Luigi Mittone, Matteo Ploner and Eugenio Verrina

When the State Doesn't Play Dice: Aggressive Audit Strategies Foster Tax Compliance.

CEEL Working Paper 2-17

Cognitive and Experimental Economics Laboratory

Via Inama, 5 38100 Trento, Italy

http://www-ceel.economia.unitn.it
tel. +39.461.282313

When the State Doesn't Play Dice: Aggressive Audit Strategies Foster Tax Compliance

Luigi Mittone* Matteo Ploner† Eugenio Verrina‡

Abstract

We experimentally test the effect of aggressive audit strategies on tax compliance. Tax payers first go through a phase of audits managed by a human tax agent who is requested to follow a rule imposed by a fair random device. However, the tax agent can freely decide to break the rule and over-inspect. Afterwards, tax payers are exposed to a genuinely random audit process governed by an algorithm, which makes compliance a strategically dominated option. Our main result is that tax payers are generally over-inspected by the human tax agents and react to this with nearly full compliance. Interestingly, these high levels of compliance persist also when controls are implemented by the algorithm. We conclude that aggressive audit strategies can effectively be used by tax authorities to raise and sustain tax compliance.

Keywords: tax evasion; audit strategies; die under the cup; endogenous inspections.

JEL classification: C91; C92; H24; H26; H83.

*CEEL-DEM, University of Trento. E-mail: luigi.mittone@unitn.it

†CEEL-DEM, University of Trento. Address: DEM-UNITN, Via Inama 5, 38122 Trento, Italy. E-mail: matteo.ploner@unitn.it.

‡Corresponding author. MPI for Research on Collective Goods and CGS-University of Cologne. E-mail: verrina@coll.mpg.de

1 Introduction

State measures to fight tax evasion are usually constrained by legal and economical considerations. Tax authorities try to optimize their strategies given these constraints. In some cases, audits are determined based on sophisticated computer algorithms which are able to detect discrepancies or construct risk profiles based on tax returns. In other cases, tax authorities target tax payers according to observable characteristics, like their profession or domicile.[1] Auditing certain categories or groups more intensively for a given period of time may help a tax authority to coordinate its efforts. Those targeted by these policies might increase their compliance temporarily to avoid sanctions. However, they might also feel unjustly treated,[2] which could lead to a backfire effect. What are the effects of these policies? Do tax payers become more compliant both while and after they are targeted?

We investigate both the dynamic evolution of tax compliance under aggressive audit policies, and their spillover effect once the pressure on tax payers is released. In our experiment, tax payers are asked to contribute repeatedly to a public project that generates indistinct benefits to the group of contributors. Tax agents are instructed to implement inspections only when the roll of a fair die delivers a certain outcome. This sets a rule they are supposed to follow, deontologically. However, tax agents are free to misreport the outcome and implement as many inspections as they like. When a tax inspection is implemented, tax payers are sanctioned with a fine proportional to the amount of taxes evaded. The reward of tax agents is

[1] As an example, this is routinely done by the Italian tax authority. Some recent cases involve dentists (Trentino, 2014), Bed & Breakfasts (Leone-Fell, 2018), or boutiques in the famous winter destination of Cortina D'Ampezzo (Il Fatto Quotidiano, 2012). In the US the IRS has also been targeting its audits, e.g., on small businesses (William, 2019). A prominent case in the US saw allegations that the Obama administration was targeting certain political groups via the IRS (CNN Library, 2018).

[2] An example are the many websites and articles dedicated to help the targets to fight back (Harrison, 2013).

experimentally manipulated and is either fixed or proportional to the taxes paid. Tax payers, then, go through a phase where fines are non-deterrent, as is mostly the case in the real world.

We find that, as expected, tax agents tend to violate the rule set by the random device and over-inspect their tax payers. However, they do so irrespective of their incentive scheme. Tax payers who are more frequently inspected display a higher degree of tax loyalty. Hence, the direct effect of being under strong fiscal pressure is positive. Crucially, compliance stays high also when audits become random and the expected impact of fines is not large enough to deter evasion,, showing a positive spillover effect and no major backfire effect. Interestingly, tax agents can be classified in distinct types based on their auditing strategies. Some of them simply implement constant inspections, while others adopt more sophisticated strategies that teach tax payers high compliance. Tax agents following the rule set by the random device never manage to discipline their tax payers.

Our experiment offers a unique test-bed to investigate the effect of targeted auditing policies on compliance. First, it allows to precisely track compliance levels. Second, it enables us to get at the causal effects of policies that would be very hard to evaluate with real world data. Our results show that aggressive audit strategies increase tax compliance and that, once high compliance is reached, tax authorities can release pressure, without the fear of a backfire effect. These findings prove that these policies constitute an effective instrument in the toolbox of tax authorities.

2 Related Literature

Tax compliance can be tackled from various perspectives, going from its behavioral determinants (see Kirchler et al., 2010, for a review) to its macroeconomic consequences. Similarly multifaceted are the methodological approaches, ranging from theoretical studies (see Allingham and Sandmo, 1972; Srinivasan, 1973; Yitzhaki, 1974, for some of the first theoretical models) to field experiments (see, e.g., Fellner et al., 2013; Hallsworth et al., 2017). Laboratory experiments on tax compliance have by now a long-standing tradition (see Friedland et al., 1978; Alm, 1991; Webley et al., 1991, for early examples).[3] Early laboratory experiments were mainly used to test theoretical models (see, e.g., Alm et al., 1992), but have now gained a more prominent role in the discipline (Alm and Jacobson, 2007). Their internal validity and flexibility can offer broad insights into tax paying behavior (Slemrod and Weber, 2012). Work by Alm et al. (1999), for example, highlighted the importance of social norms in tax compliance, while that of Fortin et al. (2007) stressed the relevance of social interactions. Hence, experiment, if well designed, can be a valuable source of knowledge to understand tax compliance behavior (Alm, 2012).

Our work is closely linked to experiments investigating the effect of different auditing schemes on compliance levels (for an early work on this topic see, Mittone, 1997). Research in this field has found two major regularities: the *echo* and the *bomb crater* effect (see Guala, 2005; Mittone, 2006). The bomb crater effect captures a drop in compliance observed after an inspection has taken place.[4] The echo effect refers to a regularity of behavior leading to higher overall compliance caused by repeated inspections placed in the early part of an individual's taxpaying life. We provide an original contribution to this literature with the introduction of endogenously generated inspection schemes by a tax agent [5].

