Masaryk University
Faculty of Economics and Administration

Tax and Regulatory Compliance: Three Experimental Studies

Habilitation thesis

Rostislav Staněk

15th April 2019

Declaration

I hereby declare that the habilitation Thesis titled *Tax and Regulatory Compliance: Three Experimental Studies* is my own work and that I have cited all of the literature and other expert resources therein in accordance with applicable legal regulations, the Internal Rules and Guidelines of Masaryk University, and the internal managing acts of Masaryk University and the Faculty of Economics and Administration.

Acknowledgements

Most of all I would like to thank Eva, Eliška and Jakub for their patience. I am also grateful to my collagues: Ondřej Krčál for fruitful discussions and invaluable help with this research project; Miloš Fišar for administration of the MUEEL laboratory; Katarína Čellárová for programming of the third experiment; Zdeněk Tomeš and Jiří Špalek for providing stable and collaborative environment supporting research.

Contents

Chapter 1

Introduction

Many key ideas in economic theory rely on the ability of public agencies and courts to enforce regulation efficiently. Perfectly enforceable regulation addresses the problem of market failures due to the excessive market power or due to the presence of externalities. In reality, many regulations may fail to increase economic and social well-being since their objectives are flawed from beginning or because there is inadequate regulatory compliance.

Regulatory compliance covers in broad sense situations when individuals have to comply with some rules or they should abstain from some behaviour. For instance, plants should not pollute excessively, taxpayers should pay their taxes properly and firms should comply with safety and health standards. Effective enforcement is vital to the successful implementation of many different types of regulations (competition policy, environmental regulation, safety regulation, tax laws etc.), and regulations that is not enforced rarely meets its social objectives. Extensive economic research examines the question of how the enforcement task might best be conducted in order to achieve compliance with regulation and be efficient in terms of doing so at least cost to both regulated entity and regulator. One of the most prominent approaches how to ensure compliance is to inspect regulated entities and impose fines for detected violations.

The basic approach for modelling compliance with tax laws and other regulations is a variant of more general model by Becker (1968). Its underlying assumption is that compliance decision can be studied by microeconomics toolbox as an example of rational decision making. Penalties for violating regulations are treated as any other costs and regulated parties choose the level of compliance in order to minimize the sum of compliance costs and expected penalties. The main result of this benchmark model is that full compliance can be reached by making detection probabilities and penalties large enough.

The empirical literature has focused predominantly on testing the effect of severity of punishment and audit probabilities on compliance behaviour. The early literature on tax and regulatory compliance was burdened with serious identification and data problems, mostly caused by endogeneity of audit probabilities and penalties (see Andreoni et al. (1998) for discussion of this issue). Later empirical research solved these problems by using quasi-experimental econometric techniques exploiting exogenous shocks (instrumental variables, difference-in-difference approach, and regression discontinuity design) or by conducting laboratory and field experiments.

One of the main advantages of studying tax and regulatory compliance in the laboratory is the control over all pertinent variables. The experimenter not only sets the levels of main variables (i.e. audit probability, penalty rate, compliance costs, tax rates etc.), but she also controls other circumstances that could affect individual compliance behaviour. For this reason, the experimental method is frequently applied in the current literature on compliance. The evidence from laboratory experiments confirmed the main theoretical predictions of the enforcement model. A higher probability of being audited (Alm et al., 2009; Cummings et al., 2009) and more severe penalties (Alm et al., 1999; Park and Hyun, 2003) lead to more compliance; although the effect of more severe penalties is limited (Alm et al., 1995; Cummings et al., 2009). Experimental results, on the other hand, demonstrated little effects of socio-demographic variables such as gender and age.

Full compliance does not have to be achievable by manipulating penalties and detection probabilities. In many settings, there will be an upper bound on the penalty that can be levied. The detection probability also cannot be manipulated freely since audits are costly. Henceforth, further research has examined how enforcement authority can utilize available information about past compliance behaviour or other parties' compliance behaviour to increase the enforcement leverage. This information can be used to make audits more efficient by targeting those who are likely least compliant (Duflo et al., 2018; Cason and Gangadharan, 2006; Gilpatric et al., 2011) or to tailor penalties according to the compliance costs (Kang and Silveira, 2018).

Besides this extension, the benchmark model includes several more simplifying assumptions which can be relaxed. For instance, the non-compliant agents may find it optimal to spend real resources in order to decrease transparency of their behaviour and hide non-compliance (Bayer, 2006). The benchmark model also assumes that non-compliance cannot happen by accident and the enforcement authority never penalizes compliant firm. Once we incorporate these features into the model, the regulated individual may find it optimal to inefficiently overcomply (Shimshack and Ward, 2008). The question arises what is the optimal enforcement strategy once we take these issues into account.

The research presented in this thesis contributes to the experimental study of regulation in general and regulatory enforcement in particular. The studies collected in the habilitation thesis investigate the links between some enforcement approaches (audit selection mechanism, enforcement discretion) and some characteristics of the environment (incomplete information, possibility to conceal non-compliance, etc.). From a methodological point of view, all studies employ the experimental research design. All experiments are conductead as laboratory experiment in the Masaryk University Experimental Economics laboratory (MUEEL).

First experimental study (chapter 4) investigates properties and efficiency of the competitive audit selection mechanism when only a very limited information is available. It has been shown in the experimental literature on tax and regulatory literature that competitive audit selection mechanism increases compliance (Alm and McKee, 2004; Gilpatric et al., 2011). However, this literature assumes that the tax authority has an unbiased observation of the actual taxpayers' income and consequently the taxpayers with the largest difference between the observed and reported income are most likely to be selected for audit. In reality, the tax authority might not have unbiased information about the taxpayers' actual incomes as these might be observed only for taxpayers who have been selected for audit. In this case, the audit selection mechanism can be based on reported incomes only.

The study designs a competitive audit selection mechanism that uses solely the reported income and experimentally compares the tax compliance under the competitive and random audit selection mechanisms. We develop a theoretical model where taxpayers have heterogeneous income and the audit selection mechanism is based only on the reported income. The model solution shows that in the symmetric Bayes-Nash equilibrium the proposed competitive audit selection mechanism entails a higher compliance than the random audit selection mechanism. The experimental test of the design confirms the theoretical prediction that taxpayers have higher compliance under the competitive audit selection mechanism than under the random audit selection mechanism. However, the comparison between competitive audit selection mechanisms with complete and incomplete information shows that tax evasion higher when the information is incomplete. The difference is driven by high income taxpayer who comply less when the information is incomplete

Second experimental study contributes to the literature on tax and regulatory compliance by investigating the interaction between concealment activities and a competitive audit selection mechanism. The basic model of regulation or tax enforcement supposes that the enforcement authority inspects the regulated party and levies a penalty against those found non-compliant. In real-world, the authority's enforcement power is not as assured as such a model would suggest, and regulated entities are able to obstruct the enforcement process and challenge regulatory decisions through various channels. They might contest the enforcement at courts or they might invest into concealment activities that reduce the transparency of their operations.

Agents subject to a regulation may choose not to comply and if possible, even invest real resources to conceal their regulatory avoidance. We develop a theoretical model that explores the effect of audit selection rules on these choices. The main predictions of the theoretical model are tested in a laboratory experiment. Namely, the experiment tests that the competitive audit selection mechanism increases compliance and at the same time reduces concealment investments. This outcome is compared with an increase in audit frequency, which raises both compliance and concealment investments. The experimental results confirm these predictions. In comparison with more extensive auditing, smart design of the selection mechanism may not only entail lower administrative costs but also discourage investment in socially wasteful concealment activities.

The third experimental study focuses on the choice between rules and discretion in regulatory enforcement. When lawmakers make legal pronouncements, they must decide not only the substance but also the form of the pronouncement. The choice of the legal form may be described as a choice between rules and discretion. Rules state a definitive legal result that follows from a triggering fact. Discretion, on the other hand, gives enforcement officials some degree of discretion to apply some set of principles to reach a legal conclusion. Granting discretion to officials permits them to take into account specific circumstances which cannot be specified precisely in a rule. On the other hand, the official's utility function does not have to correspond to the social welfare function and his personal optimal choice may deviate from social optimum (Shavell, 2007). The third experiment design tests whether the granting some discretionary power to the enforcement agency leads to higher monetary welfare. The results of the experiment suggest that the answer is positive.

Beyond this modeling framework, there might be other factors that influence the trade-off between discretion and strict rules. Although the discretion may maximize monetary welfare of the society, it might be perceived negatively by members of the society and people might be discretion averse. The discretion aversion may be justified on the ground of two distinct theories: betrayal aversion and procedural fairness. Betrayal aversion is a well-documented tendency to avoid a situation when a person, rather than nature, determines the outcome of the situation (Bohnet and Zeckhauser, 2004; Bolton and Ockenfels, 2010). In the context of legal enforcement, betrayal aversion may be manifested as an aversion towards discretionary power of the official. There is also evidence that people care not only about allocation of goods but also about procedures that leads to the allocation (Bolton et al., 2005; Sausgruber and Tyran, 2014). As there is always mistakes and subjective judgment in the discretionary regime, people might perceive this regime as unfair. Henceforth, the additional aim of the third study is to identify whether people are discretion aversion and test whether the discretion aversion is driven by betrayal aversion or by procedural fairness.

Chapter 2

Regulatory enforcement

Even if we assume that regulatory objectives have been established with clarity, it still cannot be taken for granted regulated parties will comply with the regulation. Each regulatory system must specify procedures for ensuring compliance including provision of information to the regulator, monitoring compliance and penalties for not complying with the regulation. These form the enforcement rules.

Many regulations are enforced by the so called enforced self-regulation, which involves a subcontracting of regulatory functions to regulated entities. The regulated entities are free to choose how to comply with regulatory requirements. The primary function of enforcement agency is to audit whether the regulatory targets are met and impose penalties in case of their violations. This enforcement framework is similar to tax enforcement, where the tax authority audits taxpayers in order to examine whether they declared their taxable income properly.

Broadly speaking, theories that seek to explain regulatory or tax compliance can be divided into three categories; those that see people as motivated by i) economic calculative motivations or the fear of detection of violations and imposition of sanctions; ii) social motivations, or the desire to earn respect and approval of peers; and iii) intrinsic moral motivations, or or a sense of moral obligation to comply with a particular regulation and an agreement with its legitimacy. This and the following chapter summarize current knowledge with regard to the first category.

2.1 Basic enforcement model

The basic motives for copliance are summarized in the tax evasion model of Allingham and Sandmo (1972) who adapted the model of criminal behaviour by Becker (1968) to the tax evasion setting. In the basic model, an individual is given an income I. A tax at a constant rate τ is then levied on the reported income R, being a certain part of I. The undisclosed income is $Z = I - R$. The individuals are aware that they maybe audited with a probability π. If the individual has not declared all his or her income, he or she will have to pay the penalty $\phi(I-R)$, where $\phi > \tau$. Note that the model does not contain any intrinsic or social motivation to meet any regulatory or tax obligations. The effect of possible legal penalties is no different from the effect of any other contingent costs.

The individual chooses R so as to maximize her expected utility, which in the case of proportional income tax can be written as:

$$E(U) = \pi\, u((1-\tau)R + (1-\phi)(I-R)) + (1-\pi)\, u(I-\tau R).$$

The first order condition for optimal compliance is as follows

$$\frac{U'(\overline{w})}{U'(\underline{w})} = \frac{(\phi-\tau)\pi}{(1-\pi)\tau}, \tag{2.1}$$

where $\overline{w}$ denotes the monetary wealth in the non-audited state of the world, i.e. $\overline{w} = I - \tau R$, and $\underline{w}$ denotes monetary wealth in the audited state of the world, i.e. $\underline{w} = (1-\tau)R + (1-\phi)(I-R)$. The left-hand side of the equation (2.1) is the marginal rate of substitution between the wealth in the audited state and non-audited state. Obviously, the wealth in the non-audited state is weakly less than the wealth in the audited state, i.e. $\underline{w} \leq \overline{w}$. The marginal rate of substitution for the risk-averse agent is always less than one and increasing in the reported income. It is equal to one for the risk-neutral agent.

The solution shows that the risk-neutral agent discloses all her income if the expected benefit of the evasion is less than the expected costs, i.e.

$$(1-\pi)\tau < (\phi-\tau)\pi.$$

In the opposite case, the risk-neutral agent does not comply at all and the risk-averse agent undercomplies to some degree. The solution also reveals the effect of thechange in model parameters on tax or regulatory compliance. An increase in either the probability of audit or theimposed penalty raises compliance. The higher tax rate leads to lower compliance.

In the basic enforcement model, the value of the expected penalties can be affected by manipulating either the audit probability or the fines or both. Full compliance might be achieved by making the expected penalty high enough. Taking into account that auditing is costly for the enforcement authority, it seems optimal to set fines at an arbitrarily high level. In many settings, however, there will be an upper bound on the fines that can be levied. The penalty might be exogenous to the enforcement authority, e.g. imposed by the law. Severe and rare penalties might be perceived as unfair (Harrington, 1988). The taxpayer may also be judgment proof, i.e. her wealth may not be large enough to pay the fine.

2.2 Possible applications

This section discusses possible applications of the basic enforcement model. One broad category of compliance decision-making is tax compliance. Tax compliance and tax evasion belong to the popular and frequently discussed questions of public policy. As taxes represent a basic source of public revenues and many countries have struggled to reduce public deficits in the aftermath of the economic crisis, tax enforcement has been pivotal element of the policy agenda of many countries and international organizations (OECD, 2017).

Measuring tax evasion is a difficult task due to the fact that tax evaders have a strong incentive to conceal their tax evasion. Most of the information about the magnitude of tax evasion is based on the evidence from random audits and traces of income approach (Feldman and Slemrod, 2007). The US Internal Revenue Service runs the so called National research programme (NRP) which provides estimates of tax evasion based on random audits from a stratified sample of 46 000 tax returns. The last estimate of the evaded taxes was published by IRS and covers the years 2008 to 2010. It states that the amount of evaded taxes is 16.9 percent of the overall estimated tax liability (IRS, 2016). Out of this amount, 85 percent is due to underreporting. Of course, not every income is evaded in the same proportion. Third-party reporting plays a crucial role. The estimated tax evasion for the income subject to substantial third-party reporting and withholding is 1 percent, for the income subject to substantial information reporting but not withholding is 7 percent, and for the income subject to little or no information reporting is 63 percent (IRS, 2016).

Although the majority of people pay taxes properly, the size of undisclosed income may be substantial. As we have seen, the estimated tax gap in the developed countries varies from 5.7 percent of tax liabilities in the UK in the year 2016 and to 16.9 percent in the US in the year 2010. Besides lower tax revenues, tax evasion generates also welfare losses. The most obvious are the resources taxpayers expend on concealing their non-compliance and the resources the tax authority expends on detecting and combating non-compliance. In addition, tax evasion provides a socially inefficient incentives to engage in those activities that facilitate evasion. In other words, there might be an excessive labor supply in the occupations where the income is difficult to detect for the tax authority and a shortage of labor supply elsewhere. For all these reasons, finding the motives that make people comply with the tax code is of great interest and importance.

Although the enforcement model in section 2.1 is framed in the tax compliance context, the model can be easily re-framed and applied to other regulations such as safety, health regulations or environmental regulations.

Consider the Clean Water Act as an example of such a regulation. The aim of the Clean Water Act is to protect water quality by regulating wastewater discharges from the point source of pollution. In order to fulfill this goal, the US Environmental Protection Agency issues a plant-specific permit which sets the discharge limit imposed on the regulated firms. After a limit is determined, their compliance is based only on the actual discharge which means that to comply with the imposed limit a firm may employ whatever abetment tool it deems fit. To ensure compliance with the limit, the Environmental Protection Agency inspects firms and imposes fines when non-compliance is found.

We can easily re-frame the enforcement model to describe the regulation under the Clean Water Act. Suppose a problem of a firm which is issued a limit I. The firm can conduct some abatement activities to comply with the limit to some degree R. The abatement activities are costly and the firm has to pay the costs τR where τ are the constant marginal costs of compliance. The difference between the limit and the compliance is called non-compliance and denoted as $Z = I - R$. The firm can be audited with the probability π. If the firm is audited and has not met the limit I, it has to pay a fine ϕZ which is proportional to the degree of non-compliance, the parameter ϕ represents the fine rate. The firm chooses the level of compliance in order to maximize the following expected utility

$$E(U) = \pi\, u(-\tau R - \phi(I - R)) + (1 - \pi)\, u(-\tau R).$$

Except for the constant, this is the same expected utility function as in the basic enforcement model which was framed in the tax compliance setting. Both these settings, regulatory and tax compliance, can be therefore described by the same enforcement model.

Generally, the enforcement model can be applied to all types of regulations where the regulatory authority sets some regulatory target or some standard of behaviour, and afterwards, it uses audits and fines to enforce this target. In this general setting, the variable I is interpreted as a regulatory target. In the tax compliance setting, the regulatory target is the taxable income. The variable R is understood as a degree of compliance with the target. The compliance is costly for the regulated subjects with the costs τR. The costs correspond to the tax rate in the tax compliance setting. The variables π and ϕ represent the audit probability and the fine rate. From now on in this thesis, the model will be interpreted in this general way.

2.3 Empirical evidence

This chapter summarizes the empirical evidence about the effect of the crucial variables in the enforcement model, i.e. fine magnitudes and audit probabilities, on the compliance effort. Since the core of this thesis consists of three experimental studies, special attention will be given to experimental evidence.

2.3.1 Data sources

The possibility of finding factors which affect an individual decision on compliance is critically limited by the availability of data. We can define four possible sources of data on tax or regulatory compliance: (i) audit data, (ii) survey data, (iii) administrative data, and (iv) data generated through laboratory experiments or field experiments. Measurements of real-world non-compliance levels are generally not reliable. The data from audits may not provide an unbiased estimate of compliance since the subjects most suspected from non-compliance may be audited more often. Only random audits can provide a reliable estimate of compliance. Survey data do not have to be reliable because the possibility of a punishment makes many people unwilling to respond truthfully about their non-compliance behaviour. The administrative data usually include tax returns of the entire population of the country, or at least a substantial part of it. Their disadvantage is that they include only what the subjects reported; thus it is not possible to learn the actual level of non-compliance out of them. On the other hand, these large data-samples can be successfully used to assess the effects of exogenous changes in the audit probability and fine magnitude on compliance.

There are several methodological problems to overcome when studying compliance in a real-world setting. Usually, the relationship between enforcement variables and compliance are determined jointly which results in simultaneity bias and omitted variable problems. Specifically, an increase in fine magnitude or audit probability is expected to increase compliance effort, but a change in compliance is also expected to prompt an increase in the certainty and severity of punishment, through mechanisms such as an increase in the budget of the enforcement authorities. This makes it difficult to identify the causal impact of fines or audits on compliance.

Recent research has employed three types of strategies to overcome the simultaneity problem. The first strategy is to find a natural experiment which generates a truly exogenous variation in audit probabilities and fines and use this variation as an instrument. The second strategy is to conduct a randomized control trial, i.e. to design an experiment that changes the audit probability or fine size exogenously. The most prominent form of intervention is the manipulation of the people's belief in the enforcement efficiency by sending letters (Kleven et al., 2011; Bergolo et al., 2017). The third identification strategy is to take advantage of various discontinuities in the enforcement policy and estimate the effect of fines and audit probabilities by the regression-discontinuity design. For instance, several studies (Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013) take advantage of kinks in the marginal tax rate to document the significance of tax evasion. These studies show that the income of wage earners, who do not have a chance to evade taxes, does not bunch at the points of kinks in the tax rate. On the other hand, the income of self-employed individuals bunches at the point where a lower marginal tax is applicable.

Although tax and regulatory compliance may be studied in the field in a rigorous way, there remain some important constraints. Natural experiments are rare. Randomized control trials are often limited in the scope of the interventions – it is possible to send letters, but it is difficult to imagine that different people will get a different punishment for the same wrongdoing. For these reasons, laboratory experimental research is a valuable source of knowledge about tax and regulatory compliance. There might be a serious concern regarding the external validity of laboratory tax compliance experiments with student samples. Students have little or no experience with paying taxes or a specific regulation compliance, and apparently, their socio-demographic characteristics may differ from those of the general population (Alm et al., 2015).

There are several studies that compare tax compliance among student and non-student samples (Gërxhani and Schram, 2006; Choo et al., 2016; Alm et al., 2015). Although students generally exhibit less compliant behaviour, there were little or no differences between them in the treatment effects, i.e. the student's behaviour changes in the same direction as the behaviour of non-students. Alm et al. (2015) compares a sample of university students and university staff. He studies the reaction to changes in various parameters in a standard tax compliance game, including audits and fines. Students were less compliant; however, the reaction to parameter changes was the same in both samples. The same conclusion is supported by the study by Gërxhani and Schram (2006) who conducted a tax experiment with different pools of participants: high school students, university students, high school teachers, academic staff and non-academic university stuff. The subjects were asked to choose from two different income distributions X and Y. Afterwards, they received an income from the distribution and engaged in a standard tax compliance game. The X distribution gave higher expected income but the taxpayer was audited with probability 1, the income from Y distribution is audited with probability 0.5 The student participants choose Y distribution more often, but they reactedto the possibility to evade taxes in a similar way. Choo et al. (2016) reports an experiment with three subjects samples: students, employees and the self-employed. They found a difference in tax compliance levels with students being less compliant and the self-employed being the most compliant. The differences disappeared when the researchers removed the tax framing and used neutral instructions instead. Overall, this evidence shows that student samples might not be externally valid to make conclusions about the compliance level. However, students seem to be a valid subject pool to test the effect of different treatments, especially those that do not manipulate the framing of the game.

2.3.2 The effect of random audits

More important than knowing the magnitude of non-compliance is knowing the impact of audit probability, fine magnitudes, and other similar factors. This section focuses on what we know about the impact of audit probabilities.

For the reasons discussed above, it is quite complicated to study this effect in a real world setting. Since natural experiments that exogenously change the audit probabilities are almost non-existent, the majority of the research is conducted as a randomized control trial in the form of a letter study. In a letter study, the audit probability is exogenously manipulated by sending letters to the participants. The letter contains information about the audit probability or some vague statement (e.g. "your return will be closely examined") which is expected to increase the perceived audit probability. The studies measure the effect on the reported variable (income in most case) using administrative data. The effect is estimated using a difference-in-difference approach which compares the difference between the current and the previously reported variable between the treatment group and the control group of those who received no information from the authority.

Slemrod et al. (2001) is probably the first one who conducted such a randomized control experiment in the field of tax compliance. Randomly selected taxpayers in Minnesota received letters announcing that their returns will be closely examined. The experimental results showed that the reported income of low and middle-income taxpayers increased significantly in comparison to the control group. Meiselman (2018) focuses on reporting the local taxes in Detroit. He communicates the information that the Detroit local authorities know the recipient's total federal income in order to increase the perceived audit probability. His results show that the probability of filing a return within 75 days of the intervention increased by almost 60 percent. Fellner et al. (2013) reports qualitatively similar results in the case of public television fees.

Kleven et al. (2011) reports a letter experiment from Denmark where the participants obtained exact information about the audit probability. Participants in the treatment group were informed that they would be audited with a 100 percent probability or a 50 percent probability. The participants in the control group received no letter. The participants who were informed about being audited with a 50 percent probability were roughly 1.1 percentage point more likely to increase their report when compared with the participants in the control group. The effect was 2 percentage points for the participants that were informed about being audited with a 100 percent probability. The comparison between the treatment with 50 percent and 100 percent probability addresses the inherent problem of these studies. The taxpayers' reaction depends on how the message changes the perceived audit probability. Since it is not possible to control for the perceived probability in the control group, it is not clear whether and how the intervention affects the perceived probability. By comparing the two treatments, Kleven et al. (2011) clearly documents a positive effect of audit probability on compliance.

An interesting experiment that also deals with this problem is presented by Dwenger et al. (2016). They study tax compliance in a legally binding but unenforced church tax in Germany. The fact that the tax is unenforced fixes the perception of audit probabilities to zero. Besides the treatment that informs about the audit probability, they introduce a notched treatment when the individuals who pay less or equal to 10 Euro will be audited with the probability of 0.5, and they will not be audited if they contribute more. The results show that the increased audit probability causes significant reductions in both the probability of evasion and the total evasion rate. The effect is even more pronounced for the notched treatment.

Besides letter studies, there is vast experimental evidence which shows that the frequency of audits has a strong effect on tax or regulatory compliance. The standard experimental framework used to study compliance closely follows the basic enforcement model. At the beginning of such experiment, participants are endowed with some amount of money, or in certain versions, the money can be earned in some real effort task. The participants are then asked to declare this or a smaller amount of money which will be subject to a given tax rate. The participants may be audited with a predetermined and previously known probability. If the participants are audited, they have to pay a fine which is usually a linear function of the undeclared income. Naturally, the participants pay no fine if they declared all their income. This experimental framework has been used to study how different variables (fine rate, audit probability, tax rate) and different institutions (endowed or windfall income, the presence of public good, framing) affect compliance.

Many studies have examined the relationship between audit probability and compliance using this experimental framework. Alm et al. (1995) conducts an experiment that consists of treatments with three different levels of audit probabilities: 5 percent, 30 percent, and 60 percent. The results clearly show that increasing the audit probabilities significantly increases compliance. The experiment by Alm et al. (1999) confirms this result. The audit probability in the experiment varied between 2 percent, 10 percent, and 50 percent. The compliance rates[1] were 0.23, 0.29 and 0.73 respectively. The experiment Park and Hyun (2003) uses very low levels of audit probability. The audit probability in different treatments varies between 1 percent, 10 percent, and 15 percent. However, they still observe an increase in compliance with rising

[1]The compliance rate is the reported income divided by the actual income. The compliance rate 1 represents full compliance.

audit probability. This result shows that the effect remains detectable even at low levels of audit probabilities. The same conclusion is further confirmed by Cummings et al. (2009) and Alm et al. (2009). In addition, Alm et al. (2009) show that the effect is especially strong in the cases when the income is not subject to third-party reporting.

Probably the only paper that does not find a significant effect of audit probability is Choo et al. (2016). They use two levels of audit probabilities 0.2 and 0.4. This difference did not have a significant impact on compliance in any subject pool they used (students, employees and self-employed).

In general, the evidence that audit probability has a positive impact on compliance is very strong and unambiguous. This conclusion is supported by both experimental and field studies.

2.3.3 The effect of fines

In the basic enforcement model, the penalty can always be made large enough so that the firm will always comply. In the notation used in the previous section, the firm will always comply if $\phi > \frac{\tau}{\pi}$. Despite the fact that increasing the magnitude of fines has a potential to increase compliance without any additional costs, only a minor part of the empirical tax or regulatory compliance research has been devoted to the effect of fines. This relative neglect could have been caused by a lack of natural quasi-experiments that exogenously vary the magnitude of fines[2]. Furthermore, a change in the severity of a punishment generates reactions in the behaviour of the enforcement authorities, which make the certainty of punishment endogenous to these changes (Miceli, 2008). Unlike the randomization of the audit probabilities, randomizing the magnitude of fines would imply punishing the same evasion differently, which makes randomized control trials unfeasible for legal and other reasons.

One way to tackle these problems is to use an information provision experiment as a substitute for an experiment that manipulates fines. The manipulation in the information provision experiment consists in reminding randomly chosen taxpayers what the value of existing penalties is. There is robust evidence that providing taxpayers with the information about an existing enforcement increases tax compliance. This result has been observed

[2]A notable exception is Stafford (2002) that examines an increase in the penalties for violations of hazardous waste regulations. She shows that the frequency of regulatory violations decreases significantly, even though the magnitude of the decrease is relatively small.

in a variety of contexts, e.g. income taxes (Kleven et al., 2011; Slemrod et al., 2001) VAT taxes (Pomeranz, 2015) or individual public-TV fees (Fellner et al., 2013). However, the mechanism behind these observations is not clear. The taxpayers may learn that the penalty is harsher than previously thought or the reminder may just make the penalty more salient.

Bergolo et al. (2017) conducts a field experiment in order to distinguish between these two mechanisms. They sent a letter to more than 20 000 Uruguayan firms. The firms in the control group received a letter with unspecific information about taxes. The firms in the treatment arm received a letter that informed them about audit probabilities and fines. The salience and audit probability were disentangled by varying the value of audit probabilities and fines in the treatment ar. In order to introduce this variation in a non-deceptive way, the values of audit probabilities and fines were calculated based on a random sample of firms with similar revenues. Since the tax payment did not differ among taxpayers that were randomly assigned different values of fines, they conclude that higher tax compliance is driven by the salience effect rather than by updating beliefs about audit probabilities and fines.

