windfalls such as natural resources revenues and aid, citizens demand more accountability. This in turn makes politicians more responsible and responsive to citizens' demands (Mahdavy, 1970; Ross, 2001; Karl, 2007; Bräutigam et al., 2008; Paler, 2013; Borge et al., 2015). Thus, the Rentier State Hypothesis suggests that generating public revenue through taxing citizens is a promising strategy for promoting political accountability and development.

Accountability can be defined as "(...) an obligation or willingness to accept responsibility to account for one's actions" (Merriam-Webster, 2017). According to Przeworski and Stokes (1999), "Governments are 'accountable' if voters can discern whether governments are acting in their interest and sanction them appropriately, so that those incumbents who act in the best interest of citizens win reelections and those who do not lose them". Demand for accountability can then be defined as voters' willingness to gather information about the government's actions and to enforce sanctions if these actions are not aligned with their own interests. Examples of such actions are voting in elections, participating in demonstrations for political causes, organizing advocacy campaigns, and filing lawsuits against the government. The present project will only investigate the latter aspect, willingness to enforce a sanction.

The political economy literature frequently assumes the Rentier State Hypothesis to be true, but identifying a causal effect of taxation on accountability using observational data is a challenge. In this project, I will study how taxation affects demand for accountability in a controlled experimental environment. More specifically, I will investigate whether and why individuals demand more accountability in the investment of a group endowment when this endowment is financed through taxation compared to when it is a windfall.

The project aims to answer the following research questions:

1. *Does taxation cause higher demand for accountability?*
2. *What are the underlying mechanisms?*

To address these research questions empirically, I implement a 2X2 design where the experimental treatments are intended to capture two major differences between tax and windfall revenues. First, citizens work to earn the income they pay tax from and tax revenues are therefore "Hard Earned", whereas windfall revenues are not related to citizen's effort at all ("Hard Earned" vs. "Windfall" treatments). Second, tax revenues have been in the citizens' "pockets" before they are collected, whereas windfall revenues accrue directly to the government ("Tax" vs. "Non-Tax" treatments).

In terms of the underlying mechanisms I propose that willingness to enforce sanctions (demand for accountability) when government revenues are misused can be driven by two factors. First, misuse of government revenue may cause citizens to receive less public services than what they perceive as fair.[1] Second, misuse may inflict perceived losses on citizens. The basic idea behind the Hard Earned vs. Windfall treatments is that working hard to generate the revenue the government spends makes increases the citizen utility from enforcing sanctions when the use of this revenue deviates from the fairness norm. This makes him or her more likely to enforce sanctions if receiving less public services than what is fair, compared to a citizen who has not worked hard to generate the revenue the government spends. I will refer to this as the "Hard Earned" mechanism. The Tax vs. Non-Tax treatments are based on the work of Kahneman and

[1] In this setting misuse of government revenue can be both socially inefficient investments and embezzlement/corruption.

Tversky (1979). They show that individuals are *loss averse*, meaning that they dislike losses more than they like equally sized gains. Whether misuse of government revenue is perceived as a loss or a reduction in gain is dependent on the individual's reference point. The idea is that in the Tax treatment, the tax paid gives the citizen a reference point for his or her expected payoff. If the citizen has gets a payoff higher that this reference point, she is in the gain domain, if she receives less, she is in the loss domain. I hypothesize citizens to enforce sanctions for decisions that puts them in the loss domain, and not to not enforce sanctions for decisions that puts them in the gain domain. I will refer to this as the "Loss Aversion" mechanism.

To address my research questions, I will conduct an experiment where participants are randomly assigned to the role of *citizen* or *leader*. The leader's task is to decide how much of a group endowment to invest in a common pool and how much to keep for him or herself. The amount invested in the common pool is multiplied by 1.5 and equally distributed between the citizen and the leader. The citizen decides whether or not to costly punish the investment decision of the leader. There are no strategic reasons for punishment and any punishment observed must be motivated solely by non-monetary considerations, such as reciprocity or distributional concerns. I use willingness to costly punish the leader as a proxy for a participant's demand for accountability and measure this as the citizen's "punishment threshold", the lowest share invested for which he or she does not punish the leader (Martin, 2016).

In the Windfall & Non-Tax treatment, revenues are unrelated to the citizens' effort/work and are not collected from their pockets, and this can be thought of as a stylization of the situation in a rentier state where government revenue is windfall. In the Hard Earned & Tax treatment, citizens both work to earn income and pay tax in arrears, and this can be thought of as a stylization of a state financed through a tax system. The set up allows me to answer the first research question, which intends to test the Rentier State Hypothesis, by comparing the average punishment threshold between the Windfall & Non-Tax and the Hard Earned & Tax treatment.

To answer the second research question, I first investigate the Hard Earned mechanism by comparing the punishment threshold when the group endowment is hard earned and collected through taxation to when is not hard earned, but collected through taxation. Second, I investigate the Loss Aversion mechanism by comparing the punishment threshold when the group endowment is hard earned and collected through tax to when it is hard earned but money goes directly to the group endowment. The prediction is that when the group endowment is hard earned and collected through tax (vs. windfall), demand for accountability is higher.

The project links to several literatures, but is most closely related to the experimental studies of de la Cuesta et al. (2017), Martin (2016) and Paler (2013) who studies the Loss Aversion mechanism and Bosman et al. (2005) who investigates the Hard Earned mechanism. de la Cuesta et al. (2017) and Martin (2016) use a dictator game where the receiver has the opportunity to (costly) punish the dictator's distributive choice in lab experiments in Ghana and Uganda, respectively. Both studies find evidence to support the hypothesis that taxation increases demand for accountability through loss aversion: when a group endowment is financed through taxation, the average punishment threshold is significantly higher than when it is windfall. de la Cuesta et al. (2017) additionally find the effect of taxation to be driven by a stronger feeling of ownership over tax than windfall revenue. Paler (2013) finds that framing government revenues as tax increases demand for information about the government's budget compared to framing it as oil and gas windfalls, but finds no effect on self-reported and observed political participation, in a field experiment in Indonesia. Bosman et al. (2005) use a so called "power-to-take" game that strongly resembles Martin (2016)'s design, but where punishment is continuous, its cost is proportional to the damage it does and the experimental treatment varies whether the citizen worked hard to earn his or her endowment, not whether the citizen pays tax. They find that citizens punish *less* when their endowment is hard earned compared to when it is windfall, which is contrary to my hypothesis. I suggest that this finding can be

explained by the high cost of punishment in the experiment.[2] The structure of the experimental game in the present study most closely resembles that of Martin (2016). I add experimental treatments to investigate the Hard Earned mechanism, which allows me to look at the interaction effect between the Hard Earned and Loss Aversion mechanism. To the best of my knowledge, the present project is the first to do this. I also add to Martin (2016)'s design by looking at punishment of investments in a common pool rather than a simpler dictator distributive decision, and by suggesting a test to separate the effect of taxation on the citizen's reference point from the effect on weight attached to the fairness norm. Furthermore, by looking at the Hard Earned mechanism in a setting where the cost of punishment is low as compared to Bosman et al. (2005), I shed light on the impact of cost of accountability on the relationship between taxation and demand for accountability.