[3]Muehlbacher and Kirchler (2016) review the methodological debate on the experimental approach to tax compliance, focusing on its external validity.

[4]This drop might be due to a misconception of probabilities leading experimental tax payers to evade more just after they have been detected, Another explanation is that subjects evade more to recover the losses due to fines. Kastlunger et al. (2009) found that a possible way to remove the bomb crater effect is to perform two sequential inspections.

[5]A further reason why we implement endogenous inspections is that, as Baldry (1986) points out, the decision whether to evade or not taxes should not be reduced to a gambling decision. Recent work has introduced endogenous audit schemes in a tax evasion experiment (see, e.g., Vossler et al., 2017). However, inspections were still managed by the computer and not by a human.

In a related study, Clark et al. (2004) compares random audits with different conditional audit schemes according to which participants are assigned to a "good" and a "bad" pool as a function of their past compliance. Conditional audit schemes increase compliance (on this see also, e.g., Cason et al., 2016). We analyze a different policy instrument, since agressive audit strategies are a temporary measure, which could in principle come on top of such conditional audit schemes.

To create a setup where a preset rule can be overturn by a tax agent, we rely on the die-under-the-cup paradigm (Fischbacher and Föllmi-Heusi, 2013), which has extensively been adopted in the literature about truth telling. The essential feature of this procedure is that the real outcome of the random device is known only to the participant rolling the die, leaving her completely free to choose what to report. According to the rule in the experiment, an inspection should take place whenever tax agents roll the number 3.[6] However, they are free to report whichever outcome they want. Abeler et al. (2019) review a large number of studies adopting this paradigm and find that, overall, participants do not lie as much as a purely selfish money maximizer would do. We exploit these findings and expect to observe some, but not too many, departures from the preset audit rule. At the same time, the preset rule constitutes a reference point for tax payers in our experiment, such that deviations therefrom naturally mimic a situation were they are put under pressure.

The fact that tax audits are performed by humans represents a methodological innovation and allows us to link our study to previous works on the effect of different incentive schemes on the performance of tax auditors. Performance pay is found to raise tax revenues (Kahn et al., 2001), although the net welfare gain can be undermined by an increase in bribes to tax auditors (Khan et al., 2015). We exploit these findings and manipulate

[6]Clearly, tax authorities do not inspect (only) based on random draws. However, once the pool of potential evaders has been narrowed down using risk profiles and other techniques, audits can be seen as nearly random events ceteris paribus.

the incetives of tax agents in our experiment, expecting more aggressive audit strategies from tax agents who gain from higher compliance. Surprisingly, we find that tax agents' behavior in our setting is not influenced by incentives.

3 Methodology

3.1 The Tax Game

Participants in the experiment are randomly assigned to one of two roles, Tax Agents (*TA*) and Tax Payers (*TP*)[7]. Each TA is matched with 4 TPs, and together they constitute an independent group of 5 throughout the experiment (partner matching).

In each of the 30 rounds of the experiment, participants in the role of TPs are given an endowment $E = 100$ ECU (Experimental Currency Unit) and are asked to pay a tax t that contributes to a public project (tax rate of 30%)[8]. The tax t is subtracted from E and can be freely chosen over the interval $\{\underline{T}, 1...\overline{T}\}$, with $\underline{T} = 0$ and $\overline{T} = 30$ ECU. The taxes collected in a group of 4 TPs are multiplied by an efficiency factor $\alpha = 1.5$ and the result is equally divided among the four of them. This results in a marginal per-capita return (MPCR) of the projects equal to 0.375.

Before knowing the amount of taxes collected, the TA can decide to implement an inspection, as detailed below. If she decides to do so, all TPs in her group are inspected. When a tax inspection is implemented, a fine F is paid by each TP. The size of F is given by the difference between the tax due ($\overline{T}$) and the amount actually paid t, multiplied by a penalty factor $\phi = 1.5$.

[7] The experiment was framed. See Appendix A for instructions.

[8] This procedure is equivalent to declaring an income and paying a tax according to a given tax rate. Since the ultimate aim of evading taxes is paying lower taxes and not declaring a lower income, we undertook this simplification in our experimental procedures.

Given this setting, the payoff of TPs is given by $\Pi_i^{TP} = E - t_i - C\big((\overline{T} - t_i)\phi\big) + \frac{\alpha}{4}\sum_{j=1}^{4} t_j$, where C is equal to 1 when an inspection is implemented, and to 0 when it is not.

3.2 Treatments

We experimentally manipulate two factors that affect the way inspections are implemented by the TA: in a within-subjects manipulation, we alter the mechanism governing the tax inspections (*Inspection Rule*) and in a between-subjects manipulation, we alter the payoff function of the TA (*Inspection Incentives*).

Table 1: 2x2 Factorial Design and Treatment Labels

		Inspection Rule (within)	
		Human	Machine
Inspection Incentives (between)	Flat	*Flat/Human*	*Flat/Machine*
	Contingent	*Contingent/Human*	*Contingent/Machine*

Concerning the *Inspection Rule*, in the first 20 rounds of the experiment, TAs are asked to roll a fair six-sided die and are instructed to perform a tax inspection when the outcome of the roll is equal to 3. Obviously, the probability of obtaining such an outcome is equal to 1/6. However, the roll of the die is privately performed and participants are aware that no one in the room will be able to check if the actual outcome of the roll corresponds to the self-reported one (see Fischbacher and Föllmi-Heusi, 2013).[9] This phase is labeled *Human* because the actual implementation of inspection rests upon the judgment of the TA: each TA can freely choose to report faithfully the outcome of the roll, as requested to do, or to report a different

[9]Also TPs had to roll a die and report the outcome. Although the outcome reported by TPs had no influence on the game, we decided to let them roll their die to create a fully private condition for tax agents. To this end, each cubicle was made soundproof with a bubble-wrap layer.

outcome, knowing that no sanction, either monetary or non-monetary, is associated to the misreport. In rounds 21–30, the inspection is performed by the computer (i.e., the *Machine*) in a genuinely random way, with the likelihood of inspecting equal to 1/6 in each round.[10]

Concerning the *Inspection Incentives*, in condition *Contingent* the TA earns an amount which is proportional to the sum of taxes paid by the TPs in her group. Specifically, the payoff of the TA is given by $\Pi_{TP} = \beta \sum_{j=1}^{4} t_j$, where $\beta = 1.5$. Differently, in condition *Flat*, the payoff of the TA in each round is fixed and does not depend on choices of the TPs in her group. ($\Pi_{TP} = K$, with K equal to 100 ECU).[11]

3.3 Behavioral Predictions

Given TP's payoff function, a risk-neutral TP is going to evade taxes fully, i.e., to set $t = \underline{T}$, as long as the likelihood of being inspected is $P(C) \leq 0.417$. At the other extreme, when $P(C) > 0.417$, full tax compliance, i.e., $t = \overline{T}$, should be observed. Thus, when inspections are governed by the roll of the die, i.e., $P(C) = 1/6$, full tax evasion is expected. This implies that in condition *Machine*, in the last 10 rounds, no taxes should be paid by TPs. In turn, this means that TA in condition *Contingent/Machine* should earn nothing.