Since there is an apparent lack of real-world data with sufficient exogenous variation in the magnitude of fines, the effect of fines on compliance has been studied also in a laboratory environment. To some extent, experimental studies confirm that tax compliance is increasing in fines (Alm et al., 1999; Park and Hyun, 2003). However, some studies report mixed or no results. The effect of fines is, therefore, less robust than the effect of audit frequency. Alm et al. (1992) conducts a tax compliance experiment with three different fine rates: one, two and three times the unpaid tax. He finds that the fine rates do not have any significant effect on tax compliance. In fact, the effect of higher fines is almost negligible. Tax compliance increases only by 4 percent when the fine rate doubles. Alm et al. (1995) use a very similar experimental framework. They found no effect of the fine rate on tax compliance when the audit probability was 5 percent; however they found a positive and significant effect when the audit probability was 30 percent or 60 percent. A similar pattern is reported by Cummings et al. (2009). They use a 2 x 2 experimental design with two levels of fines and two levels of audit probabilities. Thefines were one and a half and three times the evaded tax and the audit probabilities were 10 percent and 30 percent. The effect of the increased fines is almost exactly zero when the audit probability is 10 percent; however, tax compliance increases with the higher level of audit probability.

Harbaugh et al. (2013) and Schildberg-Hörisch and Strassmair (2010) study the effect of punishment in a game in which participants can anonymously steal money from other participants. Contrary to tax or regulatory compliance experiments, the subjects' behaviour in this game directly influences the other subjects' payoffs. The researchers exogenously vary the punishment scheme, i.e. the fine size and the probability of detection. The experiment by Schildberg-Hörisch and Strassmair (2010) contains 6 treatments with different fine sizes and detection probabilities. The results show that small sanction increases the violation frequency, whereas a large sanction decreases the theft frequency. Since the majority of the treatments differ in both variables, it is difficult to disentangle the effect of the fine from the effect of the detection probability. Only a pair of the treatments vary the fine size while keeping the detection probability constant. Looking only at these two treatments, the average amount taken from the others increases with the fine. The experiment by Harbaugh et al. (2013) confirms the theoretical prediction that theft frequency is decreasing in fines and detection probability. Still, the propensity to steal is slightly more responsive to changes in detection probability than to the size of the fines; the elasticities are -0.4 and -0.3 respectively.

Although the laboratory studies confirm that compliance depends positively on fines, they also reveal some limitations of using fines as an enforcement tool. The effect of increased penalties on compliance is less robust than the effect of audit frequency. In particular, the effect is small or non-existent when the audit probability is low. It seems that more severe penalties might increase compliance only when they are accompanied by medium or high levels of detection probabilities.

2.4 Enforcement limitations

The effect of audit probabilities and fine magnitudes seems to be theoretically well established and empirically confirmed. However, there are also some limitations regarding the possibility to enforce compliance through audits and fines. The obvious problem is that audits are costly and fines usually cannot exceed some socially acceptable upper bound. Moreover, some empirical evidence suggest that higher fines enforce compliance only when audit probability is sufficiently high. These limitations may be solved by using smarter mechanisms that choose subjects to be audited. These mechanisms

– so called endogenous audit selection mechanisms – will be introduced in the next chapter.

Other limitations are not so obvious. The basic enforcement model does not include many key aspects which may seriously limit the efficiency of audits and fines. This section raises two important problems which will be subject to further inquiry in the following chapters.

2.4.1 Concealment

Apparently, if an individual decides to evade taxes or not to meet the regulatory standard, she will make every effort to conceal such behaviour. The possibility to conceal raises two problems. An individual needs to spend real and scarce resources to conceal non-compliance. For example, in the tax compliance context, concealment activities may include high-end services of a tax advisor, but also various other costly actions such as keeping two sets of books, making false entries in the books and records, claiming false or overstated deductions on a return, hiding or transferring assets or income, etc. In the environmental regulation, the concealment activities may include establishing of the redundant sanitised areas or simply buying more land in order to decrease the probability that the non-compliant part of the plant will be find (Heyes, 2000). Since these activities are socially wasteful, the concealment investment itself creates welfare loss. Another problem is that concealment activities may reduce enforcement leverage. Increasing fines or audit probabilities does not have to lead to an increase in higher compliance since this can be offset by higher concealment.

Surprisingly, only limited attention has been devoted to this problem in both theoretical and empirical literature. Cremer and Gahvari (1994) present a theoretical tax compliance model which allows the taxpayers to invest some resources in order to influence the audit probability. Bayer (2006) models the concealment as a contest game between the taxpayers and the tax authority. The model is a game with two stages, a reporting decision followed by a contest game. The taxpayers, therefore, decide not only how much of their income to report[3] but also how much to invest in order to conceal tax evasion. On the other hand, the tax authority has to spend some resources in order to detect the evaded income. The concealment and the detection investment

[3] More precisely, Bayer (2006) suppose that the taxpayers have many sources of income and they are supposed to report each income separately. The decision is also binary. For each source of income, the taxpayers can be fully compliant or report zero income.

then jointly determine the probability that the evaded income will be detected. Naturally, the probability decreases with the concealment investment and increases with the detection investment. Note also that the timing of the game means that the tax authority cannot commit itself to a particular detection effort. The main result shows that higher tax rates imply higher tax evasion and more investment into wasteful concealment activities.

The following paper by Bayer and Sutter (2009) presents an experimental test of the simplified version of the model. The contribution of this paper is twofold. They model theoretically the effect of higher penalties. In particular, the model predicts that higher penalties lead to a higher concealment investment. Furthermore, they experimentally test both of these theoretical predictions. They conducted six treatments with varying tax rates and penalties. Since the decision to invest in concealment is not independent of the decision to evade, they use the Heckman sample selection model to estimate the results. Both the theoretical predictions were confirmed.

This discussion shows that strengthening enforcement can also backfire as people may react by investing more into activities that conceal non-compliance. To avoid this problem, one would need an enforcement tool that increases compliance and decreases concealment at the same time. The second experimental study presented in chapter 5 addresses this issue by examining the interaction between concealment activities and the endogenous audit selection mechanism.

2.4.2 Overcompliance

In the basic enforcement model, the rational agent will never exert compliance effort which exceeds the regulation target. However, overcompliance, i.e. the phenomenon in which firms voluntarily choose to overcomply with regulations, can be empirically observed. McClelland and Horowitz (1999) found that biochemical discharges from paper plants are only at 50 percent of the allowable limit. Shimshack and Ward (2008) use data on plant-level water pollutant discharges to empirically demonstrate that even in the cases when the compliance is almost full, an increase in fines may cause a significant reduction of discharges.

How can we explain the overcompliance? One explanation presumes that a consumer values regulatory compliance and henceforth overcompliance increases product differentiation Arora and Gangopadhyay (1995). Another theory explains overcompliance as a result of a strategic interaction between the firm and the regulator Denicolò (2008). A very simple and appealing explanation is suggested by Shimshack and Ward (2008). They argue that there are two problems that can give rise to regulatory violations. First, non-compliance can be caused by accident[4]. Second, non-compliance is the outcome of a voluntary act when the agent does not comply with the regulatory standard in order to save costs. The second problem is included in the basic enforcement model. The first problem means that the agent does not have full control over the regulated output (e.g. discharges, emissions, undisclosed income), and henceforth the overcompliance occurs because the agents want to be sure that they will not be fined if some accident happens.

In order to explain this idea more formally consider the following extension of the basic enforcement model. The agent exerts the compliance effort e. The final enforcement level R considered by the regulator is the sum of the exerted effort and a random variable ϵ with cumulative distribution function $F[-l, 0]$, $R = e + \epsilon$. The agent chooses the effort level in order to maximize his expected utility.

$$U = \pi \int u(I - \tau e - \phi \max\{(I - e - \epsilon), 0\}) dF(\epsilon) + (1 - \pi) u(I - \tau e)$$

What is the effect of these accidents? Since the expected fine is now higher, the agent chooses higher compliance effort. Even if we optimistically assume that the random variable is symmetric with mean zero, i.e. accidents are as probable as good luck events, the expected fine is at best the mean-preserving spread of the deterministic fine $\phi(I - e)$. A risk-averse decision-maker therefore increases the compliance effort which works as an insurance against the fine. It is even possible that the agent chooses an effort level that exceeds the regulatory target I. To illustrate this claim, consider an agent who is extremely risk-averse in that his payoff cannot fall below some threshold[5] (Rey and Tirole, 1986). Take zero to be this threshold; the agent clearly chooses effort level $e = I + l$.

The overcompliance problem reveals another limitation of enforcement through audits and fines. The regulation target should be chosen to maximize social welfare; i.e. social marginal benefits should be equal to the social marginal costs of the compliance effort. The enforcement should provide incentives to comply with this regulation. However, if the subjects overcomply, they invest inefficiently too much resources in the compliance effort. Chapter 6 presents

[4] Similar idea is included in the theoretical model of self-policing (Stafford, 2008).

[5] We can disregard whether this constraint is a result of the agent's preferences or high fines.

an experimental study that focuses on this problem and asks whether it can be solved by granting more discretionary power to the regulator.

Chapter 3

Moving on from basics

There is a consensus in the literature that the observed levels of tax compliance cannot be explained solely by the financial incentives captured by the basic enforcement model. Despite the fact that the levels of audit and penalty are set low so that the majority of rational individuals should evade, most people, in general, comply. This observation called compliance puzzle Torgler (2002), motivates the majority of empirical research on tax compliance. The discrepancy between the observed and the predicted levels of compliance may be explained by two distinct mechanisms. The first explanation stresses the importance of psychological phenomena such as tax morale or norm compliance. By this explanation, there is either some intrinsic utility or social pressure that provides additional incentives to comply with laws and regulations.

An alternative set of explanations is that individuals do not perceive, correctly or not, the audit probability as exogenous, or that they may overestimate the audit probability. In order to illustrate this line of reasoning, consider that the audit probability π in the basic model is not a constant but an increasing function of the undisclosed output $\pi(Z)$. The first-order condition of the model now becomes

$$\pi'(U(\overline{w}) - U(\underline{w})) + U'(\overline{w})(1-\pi)\tau = U'(\underline{w})\pi(\phi - \tau)$$

where $\overline{w}$ and $\underline{w}$ again refer to the monetary wealth in the non-audited and the audited case respectively. The assumption that the audit probability is increasing in the undisclosed outcome, i.e. decreasing in the disclosed outcome $\pi' < 0$, ensures that the compliance is higher compared to random audits.

Whatever explanation is the right one, it shows us other approaches to increasing regulatory and tax compliance without spending scarce resources on audits. The regulatory or tax authority may want to design an audit selection mechanism that explicitly defines the audit probability as a function of an undisclosed output.

3.1 Audit selection mechanism

One of the major tools of tax and regulatory agencies to increase regulatory compliance is an audit. By an audit, we mean not only the audit rate but also the mechanism used for selecting taxpayers to be audited. The simplest method of selecting taxpayers is the random audit selection rule, in which each taxpayer can be chosen for audit with a fixed and exogenously given probability. However, audits and regulatory monitoring incur substantial costs, and regulatory agencies are constrained by their budget. Therefore, regulatory and tax agencies cannot simply audit everyone and they should select the most suspicious regulated entities for audit. These audit selection mechanisms are usually called endogenous which means that the probability of being selected for an audit depends on the actual or the past behaviour of the taxpayers.

3.1.1 Dynamic audit selection mechanism

The first type of endogenous audit selection mechanism is a dynamic rule in which the probability of an audit depends solely on the past behaviour of the taxpayer. The dynamic audit selection mechanisms were proposed first by Landsberger and Meilijson (1982) in the context of tax compliance and Harrington (1988) in the context of regulatory compliance. They propose an audit selection mechanism in which the tax or the regulatory agency divides the regulated agents into two groups G_1 and G_2. The groups differ in the severity of enforcement. The audit probability in group G_i is p_i and the fine for non-compliance is F_i. The audit probability as well as the fine is lower in group 1, i.e. $p_1 < p_2$ and $F_1 < F_2$. The agents can move between the two groups based on their past compliance. Any Detected non-compliance in the group G_1 is punished by a being moved to the group G_2 and detected compliance in the group G_2 is rewarded by a chance of being transferred to the group G_1 with the probability u. The regulated agent chooses whether to comply in groups 1 and 2. The agent's optimal strategy depends on the values of the compliance costs and the expected fines. The agent does not comply if the compliance costs are low and complies in both groups if they are high. The most interesting is the case of medium compliance costs when the agent complies in group G_2 and does not comply in group G_1.

The agency's problem is to find an optimum audit policy that minimizes the number of audits while being given some compliance target and a maximum possible fine. Although Harrington (1988) does not offer a complete solution to this problem, he is able to compare this enforcement mechanism with random audits. The comparison results in two conclusions. First, a dynamic mechanism has greater leverage, which means that it can enforce some compliance even when the compliance costs are lower than maximum fines and the static mechanism is ineffective. Second, the dynamic mechanism is more cost effective, which means that it can enforce the same level of compliance with fewer audits.

There is theoretical literature that aims at enriching and extending the Harrington's model. Friesen (2003) improves the transition structure in Harrington's two group model. In the original model, the group assignment depends solely on the past compliance behaviour. Unlike this, Friesen (2003) shows that the optimal strategy is to assign the agents randomly into the group G_2 (target group with the stronger enforcement) and give them the possibility to return back into the group G_1 (group with the weaker enforcement) when being found compliant. This new audit scheme should increase compliance compared with the original model. Liu et al. (2013) considers that a regulator can make only a fixed number of audits that needs to be divided between the target and the non-target group. The optimal assignment of audits between groups has to take into account the competitive effect when firms compete with each other for a place in the target group.

Stafford (2008) provides a more interesting extension by incorporating so called self-policing into the model of a dynamic audit. Self-policing is a policy when a regulated entitiy can voluntarily undertake audits and subsequently disclose and correct any violations of the regulation. If it does so, the entity is eligible for a reduction in the punitive part of the penalty. Stafford (2008) uses the same model as Harrington (1988) with two groups G_1 and G_2. In addition to the original model, there are two types of non-compliance: intentional and unintentional. There are random events interpreted as accidents or unintentional non-compliance. Besides the decision whether to comply, the agents decide whether to undertake an audit to discover whether an accident occurred. The audit is costly. If the audit is undertaken and an accident is found, the agent returns to compliance at some costs and reports the accident. The dynamic audit selection mechanism in this setting works as follows. The regulator can observe one of four possible situations in the previous period: compliance, violation, voluntary disclosure, no information. The agents who are known to comply will move to G_1 with probability q if they begin in G_2 and stay in G_1 if they begin in G_1. Agents that were found to violate the regulation will be in group G_2. The agents that voluntarily disclose and have been found compliant would stay in G_1 if they already were in G_1 and move to G_1 with probability p if they were in G_2. The agents whose behaviour is not observed stay in their current group. A comparison of the equilibria of the model with and without this dynamic audit selection mechanism, Stafford (2008) illustrates that the dynamic audit selection mechanism provides stronger incentives to comply as well as to disclose unintentional violations.

The empirical properties of the dynamic audit selection mechanisms have been examined in the experimental literature. Alm et al. (1993) investigates two types of dynamic audits, so called future audits, and past audits. In the future audit mechanism, subjects who were found to be non-compliant are audited again after a certain period of time. The past audit rule looks backward which means that subjects who were found to be non-compliant are audited covering a specific period in the past. Alm et al. (1993) conducted an experiment with four treatments: random audit, future audit, past audit and threshold audit mechanism. The threshold mechanism sets an income threshold and each taxpayer who reports less than the threshold will be audited. The results show that compliance was always greater in each of the three endogenous treatments (future, past, and threshold) than in the random treatment. This conclusion holds even in treatments with a very high audit probability of 50 percent. The most efficient endogenous audit selection rule was the threshold mechanism. The compliance rate in the threshold mechanism was 0.808 whereas the compliance rate in random treatment was 0.492 which is 64 percent increase in compliance. It also leads to 46 percent higher compliance than the future audit and 44 percent higher compliance than the past audit. However, these comparisons are not without problems because different treatments involved a different number of audits. The seemingly most efficient threshold mechanism involved the highest number of audits. Alm et al. (1993) were trying to solve this problem by running a random treatment with several different levels of audit probabilities. Then they compared the compliance rate observed in the endogenous audit treatment with the compliance rate observed in the random treatment with the closest matching audit probability.

The dynamic audit selection mechanisms studied by Alm et al. (1993) are arbitrary and their theoretical properties are not really known. Clark et al. (2004) reports an experimental test of a dynamic audit selection mechanism that is well established in the theoretical literature: Harrington's (Harrington, 1988) and Friesen's audit selection mechanism (Friesen, 2003). They use a random audit mechanism as a control treatment. Moreover, the experimental design takes into account also the audit costs, i.e. the variable of interest is not only the compliance rate but also the number of audits (recall that the number of audits and compliance rate are endogenously and jointly determined by the agent's behaviour in these models). The audit probabilities in the experiment were 1 in the target group and 0.6 in the non-target group. The assignment probabilities are set in such a way that the theoretically predicted compliance rate is 0.5; and it remains the same for all three treatments. The results confirm that the dynamic audit selection mechanisms lead to fewer audits when compared with an equivalent random audit mechanism. However, only the difference between random audits and the Friesen's audits is significant. A more surprising result is that the observed compliance levels under random, Harrington's and Friesen's audit selection mechanisms were 0.456, 0.391 and 0.332 respectively. The compliance rate in the dynamic audit selection mechanism was significantly lower than the compliance rate under the random audits. These results suggest some serious limitations of the dynamic audit selection mechanisms. While such a mechanism is able to decrease the costs of audits, its ability to increase compliance is questionable.

The compliance result reported by Friesen (2003) is driven by the behaviour of the subjects in the target group. While the theory predicts that the subjects in the target group should be fully compliant, the compliance rate was significantly lower. Cason and Gangadharan (2006) conducted a similar tax compliance experiment that replicated these results. Each session of the experiment consisted of two within-session treatments. Subjects were sorted randomly into one of two groups at the beginning of the experiment. Subjects in the non-target group faced a low fine and a low audit probability. In the target group, subjects were audited with a higher probability and faced higher fines. If a subject in the first group was found to be non-compliant, they were switched to the target group. on the other hand, if a subject in group two was found to comply, they could switch over to the non-target group with a given probability. The switching probability and the compliance costs varied between the treatments. The results of the experiment confirmed the theoretical predications that i) a subject should be more compliant in the target group ii) compliance is higher when a probability of switching between groups is higher iii) compliance is higher when the compliance costs are lower. Similar to Friesen (2003), the the observed compliance does not change between the target and the non-target group as extremely as the theory predicted. Cason and Gangadharan (2006) shows that this result may be explained by assuming that the subjects are boundedly rational. In particular, the quantal response equilibrium model explains the observed data well.

3.1.2 Competitive audit selection mechanism

Another type of an endogenous audit selection mechanism is the so called competitive audit selection mechanism. While the dynamic audit selection mechanism is based on the past compliance history, the competitive audit selection mechanism uses observed differences in the behaviour of agents in order to select agents for audits. When such information about the other regulated individuals is available to be used for targeting audits, this naturally creates a contest amongst the regulated individuals not to be selected for audit.

The competitive audit mechanisms examined in both theoretical and experimental literature have different forms. The nature of the mechanism depends to some extent on whether the mechanism was proposed in a tax compliance context or in a regulatory compliance setting. The mechanism proposed by Alm and McKee (2004) in a tax compliance model uses the relative ranking of taxpayers based on their reported income within a group of taxpayers with the same income. Each taxpayer is assigned a DIF score which is the difference between the average reported income in the peer group and the taxpayer's reported income. The taxpayer with the highest DIF score is then selected for audit.

The competitive audit selection mechanism has a slightly different form in a regulatory context. In the audit mechanism introduced by Gilpatric et al. (2011) it is assumed that the regulator has noisy but unbiased information about the non-compliance measure. The regulator then audits the agent with the largest expected non-compliance. It is possible to describe the mechanism in more detail as follows. Each agent has some output x which is exogenously determined and they choose how much to disclose e. Obviously, you can also interpret x as a regulatory target and e as a compliance effort. The non-compliance measure is the undisclosed output $z = x - e$ The regulator has noisy but unbiased information about the undisclosed output of every individual. Specifically, the regulator observes the measured undisclosed output

$m_i = z_i + \epsilon_i$ of each agent. A random error ϵ is independently and identically distributed across all the agents with a density function $f(\epsilon)$. The agent in the peer group with the highest measured undisclosed output is selected for audit, or more generally, n agents with the highest measured undisclosed output are selected for audit.

Suppose that we have a group of two agents. The probability that agent one will be selected for audit is as follows

$$Pr(m_1 > m_2) = Pr(z_1 + \epsilon_1 > z_2 + \epsilon_2) = G(z_1 - z_2)$$

where G is the distribution function of the difference between the error terms. This audit selection can be henceforth equivalently described as a function $\Pi(z_i, z_{-i})$ which states how the audit probability depends on the agent's undisclosed output and other agent's undisclosed output. The function is increasing in the agent's undisclosed output z_i and decreasing in the undisclosed output of the other agents z_{-i}. The particular functional form depends on the probability distribution of the error term.

The efficacy of the competitive audit selection mechanism has been not only proposed theoretically but also tested in laboratory experiments. Alm and McKee (2004) conducted an experiment that tested the mechanism based on a DIF score. The experiment was designed as a repeated game in order to test whether the participants would coordinate on a low level of compliance. The results show that the DIF rule is able to achieve a higher level of compliance; the participants coordinated on low compliance levels only when cheap talk communication was allowed. Gilpatric et al. (2011) tested the efficacy of the competitive audit selection mechanism by comparing three treatments: random, tournament and a generalized version of the tournament. In the tournament treatment, the subject's undisclosed output was error-adjusted by computer and the subject with the highest adjusted undisclosed output was audited. In the generalized version, subjects were confronted with the generalized audit probability function. Unlike in Alm and McKee (2004), the subjects were rematched after each of 20 periods. The results confirmed that the compliance effort in the treatments with the competitive audit selection mechanism (tournament and generalized version) was significantly higher.

The following theoretical and experimental literature offers several extensions of the approach pioneered by Gilpatric et al. (2011). Gilpatric et al. (2015) shows that the competition between regulated agents can extend also the working of a dynamic audit selection mechanism. He suggests a novel audit selection mechanism that contains dynamic as well as competitive elements.

The mechanism again contains two groups: target and non-targeted. The transition between the groups is not given by a fixed probability conditional on compliance or non-compliance, rather it is determined by the rank order tournament within the group. The properties of this mechanism are again confirmed by a laboratory experiment. The recent studies by Dai (2016) and Kamijo et al. (2017) enrich the compliance model by considering endogenous crackdowns. A crackdown is a sudden increase in an audit probability triggered by low levels of compliance. For instance, Dai (2016) suggest a compliance game with two very different audit probabilities 0.2 and 0.9. The players are matched in groups. If the audit reveals that a given number of players did not comply, a crackdown is triggered meaning that the audit probability jumps up to a higher level and stays there until all the audited players are found compliant.

An interesting extension is presented by Cason et al. (2016). They extend the model for such situations when the output is endogenously chosen by the agent. The theoretical model predicts that the output level is not affected by the audit selection mechanism. This calim is tested by an experiment which consists of four treatments. The random and the endogenous audit selection mechanism were conducted with two levels of feedback information. In the high information treatment, subjects received information about the output and the penalties of the other peer group members. Despite the prediction, Cason et al. (2016) found significant differences in the output across the treatments. The output in the endogenous treatment is lower and closer to the social optimum. The endogenous audit selection mechanism can, therefore, pay a double dividend by reducing non-compliance and moving the output closer to the optimum.

In general, we can make two very robust conclusions out of this stream of literature. First, the experimental evidence confirms the theoretical predictions that the competitive audit selection rules always perform better in terms of compliance. Second, the compliance levels observed in experiments do not always fit those predicted by theory. This second observation is not a big surprise taking into account the usual and simplifying assumptions of the theoretical models such as risk neutrality or symmetric Nash equilibrium solutions. Despite this discrepancy, the experimental evidence suggests that the theoretical framework based on the enforcement model is a powerful instrument in terms of comparative statics and qualitative predictions.

The theory predicts that the competitive audit selection mechanisms should be successful in increasing compliance and the experimental evidence confirms this prediction. Despite the success of the competitive audit selection mechanisms, there are at least two open questions about the working of the competitive audit selection mechanisms. First, it is not clear whether the use of the competitive audit selection mechanisms does not increase the concealment investment (recall one of the problems describes in section 2.4). Second, the success of the competitive audit selection mechanisms hinges on the assumption that the regulator has noisy but unbiased information about the each individual non-compliance output z_i. This assumption may be valid in the environmental regulatory setting where the regulator may have such information about pollution. However, in many other settings, this assumption is too strong. For instance, the tax authority would need to have unbiased information about each taxpayer's taxable income in order to successfully apply the competitive audit selection mechanism. Experimental study presented in chapter 4 deals with this problem.

3.2 Discretion

Until now, we have reduced the role of regulatory agenicies to just two activities – conducting random audits and imposing strictly determined fines. However, the task of regulatory agencies is more complicated. Regulatory agencies enforce regulatory targets using imperfect information on the individuals they regulate (Laffont and Tirole, 1993). If a regulatory agency possesses some, albeit imperfect, information about the regulated party, it may be beneficial to grant the regulatory agency some flexibility in its decision making. More specifically, the regulatory authorities should be free to determine who will be audited and they may be able to choose from a variety of possible penalties when regulatory violations are detected.

There is limited empirical evidence on the value of the regulator's discretionary power. Duflo et al. (2018) conducted a field experiment in collaboration with the Indian environmental agency for emission pollution. The experiment doubles the rate of audits in the treatment group and ensures that the additional audits are assigned randomly. Surprisingly, these additional audits increased compliance only very slightly. The data on the original (non-treated) audits provides evidence that the main reason for such a negligible effect is the removal of the regulators' discretion over which facilities to audit. Therefore, they estimate a structural model that explains the incidence of original audits by audit costs and the regulator's discretion. The results show that audit decisions are driven mainly by the regulator's discretion. The regulator observes some small part of the pollution produced by the plants and uses this information to target the plants with higher pollution signals which highly improves the enforcement efficiency. Kang and Silveira (2018) explores the heterogeneity in the penalties imposed by a certain regulator. Their study takes advantage of the exogenous change in the enforcement of water quality regulations in California. The institutional changes in the dataset made it possible to to identify the dischargers' compliance cost function. Once the compliance cost function is identified, Kang and Silveira (2018) can estimate whether the heterogeneity in fines is determined by the compliance costs or the regulator's preferences. The results show that the heterogeneity is driven mainly by the compliance costs and the residents' preferences. The penalties are more severe in places where the residents value the quality of the water more. Finally, the simulation exercise shows that a one-size-fits-all enforcement policy increases both the level and dispersion of non-compliance.

There is a dark side to the regulator's discretionary power as well. Discretion may lead to enforcement decisions that reflect the regulators' personal objectives, rather than the social objectives of the regulation. The deviation from the personal and social objectives may be due to regulatory capture (Laffont and Tirole, 1991) or "minimal squawk" behaviour of public officials(Leaver, 2009). This effect states that when an informed interest group is involved, public officials may take inefficient decisions in order to keep the interest group quiet and maintain its reputation. Leaver (2009) tests the presence of this effect using the data on electricity regulation by the Public Utility Commission. The positive effect of the term length probability that an electric utility faces a new rate review rejects the regulatory capture hypothesis and provides support for the "minimal squawk" hypothesis.

When deciding whether to grant discretionary power to the regulator, both the advantages and the disadvantages should be taken into account. Shavell (2007) introduces a model where the state authority has to decide whether to impose a strict rule or to give discretionary power to an agent called adjudicator. The social welfare, as well as the adjudicator's utility, depend on the decision taken by the adjudicator and on variables which are observable only by the adjudicator and not by the state. This modeling framework reveals the basic trade-off faced by state authorities. If a state authority imposes a strict rule, it maximizes the social welfare function, but at the same time, it ignores the unobservable variable. On the other hand, if the state authority opts for discretion, it allows the adjudicator to follow his own utility function while incorporating the unobservable variable into the decision. Shavell (2007) henceforth argues that discretion is desirable if the loss caused by the inflexibility of the strict rule is larger than the loss caused by the deviation of the adjudicator's utility function form the social welfare function.

A regulator with discretionary power may tailor the enforcement to reflect the compliance costs or the possible harm caused by non-compliance (Kang and Silveira, 2018). If the regulator has proper incentives, discretion may be a powerful instrument not only to increase compliance but also to solve overcompliance. However, the disadvantages of the discretionary regime are not always generated by inappropriate incentives. In the presence of dynamic inconsistency problems, the regulator needs to get rid of discretion in order to achieve a more efficient outcome (Kotakorpi, 2006). The trade-off between dynamic inconsistency and flexibility will be studied in the third experimental design presented in chapter 6.

Chapter 4

A competitive audit selection mechanism with incomplete information

4.1 Introduction

Competitive audit selection mechanisms are powerful tools with which to increase tax and regulatory compliance (Gilpatric et al., 2011; Cason et al., 2016). The available competitive audit selection mechanisms were proposed mostly in the context of environmental regulations, and they are based on the assumption that the enforcement authority has noisy, but unbiased, information concerning the level of non-compliance. For instance, in the case of emissions regulations, the enforcement authority has available an unbiased estimate of each plant's actual emissions and observes the reported emissions. By taking advantage of this information, the enforcement authority can rank the regulated subjects according to their levels of non-compliance and then select them for audits according to this rank. Although this assumption may be valid in the environmental regulation setting, it is questionable at best in the tax compliance setting. The efficacy of the competitive audit selection mechanism in the tax compliance context where the tax authority lacks the information about each taxpayer's taxable income remains an open question.