The study also relates closely to the literature documenting a positive correlation between taxation and good governance (Ross, 2004; Brollo et al., 2013; Borge et al., 2015; Gadenne, 2017; Martínez, 2016). Because demand for accountability is likely to affect government behaviour (see for instance Dynes and Martin (2017)), shedding light on determinants of demand for accountability contributes to our understanding of how good governance can be promoted.

Further, the study adds to the literature on how source of income, i.e. whether it is hard earned earned or windfall, affects individual decision making. This literature shows that individual consumption and investment decisions are affected by both the source of income and how this income is framed/labeled, but has to my knowledge not investigated effects on willingness to punish contributions to a public good. See Hvide and Lee (2016) for an overview of this literature.

Finally, the project relates to the literature on social norms and human cooperation. Using public good and ultimatum games, this literature has shown that individuals are willing to punish norm violators, even when such punishment is costly and yields no material gains (Fehr and Gächter, 2000, 2002; Henrich et al., 2006; Henrich and Henrich, 2007). Individuals are even willing to costly punish norm violaters when they are third parties, i.e. when they are not directly affected by the norm violation. I contribute to this literature by investigating whether and why willingness to costly punish an investment decision depends on the source of income.

This pre-analysis plan describes experimental design, subject pools, hypotheses to be tested and the specifications to be employed in the empirical analysis of the research project.

2 Research strategy

I will collect experimental data using the infrastructure of the international online labor market, Amazon Mechanical Turk (mTurk henceforth) in May 2017. I plan to recruit a total of 2000 respondents, equally distributed between five treatments. The respondents will be randomized to one of two roles, resulting in 200 participants in each role in each treatment. A sample size of 200 observations per role per group allows me to identify treatment effects of 0.3 stdev at 5% significance level with power 0.86. See Appendix C for power calculations.

This pre-analysis plan was submitted to the AEA Social Science Registry prior to the data collection.

2.1 Recruitment

The participants will be recruited from mTurk by posting a so called "Human Intelligence Task" on the web site (see Appendix A). To ensure that all participants are sufficiently fluent in the English language, I will only allow individuals residing in the US to participate. Furthermore, to ensure high quality data, all participants must have a 98% or higher approval rate and have completed at least 500 HITs.

[2] See discussion in Subsection 3.3.

3 Experimental design

To answer my research questions, I will conduct a between-subject framed investment game with the opportunity to punish. This section describes the sequence of the game and the experimental treatments.

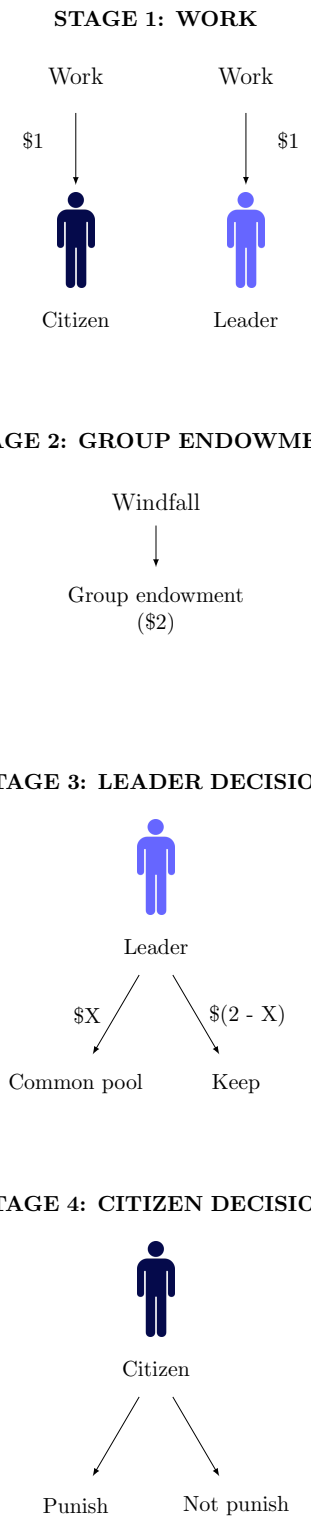
3.1 Sequence of the game

To describe the sequence of the game, I present the structure of the base treatment, Windfall & Non-Tax. The remaining experimental treatments are described in detail in Subsection 3.3.

First, individuals are randomly assigned to a treatment, to a pair and to the role as a citizen or the role as a leader. In each pair, the citizen and the leader independently decide whether they would like to perform a real effort task to earn money. Before making their decision, they are told about the payment scheme, but not told about the opportunity the citizen gets to punish the leader. In the analysis of the data, I am only interested in the citizen-leader pairs where both choose to work because I want to study the Hard Earned mechanism when both parties have contributed to the group endowment. In the following description and in the graphical illustration in Figure 1, I therefore focus on the citizen-worker pairs where both choose to work.

In the first stage of the experiment, the citizen and the leader complete the same five minute picture categorization assignment and receive $1 each. In the second stage, the group is assigned a group endowment of $2. The group receives the group endowment regardless of the citizen and leader's choice about performing the real effort task or not and is thus windfall revenue. In the third stage, the leader decides how much of the group endowment to invest in the common pool. The share invested is multiplied by a factor of 1.5 and equally divided between the citizen and the leader. The leader can invest any share of the group endowment he likes, in portions of $0.2, and gets the amount not invested for him or herself. Finally, in the fourth stage, the citizen decides whether to punish the leader's decision. To elicit willingness to punish, I use the strategy method: for every possible investment decision the leader can make, the citizen is asked whether he or she would like to take away $0.5 from the leader's payoff by paying $0.05. In line with the experimental literature on punishment, I describe the citizen's choice in neutral ("take away money") rather than loaded ("punish") language to avoid experimenter demand effects.

Figure 1: Sequence of events



3.2 Main outcome and payoffs

The experimental design provides the following main outcome of interest:

- *Demand for accountability:* lowest investment share for which the citizen does not punish.

I will refer to this is the citizen's punishment threshold. The higher a citizen's punishment threshold is, the higher is his or her demand for accountability.

The payoff for citizen c in pair i can be written as

$$y_{c,i} = 1 + \frac{3}{4} I_{l,i} - (p_{c,i} \times 0.05), \tag{1}$$

where $I_{l,i}$ is the share of the group endowment leader l in pair i invests in the common pool and $p_{c,i} \in \{0,1\}$ is an indicator variable taking the value of one if the citizen chooses to punish the leader's investment decision. Similarly, the payoff for leader l in pair i can be be expressed as:

$$y_{l,i} = 1 + (2 - I_{l,i}) + \frac{3}{4} I_{l,i} - (p_{c,i} \times 0.5). \tag{2}$$

3.3 Main experimental treatments

In the main treatments, I vary the how the group endowment is financed along two dimensions according to the mechanisms I want to test in a 2X2 design:

- **Hard Earned vs. Windfall:** group endowment is financed by citizen's and leader's work vs. group endowment is windfall revenue.
- **Tax vs. Non-Tax:** group endowment is financed by money that is first given to citizen and the leader, and then collected as tax vs. money goes directly to the group endowment.