If in condition *Human* the TA truthfully reports the outcome of the die and TPs correctly anticipate this, the same predictions as above hold. However, in case the TA is ready to misreport the outcome and the likelihood of an inspection is higher than the threshold value of 0.417, a radically different pattern emerges, with full compliance ($t = \overline{T}$) on the side of the TPs and maximum earnings for both the TA and TPs.

To understand whether TAs are going to implement inspections with a frequency inducing full compliance, we should examine the payoff in-

[10] TAs were asked to estimate the total contribution of TPs in their group in each round during condition *Machine*.

[11] This amount was chosen to avoid any ex ante inequality concerns.

centives of TAs. In condition *Flat*, where the payoff of the TA is not a function of TPs' choices, we predict that TAs are going to truthfully report the outcome. This prediction is obtained under the assumption of weak preferences for truth telling, i.e., TAs are going to report the truth as long as this comes at no material cost to them. Consequently, given that the TAs truthfully report the actual outcome of the die, full tax evasion should be observed on the side of TPs.

Viceversa, in condition *Contingent*, the payoff of the TA is linearly increasing in the size of the public project. Thus, the TA strictly prefers an outcome of full compliance to all other outcomes. It can easily be shown that in condition *Contingent/Human* two possible equilibria in pure strategies may exist: one without inspection and full evasion (no inspection equilibrium) and one with inspection and full compliance (inspection equilibrium). It is worth noting here that the latter equilibrium is payoff dominant: in the inspection equilibrium, the payoffs of TPs are equal to 115 ECU and that of the TA is equal to 180 ECU; in the no-inspection equilibrium, the payoffs are 55 ECU and 0 ECU, respectively. Hence, we argue that over rounds TAs will enforce a regime of full inspections, misreporting the outcome of the die roll. Anticipating this, TPs will fully comply in each round. Thus, in condition *Contingent/Human*, we expect to observe a large number of inspections, i.e., larger than the equilibrium threshold value of 0.417, and full compliance on the side of the TPs.

Deviations from benchmark predictions sketched above may be observed when a psychological cost is associated with lying. This cost may originate either from moral or self-image concerns.[12] A psychological cost of lying may counteract the monetary incentives to misreport the outcome in condition *Contingent*. If costs of lying are large enough, TAs may prefer to report truthfully the outcome of the die roll. The meta-study be Abeler et al. (2019) shows that people vary in how often they lie. Some refrain

[12] For a formal presentation of preference models incorporating cost of lying see, among others, Kartik (2009) and Ellingsen and Johannesson (2004).

from lie maximally or constantly, achieving suboptimal payoffs, and all possible outcomes are reported with positive probability. We expect this heterogeneity to play itself out in TAs' inspection strategies.

3.4 Participants and Procedures

We conducted a total of 8 experimental sessions, 4 for each between-subjects treatment. The computerized experiment was programmed using the z-Tree software (Fischbacher, 2007). A total of 150 subjects took part in the experiment, 80 in the *Flat* and 70 in the *Contingent* treatment.[13] When entering the laboratory, participants were randomly assigned to a cubicle and were provided with printed instructions (see Appendix A).

Participants were given some minutes go through the instructions privately. Then, a member of the staff read them aloud and answered questions. Before participants were assigned to their role, they had to answer some control questions, checking their comprehension of the instructions. Only after everybody had answered these questions correctly did the actual experiment start.

We set the exchange rate at 0.4 euro for 100 tokens (1 token = 0.004 euros). Participants were paid cumulatively over the 30 rounds of the experiment and received a fixed amount of 3 euros for having shown up. The experiment lasted between 45 minutes and 1 hour. On average, participants earned a total of 15.35 euros.

After condition *Human*, all participants had to fill in a short questionnaire to access the second part of the experiment (see Appendix B, also for results). Subjects reported their degree of agreement to some statements on a 5-point Likert scale. For TPs, we had three items investigating their motivation to comply (if for themselves, for the group, or to avoid sanctions). Moreover, four other items were dedicated to an evaluation of the

[13]This numbers are in line with previous similar experiment (see, e.g., Kastlunger et al., 2009; Gërxhani and Schram, 2006).

behavior of TAs. Concerning TAs, we had three items asking to evaluate their own behavior and an additional item asking how they thought others would have acted in their role. At the end of the experiment, TPs had to fill in a questionnaire similar to the previous one (again, see Appendix B).

4 Results

In this section, we present the results of the experiment. We, first, look at inspections and tax compliance in isolation. Then, we investigate how inspection influenced compliance to answer the main research question of our paper. We, finally, perform an exploratory analysis of different types of auditing strategies adopted by TAs.