This chapter proposes an competitive audit selection mechanism which is based solely on reported output and uses experimental methods to test whether the mechanism has a positive impact on the compliance of taxpayers with heterogeneous incomes. The experimental design addresses the problems created by the high information requirements of the audit selection mechanisms proposed in the literature: Gilpatric et al. (2011) assume that the auditors possess noisy, but unbiased, information about taxpayers' true incomes, and Alm and McKee (2004) suppose that the tax office can divide the taxpayers into subgroups having the same income. This last supposition might not be possible because true incomes can be observed only when taxpayers are audited, which means that the tax authority might have the income data for only for a few taxpayers in some sectors. The tax office may also face additional problems, even in sectors with a high number of audits. For example, the incomes of some taxpayers may vary significantly between years, or the incomes of individual taxpayers may depend on idiosyncratic factors, such as personal contacts or luck, which are not known to the tax authority. As a result, at least in some sectors, auditors do not have unbiased estimates of true incomes at the individual level, and the incomes will be heterogeneous in any group that the tax office is able to select. Given these two problems, the ranking of taxpayers reflects not only their undisclosed incomes but also the heterogeneity of their actual incomes. This chapter contributes to the current literature by examining what the effect of such a competitive audit selection mechanism is on compliance.

Competitive audit mechanisms based on a disclosed, rather than undisclosed, output were studied theoretically by Bayer and Cowell (2009) and Oestreich (2015). They both show that these competitive audit mechanisms lead to higher tax compliance. However, the output in these models is endogenously chosen by the taxpayers, and the solution is based on the symmetric Nash equilibrium concept, which means that the output is homogeneous on the equilibrium path. Our theoretical predictions follow a model which is similar to the declaration stage for the model presented by Bayer and Cowell (2009). Unlike Bayer and Cowell (2009), the output is given exogenously and is heterogeneous by assumption. In addition to this literature, this chapter examines the properties of the competitive audit selection mechanism in a situation in which the taxpayers have heterogeneous income or, more generally, when the subjects are heterogeneous in the variable they are supposed to report. The theoretical solution shows that the competitive audit selection mechanism, in which the audit probability depends on the reported incomes of taxpayers, leads to higher reported incomes than random audit selections, in which all taxpayers are selected with the same exogenously given probability.

This prediction is tested using an economic experiment. We propose a design in which all taxpayers receive income that is drawn from a uniform distribution. Their task is to choose the reported income which is then taxed at a fixed rate. They may be selected for audit with a certain audit probability. Subjects who are selected for audit and report less than their income pay a penalty. The experiment has three treatments that differ in the way the audit probability is determined: the random audit selection mechanism, the competitive audit selection mechanism based on disclosed income (incomplete information) and the competitive audit selection mechanism based on undisclosed income (complete information). In the random audit selection mechanism, the probability of audit is exogenous and the same for all taxpayers. In the treatments with a competitive audit selection mechanism with incomplete information, the subjects are divided into groups of five taxpayers and their audit probability decreases with the difference between each subject's reported income and the average reported income of the other four subjects in their group. The treatments for the competitive audit selection mechanism with incomplete information are the same, except the audit is based on undisclosed, rather than reported, income.

The rest of the chapter is structured as follows. Section 4.2 introduces the theoretical model. Next, section 4.3 provides the experimental design. The sections 4.4 and 4.5 present the data and the results of the experiment. Finally, section 4.6 concludes.

4.2 Theoretical framework

4.2.1 Model description

This section describes theoretical model that creates framework for the experiment. Taxpayer i receives an income I_i drawn form a distribution $F(I)$ with support $[\underline{I}, \bar{I}]$. The taxpayer chooses the reported income $R_i \in \langle 0, I_i \rangle$. Reported income is taxed by rate τ, so the taxpayer pays a tax τR_i. The taxpayer i is audited with a probability $\pi_i(R_i, R_{-i})$. The formula for π_i depends on the audit selection mechanism used by the tax authority. If the taxpayer is chosen for audit, she pays a fine which depends on the undisclosed income, $\phi(I_i - R_i)$ where ϕ is the fine rate and $\phi > \tau$.

We examine three different audit selection mechanisms. Under the random audit selection mechanism, the audit probability is the same for all taxpayers regardless of their reported income, $\pi_i = p$. When the competitive audit selection rule with limited information is applied, the audit probability depends on both the income reported by the taxpayer and the income reported by other taxpayers. In particular, the form of our audit selection rule resembles the generalized audit selection rule proposed by Gilpatric et al. (2011) closely. It has the following form:

$$\pi_i = p - \delta \left(R_i - \frac{\sum R_{-i}}{N-1} \right),$$

where N is the number of taxpayers in a group. This mechanism assigns higher audit probability to the taxpayers with relatively lower reported income. Parameter p defines the basic audit probability, and parameter δ defines the sensitivity of the audit selection rule to the disclosed income. The random audit selection mechanism is obviously a special case of the competitive audit selection mechanism for which $\delta = 0$.

The audit selection rule with complete information has the same functional form, but it uses undisclosed income instead of reported income:

$$\pi_i = p + \delta \left(Z_i - \frac{\sum Z_{-i}}{N-1} \right),$$

where $Z_i = I_i - R_i$ is the undisclosed income. The formula for this audit selection mechanism can be rewritten as

$$\pi_i = p - \delta \left(R_i - \frac{\sum R_{-i}}{N-1} \right) + \delta \left(I_i - \frac{\sum I_{-i}}{N-1} \right).$$

We can see that the difference between the two competitive audit selection mechanisms is the last term in the previous equation. This term shows that the audit selection mechanism with complete information acounts also for the differences in actual incomes.

The important aspect of the audit selection mechanism with limited information is that the audit probability depends only on the difference between the individual's reported income and the average reported income of other taxpayers. Therefore, this audit selection mechanism does not require information concerning each taxpayer's income or the income distribution in their peer group. On the other hand, it loses a substantial amount of information since the differences in the reported income may be caused not only by different compliance but also by differences in actual incomes. While the audit selection mechanism with complete information can differentiate between these effects, the mechanism with incomplete information cannot.

4.2.2 Equilibrium

Risk-neutral taxpayer i chooses the reported income R_i in order to maximize the expected wealth:

$$W = (1 - \pi(R_i, R_{-i}))(I_i - \tau R_i) + \pi(R_i, R_{-i})((1 - \tau)R_i + (1 - \phi)(I_i - R_i)).$$

Under the random audit selection mechanism, the first derivative of the expected wealth can be written as $p\phi - \tau$. Suppose that the tax rate τ is higher than $p\phi$. If this is the case, then the optimal choice of the risk-neutral taxpayer is to evade paying taxes on all of her income.

Now, we derive the solution for the competitive audit selection mechanism with limited information. Suppose that the other player's strategy is $R(I_i)$, where R is a non-decreasing function. The first order condition is given as follows:

$$\left(p - \delta R_i + \delta \int R(I_i) dF(I)\right)\phi - \tau + \delta\phi(I_i - R_i) = 0$$

We focus on the symmetric equilibria of the model. The solution of the first order condition is a linear function $R = a + bI$, where the reported income must be in the interval $[0, I]$. Depending on the values of the exogenous parameters, the taxpayer's optimal choice may be constrained by the upper bound I or the lower bound 0. This gives us two possible equilibrium strategies[1].

The first equilibrium occurs when the tax rate is low compared to the audit probability and the fine. In this situation, the low-income taxpayers declare their whole income, as they face a high probability of being audited. The taxpayers with higher income face lower probabilities of being audited, and they optimally react by evading being taxed on some amount of their income. Formally, the equilibrium strategy has the following form:

$$R(I_i) = \begin{cases} I_i & \text{if } I_i < \hat{I} \\ a + bI_i & \text{if } I_i \geq \hat{I}, \end{cases} \tag{4.1}$$

where $a > 0$. The equilibrium values of the parameters a, b and $\hat{I}$ are given by the solutions of the following three conditions (the last condition is derived from the equation $a + b\hat{I} = \hat{I}$), respectively:

$$a = \frac{p\phi - \tau}{\delta\phi(1 + F(\hat{I}))} + \frac{F(\hat{I})}{1 + F(\hat{I})} E(I|I < \hat{I}) + \frac{1 - F(\hat{I})}{2(1 + F(\hat{I}))} E(I|I > \hat{I})$$

[1]These strategies do not form multiple equilibria. Only one of these strategies forms an equilibrium depending on model parameters.

$$b = \frac{1}{2}$$

$$F(\hat{I})E(I|I < \hat{I}) + \frac{(1 - F(\hat{I}))}{2}E(I|I > \hat{I}) - \frac{1 + F(\hat{I})}{2}\hat{I} = \frac{\tau - p\phi}{\delta\phi}$$

The second equilibrium is applicable when the tax rate is relatively high compared to the audit probability and fine. The taxpayers with lower income evade paying taxes on all their income, as they risk losing a small amount of money if they are audited. On the other hand, taxpayers with high income have a lot of money at stake, so they will optimally disclose some income. In formal terms, the equilibrium strategy has the following form:

$$R(I_i) = \begin{cases} 0 & \text{if } I_i < \hat{I} \\ a + bI_i & \text{if } I_i \geq \hat{I} \end{cases} \tag{4.2}$$

The equilibrium values of the parameters a, b and $\hat{I}$ are given by the solution of the following three conditions (the last condition is derived from the equation $a + b\hat{I} = 0$), respectively:

$$a = \frac{p\phi - \tau}{\delta\phi(1 + F(\hat{I}))} + \frac{1 - F(\hat{I})}{2(1 + F(\hat{I}))}E(I|I > \hat{I})$$

$$b = \frac{1}{2}$$

$$\frac{(1 - F(\hat{I}))}{2}E(I|I > \hat{I}) - \frac{1 + F(\hat{I})}{2}\hat{I} = \frac{\tau - p\phi}{\delta\phi}$$

Both equilibria under the competitive audit selection mechanism with limited information result in higher tax compliance compared to using random audits. This result forms the main hypothesis of our experiment.

Still, the lack of information about taxpayer's undisclosed income affects the taxpayers' equilibrium strategies. Suppose that the audit selection rule is based on undisclosed income. The best-response function is defined implicitly by the following first-order-condition in which $Z(I)$ denotes the equilibrium strategy of the other players:

$$\tau - \left(p + \delta Z_i - \delta \int Z(I)dF(I)\right)\phi + \delta\phi Z_i = 0$$

In the symmetric Nash equilibrium, the taxpayer's strategy is to evade taxes on some fixed amount c. If thei income is lower than this amount, then they will evade paying taxes on all their income.

$$Z(I_i) = \begin{cases} I_i & \text{if } I_i < c \\ c & \text{if } I_i \geq c \end{cases} \tag{4.3}$$

The amount of evaded taxes c is determined by the following equation:

$$p + \delta F - \delta\phi(F(c)E(I|I < c)) + (1 - F(c)c) - \tau = 0.$$

In the next section, we will use this solution to obtain testable experimental predictions. Still, at this point, we can observe one stark contrast between the equilibrium strategies under complete and incomplete information. The slope of the equilibrium strategy is lower when the audit selection mechanism does not have complete information. Henceforth, higher-income taxpayers evade more taxes when estimates of individuals' incomes are missing.

4.3 Experimental design and procedures

4.3.1 Treatment and predictions

The properties of the competitive audit selection mechanism are tested in an economic experiment. Each experimental session consists of 30 rounds. At the beginning of each round, taxpayer i receives income I_i, which is drawn from a uniform distribution between 0 and 200 CZK[2], meaning that the income of each taxpayer differs in each period. The task of each taxpayer is to choose a reported income $R_i \in \langle 0, I_i \rangle$. This reported income is taxed at a rate $\tau = 0.6$. Taxpayers who are selected for audit and report less than their income pay a penalty equal to their unreported income $I_i - R_i$ (the fine rate equals $\phi = 1$)[3].

The experiment contains three treatments which differ in the way their audit probabilities are calculated. In the treatment with the random audit selection mechanism (random treatment), the audit probability of taxpayer i equals $\pi_i = p_i = 0.4$. In the treatment with the competitive audit selection mechanism with incomplete information (incomplete treatment), taxpayers

[2]This amount was equal to approximately to 7 at the time of the experiment.

[3]Setting the fine rate equal to one ensures that a taxpayer cannot have a negative payoff in any period.

are divided into groups of five taxpayers. We use partner matching in order to increase the learning effect. Each taxpayer's audit probability equals

$$\pi = 0.4 - 0.004 \left(R_i - \frac{\sum R_{-i}}{N-1} \right),$$

where R_{-i} stands for the incomes of the four remaining members of the group. The payoff of taxpayer i in a given period depends on whether he or she has been selected for audit in this period. The taxpayer receives $0.4R_i$ if audited, and $0.4R_i + (I_i - R_i) = I_i - 0.6R_i$ if not selected for audit. The treatment with the competitive audit selection mechanism with complete information (complete treatment) is similar to the incomplete treatment, with the only exception being that the audit selection mechanism is based on undisclosed income instead of disclosed income. In particular, the audit probability equals

$$\pi = 0.4 + 0.004 \left(Z_i - \frac{\sum Z_{-i}}{N-1} \right),$$

where $Z_i = I_i - R_i$.

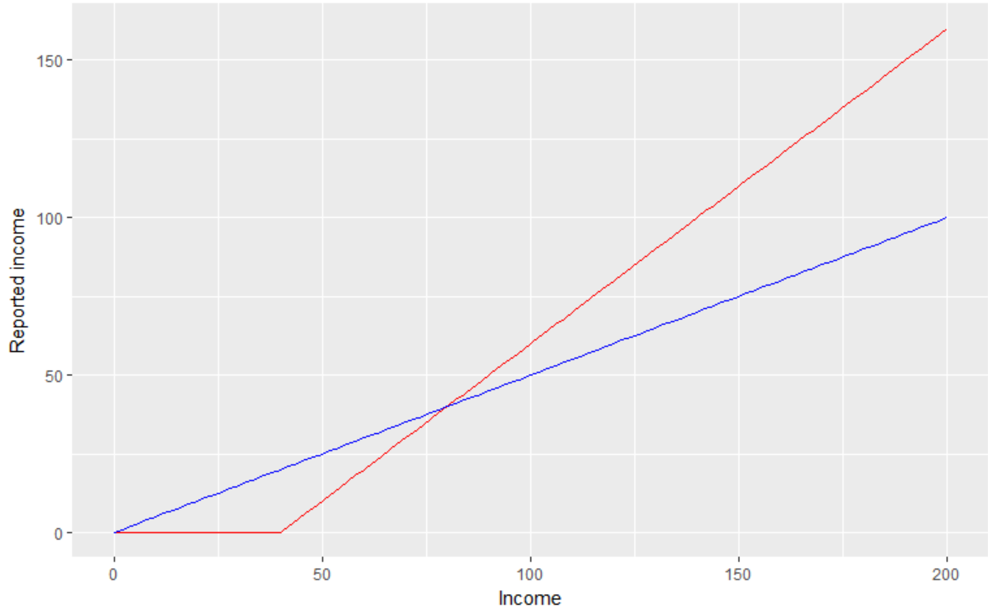
Now, it is possible to calculate equilibrium strategies for the parameters used in the experiment. The disclosed income in the random treatment is zero. In the incomplete treatment, the disclosed income should be one half of the actual income, $R_i = I_i/2$. Under the complete treatment, the disclosed income is zero when the actual income is below 40 and each additional income above 40 is fully disclosed, i.e. $R_i = \max\{0, I_i - 40\}$. The average disclosed incomes in the random, incomplete and complete treatments are 0, 50 and 128, respectively. This calculation results in the first hypothesis.

Hypothesis 1: The average disclosed income in the complete treatment is higher than in the incomplete treatment. The average disclosed income in the incomplete treatment is higher than in the random treatment.

However, the comparison of the average disclosed incomes does not tell the whole story. The figure 4.1 shows how the theoretically predicted reported income depends on the actual taxable income. It documents that the effects of different audit selection mechanisms depend on the levels of taxpayers' actual incomes. Taxpayers with incomes lower than 80 CZK are expected to comply more fully in the incomplete treatment. On the other hand, taxpayers with higher income should disclose more in the complete treatment. This generates our second hypothesis.

Hypothesis 2: For low income levels, the disclosed income will be higher in the incomplete treatment compared to the complete treatment. For medium-

Figure 4.1: The figure shows the equilibrium strategies for the uniform distribution. The red line is the tax payer's equilibrium strategy when the audit is based on undisclosed income. The blue line represents the equilibrium strategy when the audit is based on disclosed income.



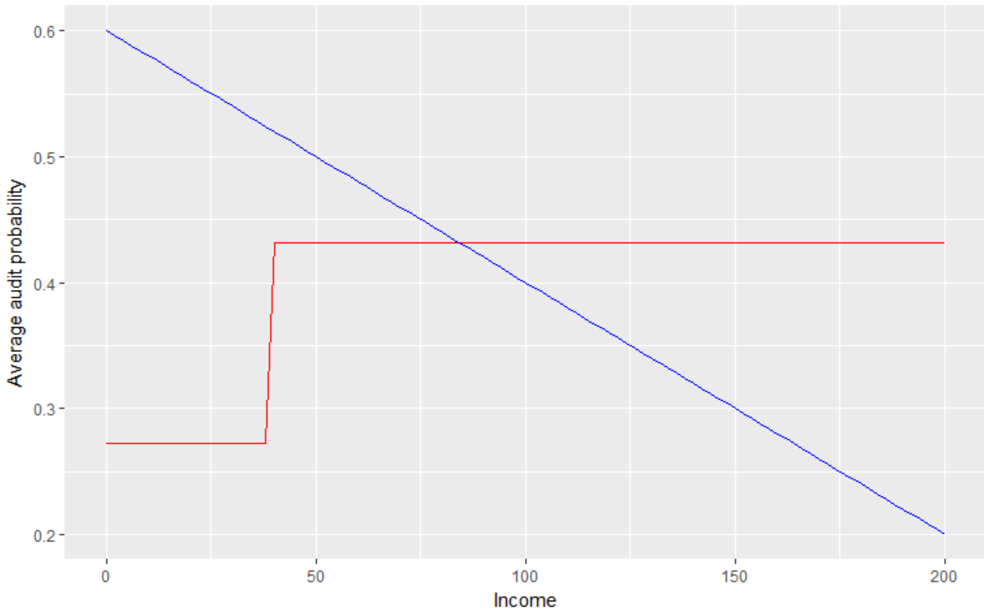
and high-income levels, the disclosed income will be lower in the incomplete treatment compared to the complete treatment.

The figure 4.2 sheds further light on the difference between low- and high-income individuals. The figure depicts the relationship between equilibrium audit probability[4] and income level. Since the audit selection mechanism is incomplete, the treatment lacks the information about each individual income, and the high-income individuals are audited with lower probability. In the complete treatment, the audit selection mechanism is based on the actual undisclosed incomes, which results into constant audit probabilities. Only the taxpayers with incomes below 40 CZK are audited with lower probability, which reflects that they are constrained by the lowest possible disclosed income of zero. The relationship in figure 4.2 generates the third hypothesis.

Hypothesis 3: The equilibrium audit probability is decreasing for the incomplete treatment and non-decreasing for the complete treatment.

[4]Recall that the audit probability depends on the taxpayer's compliance behaviour.

Figure 4.2: The figure shows the relationship between expected audit probability and income. The red line is for the case when the audit is based on undisclosed income. The blue line is for the case when the audit is based on disclosed income.



4.3.2 Procedures

At the beginning of each experimental session, an experimenter read the instructions aloud, with the subjects (taxpayers) following along with their copy. Subjects were asked to take a quiz in order to reinforce comprehension of the instructions before the experiment. To avoid the risk of anchoring, the quiz did not include any particular numbers. All numerical inputs in the quiz were entered by subjects themselves.

The experiment was conducted at MUEEL in Brno, Czech Republic, in May 2017. The subjects were mostly students and recruited through hroot (Bock et al., 2014). The experimental environment was prepared in zTree (Fischbacher, 2007). We used neutral instructions, i.e. the tax motivation of the game is not clear from the instructions (the instructions in the original Czech language are enclosed in Appendix A). We ran 11 sessions using a between-subjects design. In particular, we ran three sessions for the random treatment for which each subject was considered to be one independent observation, four sessions for the incomplete treatment where groups of five were considered to be one independent observation and four sessions for the complete treatment where each group of five was considered an independent observation. The total number of participants was 200, with no less than 15 participants in each session. The sessions lasted almost 90 minutes. The subjects received payoffs from five randomly selected rounds. The mean payoff was 240 CZK (approx. 9 EUR).

4.4 Data

A total of 200 subjects participated in the experiment. Each of them played for 30 periods, individually or as part of a group for 6,000 observations in total. We filter out the observations in which the income is $I = 0$ because, under those circumstances, the subjects had no opportunity to evade taxes. As it is standard in similar experimental literature founding their predictions on equilibrium models, we do not use data from the early rounds in the analysis (e.g. Gilpatric et al., 2011). In particular, we use the data only from the last 15 rounds. As a robustness check, all results were estimated using all 30 periods. The results remain the same, or at least very similar, in terms of statistical as well as economic significance.

Table 6.4 displays the descriptive statistics for the selected variables for the three treatments. The table includes choice variables as well as socio-demographic variables. There are fewer subjects in the random treatment since, in that case, each subject is considered as an independent observation, whereas a group of five constitute an independent observation in the other two treatments. Although our sample is not balanced in terms of gender, this should not bias the results since we control for personal characteristics in the regressions. Approximately one half of the subjects were students of economics or business. Some of our subjects had previously participated in other economics experiments, but they had not participated in a similar tax compliance experiment.

4.5 Results

In the analysis, we use parametric approaches to estimate the effect of the incomplete and random treatments first. The dependent variable in models in the table 4.2 is the disclosed income, calculated as the individual average disclosed income for the last 15 periods. The variables Incomplete and Random are treatment dummies. Complete treatment is a contrast. It shows

Table 4.1: Descriptive statistics

	Complete	Incomplete	Random
Subjects	80	75	45
Groups (Independent observations)	16	15	45
Income	98.3	101.4	101.2
Undisclosed income	26.9	37.8	47.9
Female	0.63	0.51	0.4
Age	22.2	23.0	21.5
Students of economics or business	0.53	0.55	0.58

that incomplete information decreases disclosed income by 10 CZK, which is 10 % of the average actual income. The change is significant at the 5% level. Still, the competitive audit selection mechanism with limited information performs better than the random mechanism. The disclosed income under random audits is about 20 CZK lower (20 % of the average actual income) compared to that for the competitive audit selection mechanism with complete information. The difference between the random and incomplete treatments is also statistically significant at a 5 % level (p-value 0.028). These effects are robust in terms of significance and magnitudes even if we allow for a non-linear relationship between the disclosed and actual incomes (model 2 in table 4.2) or include personal characteristics in the model (model 3 in table 4.2). Gender, age and the field of study do not have any effects on compliance.

Note that the change is not only statistically significant but also economically important. Consider an individual with an average income of 100 CZK. For this individual, the use of the competitive audit selection mechanism with incomplete information increases the level of compliance from 52 % to around 62 %, i.e. by more than 20 %. When the information about each taxpayer's individual undisclosed income exists, the level of compliance increases by another 20 % compared to compliance under random audits.

The significance of the treatment effects was also tested using group averages since these observations are independent. We calculate the mean undisclosed incomes for each group in the last 15 periods and compare those using either the t-test or the non-parametric Mann-Whitney U test. The t-test shows that there is a statistically significant difference between the incomplete and random treatments (t-test $p = 0.04$). When conducting the Mann-Whitney U test for these treatments, we face a problem in that we have group averages for the incomplete treatment and individual averages for the random treatment. In order to obtain the same units of observation and equal vari-

Table 4.2: OLS model based on individual averages

	Disclosed Income		
	(1)	(2)	(3)
Constant	0.034	−236.080	2.243
	(8.176)	(248.137)	(8.973)
Incomplete	−10.005***	−9.864**	−9.714***
	(3.835)	(3.863)	(3.643)
Random	−20.236***	−20.662***	−21.095***
	(4.659)	(4.661)	(4.712)
Income	0.726***	8.166	0.721***
	(0.083)	(7.632)	(0.083)
Income^2		−0.076	
		(0.077)	
Income^3		0.0003	
		(0.0003)	
Female			1.891
			(2.652)
Age			−0.391
			(0.431)
Econ study			−0.774
			(2.366)
Observations	200	200	200
R^2	0.327	0.334	0.331
Adjusted R^2	0.317	0.317	0.310
Note:			*p<0.1; **p<0.05; ***p<0.01
			Standard errors clustered at the group level

ances in the random and incomplete treatments, we need to compare groups of similar size. We, therefore, clustered the individual observations from the random treatments into groups of five members and performed the statistical tests using these group averages. This procedure was repeated 4,000 times with different clusterings. The test confirms a significant difference between treatments (median $p = 0.029$). The results are robust for different clusters since the rank-sum was higher in the random treatment than in the incomplete treatment for all 4,000 repetitions. The differences between them are also statistically significant (t-test $p = 0.008$, Mann-Whitney $p = 0.008$). These results confirm our first hypothesis.

The second hypothesis states how the disclosed income depends on actual income. Henceforth, we use disaggregated data in order to test this hypothesis. Model 1 in the table 4.3 just replicates previous results by using separate observations for each period. All three models reported in the table 4.3 include as controls socio-demographic characteristics, fixed-period effects and a dummy variable which takes the value 1 if the taxpayer was audited in the previous period. Standard errors are clustered at the group level.

The model in column 2 in the same table also includes interaction terms between the treatment variables and income level. We can see that the dummy variable for incomplete treatment is now positive and statistically significant, which shows that low-income individuals disclose more of their income in the incomplete treatment. The marginal effect of higher income is nearly 0.9 in the complete treatment and 0.6 in the incomplete treatment. These values are relatively close to the theoretically predicted values. More importantly, they confirm that the slope of the disclosed income function[5] is higher in the complete treatment than in the incomplete treatment. Henceforth, the rankings of the disclosed incomes in these two treatments are reversed for high income levels. Namely, the model estimates that the disclosed income in the complete treatment exceeds the disclosed income in the incomplete treatment when the income is above 70 CZK.

Model 3 in the table 4.3 allows for a non-linear relationship between the disclosed and actual incomes. In particular, the model includes dummy variables for five income brackets (variables IB1 to IB5) and the interactions of these income brackets with the treatment. The first income bracket IB1 denotes an income below 40 CZK. Variable IB2 denotes an income between 40 CZK and 80 CZK. Each successive income bracket denotes an income interval that ends 40 CZK higher than the previous one. For the sake of clarity in terms of the regression table, model 3 is estimated using only data from the complete and incomplete treatments. We can again see that the complete treatment leads to higher compliance only in high income brackets (IB3, IB4 and IB5). In order to better comprehend the relationship between income level and disclosed income, the figure 4.3 plots the estimate of this

[5]The disclosed income function assigns the equilibrium disclosed income to actual income.

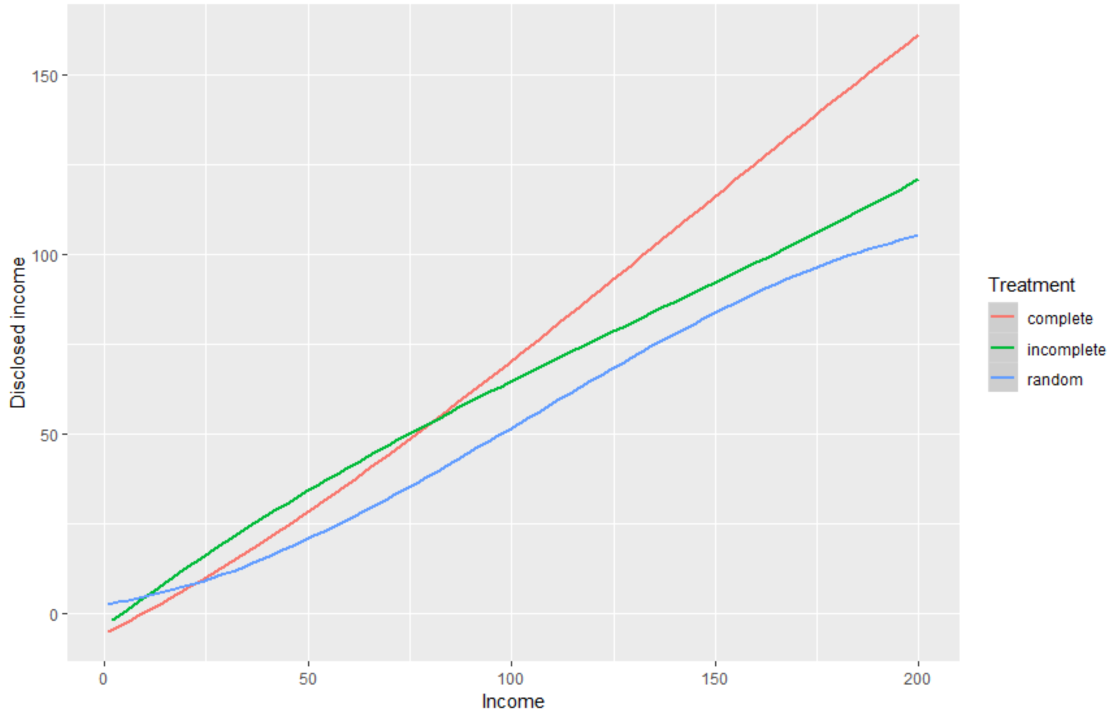
Table 4.3: OLS model explaining disclosed income

	Disclosed Income		
	(1)	(2)	(3)
Constant	10.623^{**}	-7.329^{**}	-9.268^{***}
	(4.710)	(3.591)	(1.728)
Incomplete	-9.712^{***}	17.031^{***}	5.777^{***}
	(3.553)	(1.909)	(1.356)
Random	-21.359^{***}	7.963^{**}	
	(4.638)	(3.187)	
IB2			-1.227
			(2.159)
IB3			7.011^{**}
			(3.537)
IB4			11.230^{**}
			(4.821)
IB5			17.433^{***}
			(6.158)
Income	0.571^{***}	0.863^{***}	0.742^{***}
	(0.083)	(0.024)	(0.038)
Audit_{t-1}	-2.006^{**}	-2.060^{**}	-1.455
	(0.921)	(0.869)	(1.070)
Incomplete:Income		-0.268^{***}	
		(0.038)	
Random:Income		-0.294^{***}	
		(0.056)	
Incomplete:IB2			0.296
			(2.271)
Incomplete:IB3			-14.789^{***}
			(3.747)
Incomplete:IB4			-26.794^{***}
			(4.446)
Incomplete:IB5			-38.793^{***}
			(6.454)
Period effects	Yes	Yes	Yes
Socio-demographic	Yes	Yes	Yes
Observations	2,790	2,790	2,317
R^2	0.656	0.680	0.757
Adjusted R^2	0.653	0.677	0.754

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors are clustered at the group level

relationship for all three treatments. The figure and the regression table 4.3 provide support for our second hypothesis.