Consequently, I have four main treatments: Windfall & Non-Tax (T1), Hard Earned & Non-Tax (T2), Windfall & Tax (T3) and Hard Earned & Tax (T4), as illustrated in Table 1.

Table 1: Experimental design, main treatments

	Windfall	**Hard Earned**
Non-Tax	$T1$	$T2$
Tax	$T3$	$T4$

The earnings of participants and how group endowments are financed in the four treatments are illustrated in Figure 2 and can be described as follows:

> **Windfall & Non-Tax (T1):** Citizen and leader earn 1 USD each. A windfall finances the 2 USD group endowment.
>
> **Hard Earned & Non-Tax (T2):** Citizen and leader earn 1 USD each. Their work additionally generates 1 USD each, financing the 2 USD group endowment.
>
> **Windfall & Tax (T3):** Citizen and leader earn 1 USD each. In addition they each receive a windfall of 1 USD. Their total earnings are taxed at 50%. The 1 USD tax collected from each finances the 2 USD group endowment.
>
> **Hard Earned & Tax (T4):** Citizen and leader earn 2 USD each. Their earnings are taxed at 50%. The 1 USD tax collected from each finances the 2 USD group endowment.

I theorize that there are two important non-monetary motivations for citizens to punish in the experiment: 1) the leader's decision leads to a payoff lower than the fairness norm 2) the leader's decision inflicts a loss on the citizen, i.e. she gets a payoff lower than the reference level. Since the citizen and the leader complete the same assignment in all treatments, it appears uncontroversial to assume that what citizens perceive to be the fair payoff is constant and equal to the payoff from the maximal possible leader investment in all treatments. This is

an important feature of the design and ensures that any differences in the average punishment threshold between treatments is not driven by differences in fairness norms.

To formalize these ideas, I introduce the following simple model for the utility of the citizen:

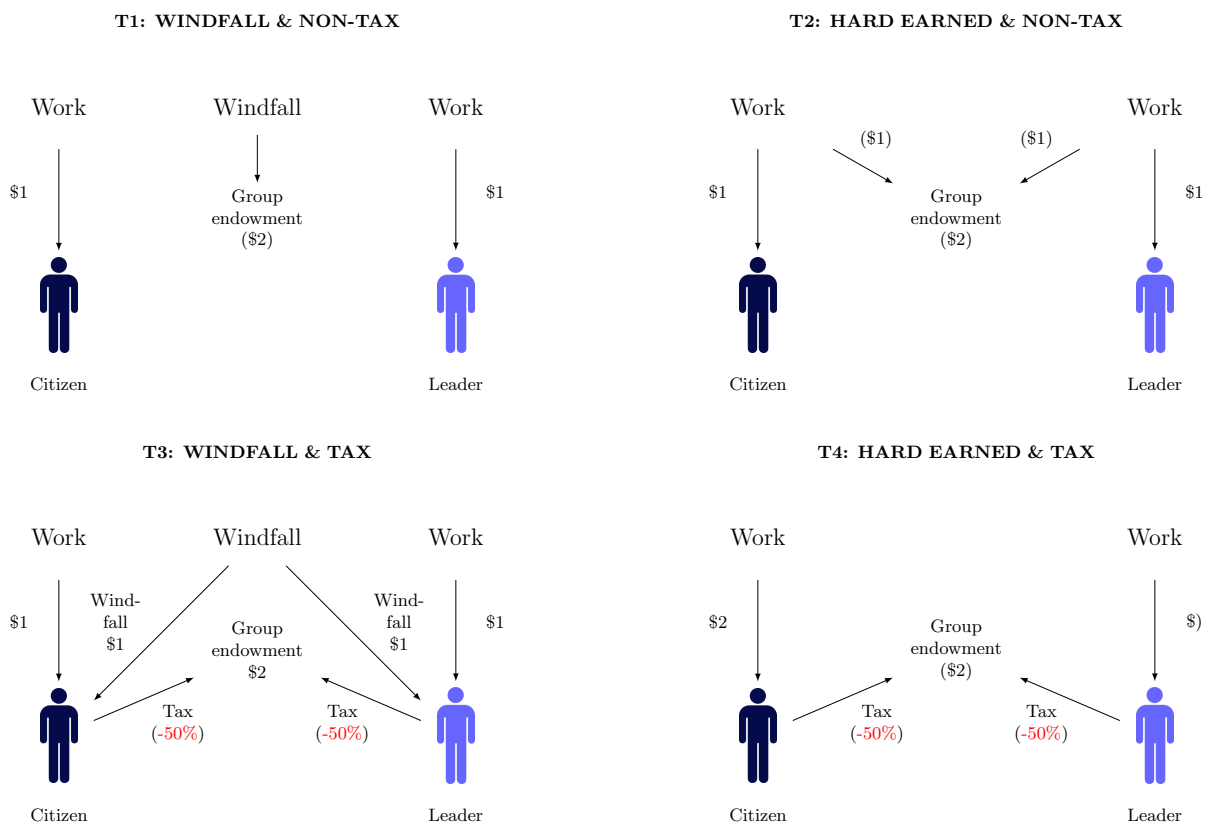
$$V_i = \beta_Y\big(Y - (p \times c)\big) + p\Big(\beta_F\big(min\{0, E(I) - F\}\big)^2 + \beta_R\big(min\{0, E(I) - R\}\big)^2\Big), \tag{3}$$

where Y is the sum of earnings from completing the assignment and the payoff from the leader's investment in the common pool, c is the cost of punishment, $p \in \{0,1\}$ is an indicator variable taking the value of one if the citizen punishes the leader, F is the fairness norm for payoff from the common pool and R is the citizen's reference payoff. $E(I)$ is the citizen's payoff from the common pool as function of the investment decision of the leader, I. If $E(I) < F$ or $E(I) < R$, the citizen derives a positive utility from punishment, and this utility is increasing in the investment's deviation from the fairness norm/reference point. β_Y, β_F and β_R are weights indicating how much utility the citizen derives from monetary pay-off, from punishing deviations from the fairness norm and from punishing payoffs lower than the reference, respectively.

The idea behind the Hard Earned treatments is that they affect how much utility the citizens derive from punishing deviations from the fairness norm for payoff. More specifically, citizens who have worked hard to generate the money financing the group endowment are hypothesised to have a higher utility from punishing deviations from the fairness norm for payoffs and be more likely to punish such deviations than citizens who receive the group endowment through a windfall. In terms of Equation 3, the hypothesis is that the Hard Earned treatments increases β_F compared to the Windfall treatments, giving rise to higher punishment thresholds.