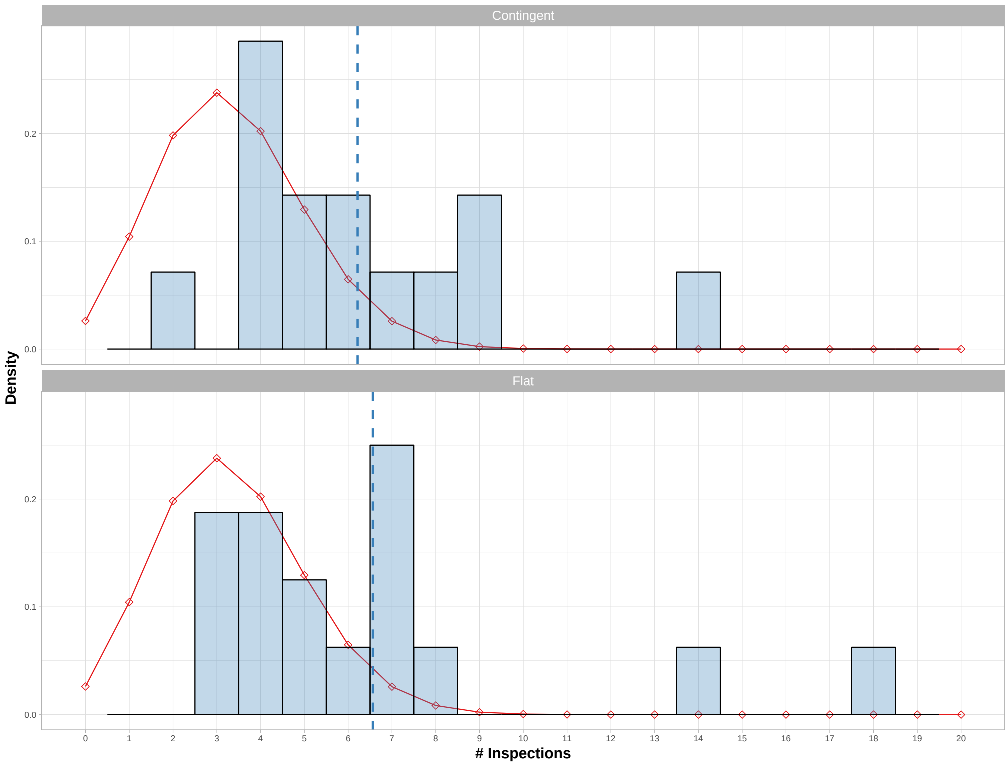
4.1 Inspections

As a first step into the analysis of results, we need to understand how TAs acted in the *Human* treatment, as this builds the stepping stone to understand TPs' behavior, in which we are ultimately interested in. TAs could influence the game by self-reporting the value from the die roll in each of the first 20 rounds, thus determining the occurence of inspections. We have a total of 30 TAs, 14 in the condition *Contingent* and 16 in condition *Flat*.

In Figure 1, we provide a representation of the distribution of the total number of inspections performed by each TAs (bars) together with the theoretical density distribution obtained from a binomial distribution with 20 random draws and a probability of success of 1/6 (solid line). The vertical dashed line captures the mean of the empirically observed distribution.

The overall frequency of inspections is equal to 32.8% and 31.1% in conditions *Flat* and *Contingent*, respectively. A binomial test shows that both frequencies are significantly different from the expected frequency of 16.7% (exact binomial test, both p-values< 0.001). Focusing on choices

Figure 1: Binomial probability distribution vs actual frequency distribution



Note: The bars in the two graphs depict the relative frequency distribution of the total number of inspections carried out by TAs in condition *Human* for incentive condition *Contingent* in the top part (14 obsevations) and *Flat* in the bottom part (16 observations); the dashed horizontal lines show their mean. As a comparison, the solid lines connecting the circles in the two graphs represent the same theoretical density distribution of inspections generated taking 20 random draws from a binomial distribution with success probability 1/6.

of TAs at the individual level, the percentage of TAs who inspect more frequently than what was theoretically predicted is equal to 35.7% and 43.7% in condition *Contingent* and *Flat*, respectively (exact binomial test, with significance level 5%).

Result 1. *Inspections are more frequent than what was predicted by the roll of the die, both in condition* Contingent *and* Flat*.*

Figure 1 also highlights a strong similarity in behavior across treatments. The average frequency of inspection is equal to 6.2 and 6.6 in conditions *Contingent* and *Flat*, respectively. Non-parametric tests show that the two distributions do not statistically differ from each other (Wilcoxon rank sum test, p-value=0.883; Kolmogorov-Smirnov test, p-value=1.000).This goes against our benchmark predictions.

Result 2. *There is no significant difference in inspection frequencies across conditions* Contingent *and* Flat.

4.2 Tax Compliance

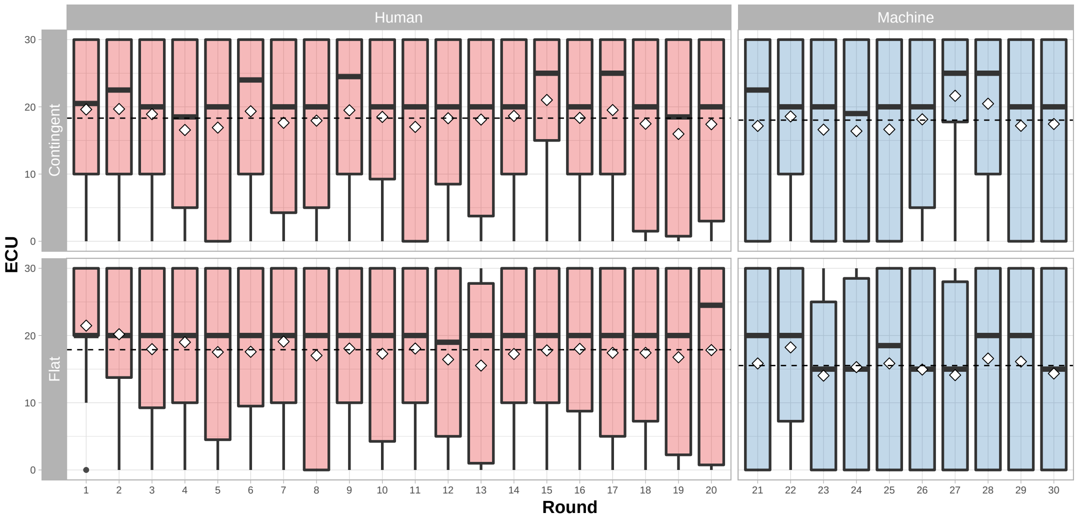
Figure 2 provides a representation of the distribution of taxes paid in the 30 rounds of the experiment. In addition to the conventional pieces of information provided by the boxplots, the diamond dots capture mean values in each round and the dashed horizontal line shows the average value for each of the four experimental conditions separately.

Considering inspection condition *Human* first, the average taxes paid are equal to 18.3 and 17.9 in conditions *Contingent* and *Flat*, respectively. As confirmed also by the boxplots, the central tendency of the distribution is definitely larger than the full evasion prediction obtained under the assumption of risk neutrality and inspections faithfully determined by the outcome of the die. Furthermore, taxes paid are quite stable throughout the 20 rounds of condition *Human*. A signed rank test shows that taxes paid in the first and in the twentieth round are not statistically different (p-value=0.294 and p-value=0.205, respectively, in *Contingent* and *Flat*).[14] Furthermore, taxes paid in the two incentive regimes of the *Human* treatment do not statistically differ (Wilcoxon Rank Sum test, p-value=0.984).

Result 3. *In condition* Human*, taxes collected are positive, quite stable over rounds, and do not differ across condition* Contingent *and* Flat.