Figure 4.3: The figure plots the estimated relationship between disclosed income and actual income. The lines are estimated by the cubic polynomial function.



The third hypothesis is also confirmed by experimental data. The table 4.4 reports on models that estimate how the audit probability depends on the income level. The logit model in column 1 shows that the audit probability is increasing slightly with income for the complete treatment. On the other hand, high-income individuals are audited less often in the incomplete information treatment. Column 2 shows the estimate of a linear probability model providing an easier interpretation of the marginal effects. We can see that the audit probability is almost 60 % for low-income individuals in the incomplete treatment and that it decreases to 20 % for individuals with the highest incomes. The figure 4.4 plots this observed relationship between audit probabilities and income levels. We can see that the estimate very much corresponds to the theoretically predicted relationship (figure 4.2). In the complete treatment, the audit probabilities are increasing for the low-income levels and relatively stable for the medium- and high-income levels.

Table 4.4: Model explaining audit probabilities

	Audits	
	Logit	OLS
	(1)	(2)
Constant	-0.614^{***}	0.350^{***}
	(0.107)	(0.025)
Incomplete	1.011^{***}	0.244^{***}
	(0.145)	(0.034)
Random	-0.020	-0.004
	(0.213)	(0.050)
Income:Complete	0.002^{**}	0.001^{**}
	(0.001)	(0.0003)
Income:Incomplete	-0.007^{***}	-0.002^{***}
	(0.001)	(0.0003)
Income:Random	0.002	0.001
	(0.001)	(0.0003)
Observations	2,790	2,790

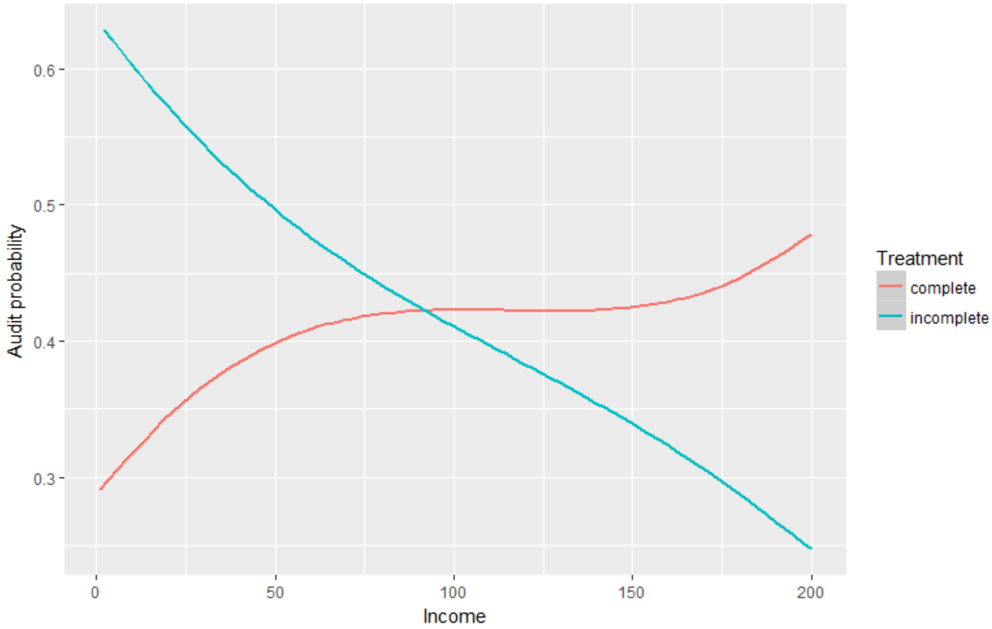
Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors are clustered at the group level

4.6 Discussion

The competitive audit selection mechanism is based on the assumption that the enforcement authority has noisy, but unbiased, information about each individual's regulated output. This assumption may be very restrictive in some settings, especially in the tax compliance setting.

In this chapter, we propose a competitive audit selection mechanism that is based only on the reported incomes of taxpayers, and we examine its properties. In this mechanism, the audit probability depends only on the reported incomes of taxpayers within a reference group. Using experimental methods, we show that the proposed mechanism leads to higher tax compliance than a random mechanism in which all taxpayers are audited with the same baseline probability. In particular, we show that the mechanism works even if the incomes of the taxpayers in the reference group differ substantially.

Figure 4.4: The figure plots the estimated dependence of audit probabilities on actual income. The lines are estimated by the logit model for which the logarithm of the odds is a cubic polynomial function of actual income



Furthermore, the mechanism is designed in such a way that the expected level of the audits is kept constant, which means that the additional cost of implementing the mechanism consist only of the administration of the more complex audit selection procedure. In sum, our paper suggests that the competitive audit selection mechanism might be an affordable and effective tool for reducing tax evasion even if the tax office does not have information about the actual income of taxpayers and is not able to place taxpayers in reference groups of taxpayers with similar incomes.

However, the efficiency of this mechanism is seriously limited. The low-income taxpayers in the reference group are audited more frequently. This may be difficult to justify from a social justice point of view. Moreover, it decreases the efficiency of the audit selection mechanism since high-income taxpayers may evade taxes and still face low audit probabilities. This limitation becomes more pronounced when the taxpayers in the reference group become more heterogeneous in their incomes.

Chapter 5

Compliance and concealment activities

5.1 Motivation

In the biggest tax evasion prosecution in the history of the United States, Walter Anderson was charged with using a complex scheme to conceal approximately \$450 million in earnings from the US authorities. The scheme involved forming offshore firms in the British Virgin Islands and Panama with the aim of concealing Mr. Anderson's investments in several telecommunication companies whose stock prices increased dramatically during the 1990s.[1] Walter Anderson was by no means the only person to be involved in such shady transactions. The recent Panama and Paradise Papers[2] suggest that people have not only avoided and evaded taxes on a scale larger than previously suspected, but also are willing to expend substantial amounts to conceal the services that these offshore firms offer. Moreover, the activities of the offshore firms represent only the tip of the iceberg. In the United States, the estimated average annual tax gap in 2008–2010 was \$458 billion, corresponding to a compliance rate of 81.7% (IRS, 2016). Similarly, concealment activities may include the high-end services of offshore companies as well as various other dubious actions such as keeping two sets of books, making false

[1] US v. Anderson, 384 F. Supp. 2d 32 (D.C. 2005).

[2] The Panama Papers refer to the 11.5 million confidential documents of the Panamanian law firm Mossack Fonseca leaked on April 15, 2016 (Harding, 2016), and the Paradise Papers are the 13.4 million documents that originated from several companies registered in offshore jurisdictions and leaked on November 5, 2017 (Zerofsky, 2017).

entries in books and records, claiming false or overstated deductions on a return, and hiding or transferring assets or income [3].

As these examples show, tax evasion represents not only the fiscal loss of governments, but also the welfare losses induced through socially wasteful activities, such as taxpayers spending real resources to conceal their taxable income.

The problem of concealment is relevant not only in tax compliance, but also for environmental and in many other realms of regulation. For instance, in the now infamous dieselgate, Volkswagen realized that they could not meet the though emission regulations, and at the same time offer sufficient performance. The solution was to create and gradually refine a software which recognized a situation when the car is tested and turned on the emission control. The non-compliance was discovered only in 2014 when data collection commissioned by the California Air Resources Board showed that the real-world-driving emissions exceed the regulatory NO_x limits up to 40 times.

In general, regulatory non-compliance creates two types of losses, the loss caused by non-compliance itself and the loss caused by spending scare resources on concealment activities. Enforcement policies usually aim at decreasing the non-compliance via audits and fines. This chapter focuses on a situation when concealment is possible, and the enforcement policy needs to follow two goals: reduce non-compliance and reduce investments into concealment activities. It examines how the audit selection mechanism influences non-compliance and concealment investments. In particular, we examine the effect of two policy changes. We change the average audit probability, and we replace the random audit selection mechanism with the competitive audit selection mechanism. Since the agents in the competitive audit selection mechanism are not randomly selected for audit, they compete in avoiding being audited by choosing the level of compliance.

From a theoretical perspective, the chapter presents an enforcement model in which costly concealment activities are possible. The introduction of competitive audit selection mechanism into the model reduces both non-compliance as well as concealment activities. This leads to better outcomes compared to the increase in basic audit probability, which reduces the non-compliance but increases the socially wasteful concealment activities. The obvious problems with measurement of non-compliance and concealment investments make it

[3] http://www.bizfilings.com/toolkit/sbg/tax-info/fed-taxes/tax-avoidance-and-tax-evasion.aspx (Accessed February 2017)

very difficult to identify these effects using naturally occurring data. We, therefore, use experimental methods to test these predictions and obtain a pattern consistent with the comparative statics of the proposed theoretical model.

A large volume of economic studies examine tax or regulatory compliance (see Slemrod, 2007; Alm, 2012; Luttmer and Singhal, 2014; Mascagni, 2018; Oestreich, 2017, for recent surveys). This paper builds on two strands of this literature, one studying the competitive audit selection mechanisms, and the other examining concealment activities.

In a competitive audit selection mechanism, the individual's audit probability depends not only on his or her compliance level but also on the compliance level of other individuals.[4] Two forms of competitive audit mechanisms have been examined in the experimental literature. The mechanism proposed by Alm and McKee (2004) uses the relative ranking of taxpayers based on their reported income in a group of taxpayers having the same income. Here, the taxpayer with the lowest reported income is selected for audit. Alm and McKee show that this mechanism results in higher tax compliance compared to the random audit selection rule. On the other hand, Gilpatric et al. (2011) propose a different competitive audit selection mechanism in the context of environmental regulation enforcement, assuming that the enforcement agency has noisy but unbiased information about the compliance effort. The enforcement agency audits the agents with the largest observed non-reporting, which is the difference between the noisy observations of the actual and reported effort. This structure can be generalized to an audit selection rule wherein the probability of being audited is a function decreasing in the agent's reported effort and increasing in the average reported effort of the other agents in the peer group. The Gilpatric et al. (2011) analysis results confirm that such a rule can result in higher tax compliance. This is in line with the findings of Cason et al. (2016).

Our approach follows Gilpatric et al. (2011) in proposing an competitive audit selection mechanism in which the audit probability of an agent is decreasing in the agent's reported effort and increasing in the average reported effort of the other agents in the peer group. This formulation of the audit selection mechanism is relevant not only in the regulatory setting as in Gilpatric et al. (2011) but also in the tax compliance setting. In this setting, we follow Alm and McKee (2004) in assuming that the tax authority can form groups

[4] A related literature examines the dynamic audit selection mechanisms in which the audit probability depends on the taxpayers' compliance histories (Alm et al., 1993; Clark et al., 2004; Cason and Gangadharan, 2006; Stafford, 2008; Liu et al., 2013).

of agents with the same (or similar) output. The agent with the lowest reported output is then audited more frequently. The stochastic nature of the audit selection mechanism may be due to purposeful design or may reflect the reality that the agents in the peer group are likely to be slightly heterogeneous in their outputs, and hence, the agent with the lowest compliance does not have to be audited with certainty.

The literature on concealment activities is usually framed in terms of tax evasion. It focuses on how the tax rate and tax penalty affects concealment activities. Bayer (2006) provides a theoretical model in which the detected tax evasion depends on the taxpayer's investment in concealment activities and the auditor's investment in detection. The audit is therefore modeled as a detection-concealment contest. Studying this situation, Bayer and Sutter (2009) experimentally show that concealment activities depend positively on the tax rate, and not on the penalty imposed when the tax evasion is detected. This chapter contributes to this literature by developing a theoretical model and an experimental design that examine how audit probability and the competitive audit selection mechanism affect compliance and concealment activities.

The rest of the chapter is structured as follows. Section 5.2 introduces the theoretical model. Section 5.3 translates the model into an experimental design. Section 5.4 describes the experimental data. Section 5.5 presents the results of the experiment. Finally, section 5.6 concludes the chapter.

5.2 Model

To obtain testable predictions, we develop a simple theoretical model. The model is framed as disclosure of some activity called output. The agent chooses what amount of output to disclose and how much to invest in concealment activity. The concealment activity determines the probability that the regulatory authority finds the undisclosed output when the agent is selected for audit. However, any concealment activity is costly with increasing marginal costs. Audits can be determined randomly or using a competitive audit selection rule, where agents who disclose lower output face a higher probability of being selected for audit. In the tax compliance setting, the output can be interpreted as a taxable income. Alternatively, the output can be understood as a required level of some regulated activity and disclosed output as an actual level of the activity.

5.2.1 Description of the model

This subsection presents the formal model. All the agents have the same output q. The probability that an agent is chosen for audit will be denoted as Π. The agent discloses output $r \in [0, q]$, and chooses the probability p that the undisclosed output $u = q - r$ will be verified by the regulatory authority when the agent is selected for audit. The reported output r is subject to a constant marginal cost τ, which may be caused for instance by taxation or environmental regulation. The choice of the verification probability p induces monetary costs $c(p)$, where $c(p)$ is a decreasing and convex function with properties $c(1) = 0$ and $c(0) \to \infty$. The audit selection rule decides whether the agent is selected for audit. If the agent is selected for audit, nature decides whether the undisclosed output will be verified. If the agent is audited and the undisclosed output is verified, the agent will have to pay a fine in proportion to the undisclosed output $\phi(I - r)$, where the parameter ϕ is the fine rate.

The audit selection rule determines the probability Π of the agent being selected for audit. We examine two audit selection rules. According to the random selection rule, every agent has the same probability $\Pi = \pi$ of being selected for audit. Under the competitive audit selection rule, the probability of an agent being audited depends on the output reported by the agent as well as that reported by other agents, $\Pi(r_i, r_{-i}, \pi, \beta)$, where π and β are parameters defining the audit selection rule. Parameter π defines the basic audit probability, while parameter β defines the sensitivity of the audit selection rule to the disclosed output. The competitive audit selection rule has the following properties:

- The audit probability is decreasing and non-concave in the agent's reported output, that is, $\Pi'_{r_i} < 0$ and $\Pi''_{r_i} > 0$, and increasing in the other agents' output, that is, $\Pi'_{r_{-i}} > 0$.
- If the reported output of all the agents is the same, the audit probability is π; that is, $\Pi = \pi$ if $r_i = r_{-i}$.
- The higher the parameter β, the higher is the sensitivity of the audit probability with respect to the undisclosed output u_i, that is, $\Pi''_{r_i\beta} < 0$ or alternatively $\Pi''_{u_i\beta} > 0$.
- The random audit selection rule is a special case of the competitive audit selection rule; that is, $\Pi = \pi$ if $\beta = 0$.

We assume that the agent is risk-neutral. Agent i chooses the reported output and verification probability that maximize his expected payoff, denoted as U_i:

$$U_i(r_i, r_{-i}, p_i) = \Pi(r_i, r_{-i})p_i(I - \tau r_i - \phi(q - r_i)) + (1 - \Pi(r_i, r_{-i})p_i)(q - \tau r_i) - c(p_i). \tag{5.1}$$

The first term is the agent's payoff when he is audited, and undisclosed output is verifiable. The second term is the agent's payoff when the undisclosed output is successfully evaded. The last term $c(p_i)$ represents the cost of the concealment activity.

In a symmetric and interior equilibrium, the optimal choice of the reported output r^* and verification probability p^* is defined by the following equilibrium conditions:

$$-\pi\phi(I - r^*) - c'(p^*) = 0 \tag{5.2}$$

$$-\Pi'_r \phi p^*(I - r^*) - \tau + \pi p^* = 0. \tag{5.3}$$

The equilibrium conditions under the random audit selection rule represent a special case of these conditions when $\Pi'_r = 0$.

5.2.2 The effect of the audit selection mechanism

In this subsection, we show the effect of the audit selection mechanism on equilibrium values of the reported output r^* and verification probability p^*. In particular, the enforcement authority can change the audit selection mechanism in two ways. The enforcement authority can invest more resources and carry out more audits by increasing the basic audit probability π, and it can make the audit selection more dependent on the comparison of different agents by increasing the sensitivity parameter β. The effects of these two policy changes are stated in Propositions 1 and 2.

Proposition 1: Suppose that condition $\tau > 2\phi p^* \pi$ holds. Then, the equilibrium reported output and the equilibrium verification probability are increasing in the sensitivity parameter.

Proof. The formal proof is presented in the section 5.7.2

This proposition establishes that a more competitive audit selection mechanism forces agents to not only disclose more output but also to invest less in concealment activities. The intuition behind this result is straightforward. The verification probability remaining constant, a more sensitive audit selection mechanism incentivizes the agent to disclose more output because it reduces the audit probability by a higher margin. Once the agent shows higher compliance, the marginal benefit from decreasing the verification probability $\pi\phi(q - r^*)$ is lower, but the marginal cost of concealment $c'(p)$ remains the same. This second-order effect results in lower concealment investment.

Proposition 2: Suppose that condition $\tau > 2\phi p^*\pi$ holds. The equilibrium reported output is increasing, but the equilibrium verification probability is decreasing, in the basic audit probability.

Proof. The formal proof is presented in the section 5.7.3

The reaction of the disclosed income to higher basic audit probability depends on two effects. While the first-order effect incentivizes the agent to disclose more output by increasing the probability of being audited, the second-order effect depends on whether the verification probability increases or decreases. Furthermore, the concealment investment changes following the two effects that can go in opposite directions. The first-order effect incentivizes the agent to invest more in concealment activities because of the higher marginal benefit of such an investment, $\pi\phi(q-r^*)$. However, the direction of the second-order effect depends on the change of the undisclosed output. Condition $\tau > 2p^*\phi\pi$ ensures that the first-order effect dominates the second-order effect.

Note that because this is a sufficient condition, the comparative static results may hold even if it is not satisfied. Moreover, this condition is likely to be satisfied in relevant empirical cases because the audit probability π is usually substantially lower than the marginal costs τ. Proposition 2 shows that the effect of a higher basic audit probability is ambiguous from a social point of view. While a higher basic audit probability increases the disclosed output, it also encourages investment in socially wasteful concealment activities.

5.3 Experimental design

We test the model predictions in an experiment. Each experimental session consists of 20 rounds. In the treatments with competitive audit selection mechanism, the subjects are divided into pairs of peer agents. To retain the one-shot incentives of the theoretical model and have more independent observations per session, each subject's partner is randomly selected from the same group of at least six subjects participating in a given session.

At the beginning of each round, all agents receive the same output $I = 100$. Their task is to choose the reported output $r \in (0, 100)$ and the verification probability $p \in (0.25, 1)$. The reported output is taxed at the rate $\tau = 0.6$. Each verification probability entails a concealment cost $c(p) = 20\left(\frac{1}{p} - 1\right)$. The lower bound of the verification probability is set to be 0.25 since $c(0.25) = 60$ and a choice of any verification probability less than 0.25 is strictly dominated by full compliance.

Next, the regulatory authority selects agent i for audit with the probability

$$\Pi_i = \pi + \beta\left(\frac{r_{-i} - r_i}{r_{-i} + r_i}\right), \tag{5.4}$$

where r_i is the reported output of agent i, r_{-i} is the reported output of the peer agent, π is the basic audit probability, and β determines the sensitivity to the reported output. We use $\pi = \{0.4, 0.6\}$ for the basic audit probability and $\beta = \{0, 0.4\}$ for the sensitivity parameter. Thus, we have four treatments, one for each combination of π and β parameters.

If agent i is selected for audit, his or her true output is verified with probability p_i. If the output is verified, the agent pays a penalty equal to her undisclosed output $u_i = 100 - r_i$. The agent i's payoff in a given round is $0.4r_i + u_i - c(p_i)$ if he or she is not selected for audit, or is selected but his or her actual output is not verified. Agent's payoff is only $0.4r_i - c(p_i)$ if he or she is selected for audit and the true output is verified.

Table 5.1: Model prediction

	Sensitivity parameter	
Audit probability	Random ($\beta = 0$)	Competitive ($\beta = 0.4$)
$\pi = 0.4$	$c^* = 8.17$	$c^* = 0$
	$u^* = 100$	$u^* = 50$
$\pi = 0.6$	$c^* = 14.48$	$c^* = 0$
	$u^* = 100$	$u^* = 0$

The predicted concealment cost c^* and unreported output u^* are summarized in Table 5.1. From this table, we can derive several hypotheses. A shift from random to competitive audit selection leads to a lower concealment cost, that is, to a lower investment into socially wasteful activities. It also reduces the undisclosed output, implying lower loss caused by non-compliance. The predicted effect of an increase in basic audit probability depends on the type of the audit selection mechanism. In the random audit selection mechanism, a rise in audit probability increases the concealment cost, with no impact on the undisclosed output. On the other hand, a higher audit probability in the competitive audit selection mechanism reduces the undisclosed output while keeping the concealment costs constant.

5.4 Data

All experimental sessions were conducted at MUEEL in Brno, the Czech Republic, using zTree (Fischbacher, 2007). Recruitment was carried out through hroot (Bock et al., 2014). The total number of participants was 280.[5] Some of them had previously participated in other economics experiments, but all were inexperienced in that they had never previously participated in a similar regulatory compliance experiment. We used neutral instructions. The instructions are enclosed in the Appendix B. We ran 16 sessions, 3 for each random and 5 for each competitive treatment using a between-subject design. The subjects received a show-up fee and payoffs from two randomly selected rounds in CZK. The mean payoff was 190 CZK. At the time of the experiment, this amount approximately equaled 8 EUR, which roughly corresponded to two times hourly wage of unqualified student labor in the Czech Republic. Descriptive statistics for each of the treatments are presented in Table 5.2.

At the beginning of each experimental session, an experimenter read the instructions aloud, with the subjects following with their copy. We did not use any quiz to reinforce the comprehension of instructions before the experiment to avoid the risk of providing different decision frameworks to the different treatments because we would have to use modified test questions for the random and competitive audit selection mechanisms. However, we used a decision environment in which all the relevant outcomes were calculated in real time as the students clicked on the verification probability or the reported output bars and provided detailed feedback after each round. The variables calculated in real time were the following: the concealment cost c for each level of verification probability, the probability of audit for a selec-

[5] A total number of 284 students took part in the experiment, but we excluded four Russian-speaking participants because it was very difficult for them to properly understand the experimental instructions in Czech.

Table 5.2: Descriptive statistics

	$\beta = 0$ $\pi = 0.4$	$\beta = 0$ $\pi = 0.6$	$\beta = 0.4$ $\pi = 0.4$	$\beta = 0.4$ $\pi = 0.6$
Subjects	49	51	92	88
Groups (Independent observations)	49	51	14	13
Disclosed output	33.84	40.65	43.00	61.46
Undisclosed output	66.16	59.35	57.00	58.54
Verification probability	0.79	0.73	0.83	0.80
Concealment cost	7.07	10.62	5.79	7.38
Frequency of paying a penalty	0.30	0.43	0.34	0.46
Tax revenues per audit including fines	98.96	81.30	113.54	90.33
Tax revenues per audit	50.77	40.65	64.49	61.46
Female	0.55	0.38	0.48	0.62
Age	21.47	21.31	21.48	22.02
Students of economics or business	0.55	0.81	0.66	0.64

ted declared output of the peer agent (only in the competitive audit selection mechanism), and the monetary outcomes for all the three possible outcomes of the game, (i) no audit, (ii) audit and no prosecution, (iii) and audit and prosecution. The feedback screen gave the outcome of the round and the corresponding payoff, the subject's own choices, the peer agent's reported output (only in the competitive audit selection mechanism), the probability of selection for audit, and the combined probability of selection for audit and prosecution. The screenshot of the decision environment can be found in the Appendix B.

To enable learning, we allowed the participants to play 20 identical rounds in each treatment. We find the learning effects, especially in the treatments with the competitive audit selection mechanism. Some of these sessions showed a clear trend in the outcome variables during the first 10 to 15 rounds (see Figure 5.1). Also, the learning effect can be illustrated by the frequency of choices violating the first-order stochastic dominance (FOSD). In the treatments with competitive audit selection mechanism, we consider as FOSD violation those choices of verification probability and disclosed output (p_i, r_i) where for any belief of the undisclosed output of the other player r_{-i}, there is always a different choice (p_i', r_i') that first-order stochastically dominates (p_i, r_i). The evolution of choices violating FOSD is shown in the Figure 5.2. We can see that the subjects made numerous choices violating FOSD in the first 10 to 15 periods, with the frequency of choices thereafter converging to 1 %. Thus, we follow the approach of Gilpatric et al. (2011) and use the data only from round 15 onward. Besides that, we discuss the results for the last 10 periods as a robustness check.

Figure 5.1: The figure shows the learning effect in the undisclosed income and concealment costs in the competitive treatments ($\beta = 0.4$). The first row shows the treatments with low audit rate ($\delta = 0.4$), and the second row shows the treatments with high audit rate ($\delta = 0.6$) in periods 1–20.

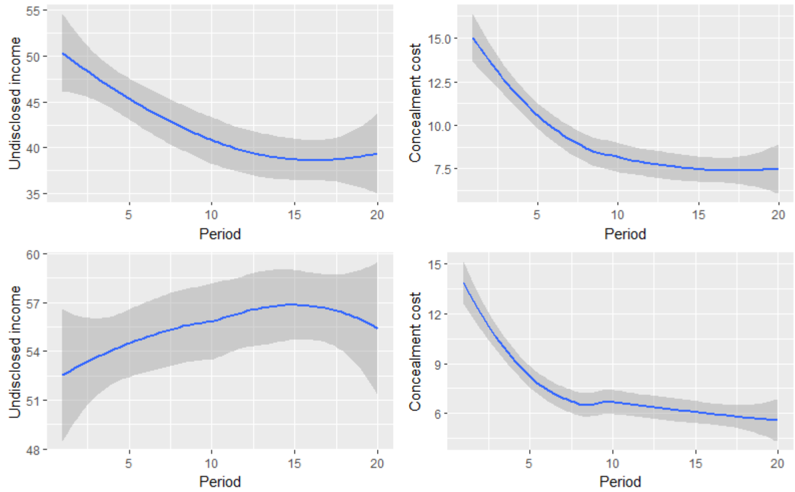
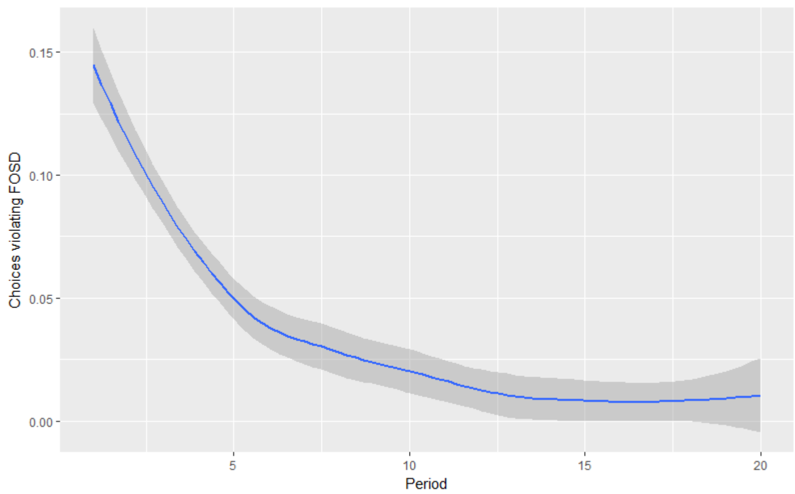


Figure 5.2: The figure shows the share of observations violating the first-order stochastic dominance in periods 1–20 in all treatments.