The idea behind the Tax treatments is that they affect citizens reference payoff, making them expect a payoff from the common pool equal to the amount they pay in tax. Thus, in the Tax treatments the reference payoff from the common pool is the \$1 tax paid, but in the Non-Tax treatments, it is \$0. According to prospect theory (Kahneman and Tversky, 1979), citizens in the Tax Treatments should perceive all payoffs from the common pool that are lower than their \$1 reference point as a loss, whereas citizens in the Non-Tax treatments should perceive all positive payoffs as a reduction in gain. I hypothesise that citizens punish decisions that give them payoffs in the loss domain, but do not punish decisions that are in the gain domain. Because citizens in the Tax treatment have a higher reference payoff, they are in the loss domain for a larger share of the leader's investment decisions and should therefore also have a higher punishment threshold than citizens in the Non-Tax treatments. In terms of Equation 3, the hypothesis is that the Tax treatments increase R compared to the Non-Tax treatments, giving positive utility from punishment of a larger share of leader investment decisions and thus to higher punishment thresholds. I also hypothesize that the Tax treatments lead to an increase in the weight attached to the fairness norm, β^F, compared to the Non-Tax treatments. To separate the Tax treatments' effect on weight attached to fairness norm and on the reference payoff, I will compare citizens' punishment decisions in the interval between \$1 and \$1.5 in the Windfall & Tax treatment to the punishment decisions made in the same interval in the Windfall & No-Tax treatment. If I find that citizens in the Windfall & Tax treatment punish amounts above \$1.5 more frequently than citizens in the Windfall & Non-Tax treatments, this can be interpreted as evidence that the Tax treatment increases both the utility from punishing deviations from fairness norms and the utility from punishing payoffs lower than the reference payoff. Another possible interpretation is that citizens also punish in the gain domain.

Figure 2: Treatments: Generation and collection of group endowment



The literature on punishment typically highlights two non-monetary motivations for punishment in experiments: reciprocity (see for instance Falk and Fischbacher (2001) and Dufwenberg and Kirchsteiger (2004)) and distributional preferences (see for instance Fehr and Schmidt (1999)). The idea behind the present design is that the motivation for punishment is reciprocity, but I cannot rule out that any differences in outcomes between treatments are partly driven by distributional preferences.

In the discussion above, I have argued that the Hard Earned and Tax treatments increase the utility the citizens derive from punishment and that they will therefore increase their punishment thresholds. However, as Bosman et al. (2005) point out, the treatments may also have a counteracting effect through increasing the perceived cost of punishment. More specifically, when a citizen has worked hard to finance the group endowment, she will be less willing to give up her payoff from the common pool to punish, compared to when the group endowment is financed by a windfall. It feels worse to give up hard earned than windfall revenue to pay for punishment.[3] Similarly, the Tax treatments may increase the perceived cost of punishment because it pushes the citizen into the loss domain and giving up some of the payoff therefore hurts more than in the Windfall treatments where the citizen is in the gain domain.

[3]Note that in the particular case that the leader invests nothing in the common pool, the perceived cost of punishment should not be affected by the Hard Earned treatments. This is because, in this case the money the citizen use to finance the punishment is the 1 USD earnings that participants in all treatments have worked to earn. In contrast, as the leader's investment increases, parts of the citizen's budget comes from the common pool. How this part of the budget is perceived is assumed to depend on the Hard Earned treatments, and there may therefore be differences in perceived cost of punishment for citizens in different treatments for leader investments larger than zero.

The net effect of the treatments on the punishment threshold is likely dependent on the cost of punishment/accountability.[4] Because the cost of punishment is very low in my experiment, I expect the the former effect to dominate for both mechanisms, meaning that I expect to see higher punishment thresholds in the Hard Earned compared to Windfall, and in Tax compared to Non-Tax, treatments.

3.4 Robustness check of Hard Earned mechanism

In the main treatments, the duration and difficulty of the real effort task is the same in the Windfall and Hard Earned treatments, but the initial earnings are higher in T4 and T2 than in T1 and T3 ($2/$1+$1 vs. $1). Thus, the implied productivity of the individuals in T4 and T3 is higher. An alternative way to investigate the Hard Earned mechanism is to keep the payment per minute constant and increase the duration of the task in the Hard Earned compared to the Windfall treatment. I expect this to produce a similar, but potentially larger effect than in the main treatments. To investigate this, I run an extra treatment, T4b that is identical to T4, but where the duration of the task is 10 minutes instead of 5 minutes.

3.5 Instructions

This section provides the main instructions for citizens in T1-T4.

Figure 3: Introduction

The results from this experiment will be used in a research project at the Norwegian School of Economics. Participation in the study is completely voluntary.

You are free to decline to participate, or to end participation at any time and for any reason. You will receive a participation fee of $1 upon completion of the experiment. Depending on the actions you and others take, you may also earn additional money.

Your will remain anonymous throughout the experiment. None of the information collected can be traced back to individual participants. We will only use your Worker ID to assign payments and to check that you have not participated in this experiment before.

Please read the instructions carefully. The duration of the experiment is approximately 15-20 minutes.

If you have any questions regarding this experiment, please contact thechoicelab@nhh.no.

---

[4] As (Anderson and Putterman, 2006) show, demand for punishment in a voluntary contributions one-shot game experiment is decreasing in its price.

Figure 4: Citizen role assignment

You have been randomly assigned to a group consisting of you and another participant who is also an Amazon Mechanical Turk worker. You have been randomly assigned to the role as citizen and the person you have been grouped with has been randomly assigned to the role as leader in the experiment. You will not get to know anything about the person you are paired with, and this person will not get to know anything about you.

Figure 5: Treatment text, T1 Citizen

You now get the opportunity to earn $1 by completing a 5-minute picture categorization assignment.

The leader you have been grouped with got the opportunity to complete the same assignment as you. **He or she chose to do so, and earned $1.**

Regardless of your and the leader's choice to complete the assignment or not, your group receives $2 in group money.

It is the task of the leader to decide how much of the group money to invest in a common pool. The amount invested in the common pool will be multiplied by 1.5 and shared equally between you and the leader. The remaining group money will be given to the leader.

Your total payment will consist of two parts:

1. Earnings ($1 if you choose to complete the assignment, $0 otherwise)
2. Half of the common pool ($0 - $1.5, depending on the leader's decision)

The leader's total payment will consist of three parts:

1. Earnings ($1 because he or she chose to complete the assignment)
2. Half of the common pool ($0 - $1.5, depending on the leader's decision)
3. The group money not invested in the common pool ($0 - $2, depending on the leader's decision)

Figure 6: Treatment text, T2 Citizen

You now get the opportunity to earn $1 by completing a 5-minute picture categorization assignment. If you work, your group receives $1 in group money.

The leader you have been grouped with got the opportunity to complete the same assignment as you. **He or she chose to do so, and earned $1. Because of the leader's work, your group receives $1 in group money.**

It is the task of the leader to decide how much of the group money to invest in a common pool. The amount invested in the common pool will be multiplied by 1.5 and shared equally between you and the leader. The remaining group money will be given to the leader.