[14]To warrant data independence, all tests are performed on averages taken at the group level, unless otherwise mentioned. In total, 16 independent observations are available for condition *Flat* and 14 for condition *Contingent*.

Figure 2: Taxes paid by the TPs



Note: We present boxplots on tax payments for each round. Incentive condition *Contigent* is shown in the top block and *Flat* in the bottom one. Diamonds represent each round's average. The dashed horizontal line depicts average tax payments for condition *Human* and *Machine* for each of the two incentive conditions.

Concerning condition *Machine*, average taxes are higher in condition *Contingent* than in *Flat*, with values equal to 18 and 15.5, respectively. However, no significant differences are registered between the two conditions (Wilcoxon Rank Sum test, p-value=0.334).

Result 4. *In condition* Machine, *taxes collected are positive and quite stable over rounds. No significant differences can be observed between condition* Contingent *and* Flat.

While there is no significant difference between condition *Machine* and *Human* across incentive conditions, we register a drop in taxes paid in incentive condition *Flat* between condition *Machine* and *Human*. This drop between *Flat/Human* and *Flat/Machine* is moderate, but statistically significant (Wilcoxon Signed Rank test, p-value=0.003). There is no significant difference between Contingent/Human and Contingent/Machine (p-value=0.808).

Table 2: Tax Compliance (Linear Mixed Models)

Taxes Paid ∼	
(Intercept)	18.815 (1.364)***
Contingent	0.432 (1.932)
Period	−0.088 (0.033)**
Machine	−3.750 (2.449)
Period:Machine	0.107 (0.099)
Contingent:Machine	2.059 (0.659)**
Num. obs.	3600
Num. groups: ID:Group	120
Num. groups: Group	30

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

To provide further support to our results, we perform a regression analysis on the taxes paid by TPs. The dependent variable *Taxes Paid* is regressed against a set of explanatory variables: *Contingent* is equal to 1 in incentive condition *Contingent* and equal to 0 in *Flat*; *Period* captures the round in which taxes were paid; *Machine* is equal to 1 in condition *Machine* and to 0 in *Human*. Table 2 presents the estimation outcomes of a linear mixed models with clustered random effects at the individual and group level.

As shown by estimates in Table 2, taxes paid are positive and stable over rounds, though slightly declining in *Human*. We confirm that there is no significant difference between incentive conditions, as the coefficient of *Contigent* shows. This supports the findings reported in Result 3. The positive interaction term *Contingent:Machine* points to a significant difference between the *Human* condition in incentive condition *Flat* and the *Machine* condition in incentive condition *Contingent*. However, a linear hypothesis test (*Contingent* + *Contingent:Machine* = 0, Chi-square test, p-value=0.206) shows no difference between *Contingent* and *Flat* in *Machine*. Furthermore, no significant drop in taxes paid between *Human* and *Machine* is observed, neither in condition *Contingent* (see coefficient of *Machine*) nor in condition *Flat* (*Machine* + *Contingent:Machine* = 0, Chi-square test, p-value=0.491). This set of findings also support Result 4 reported above. In addition, estimated parameters show that the drop in contributions observed in *Machine* for incentive condition *Flat* is not statistically significant.

4.3 Inspections and Tax Compliance

We now turn to an analysis of the relationship between inspections and tax compliance. In particular, we want to investigate the impact of inspections on tax compliance when TA could influence the occurrence of an audit, i.e., in condition *Human*, and their spillover effect in the periods where audits are purely random, i.e., in condition *Machine*. To achieve this goal we extend the regression reported in Table 2 above. Specifically, we take into account the impact of the total number of inspections in previous rounds as a potential determinant of tax compliance. Cumulated inspections are counted separately for the *Machine* and *Human* conditions. In Model 1, the overall impact of inspections is assessed, while in Model 2 we distinguish between inspections performed by a human and by the machine.[15]

[15] A priori we can not rule out that inspections performed by humans are endogenous to tax behavior. In particular, it may be that groups showing less compliance attract more controls. However, our data show that higher compliance is associated with more inspections. It seems unlikely that more compliant groups attract more controls. Thus, in light of our findings, we maintain that the causal link goes from inspections to taxes paid, and not the opposite.

Table 3: Tax Compliance and Controls (Linear Mixed Models)

Taxes Paid ∼	Model 1	Model 2
(Intercept)	18.996 (1.299)***	19.105 (1.239)***
Contingent	0.550 (1.832)	0.621 (1.743)
Period	−0.267 (0.050)***	−0.375 (0.052)***
Machine	−1.438 (2.489)	−11.728 (2.899)***
Period:Machine	0.169 (0.100)$^{\circ}$	0.753 (0.131)***
Contingent:Machine	1.824 (0.659)**	2.229 (0.657)***
Inspections(cum)	0.555 (0.115)***	0.889 (0.124)***
Machine:Inspections(cum)		−2.602 (0.382)***
AIC	26617.543	26573.542
Num. obs.	3600	3600
Num. groups: ID:Group	120	120
Num. groups: Group	30	30

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

As Model 1 shows, the coefficient of *Inspections(cum)* is positive and statistically significant. Thus, more inspections lead to overall higher tax compliance. However, when distinguishing between the source of inspection - Machine vs. Human - a composite pattern emerges. The number of cumulated inspections performed by a TA has a positive impact on taxes paid, while more inspections performed by the machine tend to decrease tax compliance (see Model 2).

Result 5. *More inspections performed by the TAs lead to higher compliance in condition* Human.

Result 6. *More inspections by the machine lead to lower compliance in condition* Machine.

The two results suggest that choices of TPs are governed by different expectations about serial correlation in inspections of humans and machines. Specifically, Result 5 suggests that TP infer a positive serial correlation in

Table 4: Tax Complaince and Spillover with Controls (Linear Mixed Models)

	Model 1
(Intercept)	10.104 (3.409)**
Contingent	3.448 (4.069)
Inspections(cum).H	0.756 (0.324)*
Period	0.018 (0.092)
Contingent :Inspections(cum).H	−0.112 (0.570)
Num. obs.	1200
Num. groups: ID:Group	120
Num. groups: Group	30

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

the inspections of TAs, i.e., those TAs who were more likely to control in the past are more likely to control in the future. In contrast, a negative serial correlation is inferred from inspections of the machine, i.e., more controls in the past make controls in the future less likely. The latter is compatible with well-known heuristics that underlie the so-called bomb-crater effect (Mittone, 2006).