5.5 Results

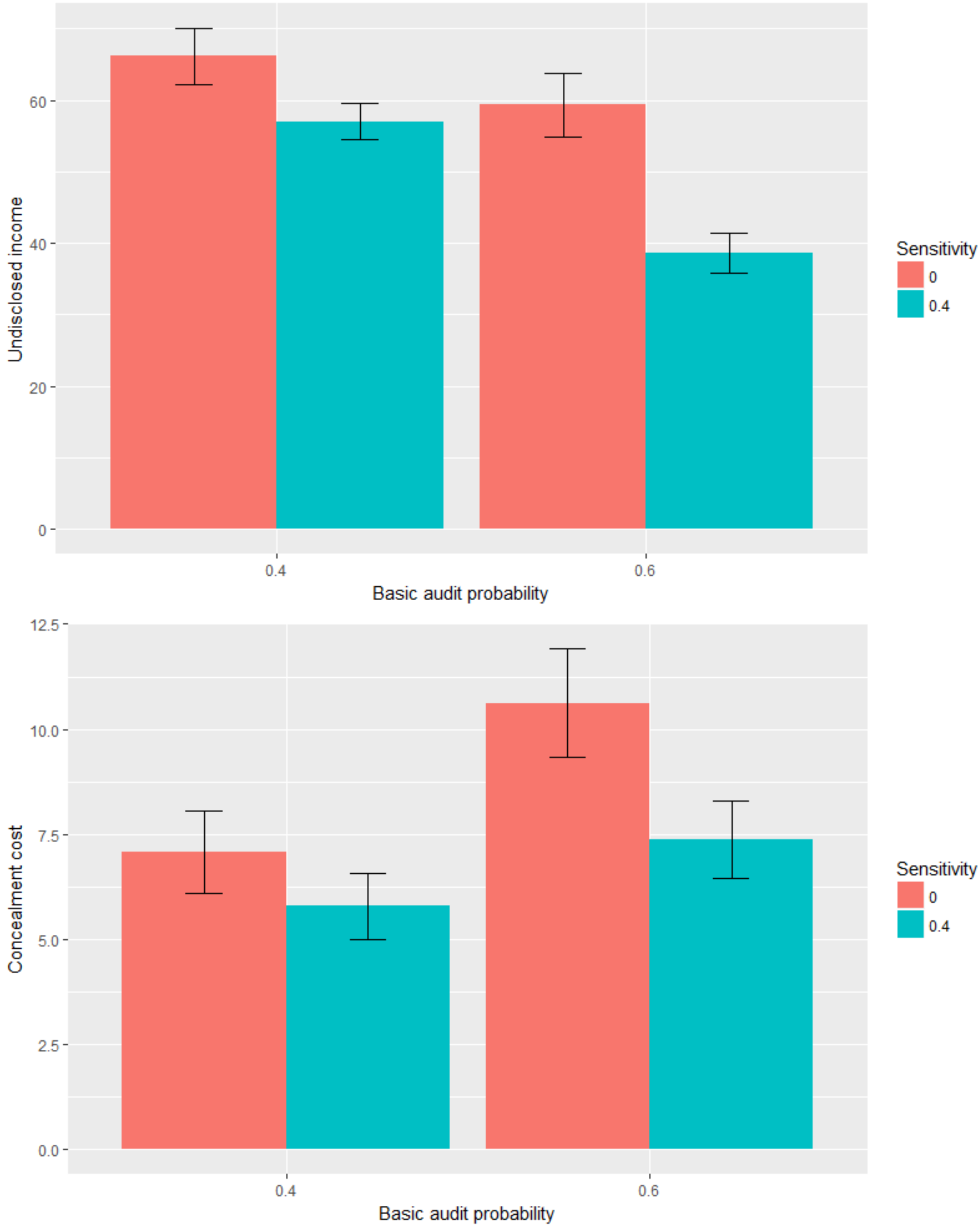
This section mainly focuses on how the treatment variables affect compliance and concealment investment. While the non-compliance is measured by the undisclosed output u, the concealment activities are measured by the concealment cost c. The group averages and standard errors for all the four treatments are presented in Figure 5.3.

When testing for statistical significance of the differences, we compare the group averages from the last 6 periods using either the t-test or the non-parametric Mann-Whitney U. We use one-sided tests if the model predicts the direction of the change (see Table 5.1), and two-sided tests if the model predicts no change. For the Mann-Whitney U test, we face a problem that we have group averages in competitive treatments and individual averages in random treatments. In order to obtain the same unit of observation and equal variances in the random and competitive treatment, we need to compare groups of similar size. We, therefore, clustered the individual observations from random treatments into groups of six or seven members and performed statistical tests using these group averages. We repeated this procedure 4,000 times, and we report the median p-value. The results are robust for different clustering since the rank-sum was higher in the random treatment than in the competitive treatment for all the 4,000 repetitions.

The hypotheses on the effect of competitive audit selection mechanism are confirmed by our data constructed from the choices in the last six rounds (rounds 15 to 20). We find that a shift from the random to competitive audit selection mechanism reduces the concealment costs (t-test $p = 0.016$, Mann-Whitney $p = 0.019$) as well as the undisclosed output (t-test $p < 0.001$, Mann-Whitney $p = 0.003$). Changes in audit probability also lead to the predicted effects. A rise in audit probability in the random audit selection mechanism increases the concealment cost (t-test $p = 0.016$, Mann-Whitney $p = 0.034$), but has no significant effect on undisclosed output (t-test $p = 0.25$, Mann-Whitney $p = 0.17$). For competitive audits ($\beta = 0.4$), a rise in audit probability does not significantly affect the concealment costs (t-test $p = 0.21$, Mann-Whitney $p = 0.19$) but reduces the undisclosed output (t-test $p = 0.002$, Mann-Whitney $p = 0.002$). All these results fit to the predictions from the Table 5.1.

The results are qualitatively similar when we use data from the last ten rounds (rounds 11 to 20). All effects are in the same direction and statistically significant at the 5% level, except for the effect of the competitive audit

Figure 5.3: The effect of sensitivity parameter (β) and audit probability (π) on the non-compliance and concealment. The measure of non-compliance is the undisclosed output (u). The measure of concealment is the concealment cost (c). The figure shows the averages of these variables, with the bars depicting the standard errors based on independent observations.



selection mechanism (β) on concealment costs, where the p-value is slightly above the 5% level (t-test $p = 0.054$, Mann-Whitney $p = 0.053$).

The model predictions are also supported by the regression models in Table 5.3, where we control for the subject's socio-demographic characteristics. We take two approaches to deal with the censoring in the data. Out of the total number of 1,680 observations, the undisclosed output u is equal to 0 for 291 observations and 100 for 209 observations and the concealment cost c is 0 for 614 observations. Models (1) to (4) present OLS regressions with individual averages of u and c in rounds 15 to 20 as dependent variables. Models (5) to (8) report Tobit regressions explaining the individual observations in the last six rounds of the experiment. In all the models, the standard errors are clustered at the group level. The results are robust to model selection. A shift from the random to competitive audit selection mechanism reduces the undisclosed output and concealment costs in both the OLS and Tobit models. Hence, a shift to competitive audit selection mechanism lowers both the non-compliance and concealment investments. Also, a rise in basic audit probability provides robust effects corresponding to the model predictions, as it reduces the undisclosed output and increases the subjects' investment in concealment activities. As an additional robustness check, we estimated the model using only the choices that do not violate FOSD. The results presented in the Table 5.5 in the appendix show that the effects are not driven by irrational choices.

Our results are robust to the inclusion of control variables, and the estimated parameter values seem reasonable. Women seem to comply more and also invest less in concealment activities. This is consistent with Torgler and Valev (2010) and Kastlunger et al. (2010), who find higher compliance among women. On the other hand, business and economics students tend to cheat more and spend more resources on hiding their undisclosed output. Models (6) and (8) control for the punishment in the previous round, to find that a participant punished in round $t-1$ discloses higher output and invests less in concealment activities in round t.

This experimental evidence confirms the advantage of the competitive audit selection mechanism over random audit selection mechanism in situations where concealment investments are a concern. While competitive audit selection motivates people to invest less in concealment activities, more frequent audits have the opposite effect. Agents for whom audit is more imminent, despite their improved compliance morale, have an incentive to spend more on concealment.

Table 5.3: Explaining undisclosed output and concealment costs

Model:	OLS model				Tobit model			
Dependent variable:	Average individual undisclosed output		Average individual concealment costs		Individual undisclosed output		Individual concealment costs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	69.994***	66.800***	7.720***	−0.164	75.797***	79.896***	4.722***	−3.932
	(3.513)	(17.556)	(0.872)	(5.808)	(5.027)	(24.600)	(1.312)	(8.150)
Competitive	−15.057***	−14.287***	−2.280**	−2.181**	−22.039***	−20.952***	−2.999**	−2.869**
	(5.747)	(5.638)	(1.476)	(1.449)	(4.028)	(3.984)	(0.999)	(1.003)
High audit probability	−14.258***	−14.607***	2.295**	2.096**	−18.609***	−18.128***	2.776**	2.783**
	(4.090)	(3.957)	(0.935)	(0.934)	(5.336)	(5.183)	(1.414)	(1.390)
Female		−6.574**		−1.817**		−10.012**		−1.214
		(3.310)		(0.900)		(4.600)		(1.323)
Age		0.089		0.349		−0.138		0.365
		(0.817)		(0.269)		(1.133)		(0.382)
Working		−0.076		0.020		−0.684		0.147
		(2.987)		(0.824)		(4.173)		(1.291)
Econ study		6.703**		1.985**		6.642		3.240**
		(3.301)		(0.944)		(4.508)		(1.377)
Punish $t-1$						−4.000*		−2.774***
						(2.419)		(0.747)
Period fixed effect					No	Yes	No	Yes
Observations	280	280	280	280	1680 (Censored 500)		1680 (Censored 614)	

Note: Standard errors clustered at the group level, * p<0.1; **p<0.05;***p<0.01

If we interpret the model in terms of tax compliance, the relative advantage of competitive audit selection mechanism is even more pronounced when the impact of policy tools on the tax revenue per audit is considered. The tax revenue per audit may be calculated either excluding fines, when $TRA_{EF} = \tau r/\pi$, or including fines, when $TRA_{IF} = (\tau r + \pi p u)/\pi$. Regression results showing the effects of the studied policy tools on tax revenue are summarized in Table 5.4. The introduction of competitive audit selection mechanism increases the revenue per audit by more than 11 CZK if fines are included, and by more than 16 CZK if fines are excluded. These effects are both statistically significant and fiscally important given the taxable output is equal to 100 CZK and on average every second tax report is audited. On the other hand, a rise in audit probability leads to a reduction in tax revenue per audit by more than 21 CZK. This outcome is no longer statistically significant when fines are not considered as tax revenues. Hence, the competitive audit selection mechanism increases the tax revenue per audit, but the impact of high audit probability is either negative or statistically insignificant.

Table 5.4: OLS estimates of tax revenue per audit

Dependent variable:	Tax revenues per audit including fines		Tax revenues per audit excluding fines	
	(1)	(2)	(3)	(4)
Constant	100.789***	93.617***	48.434***	49.983**
	(5.199)	(20.675)	(4.831)	(20.774)
Competitive	11.765**	11.476**	17.316***	16.410***
	(5.173)	(4.910)	(5.164)	(5.073)
High audit probability	−21.212***	−22.527***	−5.590	−5.624
	(5.138)	(4.665)	(5.401)	(5.320)
Punish $t-1$		12.798***		3.375*
		(4.075)		(2.012)
Female		−3.832		7.174*
		(4.049)		(3.990)
Age		0.639		−0.026
		(0.919)		(0.950)
Working		−5.315		−0.081
		(3.862)		(3.605)
Econ study		−5.436		−7.565*
		(4.546)		(4.146)
Observations	280	280	280	280

Note: Standard errors clustered at the group level, $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Interestingly, the reduction in tax revenue is larger when fines are included compared to when fines are not included in tax revenue. This is because high audit probability leads to choices with a negative impact on fine collection. Since agents spend more on concealment activities, a lower share of their tax evasion is verifiable, and they evade less, meaning that the fines imposed on them are smaller on average. The competitive dummy variable coefficients are also higher for tax revenue excluding audit, but the difference between the two tax revenue measures is not statistically significant. This is because the effects of the competitive mechanism go in both directions: agents spending less on concealment are levied more fines, while those reporting higher output are levied smaller fines.

Table 5.5: OLS model using only observations not violating FOSD.

Dependent variable:	Average individual undisclosed output		Average individual concealment costs	
	(1)	(2)	(3)	(4)
Constant	70.138***	66.722*	7.665***	−0.204
	(3.505)	(17.559)	(0.873)	(5.787)
Competitive	−15.165***	−14.418***	−2.266**	−2.164**
	(4.016)	(3.975)	(1.001)	(1.004)
High audit probability	−14.004***	−14.376***	2.304**	2.109**
	(4.071)	(3.938)	(0.936)	(0.934)
Female		−6.327*		−1.828**
		(3.299)		(0.902)
Age		0.095		0.347
		(0.816)		(0.268)
Working		−0.242		0.108
		(2.967)		(0.822)
Econ study		6.830**		1.975**
		(3.278)		(0.944)
Observations	280	280	280	280

Note: Standard errors clustered at the group level, $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

5.6 Discussion

The design of the audit selection mechanism is crucial for effective enforcement, especially when agents can conceal their non-compliance. This experimental study examines how changes in two parameters of the audit selection mechanism affect compliance and concealment activities of the agents. In our theoretical model, agents invest in activities that increase the probability that their non-compliance remains undetected. In contrast to the previous research on auditing that has focused mainly on the reporting behaviour, our model provides insight into the effect of competitive audit selection mechanism on the level of concealment activities. In particular, the model predicts that higher audit frequency motivates agents to comply more, but also leads them to spend more on concealment activities. The competitive mechanism, on the other hand, is predicted to improve agents' behaviour in both dimensions: it encourages them to comply more and invest less in concealment.

We test for these predictions in a controlled laboratory experiment. In line with theoretical predictions, we find that the competitive audit mechanism increases compliance and reduces concealment, whereas a rise in audit probability increases the compliance but also motivates the agents to invest more in concealment activities.

These findings, therefore, suggest that when concealment is a concern, the smarter approach would be to use the competitive rather than random audit selection mechanism; this is superior to the more active approach of conducting more frequent audits. The advantage seems even clearer when we consider the costs of both policies. While more frequent auditing is bound to increase the administrative costs, the frequency under competitive audit selection is the same as that under random audit selection. The only additional costs of the competitive mechanism involve selecting the agents with similar output. Hence, the competitive audit selection mechanism seems better than more frequent auditing in terms of concealment costs as well as administrative costs.

5.7 Proofs of the model

5.7.1 Equilibrium condition

Assume an interior solution. The best responses of player i are given by the following first-order conditions.

$$-\Pi(r_i, r_{-i})\Phi(q - r_i) - c'(p_i) = 0 \tag{5.5}$$

$$-\Pi'_{r_i}\Phi p_i(q - r_i) - \tau + \Pi(r_i, r_{-i})\Phi p_i = 0 \tag{5.6}$$

The following second-order conditions ensure that the payoff function is concave.

$$-c''(p^*) < 0 \tag{5.7}$$
$$\Phi p^*(-\Pi''_{rr}(q-r^*) + 2\Pi'_r) < 0 \tag{5.8}$$
$$-c''(p^*)(-\Pi''_{rr}(q-r^*) + 2\Pi'_r)p^* - \Phi(\pi - \Pi'_r(q-r^*))^2 > 0 \tag{5.9}$$

The following first-order equilibrium conditions lead to symmetric equilibrium.

$$-\pi\Phi(q-r_i) - c'(p_i) = 0 \tag{5.10}$$
$$-\Pi'_{r_i}\Phi p_i(q-r_i) - \tau + \pi\Phi p_i = 0 \tag{5.11}$$

5.7.2 Proof of proposition 1

For comparative static effect of the sensitivity parameter β, differentiate the following first-order equilibrium conditions:

$$p_\beta^{*'} = \frac{\Phi\pi r_\beta^{*'}}{c''} \tag{5.12}$$
$$r_\beta^{*'}(-\Pi''_{rr}(q-r^*)p^* + p^*\Pi'_r) + p_\beta^{*'}(\pi - \Pi'_r(q-r^*)) = \Pi''_{r\beta}p^*(q-r^*). \tag{5.13}$$

By substituting $p_\beta^{*'}$ into equation (5.13), we obtain the effect of sensitivity on the reported output,

$$r_\beta^{*'} \underbrace{''(-\Pi''_{rr}(q-r^*)p^* + p^*\Pi'_r)) + \Phi\pi(\pi - \Pi'_r(q-r^*)]}_{Z} = c''\Pi''_{r\beta}p^*(q-r^*).$$

Because the right-hand side is always negative, the effect depends on the sign of the expression Z. By substituting the expression $(\pi - \Pi'_r(q-r^*))$ from the second-order condition (5.9), we obtain the upper bound of Z.

$$Z < c''(-\Pi''_{rr}(q-r^*)p^* + p^*\Pi'_r)) + \frac{\pi p^*(-c''(p^*))(-\Pi''_{rr}(q-r^*) + 2\Pi'_r)}{\pi - \Pi'_r(q-r^*)}$$

This upper bound is negative if

$$-c''(-\Pi''_{rr}(q-r^*)p^*+p^*\Pi'_r))\underbrace{(\pi-\Pi'_r(q-r))}_{=\frac{\tau}{\Phi p^*}} > \pi((-c'')(-\Pi''_{rr}p^*(q-r^*)+2\Pi'_r p^*)).$$

By substitution from the equilibrium condition (5.11) and after some simple algebraic manipulation, we obtain the following condition:

$$\Pi''_{rr}(q-r^*)p^*(\tau-\Phi p^*\pi)+\Pi'_r p^*(2\Phi p^*\pi-\tau)>0.$$

Given that the audit probability function is decreasing, $\Pi'_r<0$, and non-concave, $\Pi''_{rr}\geq 0$, this condition is satisfied if

$$\tau > 2\Phi p^*\pi. \tag{5.14}$$

This condition is sufficient for the reported output to be increasing in the sensitivity parameter, $r^{*'}_{\beta}>0$. Now, from equation (5.12), it follows that the verification probability is also increasing in the sensitivity parameter, $p^{*'}_{\beta}>0$.

5.7.3 Proof of proposition 2

The comparative static effect of parameter π is obtained by differentiating the following first-order equilibrium conditions:

$$\begin{aligned} \Phi(q-r^*)+\pi\Phi r^{*'}_{\pi}-c''p^{*'}_{\pi} &= 0 \quad (5.15)\\ -\Pi'_r\Phi((q-r^*)p^{*'}_{\pi}-p^*r^{*'}_{\pi})+\Phi p^*+\Phi\pi p^{*'}_{\pi} &= 0. \quad (5.16)\end{aligned}$$

By substituting $r^{*'}_{\pi}$ from equation (5.15) in equation (5.16), we have the effect of parameter π on the verification probability

$$p^{*'}_{\pi}\underbrace{[c''(-\Pi''_{rr}(q-r^*)p^*+p^*\Pi'_r))+\Phi\pi(\pi-\Pi'_r(q-r^*)]}_{\tau>2p^*\Phi\pi\rightarrow<0}=$$

$$=\underbrace{(q-r^*)\Phi p^*\Pi'_r-\Phi\pi p^*}_{\tau-2p^*\pi}+(q-r^*)(\Pi''_{rr}(q-r^*)\Phi p^*).$$

The expression in the square brackets on the left-hand side is negative if $\tau>2p^*\Phi\pi$ (see the proof of Proposition 1). The right-hand side is equal to

$$\tau-2p^*\Phi\pi+(q-r^*)(\Pi''_{rr}(q-r^*)\Phi p^*).$$

(by substituting from the equilibrium condition (5.11)). The right-hand side is positive if $\tau > 2p^*\Phi\pi$. Therefore, the verification probability is decreasing in the basic audit probability $p_\pi^{*'} < 0$.

By substituting $p_\pi^{*'}$ from equation (5.15) in equation (5.16), we have the basic audit probability effect on the reported output,

$$r_\pi^{*'} \underbrace{[c''(-\Pi''_{rr}(q-r^*)p^* + p^*\Pi'_r)) + \Phi\pi(\pi - \Pi'_r(q-r^*)]}_{\tau>2p^*\Phi\pi\rightarrow<0} =$$

$$= \underbrace{-p^*c'' - \Phi(q-r^*)(\pi - \Pi'_r(q-r))}_{<0}.$$

The expression in the square brackets on the left-hand side is negative if $\tau > 2p^*\Phi\pi$ (see the proof of Proposition 1). The right-hand side is negative because $c'' > 0$ and $\Pi' r < 0$. Therefore, the reported output is increasing in the basic audit probability $r_\pi^{*'} > 0$.

Chapter 6

Discretion in regulatory enforcement

6.1 Motivation

Regulatory authorities can functionally be viewed as law enforcers. While the aim of most regulations is to achieve the most efficient outcome available, or at least a more efficient outcome, regulation needs to be also codified in some general legal rules. The generality of legal rules thus means they should be applied to a large number of subjects for a longer period of time. Given these constraints, it is impossible to design efficient regulations that are also complete, i.e. regulations that describe all possible circumstances that may arise and clearly determines the efficient outcome for all of them[1](Pistor and Xu, 2003).

When a law is incomplete, some agent needs to decide whether a particular action will fall within the scope of a regulation, whether it meets the regulatory standards, and therefore whether it will face any penalties. Granting such a discretionary power may be deemed efficient since regulatory authorities may choose from a range of possible actions, using its expert information to evaluate and considering many mitigating factors that are case-specific (Duflo et al., 2018; Kang and Silveira, 2018). On the other hand, discretion may lead to decisions that reflect regulators' personal preferences, rather than the

[1]The concept of incomplete legal rules is thus similar to the concept of incomplete contracts. In fact, it is to be noted that legal rules are likely to be more incomplete than contracts, as legal rules regulate the behaviour of more subjects and cover a much greater degree of variance with cases.

social goals of regulation (Leaver, 2009; Shavell, 2007). However, this problem may be mitigated by providing the regulator with the proper incentives. Even when the regulator's preferences align with social welfare, there might be a problem of dynamic inconsistencies, particularly when a regulator with discretionary power is not able to commit to a credible policy.

To avoid these problems that are related to regulation, the latter can be defined with a set of simple and complete legal rules which usually set a stricter regulatory standard[2]. Such legal rules do not require interpretation, thereby making them easily enforceable. On the other hand, they do not fit in many situations since they disregard the specific circumstances found in particular cases. This might result in excessive punishment for harmless actions or to ineffective over-compliance (Shimshack and Ward, 2008; Earnhart and Harrington, 2014).

This tension between flexibility and inconsistency underlies many debates on rules and discretion. This chapter will thus contribute to this debate by eliciting the subjects' choice between discretionary and complete rule regime. We propose an experiment with two within-subject treatments: discretion and complete rule. After some learning rounds, we allowed the subjects to vote on which regime thay they would prefer. Despite the fact that the discretionary regime is more efficient monetarily, we identified strong aversion towards the discretionary regime.

The rest of the chapter is thus structured as follows: Section 6.2 reviews the literature on the intrinsic value of institutions. Section 6.3 introduces the theoretical framework. Section 6.4 presents the experimental design and procedures, and formulates the hypotheses tested by the experiment. Section 6.5 describes the data generated by the experiment. Section 6.6 and 6.7 presents the results and robustness checks. Finally, Section 6.8 provides a short discussion of the results.

[2]Criminal sentence reform is an example of this. Judges have had considerable discretion over determining criminal sentences under so-called indeterminate sentencing until the mid-1980s. In 1984, the U.S. Sentencing Commission issued a sentencing guideline that imposed much tighter limits on the discretion of judges compared to the earlier system of indeterminate sentencing. Under the guidelines, the prescribed sentences are strictly determined by the seriousness of the crime and the offender's criminal history.

6.2 Related literature

6.2.1 Regulatory discretion

The decision between discretion and strict rules has been commonly modeled within the principal-agent framework (Shavell, 2007; Alonso and Matouschek, 2008; Armstrong and Vickers, 2010). Within this framework, the principal and the agent have different preferences and their preferred decision depends on the state of the world. The principal can be viewed as a benevolent state that follows the social welfare function. The agent is the regulator with his own preferences. The problem arises from the fact that only the agent is informed about the state of the world. The choice between discretion and rules is thus a trade-off between flexibility and the agent's opportunity to follow his private objectives instead of social welfare.[3] A discretion is desirable if the loss given by the inflexibility of the strict rule is larger than the loss given by the deviation of the agent's utility function from the social welfare function (Shavell, 2007). More generally, the principal may want to restrict the set of agent's choices (Alonso and Matouschek, 2008). However, this solution assumes that there cannot be any contingent monetary transfers between the principal and the agent. If such transfers are possible, the principal can provide the agent with monetary incentives to follow social welfare function (e.g. Sappington, 1991; Bester and Krähmer, 2008).

Beyond this basic trade-off, there are other considerations which might influence the choice between discretion and rule. One of the most profound problems would be dynamic inconsistency (Kydland and Prescott, 1977). The dynamic inconsistency problem can arise if the official only has a single instrument to achieve multiple goals. In the compliance context, the official with discretionary power would thus decide whether the defendant (subject to regulation) complied with the regulation. If the official decides that the defendant did not comply, he or she will impose some costs on the defendant. On the other hand, if the official should decide that the defendant complied with the regulation, he or she might impose some external costs on others. As such, the official might want to decide on the costs imposed on the defendant or other members of the society according. Simultaneously, the official would also want to enforce some level of care so that non-compliance does not occur to often.

[3]There is also an empirical evidence supporting relevance of this framework (Leaver, 2009; Duflo et al., 2018; Kang and Silveira, 2018).

Typically, the official wants to convince regulated subjects that the instrument will be used to enforce some effort level, but ex post he or she has an incentive to use it, in order to allocate the cost burden efficiently. In other words, the official wants to persuade the regulated entities that compliance standards will be high enough to enforce sufficient levels of care. However, once the effort investment is made, the official has an incentive to reduce the standard to minimize social costs.

6.2.2 Discretion aversion

The traditional economic modelling framework considers institutions and procedures according to their instrumental values, i.e. according to the expected utility associated with the outcomes generated by the procedures. This view has guided research on how to design procedures in order to achieve efficient outcomes. However, recent studies suggest that individuals value institutions and procedures for their intrinsic value, i.e. beyond the expected utility associated with the achieved outcome. (Bohnet and Zeckhauser, 2004; Bolton and Ockenfels, 2010; Owens et al., 2014; Bolton et al., 2005; Sausgruber and Tyran, 2014). From this observation, it would follow that while some procedures are designed to generate efficient outcomes in monetary terms, they do not have to generate efficient outcome when the intrinsic utility is taken into account.

In the context of this experiment, we present a hypothesis that people might have an aversion towards the discretionary power of someone else. Consequently, it is our belief that this hypothesis might be justified along several lines.

Discretionary decision-making always involves some form of ambiguity regarding the decisions to be made. As people generally dislike ambiguity (Dimmock et al., 2015), the ambiguity aversion may manifest as discretion aversion. This aversion may be even stronger when the human factor is present. There is a well-documented tendency to avoid a situation when a person, rather than nature, determines the outcome of the situation. This phenomenon is known as "betrayal aversion" (Bohnet and Zeckhauser, 2004; Bolton and Ockenfels, 2010; Hong and Bohnet, 2007; Aimone and Houser, 2012).

The betrayal aversion is normally studied in the trust game experimental framework. In the first stage of the game, the proposer decides whether to trust to the responder. If he or she does not trust the responder, both players obtain a fixed fee. In the second stage, the responder (if trusted) decides whether they want to reciprocate or betray the other. The proposer's payoff is highest if the responder reciprocates and lowest in case of betrayal. On the other hand, the responder's payoff is highest in case of betrayal. As a control, the same game is used, but human responders are replaced by a computer that decides according to the predetermined probabilities. Indeed, experimental studies of betrayal aversion have used a common elicitation procedure. The participants of the experiment were asked to report a "minimum acceptable probability" at which they would choose to trust or risk. In the trust game, with the humans, if the probability is lower than the fraction of reciprocating responders, then the proposer is matched with some responder and is paid accordingly, based on the responder's decision. If the reported probability is larger than the fraction of reciprocating responders, the proposer receives a fixed payoff. In the trust game with humans, if the minimum acceptance probability is lower than some unknown probability, the proposer and the responder are paid according to the risky lottery. Otherwise, they receive a fixed payoff. This procedure is an incentive compatible mechanism for an elicitation of the players' preferences. Many studies have consistently found that proposers report higher minimum acceptance probability for games with humans than games with nature. This result suggests that people suffer some disutility if they get a lower payoff because of the other player's decision.