Your total payment will consist of two parts:

1. Earnings ($1 if you choose to complete the assignment, $0 otherwise)
2. Half of the common pool ($0 - $1.5, depending on the leader's decision and the group money available)

The leader's total payment will consist of three parts:

1. Earnings ($1 because he or she chose to complete the assignment)
2. Half of the common pool ($0 - $1.5, depending on the leader's decision and the group money available)
3. The group money not invested in the common pool ($0 - $2, depending on the leader's decision and the group money available)

Figure 7: Treatment text, T3 Citizen

You now get the opportunity to earn $1 by completing a 5-minute picture categorization assignment. Regardless of whether you chose to complete the assignment, you additionally receive $1.

The leader you have been grouped with got the opportunity to complete the same assignment as you. **He or she chose to do so, and earned $1. In addition, the leader received $1, independent of his or her choice to complete the assignment.**

Your and the leader's total earnings (earnings from assignment + additional dollar received) will be taxed at 50%. It is the task of the leader to decide how much of the tax revenues collected to invest in a common pool. The amount invested in the common pool will be multiplied by 1.5 and shared equally between you and the leader. The remaining tax revenues will be given to the leader.

Your total payment will consist of two parts:

1. Post-tax earnings ($1 if you choose to complete the assignment, $0 otherwise)
2. Half of the common pool ($0 - $1.5, depending on the leader's decision and the tax revenues available)

The leader's total payment will consist of three parts:

1. Post-tax earnings ($1 because he or she chose to complete the assignment
2. Half of the common pool ($0 - $1.5, depending on the leader's decision and the tax revenues available)
3. The tax revenues not invested in the common pool ($0 - $2, depending on the leader's decision and the tax revenues available)

Figure 8: Treatment text, T4 Citizen

You now get the opportunity to earn $2 by completing a 5-minute picture categorization assignment.

The leader you have been grouped with got the opportunity to complete the same assignment as you. **He or she chose to do so, and earned $2.**

Your and the leader's earnings will be taxed at 50%. It is the task of the leader to decide how much of the tax revenues collected to invest in a common pool. The amount invested in the common pool will be multiplied by 1.5 and shared equally between you and the leader. The remaining tax revenues will be given to the leader.

Your total payment will consist of two parts:

1. Post-tax earnings ($1 if you choose to complete the assignment, $0 otherwise)
2. Half of the common pool ($0 - $1.5, depending on the leader's decision and the tax revenues available)

The leader's total payment will consist of three parts:

1. Post-tax earnings ($1 because he or she chose to complete the assignment)
2. Half of the common pool ($0 - $1.5, depending on the leader's decision and the tax revenues available)
3. The tax revenues not invested in the common pool ($0 - $2, depending on the leader's decision and the tax revenues available)

3.6 Non-incentivized questions

After making the incentivized choices, the participants will answer a set of non-incentivized questions to shed more light on why the treatments have the effects they do, and whether there are any heterogeneities in effects.

3.6.1 Manipulation check

To check that the research design worked as intended, and that the treatments did not affect the fairness norm, I ask:

- *What would you consider to be the fair decision of the leader, when both of you completed the assignment?*
  - *$0, $0.2, $0.4, ..., $2*

If I find no differences in fairness norms across the four treatments, the experimental design worked as intended.

3.6.2 Mechanims

To shed light on motivation for punishment, I ask the following questions:

- *How upset would you feel if the leader invests less than what you think is fair in the common pool?*
  - *0 - Not upset at all, ..., 10 - Very upset*
- *Why did you make the decisions the way you did? Please explain briefly in the box below.*

I will use the first question both to shed light on why the treatments worked, by checking if participants in Hard Earned and Tax treatments are significantly less upset than participants in Windfall and Non-Tax treatments, and to investigate heterogeneity in treatment effects. I will use the second question for an explorative analysis of the motivation behind punishment.

3.6.3 Preferences

I also ask a number of questions to measure the participants' preferences. I will use these measures to investigate a) the correlation between preferences and the citizens' punishment thresholds, and b) heterogeneity in treatment effects between different types of participants.

The reason for posing the preference questions after the treatment is to avoid priming effects. This comes at the cost that the outcomes may be affected by treatments. However, this can be tested. In the pilot, I found no significant differences in preferences between T1 and T4.

I use the following (slightly modified) formulations provided by Falk et al. (2015) to measure negative and positive reciprocity and risk aversion:

Negative reciprocity

- *If I am treated very unjustly, I will take revenge at the first occasion, even if it is a cost to do so* (Strongly disagree (0) - strongly agree (10))
- *How willing are you to punish someone who treats you unfairly, even if there may be costs for you?* (Completely unwilling (0) - completely willing (10))
- *How willing are you to punish someone who treats others unfairly, even if there may be costs for you?*(Completely unwilling (0) - completely willing (10))

Positive reciprocity

- *When someone does me a favor, I am willing to return it* (Strongly disagree (0) – strongly agree (10))

Risk aversion

- *How willing or unwilling are you to take risks?* (Completely unwilling (0) - completely willing (10))

Loss aversion
To measure loss aversion, I formulate the following hypothetical scenario inspired by Fehr and Goette (2007):

- *If you could chose between the following hypothetical scenarios, which would you chose?*
  - *Lottery: win $80 with probability 1/2, lose $50 with probability 1/2*
  - *Receive $0 for sure*

Political engagement
To measure political engagement, I formulate the following items:

- *During the last year, have you done any of the following activities*
  - *Taken part in a demonstration or protest march (yes - no)*
  - *Joined others in your community to request action from the government (yes - no)*
  - *Voted in a national election (yes - no)*
  - *Attended a campaign rally (yes - no)*
  - *Tried to persuade others to vote for a particular political party or candidate (yes - no)*
  - *Discussed political matters with family and friends (yes - no)*
- *How would you describe your political views?*
  - *Strongly liberal (0) - strongly conservative (10)*

3.6.4 Background variables

The respondents will also answer a set of background questions about gender, age, level of education, employment status and income (see Appendix B for question formulations).

3.7 Leader response

The experimental design also allows me to study whether and why the way the group endowment is financed (hard earned vs. windfall and through the citizen's pocket or not) affect the leaders' investments. I define:

- *Leader accountability:* the share of the group endowment the leader invests in the common pool.

The larger the share of the group endowment the leader allocates to the common pool, the more responsive the to the citizen's demands the leader is considered to be.

To shed light on the leaders' motivation for investment in the common pool, I also elicit beliefs about citizens' behavior by posing the following incentivized question:

- *We now ask you to tell us what you think the citizen did. For each of the possible decisions you could make, please indicate what you think is the most likely choice of the citizen. If you guess correctly for 8 or more choices, you will receive an additional bonus of $0.25.*

3.8 Selection bias

A potential problem with announcing the payment scheme before the participants make the decision about performing the task is that it might lead different types of participants to choose to work in the different treatments, giving rise to a selection bias. To assess the problem and calibrate the size of the earnings, I conducted a pilot recruiting 50 citizens and 50 leaders to each of T1 and T4, in March 2017. Of the 200 participants recruited, 178 chose to perform the task for \$0.75. The share of citizens choosing not to work was the same in the two treatments (12%), but fewer leaders chose not to work in T4 (4%) compared to T1 (13%). Thus, selection of different types of respondents to different treatments does not seem to be a large problem with the design, particularly not for the citizens. However, in order to further reduce the chances of a selection issue arising, I decided to increase the earnings for work from \$0.75 to \$1. An additional selection issue might arise if citizens' beliefs about their leader's type differ between treatments. However, since I will use the strategy method to elicit the punishment threshold, the citizen's beliefs about the leader is not likely to affect the punishment threshold.