In Table 4 ,we estimate the spillover effect of cumulated inspections in condition *Human* on taxes paid in condition *Machine*. The effect is captured by the coefficient *Inspections(cum).H* which counts the number of inspections performed by the TA.

As the estimated coefficient of *Inspections(cum).H* shows, more inspections performed by a TA determine higher compliance levels in the machine condition, when inspections are fully automatized. No significant difference between the two incentive schemes is registered.

Result 7. *More inspections performed by the TAs lead to higher compliance in condition* Machine*.*

4.4 Types of audit strategies

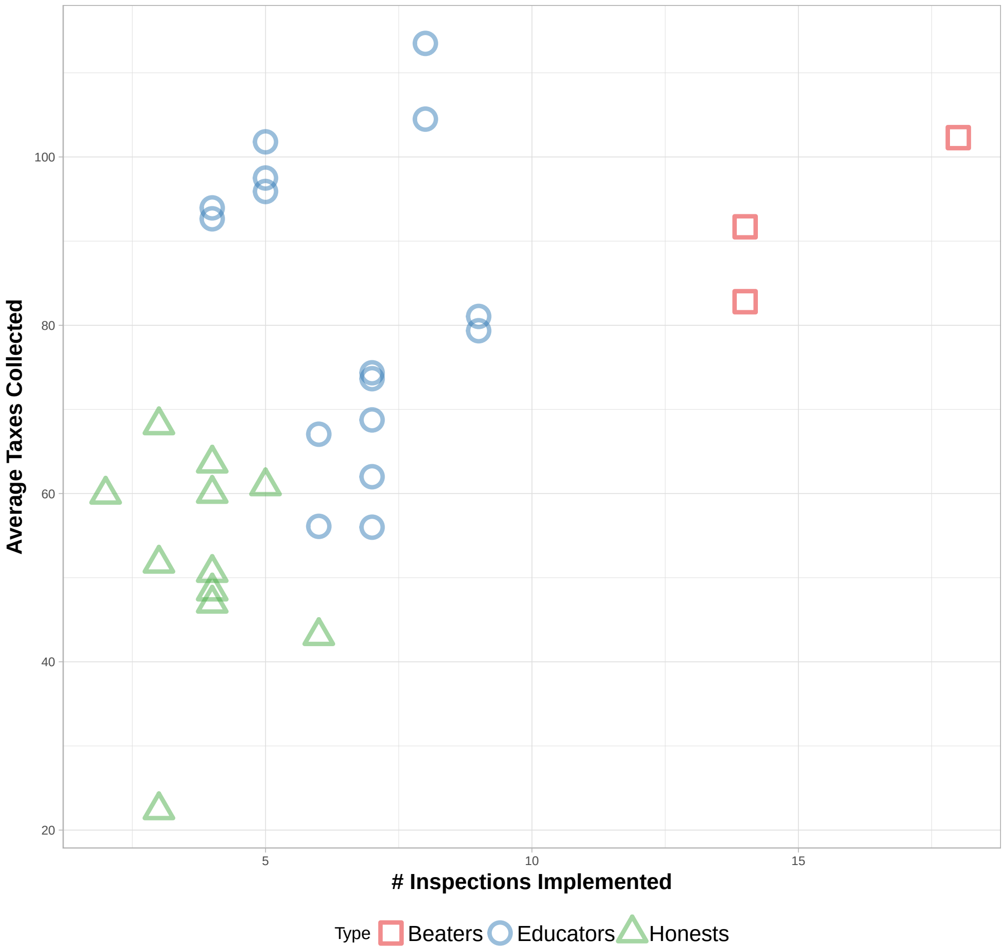
We conclude our results with a qualitative analysis of the inspection strategies employed by the TAs. TAs implement very different inspection strategies, which in turn lead to different compliance levels on the TPs' side. To gain insight into this, we categorize TAs according to the total number of inspections performed in condition *Human* and the average compliance reached therein (see Appendix C, where we display group compliance patterns together with audit strategies). A hierarchical cluster analysis leads to the isolation of three major groups,[16] as displayed in Figure 3. We label the three groups in the following way: *Honests* (triangles), *Beaters* (squares), and *Educators* (circles).

As the figure shows, the *Beaters* perform a large number of inspections and the TPs to whom they are associated display high levels of tax compliance. The *Educators* perform fewer inspections, but register high tax compliance as well. The *Honests* conduct the lowest number of inspections and register the lowest level of compliance. Non-parametric tests show that *Beaters* perform significantly more inspections than *Educators* and *Honests* (Wilcoxon rank sum test, p-value=0.008 and p-value=0.010, respectively). At the same time, *Educators* inspect more often than *Honests* (Wilcoxon rank sum test, p-value$<$ 0.001). The frequency of inspection of the latter does not significantly differ from the truthful frequency of 1/6 (Wilcoxon signed rak test, p-value=0.115).

Concerning the effects on tax compliance, *Educators* and *Beaters* reach significantly higher levels *Honests* (Wilcoxon Rank Sum test, p-value$<$ 0.001 and p-value=0.005, respectively). In contrast, no significant difference in taxes collected is registered when comparing *Educators* and *Beaters*

[16] We perform a hierarchical cluster analysis with complete linkage using Euclidean distances based on two dimensions: total number of inspections performed by the TA in condition *Human* and the average compliance level reached by his group in the same condition. Using a divisive approach, this analysis leads to the identification of 3 main clusters, since for any further splits gains are negligible.

Figure 3: Categorization of audit strategies by TAs



Note: The graph depicts TAs as points on the graph displaying the relationship between the total number of inspection implemented during condtion *Human* and the average taxes collected within that period. A hierarchical cluster analysis provides a categorization of TAs in three distinct types: *Beaters* (squares), *Educators* (circles) and *Honests* (triangles).

(Wilcoxon Rank Sum test, p-value= 0.421).

Result 8. *Three alternative styles of inspection strategies can be identified:* Honests, Beaters, *and* Educators. Honests *perform inspections in line with the preset inspection rule, but obtain low compliance. The other two obtain* *high compliance. However,* Educators *perform significantly less inspection than* Beaters*.*

5 Discussion

As predicted, assigning real participants to the role of TA and leaving them some discretion, led to more inspections than prescribed by the preset rule (Result 1). Surprisingly, the behavior of TAs did not differ across the incentive conditions *Flat* and *Contingent* (Result 2). This suggests that TAs might have had other, non-monetary, interests in reaching and keeping a norm of high compliance among their TPs. Questionnaire results support this: TAs state to be guided by the intention to enforce the rules in their audit strategies (see Appendix B). Thus, we can conjecture that some TAs interpreted their role more broadly and were motivated to increase tax compliance through "efficient lies". This is in line with the wide literature about third party punishment showing that people are willing to intervene to sustain fairness norms, even if not directly affected by consequences (see, e.g., Fehr and Fischbacher, 2004).