Butler and Miller (2017) focuses on the context in which the betrayal aversion might be observed. They use a modification of the betrayal aversion experimental paradigm (Bohnet and Zeckhauser, 2004) in order to evaluate the hypothesis that the social risk premium is dependent on the agent's ability to behave intentionally. They implement between-subject treatments where the responders cannot behave intentionally as they will not know the consequences (i.e. payoffs) of the particular action. The results show that acceptable probability differs to a minimum between these treatments. Henceforth, they argue that betrayal aversion occurs only in the context where other people can make intentional decisions.

Along these similar lines, there has been evidence demonstrating that individuals have a tendency to incur a cost in order to keep control over their own outcome. This tendency is usually interpreted as a preference for payoff autonomy. Owens et al. (2014) present a laboratory experiment that documents the player's willingness to control his own payoff. Participants in their experiment chose whether they wanted to bet on themselves or on a partner in answering a trivia quiz question correctly. Given the elicited beliefs, participants bet on themselves more than expected money maximizers would do.

Fehr et al. (2013) studies valuation of decision rights in a principal-agent experiment. This is a two-player game where the principal decides whether he should delegate the right to decide a project. The project will then determine the payoff for both players. Among the projects, there are three projects of particular interest: project with a known payoff π_s for both players, project with an unknown but high payoff π_h for the principal and a lower payoff π_l for the agent, and one project with a high payoff π_h for agent and a lower payoff π_l for principal. It holds that $\pi_h > \pi_l > \pi_s$. All other projects would give zero payoffs. The projects are ex-ante identical, and players have to exert a costly effort to observe their preferred project. After that, the player without decision rights would recommend a project to the player with decision rights and this player would then choose a project. The delegation of decision rights has two effects: (i) it decreases the probability that the principal's preferred projected will be chosen, (ii) it reduces equilibrium effort and saves costs. Fehr et al. (2013) found that the principals tend to keep the decision rights when predicted by theory authority, but they observed strong under-delegation in the treatment where theory predicts delegation. They interpret this result as evidence for a desire to retain control.

Bartling et al. (2014) further explores the robustness of this claim. They use the same principal-agent experiment with different payoff structures to show that the value of decision rights would not be driven by risk aversion, prosocial preferences, ambiguity aversion, loss aversion, the illusion of control, reciprocity or learning effects. Therefore, they argue that the value attributed with decision right originate from the intrinsic preferences for decision rights. These studies differ from our experimental environment in one main feature:; in these experiments, the participants chose whether to delegate the decision right or to keep full control. In our experimental design, the choice is not between delegating or keeping full control. Instead, the choice is whether to guarantee a discretionary power to someone else or to be a subject to a rule. Still, the preference for decision rights, which is documented by these studies, might be interpreted within our framework as a source of discretion aversion, as the player's payoff under the rule regime is influenced only by his actions and random moves.

Beyond the experimental evidence, Frey and Stutzer (2005) also used survey data to show that people obtain intrinsic utility from participating in the political voting, irrespective of the outcome. They use self-reported life satisfaction as a proxy for overall utility. Simultaneously, they take advantage of the unique political system in Switzerland, which assigns different rights to participate in the political process at the canton level. By using this exogenous variation, they show that life satisfaction was found to be higher when people have more opportunities to participate in public decision making. They interpret this observation as evidential in the existence of an intrinsic preference for decision rights.

Our experiment is based on the choice between two allocation procedures. When deciding how to allocate goods or property rights, people often care about the fairness of the allocation procedure as much as about the fairness of the outcome itself. Bolton et al. (2005) provides experimental support for this claim. The participants in their experiment play a mini-ultimatum game in one of three treatments: i) the proposer chooses between a highly unequal offer favouring himself and similar offer favouring the responder, ii) the proposer can additionally choose equal division, iii) the proposer can choose a lottery which randomly selects one of the two unequal offers. The results show that the possibility of choosing a fair procedure has the same effect on the subjects' behaviour as the possibility of choosing equal allocation. It thus indicates that people have a strong preference for impartial procedures that treats all players in the same way. In our experiment, the two allocation procedures generate different outcomes ex-post. However, since both procedures in our experiment are invariably unbiased and fair, this approach should deem both procedures to be equally fair.

6.3 Theoretical framework

We are interested in the relationship between the effort of compliance effort, the enforcement regime, and the efficiency of the outcome. The experiment is structured with the use of the following model.

Description of the model

The decision maker chooses effort level e. The effort is costly and the monetary costs are given by the function $c(e)$, which is increasing and non-concave, i.e. $c'(e) > 0$ and $c''(e) \geq 0$. Bad luck, b, is a random variable with support $[-\underline{b}, 0]$ and probability distribution functions $f(b)$ which is increasing, $f'(b) > 0$. The effort level and bad luck determine whether bad event occur. The bad event occurs if the sum of the effort and bad luck falls below some threshold T, i.e. $b + e < T$. Only the sum $b + e$ is observable to the third party, while the effort e itself is not observable. The bad event causes monetary harm which can then be paid by the decision maker or by society. In the case of the former, the monetary harm is H_{DM}; in the latter case, it is H_S.

There are two regimes that assign responsibility for the harm. Under the *strict rule* regime, the decision maker always pays for the harm. Under the *discretionary regime* there is a benevolent official who decides who will pay for the harm. The official can observe $e + b$, which we will call observable effort.

The official's task is trivial when the harm for the society is at least as large as the harm for the individual $H_S \geq H_{DM}$. In this case, the official should always mimic the strict rule by deciding that the harm will be paid by the decision maker. Henceforth, there is no reason to grant the official a discretionary power. As such, we focus on a more problematic case, particularly one where the costs paid by the decision maker exceeds the costs paid by the society $H_{DM} > H_S$.

If the decision maker's cost exceeds the cost for the society, then the official faces a dynamic inconsistency problem. The dynamic inconsistency problem in this context stems from the fact that the official would like to forgive the harm whenever possible, while simultaneously needing to enforce some effort level. Ideally, the official would like to commit ex-ante to a rule that defines some threshold value of D. If the decision maker's observable effort is above this threshold, then the official decides that the decision maker is not responsible for the harm. The problem is that such a rule is dynamically inconsistent in the one-shot game. Once the effort's decision is made, and effort costs are sunk, the official is tempted to deviate from this rule and forgive even if the observable effort falls under this threshold. The decision maker thus realizes this dynamic inconsistency, resulting in an under-provision of effort.

On the other hand, if the officials are able to develop and retain a reputation for fulfilling its promises, they could avoid the dynamic inconsistency. The concern for its reputation provides the official with an incentive to implement a commitment to a particular decision rule. If the official's utility function coincides with the social welfare function, then the decision rule generates a socially optimal outcome from the long-run perspective. The underlying argument would thus be that the officials are engaged in repeated interactions, they have an incentive to develop and retain his reputation.

The rest of this chapter will present the solution of this model under different regimes.

First-best solution

The welfare function in this modelling framework is given by the negative of the total monetary costs. It comprises of three elements: i) the probability that the total effort falls below the threshold $Pr(b+e<T)$, ii) the cost that is paid by the decision maker H_DM or by the society H_S, iii) the costs of exerting effort $c(e)$.

The first-best solution is given as a solution of the following problem - a benevolent dictator maximizes the welfare function by choosing the effort level e and variable $h \in \{H_S, H_{DM}\}$ from which he determines who is responsible for the harm

$$\max_{e,r} -Pr(b+e<T)h - c(e)$$

In the first-best solution, the harm is always paid by the society, i.e. $h = H_S$ and the first-best effort level is given by the following first-order condition

$$f(T-e^{FB})H_S = c'(e^{FB}), \tag{6.1}$$

which states that the marginal costs of exerting the effort $c'(e)$ are equal to the marginal benefits which are given by the expected reduction of the amount paid by the society $f(T-e)H_S$. The second-order conditions are also satisfied because the costs are non-concave and the probability function is increasing. These assumptions will thus ensure that marginal costs are non-decreasing and that the marginal benefits are decreasing in the effort.

Strict-rule regime

In the strict-rule regime, the decision maker always pays the costs whenever the sum of his effort and bad luck falls below the threshold T. Hence, the decision-maker chooses effort levels that maximizes his own payoff

$$\max_{e} -Pr(b+e<T)H_{DM} - c(e).$$

The solution of the problem is given by the following first-order condition that implicitly defines the optimal effort under the strict-rule regime e^{SR}.

$$f(T-e^{SR})H_{DM} = c'(e^{SR}). \tag{6.2}$$

The second-order conditions are satisfied by the same reasoning as in the case of the first best solution.

We further impose the assumption 6.3 to assure that strict rule is still better than a situation when the decision maker does not exert any effort and the externality is always covered by the society.

$$-Pr(b < T)H_S < -Pr(b + e^{SR} < T) - c(e^{SR}) \tag{6.3}$$

Discretionary regime

Suppose that the official retains the discretionary power to decide who pays the cost in order to maximize the social welfare function. Simultaneously, the offcial is not able to make ex-ante commitment to some particular decision rule. In this case, the situation can be described as a sequential game in which the decision-maker moves first and chooses the effort level e. The official moves second. He or she observes the total effort and decides who will have to pay the cost in the case when the total effort falls below the threshold. By solving through backward induction, the official's reaction function is to always forgive $h(t) = H_S$. Consequently, the decision maker's optimal effort level is equal to zero, $e^D = 0$

The officials could be asked to maximize the social welfare function, provided that they are able to develop a good reputation. If so, concern for their reputation would provide the officials with an incentive to implement the commitment with some decision rule, which would then determine the officials' decisions as a function of the total effort.

When the official uses discretion to maintain a credible reputation, there are different sequences of moves that would be appropriate. In this case, the official moves first and the decision maker moves second. The official's task is to decide when to forgive and when to impose an externality on other members of the society based on the observable effort. The official chooses the discretionary threshold D. If the bad event occus and the observable effort is above or equal to D, then the official forgives, and the harm is transferred to society. If the bad event occurs and the observable effort is below D, then the decision maker has to incur the cost. In the second stage, the decision maker chooses the effort level e^{DC}, given the official's choice of threshold D.

The equilibrium is found by backward induction. In the second stage, the decision maker chooses an effort level that maximizes the following payoff

$$\max_e -Pr(b + e < D)H_{DM} - c(e)$$

The solution of the problem is given by the condition which states that marginal costs of exerting effort are equal to the expected marginal reduction in the amount paid by the decision maker

$$f(D - e^*)H_{DM} = c'(e^*).$$

As such, the condition implicitly defines the decision maker's best-response function $e^*(D)$. By applying the implicit function theorem, we can derive the slope of this best-response function

$$e^{*\prime}(D) = \frac{f'(D-e)H_{DM}}{f'(D-e)H_{DM} + c''}.$$

As the cost function is non-concave, i.e. $c'' \geq 0$, the slope is positive but less or equal to one. It would be equal to one if and only if the effort costs are linear.

In the second stage of the game, the official chooses the threshold D in order to maximize the welfare function, given the decision maker's best-response function

$$\max_D -Pr(b + e^* < D)H_{DM} - Pr(D < b + e^* < T)H_S - c(e^*)$$

The solution of this problem is given by the following first-order condition

$$f(T-e)\, e^{*\prime}\, H_S - f(D-e)(1 - e^{*\prime})(H_{DM} - H_S) = c'(e)\, e^{*\prime}. \tag{6.4}$$

The solution of the condition gives the optimal threshold D. This condition shows that increasing the threshold induces three effects: i) the probability of non-compliance decreasing because of the higher effort; ii) higher effort producing higher costs; iii) by increasing the threshold, the official imposes the costs more frequently on the decision maker than on society. This last effect is partly compensated by the increased effort. However, it disappears completely if the effort fully offsets the threshold increase, i.e. when the slope of the best response function $e^*(D)$ is one. Note that this happens in the linear cost case. The condition (6.4) is then equivalent to the condition (6.1) which means that the official sets the threshold such that it induces the first-best effort level.

Comparison

This section provides a comparison of different regimes in terms of effort and welfare. The first proposition summarizes the comparison of effort levels.

Proposition 6.3.1. *The effort levels under the different regimes rank as follows* $e^D < e^{DC} \leq e^{FB} < e^{SR}$.

Proof. It follows from the equilibrium description under discretion without commitment that $e^D = 0$ and the condition (6.3) ensures that $e^{DC} > 0$. This proves the first inequality. By comparing the first order conditions 6.1 and (6.2), we can see that the last inequality $e^{FB} < e^{SR}$ holds. The optimal effort level e^{DC} satisfies the condition (6.4), which can be rewritten as $-\frac{1-e'}{e'} f(D - e^{DC})(H_{DM} - H_S) + f(T - e^{DC})H_s = c'$. Now, suppose, by contradiction, that $e^{DC} > e^{FB}$. The condition (6.1) together with the assumptions that marginal cost are non-decreasing $c' \geq 0$ and the probability function is increasing $f' > 0$, imply that $f(T - e^{DC})H_s < c'$. For the condition (6.4) to be satisfied, it has to be the case that $\frac{1-e'}{e'} f(D - e^{DC})(H_{DM} - H_S)$ is negative. This cannot be true since the slope of the best-response function $e' \in (0, 1]$. Hence, we have a contradiction that would prove the second inequality $e^{DC} \leq e^{FB}$.

Second proposition shows the welfare ranking of different regimes. The strict rule regime is a special case within the discretionary regime with commitment when $D = T$.

Proposition 6.3.2. *The welfare levels under the different regimes rank as follows* $W^D < W^{SR} < W^{DC} < W^{FB}$.

Proof. The first inequality $W^D < W^{SR}$ holds by assumption (6.4). The proof the second inequality $W^{SR} < W^{DC}$ consider the welfare in a discretionary regime with commitment as a function of the discretionary threshold $W(D) = -Pr(b + e^* < D)H_{DM} - Pr(D < b + e^* < T)H_S - c(e^*)$ where e^* is given by the condition (6.4). The welfare W^{DC} is the maximum value of this welfare function. The welfare in strict rule regime is a equal to this welfare evaluated at T, i.e. $W^{SR} = W(T)$. Therefore, we only need to show that the inequality is strict. When we calculate the first derivative of the welfare and substitute for c' from condition (6.4) we get $f(T - e)\, e^{*\prime}\, H_S - f(D - e)(1 - e^{*\prime})(H_{DM} - H_S) = f(D - e)H_D M\, e^{*\prime}$. By evaluating the first derivative as point T, we have $-f(T - e)\,(H_{DM} - H_S) < 0$. Since the welfare function is decreasing at T, it holds that $W^{DC} > W^{SR}$. The third inequality $W^{DC} < W^{FB}$ also holds as strict because it cannot simultaneously be the case that $D = 0$ and $e^{DC} = e^{FB}$.

6.4 Experimental design and procedures

The experimental design combines a voting experiment and an effort provision experiment that is similar to the model described in the preceding section. In the effort provision part, the subjects choose how much effort to exert in order to prevent a loss. In the voting part, they vote, by majority rule, on the procedure that determines who will pay for the loss. Section 6.4.1 describes the rules, the sequence of events and the feedback that subjects receive during the experiment. Section 6.4.2 describes the questionnaire used in the experiment and section 6.4.3 explains the experimental treatments, along with a discussion of the predictions.

6.4.1 Experimental procedure

The experiment consists of four stages: discretionary regime, strict-rule regime, the voting stage, and the final stage. At the beginning of each stage, an experimenter reads the instructions aloud, with the subjects follow with their own copy. Instructions use a neutral language, and the subjects receive instructions for each stage separately (the example of instructions in the original Czech language is in the Appendix C). At any particular stage, the subjects would have yet to be informed about what will happen in subsequent stages. The order of the strict-rule-regime stage and discretionary-regime stage was randomized. At the end of the experiment, subjects were to answer a questionnaire.

Subjects were randomly matched into groups of five. Four subjects are given the role of *players*, and one subject has the role of the *official*. The matching remains fixed during the whole experiment in order to strengthen the learning effect.

The strict rule regime consists of 15 periods. The officials are inactive in the strict rule regime: they do not make any decisions, and they do not get any feedback about the other players' behaviour. At the beginning of each period, subjects are endowed with 140 CZK[4] and 6 tokens. The subjects know that zero to six tokens can be lost according to a predetermined probability distribution. The probability distribution is presented in Table 6.1. Subjects have the opportunity to buy 0 to 6 additional tokens. Each token costs 10

[4]At the time of the experiment, 1 USD is equivalent to 23 Czech Crown (CZK) and 1 EUR is equivalent 26 CZK. A standard wage of an hour of unqualified student labour was approx. 100 CZK.

CZK. After the buying decision is made, a random draw determines how many tokens are lost. If the number of tokens remaining is less than six, than the player suffers a loss and he or she has to pay 100 CZK.

Table 6.1: Descriptive statistics

Number of tokens lost	0	1	2	3	4	5	6
Probability	0.26	0.20	0.16	0.12	0.12	0.08	0.06

The discretionary regime also consists of 15 periods. The only difference between the discretionary regime and the strict rule regime is the active role of the officials. The officials observe the remaining number of tokens, but they do not observe how many tokens were lost. If the remaining number of tokens falls below six, the official will then decide whether the loss is paid by the player with an insufficient number of tokens or by the other three players in the group. In the case of the former, the player will pay 100 CZK; in the latter case, each of the other three group members pays 25 CZK. After each round, subjects receive feedback about the remaining number of tokens and their own payoff. For each of the other players, they learn about whether the remaining number of tokens was below the threshold and the decision of the official. The information about other players is displayed in random order, so the players and the official are not able to track the identity of other players during the subsequent periods. The purpose of the first two stages is twofold. First, we can test whether the discretionary regime is more efficient and players get higher monetary payoffs. Second, players become familiar with the discretionary regime and the strict-rule regime. They learn what monetary payoffs can be gained in both regimes, allowing them to make a competent voting decision.

In the voting stage, players in each group vote on which of the regimes should be played in the final stage. The regime that receives a majority of votes in each group is chosen. If two players vote for each regime, one of the regimes is chosen randomly (each with a 50 % probability). In the final stage, the participants play according to the rules of a regime that was chosen in the voting stage. The number of periods in the final stage is random. After each period, the game ends with a probability 0.3. The random number of periods ensures that the final stage will not take too much time and the officials, if active, will simultaneously still face a trade-off between enforcing sufficient level of effort and capturing gains by letting the group members pay for the loss.

At the end of the experiment, one randomly-selected period from each of the first two stages (discretionary regime, strict-rule regime), and the last period from the final stage are selected to be paid. The officials are also paid by the average one-period payoff of the players in their group. Indeed, the average was calculated separately for the discretionary regime, the strict-rule regime and the final stage. Note that this payment scheme provides incentives for the official to maximize the monetary wealth of the group.

6.4.2 Questionnaire

The final questionnaire includes questions on socio-demographic variables, self-reported risk attitude and measures of tolerance to ambiguity (Budner, 1962), personality traits (Rammstedt and John, 2007) and tendency to make moral judgments based on utilitarian or deontological principles (Robinson et al., 2015). At the end, players were asked about the reason behind their voting decision.

- *Risk* Risk preferences were measured by asking subjects for their self-assessment on their willingness to take risks on an 11-point scale (In general, how willing are you to take risks?). Higher scores are associated with higher risk tolerance. This self-assessment question has been shown to previously correlate with risk-taking behaviour in real-world situations (Dohmen et al., 2011) and with incentivized experimental measures of risk-taking across countries in student samples (Vieider et al., 2015). [5]

- *Ambiguity scale* Tolerance to ambiguity was measured by using 4 item scale ("People who insist upon a yes or no answer just don't know how complicated things really are".; "Many of our most important decisions are based on insufficient information".; "An expert who does not come up with a definite answer probably does not know too much".; "Teachers who hand out vague assignments gives one a chance to show initiative and originality."). The questions are a subset of an original 16 item scale developed by Budner (1962). The subjects report their agreement with each statement through the use of a 5-point Likert scale, and people with higher scores have a tendency to perceive ambiguous situations as desirable.

[5] We have not used incentivized risk elicitation in order to save time. The experiment took nearly two hours and it would be too long with a risk-taking elicitation procedure.

- *Consequentialist Scale* The decision to vote in favour of discretion or strict rules might be guided by a moral principle to adherence to certain universal rules or by the aim to maximize social benefit. Therefore, the questionnaire includes a measure of deontological and utilitarian moral tendencies. Participants also completed a shorter version of the Consequentialist Scale (Robinson et al., 2015). This short version contained 4 questions, two that assessed endorsement of utilitarian beliefs (Rules and laws should only be followed when they maximize happiness; When deciding what action to take, the only relevant factor to consider would be the outcome of the action) and two that assess deontological beliefs (Some rules should never be broken; It is never morally justified to cause someone harm). Participants indicate how much they agreed with each statement on a 5-point Likert scale. The total score ranged between 4 and 20, with higher scores showing a tendency towards a more deontological attitude and lower score points towards a more utilitarian attitude.
- *Big Five personality traits* The personality traits were measured using a short version of the Big Five personality test (Rammstedt and John, 2007).The Big Five model is a widely accepted framework for the description of one's personality. It consists of five subscales: extroversion, agreeableness, conscientiousness, neuroticims, and openness. Agreeableness and conscientiousness was thus viewed to be of special interest in our experiment. Agreeableness measures the tendency to be cooperative towards others. People who achieve a high score on this scale are empathetic and altruistic, while a low agreeableness score relates to selfish behaviour and a lack of empathy. Conscientiousness measures a trade-off between flexibility and reliability. People with a high score tend to be self-disciplined and stubborn, while a low score relates to flexibility and a lack of reliability. All Big Five questions are measured on a 5-point Likert scale, and based on how much they agreed with each statement. Each subscale score ranges from 2 to 10.

6.4.3 Treatments and predictions

The design has two treatments. The treatment HUMAN implements the procedure described in the previous section where the official is a human subject. The treatment NATURE is the same, but the official is played by a computer program. This treatment was implemented to test whether the choices of the players were in fact influenced by the fact that another player has discretionary powers to transfer the cost. The computer decides according to a function that defines the probability that the loss is paid by other members of the group conditional on the number of remaining tokens. The value of probabilities was established as the fraction of choices in which the group members had to pay for the loss in the six HUMAN treatment sessions. The decisions made by nature therefore mimicked the decisions of our human officials. The probabilities are presented in Table 6.2, and the subjects were informed about the value of these probabilities in the instructions.

Table 6.2: Nature behaviour

Number of remaining tokens	5	4	3	2	1	0
Probability the loss is be paid by others	0.88	0.66	0.38	0.25	0.24	0.0

Table 6.3 shows the equilibrium values and payoff consequences of the two regimes. The numbers in the cells are the equilibrium number of purchased tokens and the expected Nash-equilibrium payoffs. In the discretionary regime, the optimal behaviour of the official is to move the payment on other group members whenever the remaining number of tokens totals to at least four. This follows from the condition (6.4). Given this behaviour of the official, the player's best response would thus be to invest into two additional tokens (condition (6.3)). In the strict rule regime, the player's optimal behaviour would be to purchase four additional tokens as the expected marginal benefit for the fourth token is still 12 CZK, but the marginal benefit of the fifth token 8 CZK, which is less than the token price. The expected payoff is calculated as the endowment minus the price of the additional tokens and expected loss.

Table 6.3: Equilibrium prediction

	Regime	
	Discretion	Strict rule
Purchased tokens	2	4
Frequency of losses	0.38	0.14
Expected payoff	88	86

The theoretical framework and the parameterization of the experiment allow us to test the following hypotheses in relation to the efficiency of the discretionary regime. Since the experiment is played in multiple rounds with a partner-matching protocol, we expect that the official will be able to solve the dynamic inconsistency problem. Propositions 6.3.1 and 6.3.2 from the theoretical model then lead to the following hypotheses with regards to the effect of the discretionary regime.

Hypothesis 1: The effort measured by purchased tokens is lower in the discretionary regime compared to the strict rule regime

Hypothesis 2: The average monetary payoff is higher in the discretionary regime compared to the strict rule regime.

Our next interest is in the voting decision. While we do not formally model the impact of discretion aversion, the discussion of betrayal aversion and valuation of decision rights identifies several channels as to why people might be biased against the discretionary regime. The bias can operate via two channels. Firstly, and in the most direct sense, ambiguity-averse subjects might prefer some clear and foreseeable rule to the official's discretionary power. Secondly, subjects might dislike the fact that their payoff is dependent on the behaviour of other group members. This leads to the following conjecture.

Conjecture 1: The subjects will be biased against the discretionary regime.

In order to identify discretion-averse or discretion-loving preferences, we use the following identification strategy. According to the standard assumption that people care about their own material payoff, the strict-rule regime should be rejected in both treatments. However, it is easy to think of non-standard preferences where people have some preference towards the discretionary or the strict-rule regime. Assume that players have a utility function

$$U_i(m_i, D) = \alpha_0 D + \alpha_1 m_i + \epsilon_i,$$

where m is the monetary payoff of the player, D is a dummy variable which takes the value 1 if the regime is discretionary and ϵ_i is the unobserved portion of the utility. Based on the voting decision and actual payoffs in the first two stages of the experiment, we can use discrete choice techniques (Train, 2009) to identify the parameters α_0 and α_1. The parameter α_0 is interpreted as an alternative-specific constant indicationg the utility of discretionary regime not explained by monetary payoff average utility. Negative values of this parameter suggest discretion aversion and vice versa. Ratio α_0/α_1 further measures the monetary premium for being in the discretionary regime.

The purpose of HUMAN and NATURE treatment is to test whether the presumed unpopularity of discretionary regime stems from the ambiguity related to the official's decision. If so, we should observe that the mere presence of a human official makes the discretionary regime less popular and therefore players would vote less for the discretionary regime in the HUMAN treatment. On the other hand, if the discretionary regime also remains unpopular in the NATURE treatment, the reason should be the unpopularity of the procedure i.e. due to that notion that my payoff depends on the actions of other members of the group.

6.5 Data

The experiment was conducted in October 2018 at the Masaryk University Experimental Economics Laboratory in Brno, Czech Republic. In total, we recruited 212 student subjects using hroot (Bock et al., 2014). The experiment environment was programmed in zTree (Fischbacher, 2007). There were 12 experimental sessions in total, with 6 sessions of the HUMAN treatment and 6 sessions of NATURE treatment. The experiment took about two hours and participants received 254 CZK on average.

Table 6.4 shows a mean of the selected variables in the HUMAN and NATURE treatment. The table includes choice variables, socio-demographic variables and psychological scales.

Figure 6.1 plots the histograms of purchased tokens and monetary payoffs, showing that the treatment does not influence the behaviours of players. In particular, there are no statistical differences between the HUMAN and NATURE treatment in the number of tokens bought by the participants (Mean NATURE=2.73, Mean HUMAN=2.86, t-test $p = 0.396$). There are also no statistical differences in payment difference between the discretionary regime and the strict-rule regime (Mean NATURE=2.86, Mean HUMAN=3.08, t-test $p = 0.911$). The same conclusions hold true not only for the mean number of tokens and payment differences, but also for the whole distributions (K-S test $p = 0.636$ and $p = 0.895$, respectively). These observations confirm that the NATURE treatment mimics the decision of human officials and there are henceforth no differences between these treatments in terms of the player's behaviour or monetary payoffs.

Figure 6.1: The effect of HUMAN and NATURE treatment on the number of tokens and payment difference. The figure shows the histograms of these variables.