The argument for announcing the payment scheme in advance is twofold. First, it is important that that the participants feel that the earnings (and the group endowment in the hard earned treatments) are financed by their work, and they are more likely to do so if they know that their work will be rewarded, and by how much, in advance. Second, it gives the participants time to internalize their earnings before paying tax, making it easier to identify the Loss Aversion mechanism.

4 Empirical strategy

The experiment is designed to answer the two research questions posed in the introduction. This section formulates the hypotheses to be tested and describes the empirical strategy for testing them.

4.1 Main hypotheses

To answer the first research question, I formulate the following main (alternative) one-sided hypothesis:

Hypothesis 1 *Taxed citizens have a higher punishment threshold than non-taxed citizens.*

In this formulation, "taxed" means that the group endowment is both financed by hard work and is collected through taxation. This hypothesis tests the Rentier State Hypothesis

To answer the second research question, I formulate the following main (alternative) hypotheses:

Hypothesis 2 *Citizens have a higher punishment threshold when the group endowment is financed by their hard work than when it is windfall. (One-sided)*

Hypothesis 3 *Citizens have a higher punishment threshold when the group endowment is financed by tax collected from the their pocket than when the money goes directly to the group endowment. (One-sided)*

Hypothesis 4 *The Hard Earned and Loss Aversions mechanisms interact positively. (Two-sided)*

Hypothesis 5 *When the group endowment is tax collected from the citizen's pocket, it does not only increase the reference point, but also the weight attached to the fairness norm. (One-sided)*

The main alternative hypothesis for Hypothesis 4 is that there is no interaction effect. However, since the effect has not, to my knowledge, been investigated before and because the two mechanisms may potentially crowd each other out, I will use a two-sided test for the hypothesis.

4.2 Hypotheses for leader response

I formulate the following main (alternative) hypotheses for the leaders' investments:

Hypothesis 6 *Leaders invest a larger share in the common pool when the group endowment is financed by the citizen's hard work than when it is windfall. (Two-sided)*

I expect leaders to make more accountable investment choices when the group endowment is financed by hard work because they are likely to recognize that the leader will attach a higher weight to the fairness norm when he or she has worked to generate the group endowment. However, because the leader is also working hard to generate the group endowment he might feel stronger about keeping more money to himself, and it is therefore also plausible that the effect could go in the opposite direction. I will therefore use a two-sided test for Hypothesis 6.

The investment choice of the leader is likely to be affected by two factors; their own fairness views and their beliefs about the citizen's punishment decisions. To shed light on what is driving the choice of leaders, I formulate the following hypothesis:

Hypothesis 7 *Leaders expect citizens to have a higher demand for accountability when the group endowment is financed by the citizen's (and the leader's) hard work (as opposed to when it is financed by a windfall). (Two-sided)*

4.3 Hypothesis for robustness check

Hypothesis 8 *The effect of increasing duration of the real effort task on the citizen's punishment threshold is not statistically different from the effect of increasing payment for the real effort task. (One-sided)*

The main alternative hypothesis is that increasing the duration of the real effort task has a larger positive effect on demand for accountability than increasing the payment for the real effort task.

4.4 Heterogeneity

I will also study whether the treatments affect subgroups of the sample differently. I will focus on the following dimensions:

- Male (indicator variable = 1 if respondent is male)
- Age (indicator variable = 1 if age above sample median)
- Education (indicator variable = 1 if education level above sample median)
- Income (indicator variable = 1 if income is above sample median)
- Negative reciprocity (indicator variable = 1 if index[5] is above median)
- Positive reciprocity (indicator variable = 1 if positive reciprocity is above median)
- Risk preferences (indicator variable = 1 if risk willingness is above median)
- Loss aversion (indicator variable = 1 if respondent chooses \$0 for sure)
- Political orientation and engagement

[5]Following Falk et al. (2015), I will construct an index for negative reciprocity = 0.2631× Willingness to punish if oneself is treated unfairly + 0.2631× Willingness punish if other is treated unfairly + 0.3738× Willingness to take revenge. These weights were obtained by running a regression of observed risk behaviour in the lab on responses to these survey questions.

- Liberal (indicator variable = 1 if higher than sample median)
- Politically engaged (indicator variable = 1 if index[6] is higher than sample median)

- Emotions (indicator variable = 1 for citizen with higher than median score on question about how upset he or she would be if receiving less than fairs share)

I do not have any directional hypotheses about heterogenous treatment effects.

4.5 Empirical specifications

In this section, I provide the main specifications for the robust OLS regressions that will be estimated in my analysis. The main outcome variable of interest is citizens' elicited punishment threshold. In the analysis, I will drop citizens that switch the "wrong" way (i.e. citizens that do not punish low investments, but do punish high investments) because there is no good theoretical reason for such behavior and it is likely to be caused by careless responding or misunderstandings. Furthermore, citizens that punish even when the leader invests everything in the common pool will be analysed as having the highest possible punishment threshold (1), even though they strictly speaking never switch from punishing to not punishing.

In addition to the citizens' punishment thresholds, I will look at the leader's investment choice, measured as the share invested in the common pool.

4.5.1 Hypothesis 1 - 4

To test Hypothesis 1 - 4, I estimate the following regression on the whole sample:

$$y_i = \alpha + \beta^{\mathrm{H}}\mathrm{H}_i + \beta^{\mathrm{T}}\mathrm{T}_i + \theta\mathrm{T}_i \times \mathrm{H}_i + \beta^X X_i + \beta^Z Z_i + \varepsilon_i. \tag{4}$$

y_i is the citizen in pair i's punishment threshold (lowest share invested in the common pool the citizen does not punish), α is a constant, H_i is an indicator variable taking the value of one for individuals in the Hard Earned treatments, T2 and T4, T_i is an indicator variable taking the value of one for individuals in the Tax treatments, T3 and T4, and $\mathrm{H}_i \times \mathrm{T}_i$ is and interaction term between H_i and T_i. $\mathrm{H}_i \times \mathrm{T}_i$ takes the value of one for individuals in the Hard Earned & Tax treatment, T4. X_i is a vector of the background variables of the citizen (indicator variables for gender, age, education, employment status and income), β^Z is a vector of preferences (negative reciprocity, positive reciprocity, risk aversion and political engagement) and ε_i is an error term. I will estimate three specifications for all equations: one without X_i and Z_i, one with only X_i and one with both X_i and Z_i.