Moving to the main focus of this paper, we find that in condition *Human*, i.e., when TAs determine the occurrence of audits, taxes paid are high, quite stable over rounds and not influenced by the incentive condition TPs were in (Result 3). We also find that compliance in condition *Machine* does not differ across incentive conditions, although, given the inspection probability, selfish utility maximizing TPs should pay no taxes at all (Results 4). We show that these results are caused by the intensity with which TAs inspect TPs in their group. More inspections lead to more taxes paid in condition *Human* (Result 5) and in condition *Machine* (Result 6). This establishes a positive effect of aggressive audit strategies on tax compliance, while these policies are still in place. Moreover, we provide direct causal evidence for a positive spillover effect between inspections in condition *Human* and compliance in condition *Machine* (Result 7).

Taken together, this evidence points to the fact that TPs internalized (knowingly or not) the high compliance norm, if subject to its enforcement, and stuck to the virtuous high compliance equilibrium. This finding is at odds with results on the effect of peer punishment on cooperation, where subjects start misbehaving as soon as such opportunities are taken away (Fehr and Gachter, 2000). However, it is in line with recent findings by Galbiati et al. (2018), who show that fines can have an effect on cooperation even when no longer in force. In this sense, tax agents in our experiment managed to establish a persistent norm of compliance, which outlived their existence.

Finally, our analysis of types shows that TAs reached this high compliance norm with altogether different strategies. While adhering to the preset rule did not benefit *Honests*, bluntly ignoring it obviously led to higher compliance for *Beaters* in the *Human* and even in condition *Machine*. In comparison to the *Beaters*, *Educators* obtained similar levels of compliance by adopting much fewer inspections. By departing less evidently from the rule, they manage to educate their TPs to the high compliance norm without beating them too hard. These differences in audit patterns are particularly relevant in the evaluation of results reached by *Educators* and *Beaters*. In fact, *Educators* adopt a more efficient strategy for two reasons, a direct and an indirect one. First, by implementing fewer inspections their strategy is clearly more cost effective. Indeed, if one were to introduce a cost of inspection, *Educators* would arguably fare better than *Beaters* in a cost-benefit comparison. Second, more subtle deviations from the preset rule might positively influence how TPs perceive the use of TAs' power. In turn, this can lead to more compliant behavior in general, especially in areas where the use of raw power by tax authorities is more difficult.

6 Conclusion

We present an innovative design which allows us to test the effect of aggressive audit policies on tax compliance. We find that participants in the role of the tax authority bend the rules to reach high compliance, even if they have no direct interest in the level of taxes paid. This behavior puts participants in the role of tax payers under pressure and leads to a higher tax compliance. Once a norm of high compliance is reached, it is carried over in a setting in which fines are non deterrent. Our results provide evidence in favor of a positive effect of aggressive audit strategies both when these policies are in place and after they ended. Tax payers reach a high compliance equilibrium and stick to it even when not forced to do so. Our study also provides an original methodological contribution, which offers new ways to investigate how the relationship between tax authorities and tax payers influences compliance in a controlled setting.

Using experiments to study real-life phenomena like tax compliance which involve policy interventions, can paint only a partial picture. Our results show that tax authorities can apply more pressure to fight tax evasion without fearing backfire effects. However, this evidence has to be complemented with further empirical data and validated with different methods. A great advantage of lab experiments is that they can establish clear causality relationships and open the way to controlled manipulations in the field. Evidence presented here may encourage this venture.

References

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

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

Alm, J. (1991). A perspective on the experimental analysis of taxpayer reporting. *The Accounting Review*, 66(3):577–593.

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. and Jacobson, S. (2007). Using laboratory experiments in public economics. *National Tax Journal*, pages 129–152.

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

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

Baldry, J. C. (1986). Tax evasion is not a gamble: A report on two experiments. *Economics Letters*, 22(4):333–335.

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

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.

Ellingsen, T. and Johannesson, M. (2004). Promises, Threats and Fairness*. *The Economic Journal*, 114(495):397–420.

Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and human behavior*, 25(2):63–87.

Fehr, E. and Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994.

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.

Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise – an experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547.

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

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

Galbiati, R., Henry, E., and Jacquemet, N. (2018). Dynamic effects of enforcement on cooperation. *Proceedings of the National Academy of Sciences*, 115(49):12425–12428.

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

Guala, F. (2005). The Methodology of Experimental Economics. chapter 3. Cambridge University Press.

Hallsworth, M., List, J. A., Metcalfe, R. D., and Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14–31.

Kahn, C. M., Silva, E. C., and Ziliak, J. P. (2001). Performance-based wages in tax collection: The brazilian tax collection reform and its effects. *The Economic Journal*, 111(468):188–205.

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

Kastlunger, B., Kirchler, E., Mittone, L., and Pitters, J. (2009). Sequences of audits, tax compliance, and taxpaying strategies. *Journal of Economic Psychology*, 30(3).

Khan, A. Q., Khwaja, A. I., and Olken, B. A. (2015). Tax farming redux: Experimental evidence on performance pay for tax collectors. *The Quarterly Journal of Economics*, 131(1):219–271.

Kirchler, E., Muehlbacher, S., Kastlunger, B., Wahl, I., and Others (2010). Why pay taxes? A review of tax compliance decisions. *Developing alternative frameworks for explaining tax compliance*, pages 15–31.

Mittone, L. (1997). Subjective versus objective probability: Results from seven experiments on fiscal evasion.

Mittone, L. (2006). Dynamic behaviour in tax evasion: An experimental approach. *The Journal of Socio-Economics*, 35(35):813–835.

Muehlbacher, S. and Kirchler, E. (2016). Taxperiments: About the external validity of laboratory experiments in tax compliance research. *Die Betriebswirtschaft*, 76(1):7–19.

Slemrod, J. and Weber, C. (2012). Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance*, 19(1):25–53.

Srinivasan, T. N. (1973). Tax evasion: A model. *Journal of Public Economics*, 2(4):339–346.

Vossler, C. A., Gilpatric, S. M., and Others (2017). Endogenous Tax Audits and Taxpayer Assistance Services: Theory and Experiments. Technical report.