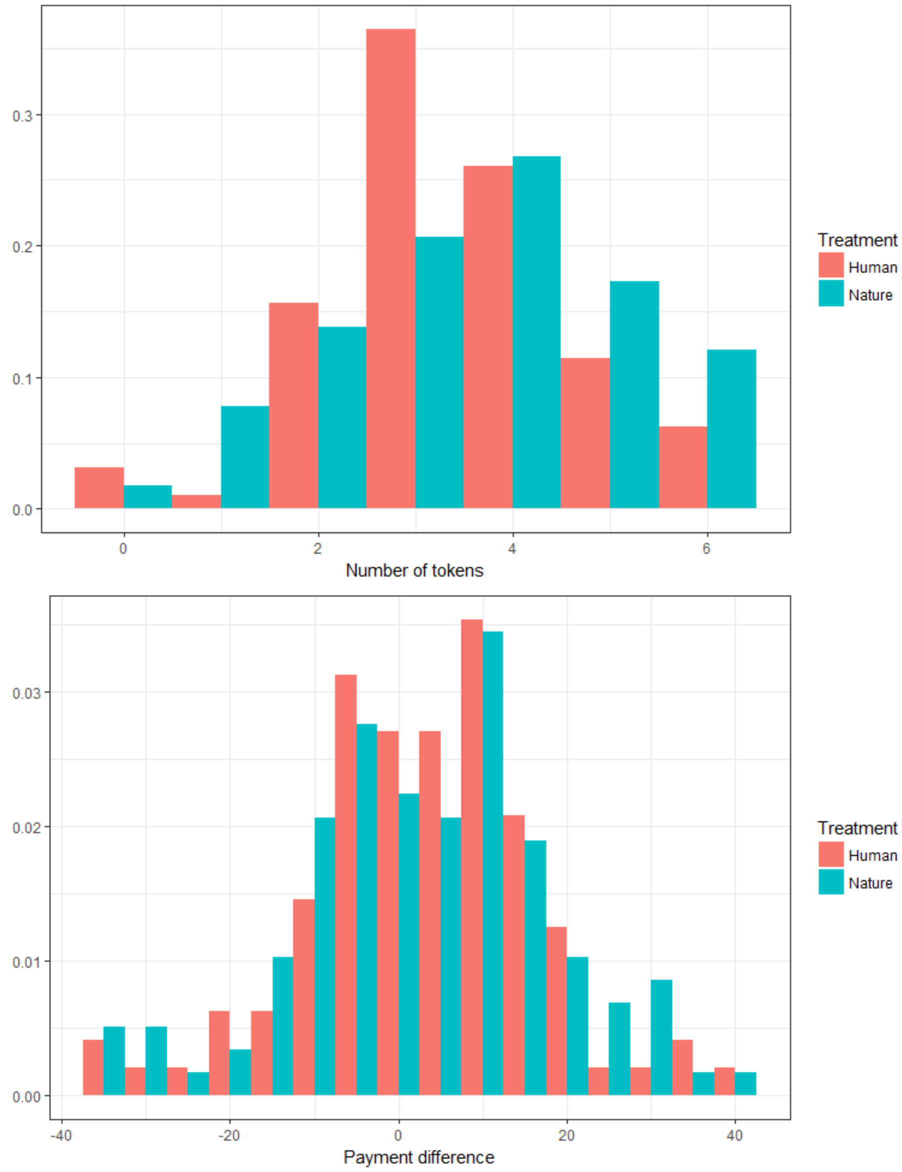


Table 6.4: Descriptive statistics

	Nature treatment	Human treatment
Subjects	116	96
Groups (Independent observations)	29	24
Discretion payoff	86.8	86.6
Rule payoff	83.94	83.52
Discretion tokens	2.73	2.86
Rule tokens	3.84	3.61
Frequency of harm in discretion	0.3	0.29
Frequency of harm in rule	0.18	0.2
Female	0.56	0.51
Age	22.04	21.34
Students of economics or business	0.69	0.63
Working	0.56	0.45
Risk	5.43	5.28
Ambigutity scale	11.17	11.33
Consequentialist scale	8.89	8.31
BF extraversion	5.3	5.46
BF agreeableness	5.11	5.16
BF conscientiousness	5.69	5.69
BF neuroticims	5.25	5.5
BF openness	6.27	6.4

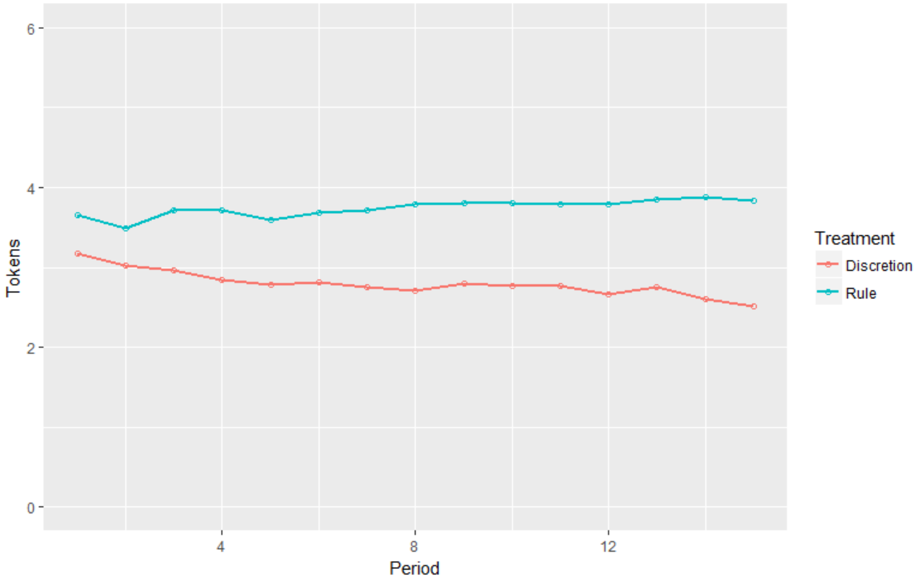
6.6 Results

This section presents the results of the experiment, and we focus on two separate questions. First, we test whether the discretionary regime incentivized subjects to exert sufficient effort and whether the discretionary regime is more efficient than the strict-rule regime. Second, we analyse the subjects' preferences, which were elicited via voting.

6.6.1 Efficiency

This section mainly focuses on how the enforcement regime affects effort expenditures and efficiency. Recall the two hypotheses to be tested. The first hypothesis states that the effort expenditures under the discretionary regime are lower and therefore closer to the first-best effort level. The effort expenditures are measured by the number of tokens that are purchased by the player. Figure 6.2 plots the evolution of the purchased tokens over 15 periods. The figure shows that there is a clear difference between the two regimes with the average number of tokens higher in the strict rule regime. The

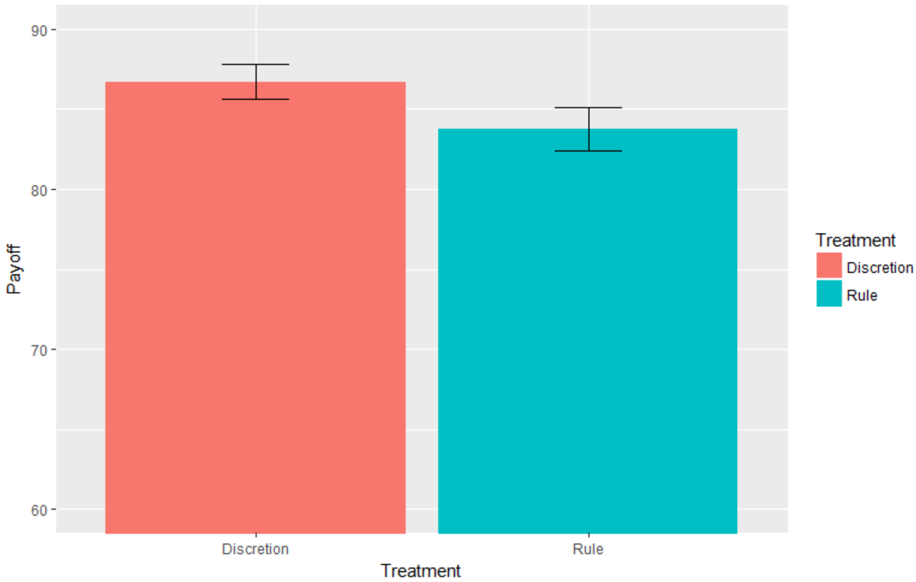
Figure 6.2: The evolution of the number of purchased tokens in strict rule and discretionary regime.



difference tends to be modestly higher in later periods. While the number of purchased tokens is fairly stable in the strict rule regime, it is slowly decreasing in the discretionary regime. It seems like increased experience and learning about official's behaviour leads the number of purchased tokens skewing towards the predicted value in the first 4 periods. The modest decrease in the last two periods may be attributable to the end-game effects, when the players predict that the officials will be more willing to transfer the cost. The second hypothesis states that the players' average monetary payoff is higher in discretionary regime. Figure 6.3 plots the monetary payoffs in both regimes, together with a 95% confidence intervals.

When testing for a statistical significance of the differences, we compared the group averages using the t-test and the non-parametric Mann-Whitney test. When conducting these tests, we pooled the data from HUMAN and NATURE treatments, since there are no differences between these treatments in the number purchased tokens or payoffs. Both hypotheses were confirmed by our data. We find that a shift from strict rule regime to discretionary regime reduces the effort level (t-test $p < 0.001$, Mann-Whitney $p < 0.001$) and increases the monetary payoff (t-test $p = 0.001$, Mann-Whitney $p = 0.001$).

Figure 6.3: The figure shows the average payoff in discretionary and strict rule regime, with the bars depict the 95% confidence intervals based on independent observations.



The theoretical prediction states that players should purchase 4 tokens in the strict-rule regime and 2 tokens in the discretionary regime. To evaluate this point prediction, we tested whether group averages are different from this prediction. The results show that the players have a tendency to over-invest in the effort in the discretionary regime (t-test $p < 0.001$, Mann-Whitney $p < 0.001$) and under-invest in strict rule regime (t-test $p < 0.001$, Mann-Whitney $p < 0.001$). In a similar way, we test the deviation of actual monetary payoffs from the equilibrium payoffs. Since the player's and official's behaviour deviate from optimal behaviour, the actual payoffs are significantly lower than the predicted payoffs in discretionary (t-test $p = 0.003$, Mann-Whitney $p < 0.001$) as well as strict rule regime (t-test $p = 0.002$, Mann-Whitney $p < 0.001$).

The hypotheses are also supported by the regression models in Table 6.5, where we control for age, gender, self-reported risk attitude, order, and number of lost tokens. The estimates are based on data from all 15 periods. The standard errors are clustered at the group level. Models in columns (1) and (3) contain only the treatment variables for the strict-rule regime (rule regime), and for the discretionary regime with a human official (HUMAN treatment). The table thus confirms the previous results. The strict rule regime leads to higher effort and lower payoffs, and the presence of human official does not have any effect on the players' behaviour. Models (2), (4) and (5) clearly demonstrate that the results remain robust when additional control variables are included in the model. We can also see that more risk-averse subjects tend to purchase more tokens. The order of the regimes does not have any effect.

Table 6.5: Efficiency of the discretionary regime

	Number of tokens		*Monetary payoff*		
	(1)	(2)	(3)	(4)	(5)
Constant	3.896***	3.856***	86.793***	88.706***	111.451***
	(0.315)	(0.318)	(0.821)	(2.466)	(2.030)
Rule regime	1.000***	1.076***	−3.042**	−3.168**	−3.220***
	(0.117)	(0.142)	(1.217)	(1.238)	(0.788)
HUMAN treatment	0.125	0.132	−0.185	−0.464	−0.468
	(0.178)	(0.181)	(1.067)	(1.151)	(0.872)
Age		−0.065**		−0.062	−0.092
		(0.029)		(0.231)	(0.194)
Female		0.155		−2.218**	−2.497***
		(0.135)		(1.038)	(0.855)
Risk		−0.193***		0.032	0.012
		(0.031)		(0.327)	(0.237)
Order		0.067		−0.934	−0.721
		(0.149)		(0.772)	(0.616)
Lost tokens					−10.688***
					(0.368)
Observations	6,360	6,360	6,360	6,360	6,360
R^2	0.100	0.178	0.019	0.027	0.327
Adjusted R^2	0.100	0.177	0.015	0.019	0.327
Note:	Standard errors clustered at the group level, **p<0.05; ***p<0.01				

6.6.2 Voting

Despite the fact that the discretionary regime is more efficient than the rule regime, there was a higher tendency for participants to vote for the rule based regime. In the HUMAN treatment, 60 % of participants have higher payoffs in discretionary regime but only 38 % voted for the discretionary regime. In the NATURE treatments, 60 % of participants have higher payoffs in the discretionary regime but only 45 % voted for the discretionary regime.

Table 6.6 shows a more rigorous analysis of voting decision. The dependent variable *Voting* takes value of one if the player voted for discretionary regime and zero if he or she voted for the complete rule regime. The variable *Payoff difference* is the difference between subject's average payoff in the discretionary regime and average payoff in the rule regime. The results should thus be interpreted as an estimate of a utility function in a discrete choice model. The *constant* therefore represents an alternative-specific variable that measure intrinsic utility of the discretionary regime in the NATURE treatment. The treatment dummy for HUMAN treatment is an additional utility of the discretionary regime when the official is human.

Two main results stand out from model (1) in the Table 6.6. First, there is a significant and substantial bias against the discretionary regime. The average willingness to pay in order to avoid the discretion is around 12 CZK, which is approximately four times the average payment difference between the regimes. Second, the bias against the discretionary regime is not driven solely by the presence of a human official. The average marginal effect of the human official is 0.061 ($p = 0.332$). Subjects in the NATURE treatment are, on average, willing to pay 7 CZK, which is 2.5 times the average payment difference, in order to participate in the strict-rule regime. Although the presence of a human official makes the bias more profound, there are no statistical differences between the treatments. The models in columns (2) to (6) show whether the bias is correlated with some survey questions. All surveys measures are standardized to have a mean 0 and standard deviation 1. We can see that only the consequentialist scale is related to discretion aversion. People who are more sympathetic with the position that some rules needs to be honoured in all circumstances were more likely to vote for the rule regime. People who are extreme consequentialists (i.e. two standard deviations from the mean) did no manifest the bias.

Table 6.6: Logit model explaining voting decision

	Voting					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-0.386^{**}	-0.373^{**}	-0.382^{**}	-0.385^{**}	-0.382^{**}	-0.411^{**}
	(0.166)	(0.166)	(0.165)	(0.166)	(0.169)	(0.162)
Payoff difference	0.061^{***}	0.059^{***}	0.061^{***}	0.060^{***}	0.061^{***}	0.061^{***}
	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)
Human treatment	-0.290	-0.326	-0.297	-0.289	-0.293	-0.242
	(0.271)	(0.278)	(0.266)	(0.269)	(0.273)	(0.259)
Risk		0.237				
		(0.156)				
Ambiguity			-0.289			
			(0.168)			
BF agreeableness				-0.058		
				(0.150)		
BF conscientiousness					0.133	
					(0.133)	
Consequentialist						0.252^{**}
						(0.124)
Observations	212	212	212	212	212	212

Note: Standard errors clustered at the group level, **p<0.05; ***p<0.01

6.7 Robustness checks

The previous section presented evidence supporting the idea of discretion aversion: subjects vote for the rule-based regime despite the fact that this regime is less efficient and leads to lower monetary payoffs. This section provides a robustness check as to whether other motivations could have played a role in generating the observed result.

It is conceivable that voting decisions are biased against discretions because the decision makers are risk averse or loss averse. The distribution of monetary payoffs in the discretionary regime has not only higher mean but also have larger support. The minimum possible values of monetary payoff in the rule and discretionary regime are -10 (agent purchase 5 additional tokens and lose 6 tokens) and -85 (agent purchase 5 additional tokens, lose 6 tokens and pays 75 extra in external costs), respectively. Figure 6.4 shows the actual distribution of monetary payments in discretionary and rule regimes, and confirms that the payoff distributions are different (K-S test $p < 0.001$) with discretionary regime having a larger support. Although we control the payment difference between the regimes, this does not have to be sufficient since risk-averse or loss-averse agents can take into account the whole distributions of monetary payoffs when making the voting decisions. Model 2 in the table 6.6 shows that the bias against discretion is present even if we control for self-assessed risk attitudes. However, this measure is still far from perfect, and it does not rule out the possibility that discretion bias is caused by loss-aversion.

In order to provide conclusive evidence, we conducted an additional treatment in four sessions with 92 subjects. The treatment was conducted in the MUEEL in February 2019, using the same procedures as in the NATURE and HUMAN treatment. The treatment is similar to the NATURE treatment with one difference. Subjects do not choose the compliance effort. Instead, the number of purchased tokens is generated by computer from the empirical distribution function of purchased tokens in NATURE and HUMAN treatments. This procedure exogenously generates the same distribution of as in the other treatments (K-S test $p = 0.976$). Henceforth, subjects in the voting stage simply reveal their preference for the payoff distribution generated by the discretionary or strict-rule regime. If players only care about their own monetary payoffs and the probabilities of securing it, the voting results in this additional treatment should be the same as those found in the NATURE and HUMAN treatment. However, the presence of discretion aversion may create a wedge between the voting decisions.

Figure 6.4: The histogram of payments in discretionary and rule regime.

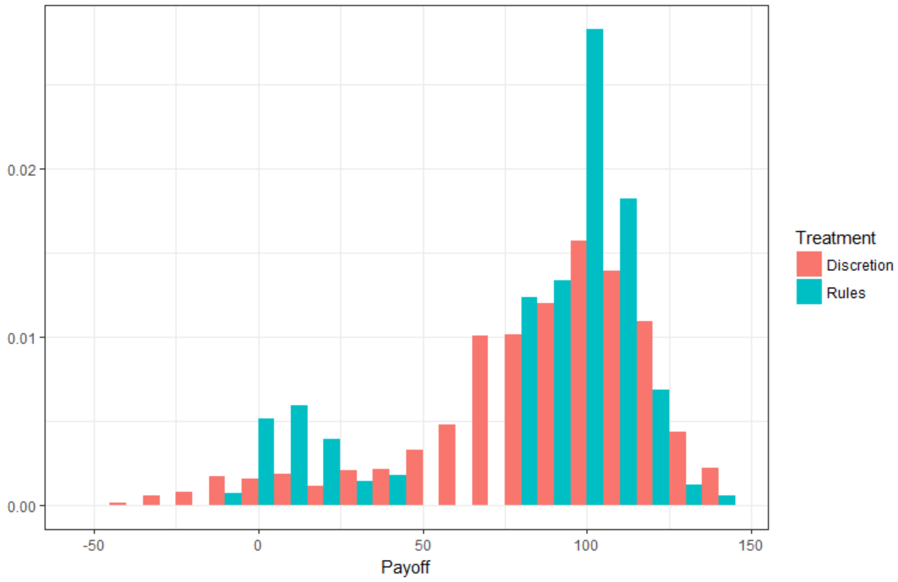


Table 6.7: Logit model explaining voting decision

	Voting		
	(1)	(2)	(3)
Constant	0.091	0.129	0.060
	(0.305)	(0.264)	(0.292)
Payment difference	0.081***	0.066***	0.066***
	(0.026)	(0.012)	(0.012)
NATURE treatment		−0.537*	−0.534*
		(0.307)	(0.307)
HUMAN treatment		−0.833**	−0.828**
		(0.327)	(0.328)
Order			0.129
			(0.244)
Observations	92	304	304

Note: Standard errors clustered at the group level; *p<0.1; **p<0.05; ***p<0.01

Table 6.7 reports the results. The model in column (1) is based on 92 observation in the additional treatment, and the bias against the discretionary regime completely disappears in this treatment. Based on the 304 observations from all three treatments, Models (2) and (3) confirm that subjects voted significantly more often for complete rule regime in NATURE and HUMAN treatment. The average marginal effect is 0.12 in NATURE treatment and 0.18 in HUMAN treatment. We did not observe any order effect. Overall, the results show that discretion aversion is not driven by risk-aversion, loss-aversion or any other preferences for different payoff distribution.

6.8 Discussion

The chapter compares two types of regimes: strict rules and incomplete rules. Strict rules are unambiguously defined but they are not tailored to each particular case. Incomplete rule only articulates the general principle and grant a discretionary power to some enforcement authority to decide whether particular behaviour complies with the rule. In our experimental design, the enforcement authority has certain imperfect information about the compliance effort and faces a dynamic inconsistency problem. We observe that the discretion was more efficient with higher monetary payoffs and had compliance efforts closer to the first-best solution. This result suggests that dynamic inconsistency may not be a relevant concern when assessing the discretion in regulatory and law enforcement.

Furthermore, the experiment provides evidence, based on the revealed preferences, that people preferred complete but inefficient rules to efficient discretion. We call this phenomenon discretion bias or discretion aversion. There are several possible mechanisms behind discretion aversion. Discretion aversion stems from the aversion of being affected by someone else's decision (Bohnet and Zeckhauser, 2004; Bolton and Ockenfels, 2010; Owens et al., 2014) or from the distaste for an incomplete rule per se, or both. While our results do not allow a complete disentanglement from these mechanisms, they suggest that a major part of discretion bias stems from the aversion to a situation when someone is responsible for the harm but does not bear the cost on his own. This conjecture is supported by the fact that discretion bias also occurs in the NATURE treatment and by the observed correlation between the bias and a measure of the consequentialist's attitudes. This result contributes to the literature on the value of autonomy and fair procedures since it suggest that the match or mismatch between decision powers and responsibility is of great importance.

Chapter 7

Conclusion

Recent literature on tax and regulatory compliance has moved on from the benchmark enforcement model (Allingham and Sandmo, 1972). It extends our understanding of how to design the optimal enforcement policy by examining the effects of more sophisticated policy instruments such as targeted audits (Duflo et al., 2018), competitive audit selection mechanisms (Gilpatric et al., 2011, 2015) or flexible fines (Kang and Silveira, 2018). The submitted thesis contributes to this literature by conducting three experimental studies.

The first experimental study investigates whether the competitive audit selection mechanism that was designed for the setting of environmental compliance may be employed also in the tax compliance setting. The crucial difference between these two settings is the availability of information about the individual undisclosed output or income. It is arguable that the tax authority lacks this information and the competitive audit selection mechanism has to be based solely on the reported income. This creates a problem when the tax authority is not able to create groups of taxpayers with homogeneous income since the audit selection mechanism is not able to distinguish between tax evasion and lower actual income. Results of our study reveal strength as well as weakness of this audit selection mechanism. The audit selection mechanism still performs better than random audits but on the other hand the low-income taxpayers are audited more frequently than medium- and high-income taxpayers. Henceforth, there arises a question whether this feature of the mechanism would not be perceived as unfair.

Our second experimental study addresses the question of optimal enforcement policy when it is possible to conceal the non-compliance (Bayer and

Sutter, 2009; Bayer, 2006). The concealment investments are undesirable for two reasons; i) it reduces the leverage of the regulatory enforcement ii) it consumes real resources that could be used more productively. The experimental design is based on a novel theoretical result which shows that there is stark difference between conducting more audits and conducting audits in a more competitive way. The first policy reduces non-compliance; however it also leads to higher investment into the concealment activities. On the other hand, making the audit selection mechanism more competitive reduces both non-compliance and concealment. This prediction was confirmed by the experimental results. This study therefore suggests that the advantages of the competitive audit selection mechanism may go beyond just compliance enforcement.

The last study presented in the thesis contributes to the discussion on discretion and rules in regulatory enforcement. There are arguments in favor of discretion (account for specific circumstances) as well as against (dynamic inconsistency, regulator may follow its own goals). Since many arguments again discretion may be solved by proper design of regulator's incentive, we focus on dynamic inconsistency which stems from the absence of commitment. We design an experiment where the regulator has noisy information about the compliance effort and faces a dynamic inconsistency problem. In the experiment, the regulator can overcome the dynamic inconsistency problem and discretion is more efficient than regulation based on strict rules. Surprisingly, the participants of the experiment still prefer the rule based regulatory regime. This result shows a limitation of considerations based solely on monetary efficiency since people may have aversion against some procedures. In particular, it seems that people have aversion against the mismatch between decision power and responsibility.

Chapter 8

Appendix

Appendix A

Instructions for incomplete treatment (in Czech original)

Úvod

V tomto experimentu zkoumáme vaše rozhodování jednotlivců i jako skupiny. Na vašich rozhodnutích závisí, kolik si vyděláte peněz. Proto vám doporučujeme si následující instrukce důkladně prostudovat. Vydělané peníze vám vyplatíme na konci experimentu v hotovosti a v soukromí. Budete se rozhodovat samostatně bez komunikace s ostatními účastníky experimentu. Pokud vás při čtení instrukcí nebo později při samotné hře napadne nějaký dotaz, prosíme, zvedněte ruku a moderátor experimentu k vám přijde a dotaz zodpoví. Během celého experimentu, prosíme, nekomunikujte s ostatními účastníky, nepoužívejte mobilní telefon ani jiná elektronická zařízení vyjma počítače, u kterého jste usazeni a věnujte svoji pozornost výhradně experimentu. V případě neuposlechnutí budete vyloučeni z experimentu bez nároku na odměnu.

Průběh experimentu

Experiment bude probíhat ve skupině po pěti lidech: vy a další čtyři účastníci. Hráči ve vaší skupině sedí v této místnosti, ale neřekneme vám, kdo do vaší skupiny náleží. Do skupiny budete rozlosováni náhodně na začátku experimentu a v průběhu experimentu se složení vaší skupiny nebude měnit.

Experiment sestává z 30 identických kol. Na začátku každého kola od nás obdržíte částku, která bude vždy znovu náhodně vybrána z rozmezí od 0 do 200 Kč. Každá korunová částka může být vybrána se stejnou pravděpodobností. Výběr této částky si tedy lze představit jako jedno tažení z klobouku, který obsahuje 201 míčků s čísly od 0 do 200. V dalším kroku budete požádáni, abyste nám nahlásili určitou částku. Tato nahlášená částka může být stejná nebo nižší než částka, kterou jste od nás obdrželi. Např. pokud od nás obdržíte částku 102 Kč, můžete nahlásit libovolnou celočíselnou částku mezi 0 a 102 Kč. Z nahlášené částky přijdete s jistotou o 60 %, zbyde vám tedy 40 % z nahlášené částky. Části peněz, které jste od nás obdrželi na začátku kola a nenahlásili je, budeme říkat nenahlášená částka. Osud nenahlášené částky závisí na náhodě. S určitou pravděpodobností nastane nepříznivá událost a vy přijdete o celou nenahlášenou částku. V opačném případě si celou nenahlášenou částku ponecháte.

Základní pravděpodobnost, že nastane nepříznivá událost a vy přijdete o nenahlášenou částku, je 40 %. Navíc se za každých 10 Kč, o které bude vaše nahlášená částka nižší než průměr částek nahlášených ostatními hráči z vaší skupiny, zvýší pravděpodobnost nepříznivé události o 4 procentní body. Naopak každých 10 Kč nad průměrem ostatních hráčů ve skupině znamená snížení pravděpodobnosti nepříznivé události o 4 procentní body. Pokud např. nahlásíte o 81 Kč vyšší částku než je průměr ostatních hráčů ve vaší skupině, je pravděpodobnost nepříznivé události o 81 x 0,4 = 32,4 procentních bodů nižší než základ, to znamená 40 – 32,4 = 7,6 %. Pokud byste naopak nahlásili o 35 Kč nižší částku, bude pravděpodobnost 40 + (35 x 0,4) = 54 %. Zda nastane nepříznivá událost, se losuje v každém kole znovu.

Výplaty

V každém kole tohoto experimentu mohou nastat dvě situace: 1. Nastane nepříznivá událost. Vaše výplata v daném kole bude 0,4*nahlášená částka. 2. Nenastane nepříznivá událost. Vaše výplata bude 0,4*nahlášená částka + nenahlášená částka. Na konci každého kola dostanete informaci o tom, zda nastala nepříznivá událost a jaká byla v daném kole vaše výplata. Také vás budeme informovat o tom, kolik v průměru nahlásili ostatní hráči ve skupině a jaká byla pravděpodobnost nepříznivé události. Na konci experimentu od nás dostanete v Kč výplaty z pěti náhodně vybraných kol experimentu.

Appendix B

Instructions for endogenous treatment (in Czech original)

V tomto experimentu zkoumáme vaše rozhodování jednotlivců i jako skupiny. Na vašich rozhodnutích závisí, kolik si vyděláte peněz. Proto vám doporučujeme si následující instrukce důkladně prostudovat. Vydělané peníze vám vyplatíme na konci experimentu v hotovosti a v soukromí. Budete se rozhodovat samostatně bez komunikace s ostatními účastníky experimentu. Pokud vás při čtení instrukcí nebo později při samotné hře napadne nějaký dotaz, prosíme, zvedněte ruku a moderátor experimentu k vám přijde a dotaz zodpoví.

Během celého experimentu, prosíme, nekomunikujte s ostatními účastníky, nepoužívejte mobilní telefon ani jiná elektronická zařízení vyjma počítače, u kterého jste usazeni a věnujte svoji pozornost výhradně experimentu. V případě neuposlechnutí budete vyloučeni z experimentu bez nároku na odměnu. Průběh experimentu

Experiment bude probíhat ve skupině po pěti lidech: vy a další čtyři účastníci. Hráči ve vaší skupině sedí v této místnosti, ale neřekneme vám, kdo do vaší skupiny náleží. Do skupiny budete rozlosováni náhodně na začátku experimentu a v průběhu experimentu se složení vaší skupiny nebude měnit.

Experiment sestává z 30 identických kol. Na začátku každého kola od nás obdržíte částku, která bude vždy znovu náhodně vybrána z rozmezí od 0 do 200 Kč. Každá korunová částka může být vybrána se stejnou pravděpodobností. Výběr této částky si tedy lze představit jako jedno tažení z klobouku, který obsahuje 201 míčků s čísly od 0 do 200.