If $\beta^H + \beta^T + \theta > 0$, meaning that the lowest acceptable share invested in the common pool is higher when the group endowment is financed through taxation and hard work compared to when it is windfall revenue, this can be interpreted as evidence in support of the Rentier State Hypothesis (Hypothesis 1).

If $\beta^{\mathrm{H}} > 0$, it shows that Hypothesis 2 (Hard Earned mechanism) holds when the group endowment is not collected through tax (comparing T2 to T1) and if $\beta^{\mathrm{H}} + \theta > 0$, it shows that Hypothesis 2 holds when the group endowment is collected through tax (comparing T4 to T2).

If $\beta^{\mathrm{T}} > 0$, it shows that Hypothesis 3 (Loss Aversion mechanism) holds when the group endowment is windfall (comparing T3 to T1), and if $\beta^{\mathrm{T}} + \theta > 0$ it shows that Hypothesis 3 holds when group endowment is financed by hard work (comparing T4 to T3).

θ indicates whether the effect of going from Windfall to Hard Earned is significantly different between Non-tax and Tax treatments. If $\theta >$, this shows that Hypothesis 4 holds.

[6]The index is equal to the number of sub questions where the respondent answers "yes".

4.5.2 Hypothesis 5

To test Hypothesis 5, I estimate the following regression on a sample restricted to T1 and T3:

$$y_i = \alpha + \beta^{\mathrm{T}}\mathrm{T}_i + \beta^X X_i + + \beta^Z Z_i + \varepsilon_i. \tag{5}$$

where y_i is the punishment threshold restricted to investments that gives the citizen a payoff higher than \$2.

If $\beta^T > 0$, meaning that citizens have a higher punishment threshold for payoffs higher than their reference point, this means that the Windfall & Non-Tax treatment does not only increase the reference payoff, put also increases the weight attached to receiving a fair outcome.

4.5.3 Hypothesis 6 and 7

To test Hypothesis 6, I estimate the following equation:

$$y_i = \alpha + \beta^{\mathrm{H}}\mathrm{H}_i + \beta^X X_i + \beta^Z Z_i + \varepsilon_i, \tag{6}$$

where y_i = share of the group endowment allocated to the common pool and H is an indicator variable taking the value of one for respondents in the Hard Earned treatments (T2 and T4) and zero otherwise. If β^{H} positive and significant, this shows that Hypothesis 6 holds.

To test Hypothesis 7, I estimate Equation 6 with y_i the leader's beliefs about the citizen's punishment threshold. If β^{HE} positive and significant, this shows that Hypothesis 7 holds. If Hypothesis 6 and Hypothesis 7 holds simultaneously, this indicates that treatment effect on leader behaviour is partly driven by the treatments effects on leader beliefs about citizens punishment behaviour.

4.5.4 Hypothesis 8

To test Hypothesis 8 i estimate on a sample restricted to to T4 and T4b:

$$y_i = \alpha + \beta^{\mathrm{T4b}}\mathrm{T4b}_i + \beta^X X_i + \beta^Z Z_i + \varepsilon_i, \tag{7}$$

where $\mathrm{T4b}_i$ and an indicator variable taking the value of one for citizens in the T4b treatment. If $\beta^{\mathrm{T4b}} > 0$ this indicates that the Hard Earned mechanism is stronger when the citizen actually has to work harder.

4.5.5 Heterogeneity

To investigate whether the treatments affect different subgroups of the sample differently, I will estimate the following regression:

$$\begin{aligned} y_i = \alpha + \beta^{\mathrm{H}}\mathrm{H}_i + \beta^{\mathrm{T}}\mathrm{T}_i + \theta^1\mathrm{H}_i \times \mathrm{T}_i + \beta^{\mathrm{Var}}\mathrm{Var}_i + \theta^2\mathrm{H}_i \times \mathrm{Var}_i + \\ \theta^3\mathrm{T}_i \times \mathrm{Var}_i + \gamma\mathrm{H}_i \times \mathrm{T}_i \times \mathrm{Var}_i + \beta^X X_i + \beta^Z Z_i + \varepsilon_i. \end{aligned} \tag{8}$$

Where Var_i (Male, Gender, Age, Negative Reciprocity, Positive Reciprocity, Loss Aversion Political Engagement) are indicator variables for the subgroups of interest and $\mathrm{H}_i \times \mathrm{Var}_i$ is an interaction term between H_i and Var_i, $\mathrm{T}_i \times \mathrm{Var}_i$ is an interaction term between T_i and Var_i, and $\mathrm{H}_i \times \mathrm{T}_i \times \mathrm{Var}_i$ is an interaction term between H_i, T_i and Var_i.

References

Anderson, Christopher M. and Louis Putterman (2006). "Do non-strategic sanctions obey the law of demand? the demand for punishment in the voluntary contribution mechanism," *Games and Economic Behavior*, 54(1): 1 – 24.

Borge, Lars-Erik, Pernille Parmer, and Ragnar Torvik (2015). "Local natural resource curse?" *Journal of Public Economics*, 131: 101 – 114.

Bosman, Ronald, Matthias Sutter, and Frans van Winden (2005). "The impact of real effort and emotions in the power-to-take game," *Journal of Economic Psychology*, 26: 407–429.

Bräutigam, Deborah, Odd-Helge Fjeldstad, and Mick Moore (2008). *Taxation and state-building in developing countries: Capacity and consent*, Cambridge University Press.

Brollo, Fernanda, Tommaso Nannicini, Roberto Perotti, and Guido Tabellini (2013). "The political resource curse," *American Economic Review*, 103(5): 1759–96.

de la Cuesta, Brandon, Lucy Martin, Helen Milner, and Dan Nielson (2017). "When non-earned revenue looks like tax: An experimental test of the differential accountability theis in Ghana," .

Dufwenberg, Martin and G. Kirchsteiger (2004). "A theory of sequential reciprocity," *Games and Economic Behavior*, 47(2): 268–298.

Dynes, Adam and Lucy Martin (2017). "Revenue source and electoral accountability: Experimental evidence from local U.S. policy makers," .

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2015). "The nature and predictive power of preferences: Global evidence," IZA DP 9504, IZA.

Falk, Armin and Urs Fischbacher (2001). "Distributional consequences and intentions in a model of reciprocity," *Annales d'Économie et de Statistique*, 63: 111–129.

Fehr, Ernst and Simon Gächter (2000). "Fairness and retaliation: The economics of reciprocity," *Journal of Economic Perspectives*, 14(3): 159–181.

Fehr, Ernst and Simon Gächter (2002). "Altruistic punishment in humans," *Nature*, 415: 137–140.

Fehr, Ernst and Lorenz Goette (2007). "Do workers work more if wages are high? evidence from a randomized field experiment." *American Economic Review*, 97(1): 298 – 317.

Fehr, Ernst and Klaus M. Schmidt (1999). "A theory of fairness, competition and cooperation," *Quarterly Journal of Economics*, 114(3): 817–868.

Ferraz, Claudio and Frederico Finan (2011). "Electoral accountability and corruption: Evidence from the audits of local governments," *The American Economic Review*, 101(4): 1274–1311.

Gadenne, Lucie (2017). "Tax me, but spend wisely? sources of public finance and government accountability," *American Economic Journal: Applied Economics*, 9(1): 274–314.

Henrich, Joseph, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank Marlowe, David Tracer, and John Ziker (2006). "Costly punishment across human societies," *Science*, 312(5781): 1767–1770.

Henrich, Natalie and Joseph Patrick Henrich (2007). *Why humans cooperate: A cultural and evolutionary explanation*, Oxford University Press.

Hvide, Hans K. and Jae Ho Lee (2016). "Windfall gains and investment choices," .

Kahneman, Daniel and Amos Tversky (1979). "Prospect theory: An analysis of decision under risk," *Econometrica*, 47(2): 263–291.

Karl, Terry Lynn (2007). chapter Ensuring fairness: The case for a transperant fiscal social contract., Columbia University Press, pp. 258 – 293.

Lederman, Daniel, Norman V. Loayza, and Rodrigo R. Soares (2005). "Accountability and corruption: Political institutions matter," *Economics & Politics*, 17(1): 1 – 35.

Mahdavy, Hossein (1970). *Studies in the history of the Middle East from the rise of Islam to the present day*, chapter The patterns and problems of economic development in rentier states: the case of Iran., Oxford University Press, pp. 428 – 467.

Martin, Lucy (2016). "Taxation, loss aversion and accountability: Theory and exexperiment evidence for taxation's effect on citizen behavior," Technical report, Working paper. Available: https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxsdWN5bWFydGlufGd

Martínez, Luis R. (2016). "Sources of revenue and government performance: evidence from Colombia," .

Merriam-Webster (2017). *Merriam-Webster Dictionary*, chapter Accountability, https://www.merriam-webster.com/dictionary/accountability.

Paler, Laura (2013). "Keeping the public purse: An experiment in windfalls, taxes, and the incentives to restrain government," *American Political Science Review*, 107(4): 706–725.

Przeworski, Adam and Susan C Stokes (1999). *Democracy, accountability, and representation*, volume 2, Cambridge University Press.

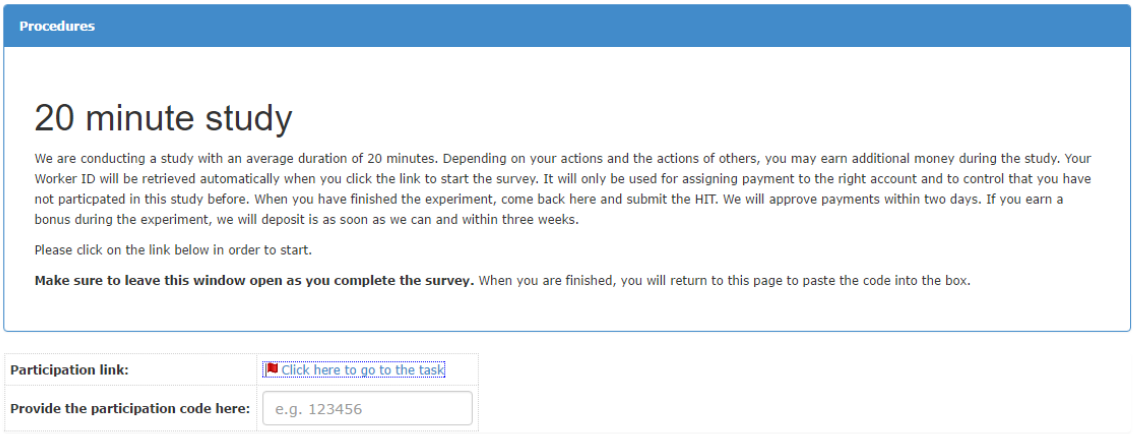
Ross, Michael (2004). "Does taxation lead to representation?" *British Journal of Political Science*, 34: 229–249.

Ross, Michael L. (2001). "Does oil hinder democracy?" *World Politics*, 53(3): 325–361.

UNDP (2008). "Primer on corruption and development," Technical report, United Nations Development Programme, December.

Appendix A HIT announcement

Figure A.1: HIT announcement



Appendix B Background questions

- What is your gender?
  - Male
  - Female
  - I prefer not to answer
- What is your age?
- What is the highest level of education you have completed?
  - Less than High School
  - High School/GED
  - Some College
  - 2 year College degree
  - 4 year College degree
  - Master's degree
  - Professional degree
  - Doctoral degree
  - Professional degree (JD, MD)
  - I prefer not to answer
- What is your current employment status?
  - Employed full time (30+ hours per week)
  - Employed part time (less than 30 hours per week)
  - Self-employed
  - Retired/unable to work/disabled
  - Still in school
  - In full time higher education
  - Unemployed and seeking work
  - Unemployed and not seeking work
  - I prefer not to answer
- What is your individual yearly income (gross income before taxes are deducted)?
  - $0-$4,999
  - $5,000-$9,999
  - $10,000-$14,999
  - $15,000-$19,999
  - $20,000-$24,999
  - $25,000-$29,999
  - $30,000-$39,999
  - $40,000-$49,999
  - $50,000-$74,999
  - $75,000 or more

- I prefer not to answer

- What is your household's combined yearly income (gross income before taxes are deducted)?
  - Less than $20,000
  - $20,000 - $29,999
  - $30,000 - $39,999
  - $40,000 - $49,999
  - $50,000 - $59,999
  - $60,000 - $74,999
  - $75,000 - $99,999
  - $100,000 - $149,999
  - $150,000 or more
  - prefer not to state

Appendix C Power calculation

To do the power calculation, I first use Martin (2016)'s estimated standard error of the effect of the tax treatment on punishment threshold, 19.10, and the total number of observations, 296, reported in Table 2 to calculate the standard deviation.

$$var(\beta) = \sigma^2\left(\frac{1}{n_{tax}} + \frac{1}{n_{grant}}\right) \tag{9}$$

assume $n_{tax} = n_{grant}$

$$var(\beta) = \sigma^2\left(\frac{1}{\frac{n}{2}} + \frac{1}{\frac{n}{2}}\right) = \sigma^2\left(\frac{4}{n}\right) \tag{10}$$

$$se = \sqrt{\sigma^2\left(\frac{4}{n}\right)} \tag{11}$$

$$\frac{19.1 \times \sqrt{296}}{2} = \sigma^2$$

$$164.3 = \sigma^2$$

Given the regression coefficient of 54.8, a σ^2 of 164.3 equals an estimated effect size of the tax treatment of 0.317.

Assuming $\sigma^2 = 164$, I will need 170 observations in each group to have a power of 0.8 and 228 observations to have a power of 0.90 to identify a treatment effect of 0.3. A sample size of 200 observations per group allows me to identify a treatment effect of 0.3 with a power of 0.86. Figure A.2 below illustrates the relationship between power and sample size for three different effect sizes.

Figure A.2: Sample size and power for given effect sizes