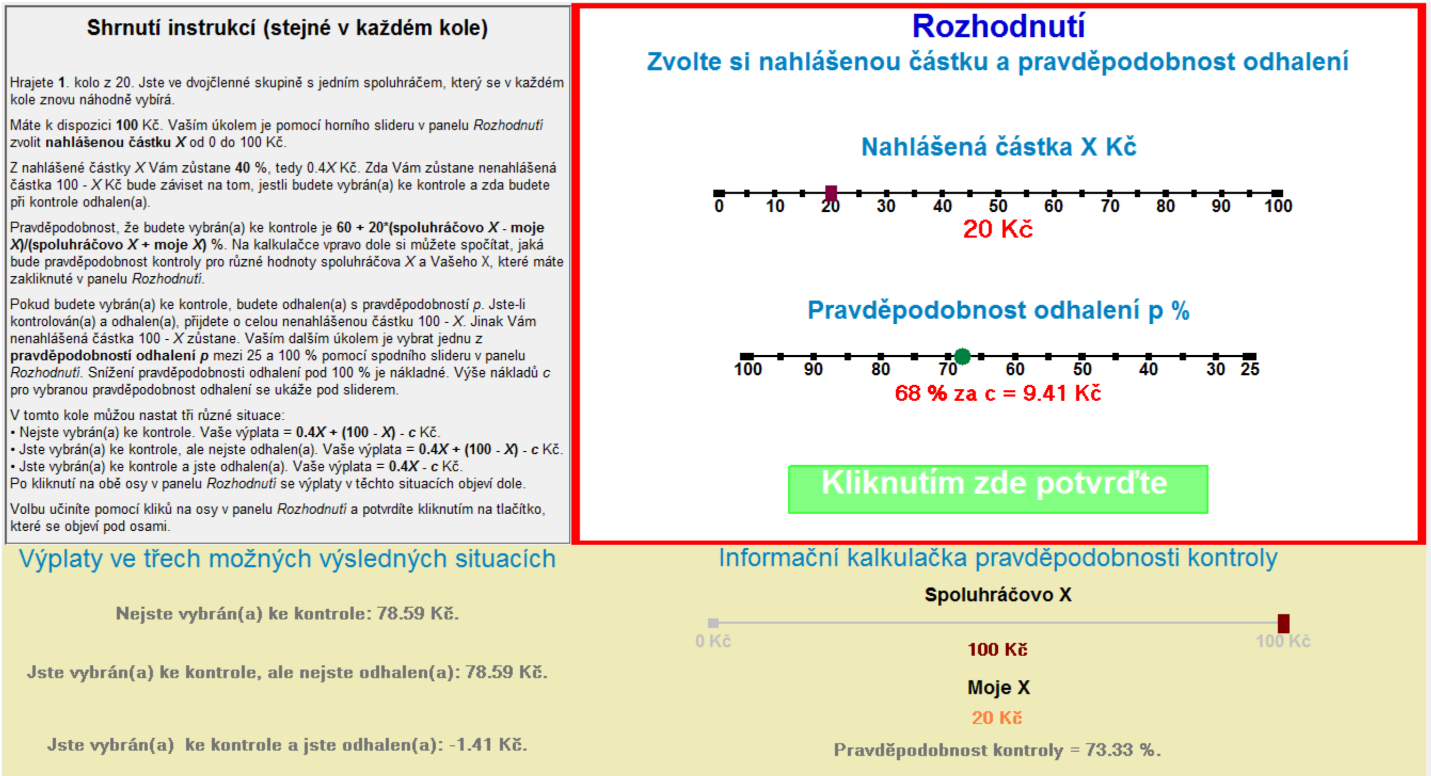
V dalším kroku budete požádáni, abyste nám nahlásili určitou částku. Tato nahlášená částka může být stejná nebo nižší než částka, kterou jste od nás obdrželi. Např. pokud od nás obdržíte částku 102 Kč, můžete nahlásit libovolnou celočíselnou částku mezi 0 a 102 Kč. Z nahlášené částky přijdete s jistotou o 60 %, zbyde vám tedy 40 % z nahlášené částky. Části peněz, které jste od nás obdrželi na začátku kola a nenahlásili je, budeme říkat nenahlášená částka. Osud nenahlášené částky závisí na náhodě. S určitou pravděpodobností nastane nepříznivá událost a vy přijdete o celou nenahlášenou částku. V opačném případě si celou nenahlášenou částku ponecháte. Základní pravděpodobnost, že nastane nepříznivá událost a vy přijdete o nenahlášenou částku, je 40 %. Navíc se za každých 10 Kč, o které bude vaše nahlášená částka nižší než průměr částek nahlášených ostatními hráči z vaší skupiny, zvýší pravděpodobnost nepříznivé události o 4 procentní body. Naopak každých 10 Kč nad průměrem ostatních hráčů ve skupině znamená snížení pravděpodobnosti nepříznivé události o 4 procentní body. Pokud např. nahlásíte o 81 Kč vyšší částku než je průměr ostatních hráčů ve vaší skupině, je pravděpodobnost nepříznivé události o 81 x 0,4 = 32,4 procentních bodů nižší než základ, to znamená 40 − 32,4 = 7,6 %. Pokud byste naopak nahlásili o 35 Kč nižší částku, bude pravděpodobnost 40 + (35 x 0,4) = 54 %. Zda nastane nepříznivá událost, se losuje v každém kole znovu.

V každém kole tohoto experimentu mohou nastat dvě situace:

1. Nastane nepříznivá událost. Vaše výplata v daném kole bude *0,4*nahlášená částka.*
2. Nenastane nepříznivá událost. Vaše výplata bude *0,4*nahlášená částka + nenahlášená částka.*

Na konci každého kola dostanete informaci o tom, zda nastala nepříznivá událost a jaká byla v daném kole vaše výplata. Také vás budeme informovat o tom, kolik v průměru nahlásili ostatní hráči ve skupině a jaká byla pravděpodobnost nepříznivé události. Na konci experimentu od nás dostanete v Kč výplaty z pěti náhodně vybraných kol experimentu.

Figure 8.1: Experimental environment



Appendix C

Instructions for human treatment (in Czech original)

Vítejte na experimentu. Cílem studie je pochopit, jak se rozhodují lidé v určitých situacích. Za Vaši účast na experimentu si budete moci vydělat peníze v závislosti na Vašich rozhodnutích a na rozhodnutích ostatních účastníků experimentu. Vaše příjmy budou vyjádřeny v českých korunách (Kč). Výplatu dostanete na konci tohoto sezení v hotovosti a v soukromí. Ostatní účastníci nebudou o Vaší výplatě informováni.

Během celého experimentu nekomunikujte s ostatními účastníky, nepoužívejte mobilní telefon ani jiná elektronická zařízení vyjma počítače, u kterého jste usazeni, a věnujte svoji pozornost výhradně experimentu. V případě neuposlechnutí budete vyloučeni z experimentu bez nároku na odměnu. Pokud Vás při čtení instrukcí nebo později při samotné hře napadne nějaký dotaz, prosíme, zvedněte ruku a moderátor experimentu k Vám přijde a dotaz zodpoví.

Tento experiment se skládá ze tří částí. Nyní Vám přečtu instrukce k části 1. Prosím, pozorně poslouchejte.

Část 1

Účastníkům experimentu v této místnosti náhodně přiřadíme jednu ze dvou rolí: hráč nebo kontrolor. Na každého kontrolora připadají čtyři hráči. Následně náhodně utvoříme skupiny po 5 účastnících tak, aby obsahovaly čtyři hráče a jednoho kontrolora. V průběhu celého experimentu nedostanete žádnou informaci o identitě lidí, kteří jsou s Vámi ve skupině.

Tato část experimentu se skládá z 20 identických kol. Na začátku každého kola dostane každý hráč 6 žetonů a 150 Kč, za které si může koupit dalších 0 až 6 žetonů. Cena každého dalšího žetonu je 10 Kč. Po nákupu bude mít každý hráč na začátku kola počet zakoupených žetonů + 6 žetonů a 150 – 10 * počet zakoupených žetonů Kč.

V průběhu kola může hráč 0 až 6 žetonů ztratit. Následující graf ukazuje pravděpodobnostní rozdělení ztráty žetonů. S pravděpodobností 26 % hráč neztratí žádný žeton (viz první sloupec grafu). Na konci kola tedy bude mít všechny žetony, které dostal a které si přikoupil. S pravděpodobností 20 % ztratí 1 žeton (viz druhý sloupec grafu); 2 žetony ztratí s pravděpodobností 16 %; 3 žetony s pravděpodobností 12 %; 4 žetony s pravděpodobností 12 %; 5 žetonů s pravděpodobností 8 % a 6 žetonů s pravděpodobností 6 %. Počet ztracených žetonů se losuje pro každého hráče zvlášť. V jednom kole tedy může každý hráč ve skupině ztratit jiný počet žetonů. Losuje se také v každém kole znovu. Jeden hráč tedy může v různých kolech ztratit různé množství žetonů.

Na konci kola pak proběhne kontrola všech hráčů. Pokud některý z nich má méně než 6 žetonů, zaplatí náklad ve výši 100 Kč. Kontrola probíhá automaticky. Kontrolor tedy v této části experimentu nemá žádnou úlohu. V průběhu celé části 1 tedy uvidí pouze obrazovku s instrukcí, že má čekat, až tato část skončí. Nedostane žádné informace o tom, jak se rozhodují a jak dopadli hráči v jeho skupině. Jeho odměna z této části se však odvíjí od výsledků hráčů v jeho skupině.

Každé kolo bude mít tyto čtyři fáze:

1. Nakupování žetonů: Hráči si na slideru vyberou, kolik si koupí dalších žetonů z rozmezí 0 až 6. Pro zvolený počet žetonů se jim ukáže, kolik budou mít celkem žetonů, kolik jim zbyde peněz a jaká je celková pravděpodobnost, že budou mít na konci kola méně než 6 žetonů. Poté, co se hráč rozhodne, kolik dalších žetonů chce koupit, potvrdí své rozhodnutí pomocí tlačítka, které se objeví v pravém dolním rohu obrazovky.

2. Mizení žetonů: Hráčům zmizí 0 až 6 žetonů podle výše uvedeného pravděpodobnostního rozdělení.

3. Kontrola: Na všechny hráče, kteří mají méně než 6 žetonů, dopadne náklad 100 Kč.

4. Výsledky: Na konci kola hráči uvidí, kolik jim zůstalo žetonů, zda na ně dopadl náklad a celkovou výplatu. O ostatních hráčích ve skupině se dozví, zda měli méně než 6 žetonů a zda na ně dopadl náklad. Hráči neuvidí identity ostatních hráčů, uvidí pouze výsledky hráčů v náhodném pořadí, které se bude v každém kole měnit. Nebudou tedy schopni sledovat, jak si vede konkrétní hráč v jednotlivých kolech.

Výplaty: Z části 1 bude k výplatě náhodně vybráno jedno z celkového počtu 20 kol. V jednotlivých kolech může hráč získat některou z těchto výplat:

- Pokud bude mít na konci kola více než 6 žetonů, bude jeho výdělek na konci kola *150 – 10 * počet zakoupených žetonů Kč.*

- Pokud bude mít na konci kola méně než 6 žetonů, bude jeho výdělek na konci kola o 100 Kč nižší než v předchozím případě, tedy *50 – 10 * počet zakoupených žetonů Kč.* Kontrolor dostane průměrné výplaty všech hráčů své skupiny ze všech kol části 1. Jeho výplata se tedy spočítá jako součet všech 80 výplat části 1 (4 hráči krát 20 kol) děleno 80.

Část 2

Složení skupin i role hráčů zůstávají stejné jako v části 1. Část 2 má celkem 20 identických kol. Hlavní rozdíl oproti části 1 spočívá v aktivní roli kontrolora, který se rozhoduje, zda hráči, kteří mají na konci méně než 6 žetonů, ponesou náklad 100 Kč, nebo zda každý ze zbylých tří hráčů zaplatí 20 Kč. Jinak je struktura kola stejná jako v části 1.

Na začátku každého kola dostane každý hráč 6 žetonů a 150 Kč, za které si může koupit dalších 0 až 6 žetonů. Cena každého dalšího žetonu je 10 Kč. V průběhu kola může hráč 0 až 6 žetonů ztratit. Pravděpodobnostní rozdělení, podle kterého se budou žetony ztrácet, je stejné jako v části 1.

Na konci kola pak proběhne kontrola všech hráčů. Pokud některý z nich má méně než 6 žetonů, pak vzniká náklad. Kontrolor se rozhodne, zda ponese daný hráč náklad 100 Kč, či zda přenese menší náklad v celkové výši 60 Kč na ostatní tři hráče. Pokud se rozhodne náklad přenést, rozdělí se tento menší náklad mezi hráče rovnoměrně; každý z nich tedy zaplatí 20 Kč. Hráč, který má méně než 6 žetonů, v tomto případě neplatí nic.

Každé kolo bude mít tyto čtyři fáze:

1. Nakupování žetonů: Hráči si na slideru vyberou, kolik si koupí dalších žetonů z rozmezí 0 až 6. Pro zvolený počet žetonů se jim ukáže, kolik budou mít celkem žetonů, kolik jim zbyde peněz a jaká je celková pravděpodobnost, že budou mít na konci kola méně než 6 žetonů. Poté, co se hráč rozhodne, kolik dalších žetonů chce koupit, potvrdí své rozhodnutí pomocí tlačítka, které se objeví v pravém dolním rohu obrazovky.

2. Mizení žetonů: Hráčům zmizí 0 až 6 žetonů podle výše uvedeného pravděpodobnostního rozdělení.

3. Kontrola: Kontrolor se rozhodne, zda hráč, který má méně než 6 žetonů, ponese náklad ve výši 100 Kč, nebo zda přenese menší náklad na ostatní hráče, takže každý ze tří zbývajících hráčů zaplatí 20 Kč. Kontrolor u každého hráče uvidí, kolik mu zbylo žetonů. Neuvidí ale identity hráčů; uvidí pouze výsledky hráčů v náhodném pořadí, které se bude v každém kole měnit.

4. Výsledky: Na konci kola každý hráč uvidí, kolik mu zůstalo žetonů, zda na něj dopadly nějaké náklady, a svoji výslednou výplatu z daného kola. O ostatních hráčích ve skupině se dozví, zda měli méně než 6 žetonů a zda na ně dopadl náklad. Hráči neuvidí identity ostatních hráčů, uvidí pouze výsledky hráčů v náhodném pořadí, které se bude v každém kole měnit. Nebudou tedy schopni sledovat, jak si vede konkrétní hráč v jednotlivých kolech. Kontrolor na konci kola uvidí pouze informace o hráčích ve své skupině ve stejné struktuře a pořadí, jak to viděl při kontrole. Informace o své výplatě uvidí až na konci dnešního experimentálního sezení.

Výplaty: Z části 2 bude k výplatě náhodně vybráno jedno z celkového počtu 20 kol. V jednotlivých kolech může hráč získat některou z těchto výplat:

- Pokud bude mít na konci kola 6 žetonů a více, bude jeho výdělek *150 – 10 * počet zakoupených žetonů – 20 * počet přenesených nákladů jiných hráčů Kč.*
- Pokud bude mít na konci kola méně než 6 žetonů a kontrolor na něj nenechá dopadnout náklad, bude jeho výdělek na konci kola *150 – 10 * počet zakoupených žetonů – 20 * počet přenesených nákladů jiných hráčů Kč*
- Pokud bude mít na konci kola méně než 6 žetonů a kontrolor na něj nechá dopadnout náklad, bude jeho výdělek na konci kola o 100 Kč nižší než v předchozím případě, tedy *50 – 10 * počet zakoupených žetonů – 20 * počet přenesených nákladů jiných hráčů Kč.* Kontrolor dostane průměrné výplaty všech hráčů své skupiny ze všech kol části 2. Jeho výplata se tedy spočítá jako součet všech 80 výplat části 2 (4 hráči krát 20 kol) děleno 80.

Část 3

Složení skupin i role hráčů zůstávají stejné jako v části 1 a 2. Počet kol v části 3 bude náhodný. Po každém kole hra skončí s pravděpodobností 0,2 a bude pokračovat s pravděpodobností 0,8.

Hráči ve skupině určí hlasováním, zda budou kola v části 3 stejná jako v části 1, nebo jako v části 2. Kontrolor nehlasuje. Každý hráč zvolí jednu ze dvou možností:

1. „Hra bez kontrolora (jako v části 1)“, ve které na každého hráče s méně než 6 žetony dopadne náklad ve výši 100 Kč.

2. „Hra s kontrolorem (jako v části 2)“, ve které může kontrolor nechat na hráče s méně než 6 žetony dopadnout náklad 100 Kč, nebo náklad přenese a ostatní tři hráči ve skupině zaplatí každý 20 Kč.

Hrát se bude hra, která získá vyšší počet hlasů. Pokud obě hry získají shodně 2 hlasy, počítač vybere hru, která se bude hrát tak, že si hodí korunou (šance 50 na 50).

Výplaty: Z části 3 bude vyplaceno poslední kolo. Výplaty hráčů se počítají dle pravidel části, která vyhraje v hlasování. Kontrolor dostane průměrné výplaty všech hráčů své skupiny ze všech kol části 3.

Bibliography

Aimone, J. A. and Houser, D. (2012). What you don't know won't hurt you: a laboratory analysis of betrayal aversion. *Experimental Economics*, 15(4):571–588.

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

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance*, 19(1):54–77.

Alm, J., Bloomquist, K. M., and McKee, M. (2015). On the external validity of laboratory tax compliance experiments. *Economic Inquiry*, 53(2):1170–1186.

Alm, J., Cronshaw, M. B., and McKee, M. (1993). Tax compliance with endogenous audit selection rules. *Kyklos*, 46(1):27–45.

Alm, J., Deskins, J., and McKee, M. (2009). Do individuals comply on income not reported by their employer? *Public Finance Review*, 37(2):120–141.

Alm, J., Jackson, B., McKee, M. J., et al. (1992). Estimating the determinants of taxpayer compliance with experimental data. *National Tax Journal*, 45(1):107–114.

Alm, J., McClelland, G. H., and Schulze, W. D. (1999). Changing the social norm of tax compliance by voting. *Kyklos*, 52(2):141–171.

Alm, J. and McKee, M. (2004). Tax compliance as a coordination game. *Journal of Economic Behavior & Organization*, 54(3):297–312.

Alm, J., Sanchez, I., De Juan, A., et al. (1995). Economic and noneconomic factors in tax compliance. *Kyklos*, 48:1–18.

Alonso, R. and Matouschek, N. (2008). Optimal delegation. *The Review of Economic Studies*, 75(1):259–293.

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

Armstrong, M. and Vickers, J. (2010). A model of delegated project choice. *Econometrica*, 78(1):213–244.

Arora, S. and Gangopadhyay, S. (1995). Toward a theoretical model of voluntary overcompliance. *Journal of economic behavior & organization*, 28(3):289–309.

Bartling, B., Fehr, E., and Herz, H. (2014). The intrinsic value of decision rights. *Econometrica*, 82(6):2005–2039.

Bayer, R. and Cowell, F. (2009). Tax compliance and firms' strategic interdependence. *Journal of Public Economics*, 93(11):1131–1143.

Bayer, R.-C. (2006). A contest with the taxman–the impact of tax rates on tax evasion and wastefully invested resources. *European Economic Review*, 50(5):1071–1104.

Bayer, R.-C. and Sutter, M. (2009). The excess burden of tax evasion—an experimental detection–concealment contest. *European Economic Review*, 53(5):527–543.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pages 13–68. Springer.

Bergolo, M., Ceni, R., Cruces, G., Giaccobasso, M., and Perez-Truglia, R. (2017). Tax audits as scarecrows: Evidence from a large-scale field experiment. *NBER Working Paper*, (w23631).

Bester, H. and Krähmer, D. (2008). Delegation and incentives. *The RAND Journal of Economics*, 39(3):664–682.

Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71:117–120.

Bohnet, I. and Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, 55(4):467–484.

Bolton, G. E., Brandts, J., and Ockenfels, A. (2005). Fair procedures: Evidence from games involving lotteries. *The Economic Journal*, 115(506):1054–1076.

Bolton, G. E. and Ockenfels, A. (2010). Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey, and the united states: Comment. *American Economic Review*, 100(1):628–33.

Budner, S. (1962). Intolerance of ambiguity as a personality variable 1. *Journal of personality*, 30(1):29–50.

Butler, J. V. and Miller, J. B. (2017). Social risk and the dimensionality of intentions. *Management Science*, 64(6):2787–2796.

Cason, T. N., Friesen, L., and Gangadharan, L. (2016). Regulatory performance of audit tournaments and compliance observability. *European Economic Review*, 85:288–306.

Cason, T. N. and Gangadharan, L. (2006). An experimental study of compliance and leverage in auditing and regulatory enforcement. *Economic Inquiry*, 44(2):352–366.

Chetty, R., Friedman, J. N., Olsen, T., and Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *The quarterly journal of economics*, 126(2):749–804.

Choo, C. L., Fonseca, M. A., and Myles, G. D. (2016). Do students behave like real taxpayers in the lab? evidence from a real effort tax compliance experiment. *Journal of Economic Behavior & Organization*, 124:102–114.

Clark, J., Friesen, L., and Muller, A. (2004). The good, the bad, and the regulator: An experimental test of two conditional audit schemes. *Economic Inquiry*, 42(1):69–87.

Cremer, H. and Gahvari, F. (1994). Tax evasion, concealment and the optimal linear income tax. *Scandinavian Journal of Economics*, 96(2):219–39.

Cummings, R. G., Martinez-Vazquez, J., McKee, M., and Torgler, B. (2009). Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment. *Journal of Economic Behavior & Organization*, 70(3):447–457.

Dai, Z. (2016). Endogenous crackdowns: Theory and experimental evidence. *Working paper*, Available at SSRN 2799955.

Denicolò, V. (2008). A signaling model of environmental overcompliance. *Journal of Economic Behavior & Organization*, 68(1):293–303.

Dimmock, S. G., Kouwenberg, R., and Wakker, P. P. (2015). Ambiguity attitudes in a large representative sample. *Management Science*, 62(5):1363–1380.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2018). The value of regulatory discretion: Estimates from environmental inspections in india. *Econometrica*, 86(6):2123–2160.

Dwenger, N., Kleven, H., Rasul, I., and Rincke, J. (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in germany. *American Economic Journal: Economic Policy*, 8(3):203–32.

Earnhart, D. and Harrington, D. R. (2014). Effect of audits on the extent of compliance with wastewater discharge limits. *Journal of Environmental Economics and Management*, 68(2):243–261.

Fehr, E., Herz, H., and Wilkening, T. (2013). The lure of authority: Motivation and incentive effects of power. *American Economic Review*, 103(4):1325–59.

Feldman, N. E. and Slemrod, J. (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *The Economic Journal*, 117(518):327–352.

Fellner, G., Sausgruber, R., and Traxler, C. (2013). Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association*, 11(3):634–660.

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

Frey, B. S. and Stutzer, A. (2005). Beyond outcomes: measuring procedural utility. *Oxford Economic Papers*, 57(1):90–111.

Friesen, L. (2003). Targeting enforcement to improve compliance with environmental regulations. *Journal of Environmental Economics and Management*, 46(1):72–85.

Gërxhani, K. and Schram, A. (2006). Tax evasion and income source: A comparative experimental study. *Journal of Economic Psychology*, 27(3):402–422.

Gilpatric, S. M., Vossler, C. A., and Liu, L. (2015). Using competition to stimulate regulatory compliance: a tournament-based dynamic targeting mechanism. *Journal of Economic Behavior & Organization*, 119:182–196.

Gilpatric, S. M., Vossler, C. A., and McKee, M. (2011). Regulatory enforcement with competitive endogenous audit mechanisms. *The RAND Journal of Economics*, 42(2):292–312.

Harbaugh, W. T., Mocan, N., and Visser, M. S. (2013). Theft and deterrence. *Journal of Labor Research*, 34(4):389–407.

Harding, L. (2016). What are the panama papers? a guide to history's biggest data leak. *Guardian*, Retrieved from https://www.guardian.com/.

Harrington, W. (1988). Enforcement leverage when penalties are restricted. *Journal of Public Economics*, 37(1):29–53.

Heyes, A. (2000). Implementing environmental regulation: enforcement and compliance. *Journal of regulatory economics*, 17(2):107–129.

Hong, K. and Bohnet, I. (2007). Status and distrust: The relevance of inequality and betrayal aversion. *Journal of Economic Psychology*, 28(2):197–213.

IRS (2016). Federal tax compliance research: Tax gap estimates for tax years 2008–2010. Technical report, Internal Revenue Service.

Kamijo, Y., Masuda, T., and Uemura, H. (2017). Who is audited? experimental study on rule-based tax auditing schemes. *Working paper*, Institute of Economic Research, Kyoto University.

Kang, K. and Silveira, B. S. (2018). Understanding disparities in punishment: Regulator preferences and expertise. *Working paper 2018.*

Kastlunger, B., Dressler, S. G., Kirchler, E., Mittone, L., and Voracek, M. (2010). Sex differences in tax compliance: Differentiating between demographic sex, gender-role orientation, and prenatal masculinization (2d: 4d). *Journal of economic psychology*, 31(4):542–552.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., and Saez, E. (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica*, 79(3):651–692.

Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *The Quarterly Journal of Economics*, 128(2):669–723.

Kotakorpi, K. (2006). Access price regulation, investment and entry in telecommunications. *International Journal of Industrial Organization*, 24(5):1013–1020.

Kydland, F. E. and Prescott, E. C. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of political economy*, 85(3):473–491.

Laffont, J.-J. and Tirole, J. (1991). The politics of government decision-making: A theory of regulatory capture. *The quarterly journal of economics*, 106(4):1089–1127.

Laffont, J.-J. and Tirole, J. (1993). *A theory of incentives in procurement and regulation*. MIT press.

Landsberger, M. and Meilijson, I. (1982). Incentive generating state dependent penalty system: The case of income tax evasion. *Journal of Public Economics*, 19(3):333–352.

Leaver, C. (2009). Bureaucratic minimal squawk behavior: Theory and evidence from regulatory agencies. *American Economic Review*, 99(3):572–607.

Liu, L., Neilson, W., et al. (2013). Enforcement leverage with fixed inspection capacity. *Strategic Behavior and the Environment*, 3(4):305–328.

Luttmer, E. F. and Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives*, 28(4):149–68.

Mascagni, G. (2018). From the lab to the field: A review of tax experiments. *Journal of Economic Surveys*, 32(2):273–301.

McClelland, J. D. and Horowitz, J. K. (1999). The costs of water pollution regulation in the pulp and paper industry. *Land Economics*, 75(2):220–232.

Meiselman, B. S. (2018). Ghostbusting in detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics*, 158:180–193.

Miceli, T. J. (2008). Criminal sentencing guidelines and judicial discretion. *Contemporary Economic Policy*, 26(2):207–215.

OECD (2017). *Tax Administration 2017-Comparative Information on OECD and Other Advanced and Emerging Economies*. OECD Publishing.

Oestreich, A. M. (2015). Firms' emissions and self-reporting under competitive audit mechanisms. *Environmental and Resource Economics*, 62(4):949–978.

Oestreich, A. M. (2017). On optimal audit mechanisms for environmental taxes. *Journal of Environmental Economics and Management*, 84:62–83.

Owens, D., Grossman, Z., and Fackler, R. (2014). The control premium: A preference for payoff autonomy. *American Economic Journal: Microeconomics*, 6(4):138–61.

Park, C.-G. and Hyun, J. K. (2003). Examining the determinants of tax compliance by experimental data: A case of korea. *Journal of Policy Modeling*, 25(8):673–684.

Pistor, K. and Xu, C. (2003). Incomplete law. *New York University Journal of International Law and Politics*, 35(4):931–1013.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 105(8):2539–69.

Rammstedt, B. and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality*, 41(1):203–212.

Rey, P. and Tirole, J. (1986). The logic of vertical restraints. *American Economic Review*, 76(5):921–39.

Robinson, J. S., Joel, S., and Plaks, J. E. (2015). Empathy for the group versus indifference toward the victim: Effects of anxious and avoidant attachment on moral judgment. *Journal of Experimental Social Psychology*, 56:139–152.

Saez, E. (2010). Do taxpayers bunch at kink points? *American economic Journal: economic policy*, 2(3):180–212.

Sappington, D. E. (1991). Incentives in principal-agent relationships. *Journal of economic Perspectives*, 5(2):45–66.

Sausgruber, R. and Tyran, J.-R. (2014). Discriminatory taxes are unpopular—even when they are efficient and distributionally fair. *Journal of Economic Behavior & Organization*, 108:463–476.

Schildberg-Hörisch, H. and Strassmair, C. (2010). An experimental test of the deterrence hypothesis. *The Journal of Law, Economics, & Organization*, 28(3):447–459.

Shavell, S. (2007). Optimal discretion in the application of rules. *American Law and Economics Review*, 9(1):175–194.

Shimshack, J. P. and Ward, M. B. (2008). Enforcement and over-compliance. *Journal of Environmental Economics and Management*, 55(1):90–105.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *The Journal of Economic Perspectives*, 21(1):25–48.

Slemrod, J., Blumenthal, M., and Christian, C. (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in minnesota. *Journal of public economics*, 79(3):455–483.

Stafford, S. L. (2002). The effect of punishment on firm compliance with hazardous waste regulations. *Journal of Environmental Economics and Management*, 44(2):290–308.

Stafford, S. L. (2008). Self-policing in a targeted enforcement regime. *Southern Economic Journal*, 74(4):934–952.

Torgler, B. (2002). Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. *Journal of Economic Surveys*, 16(5):657–683.

Torgler, B. and Valev, N. T. (2010). Gender and public attitudes toward corruption and tax evasion. *Contemporary Economic Policy*, 28(4):554–568.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Vieider, F. M., Chmura, T., Fisher, T., Kusakawa, T., Martinsson, P., Thompson, F. M., and Sunday, A. (2015). Within-versus between-country differences in risk attitudes: implications for cultural comparisons. *Theory and Decision*, 78(2):209–218.

Zerofsky, E. (2017). How a german newspaper became the go-to place for leaks like the paradise papers. *The New Yorker*, Retrieved from https://www.newyorker.com/.