Webley, P., Robben, H. S. J., Elffers, H., and Hessing, D. J., editors (1991). *Tax evasion: An experimental approach*. Cambridge: Cambridge University Press.

Yitzhaki, S. (1974). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 3(2):201–202.

Web References

CNN Library (2018). IRS Scandal Fast Facts. [online] Available at: https://edition.cnn.com/2014/07/18/politics/irs-scandal-fast-facts/index.html [Accessed 30 Aug. 2019].

Harrison, J. (2013). IRS auditors targeting small business owners — but there are ways to protect your company. The Washington Post. [online] Available at: https://www.washingtonpost.com/blogs/on-small-business/post/irs-auditors-targeting-small-business-owners–but-there-are-ways-to-protect-your-company/2013/04/16/32ac68aa-a697-11e2-8302-3c7e0ea97057$_{b}log.html?noredirect = on$ [Accessed 30 Aug. 2019].

Il Fatto Quotidiano (2012). Controlli fiscali a Cortina Blitz di 80 agenti per stanare gli evasori. [online] Available at: https://www.ilfattoquotidiano.it/2012/01/03/controlli-fiscali-cortina-blitz-agenti-stanare-evasori/181286/ [Accessed 30 Aug. 2019].

Studio Legale Leone-Fell & Associati (2018). BB e case vacanze, controlli a tappeto: cosa fare in caso di verifica fiscale. [online] Available at: https://www.avvocatoleone.com/bb-e-case-vacanze-controlli-a-tappeto-cosa-fare-in-caso-di-verifica-fiscale/ [Accessed 30 Aug. 2019].

Trentino (2014). I dentisti nel mirino della Guardia di Finanza. [online] Available at: http://www.giornaletrentino.it/cronaca/trento/i-dentisti-nel-mirino-della-guardia-di-finanza-1.990110 [Accessed 30 Aug. 2019].

William, D. (2019). Is the IRS Targeting Small Businesses? Yep, Says One Expert. Small Business Trends. [online] Available at: https://smallbiztrends.com/2016/09/irs-targeting-small-businesses.html [Accessed 30 Aug. 2019].

Appendix A

Here we report the instructions given to participants at the experiment translated from Italian. We insert only one version and specify which parts were different according to the incentive condition. Of course participants received different instructions for condition Flat *and* Contigent.

General instructions

Welcome,
You are about to take part in an experiment on economic decision-making. For having shown up on time you will receive 3 Euros at the end of the experiment. If you have any doubt during the experiment, please address a lab assistant by raising your hand. In case you use the computer for activities not strictly tied to the experiment, you will be excluded from taking part in the experiment and from any payment.
Hereunder, you will find the instructions regarding the exact proceeding of the experiment. Please read the instructions carefully. Before the experiment starts, you will have to answer some questions to verify your comprehension of the instructions.

Instructions

The experiment is consists of a total of 30 rounds and is divided into two phases. Phase 1 consists of 20 rounds and Phase 2 consists of 10 rounds. Participants will be randomly assigned to the role of tax payer or to the role of tax agent. Participants assigned to the role of tax payer will be divided into groups of four participants each. The composition of these groups will be the same for all 30 rounds. The identity of the other group members will not be revealed to any participant. A single participant with the role of tax agent will be assigned to each of these groups and will remain the same for all 30 rounds. The identity of the tax agent will not be revealed to any participant; likewise, the tax agent will not know the identity of any other participant. There will, thus, be 4 distinct groups each made up of 4 tax payers and 1 tax agent. Groups will never interact among each other. During the experiment participants will use experimental currency tokens, which will be converted in Euros at the end of the experiment at an exchange rate of 100 tokens = 0.40 Euros.

Phase 1

Tax payer

At the beginning of each round, tax payers will receive 100 tokens and will decide how much to pay in taxes. The tax rate is set at 30%; on an amount of 100 tokens, taxes due are, thus, 30 tokens. Each tax payer will freely choose the amount of tokens he or she wants to pay, between 0 and 30 tokens. Inside each group, taxes will be gathered in a unique account, the total amount will be multiplied by a factor of 1.5. The total amount of tokens obtained this way will be divided in equal parts among all four tax payers belonging to the same group, independently of the contribution of each single tax payer.
EXAMPLE:
Tax payer T1 pays 20 tokens in taxes, T2 10 tokens, T3 30 tokens, and T4 0 tokens. The total amount 20+10+30+0=60 is multiplied by 1.5; thus, 60*1.5=90; and divided in equal parts among each tax payer, hence 90/4=22.5. Each tax payer will receive 22.5 tokens, which will be added to those he or she still owns. In the example, T1 will receive 80+22.5=102.5, T2 90+22.5=112.5, T3 70+22.5= 92.5, and T4 100+22.5=122.5.

Tax agent

Each tax agent is in charge of carrying out the auditing process of tax payers inside his or her group. In case an inspection takes place, each tax payer of the group who has not paid the entire amount due will be fined. The fine is composed by the rest of the tax due multiplied by 2.
Flat: "The payment of the tax agent is fixed and independent from the proceeding of the experiment. The exact amount of the payment will be revealed only at the end of the experiment."
Contingent: "The payment of the tax agent in each round is equal to the total taxes paid in his or her group multiplied by 1.5.
EXAMPLE:
The total amount of taxes paid by the tax payers, 20+10+30+0=60, is multiplied by 1.5, hence 60*1.5=90. The tax agent has earned 90 tokens in this round."

Die roll

During each round, participants will be asked to roll privately a die inside their cubicle on the bubble wrap layer a single time and report the outcome via the computer interface.
The outcome reported by tax payers will have no consequence on the experiment.
The outcome reported by the tax agents, on the contrary, will determine the implementation of the fiscal inspection. The inspection will take place each time the tax agent declares via the computer interface that the outcome of his or her die roll is equal to 3. The probability of obtaining this result is 1/6 (16.67%). When the number 3 is reported, and the inspections takes place, all tax payers of the same group are audited.
EXAMPLE:
If the result reported by the tax agent was equal to 3, an inspection would take place. T1 would be fined for 20 tokens, T2 for 40, T3 for 0, and T4 for 60. The fines would diminish the total earnings in that round. T1 would remain with 82.5, T2 with 72.5, T3 with 92.5, and T4 with 62.5.
At the end of each round, each tax payer will get to know his or her earnings in that round and be told whether an inspection took place. Each tax agent will get to know the taxes paid inside his or her group (*Contingent*: and the total amount earned in that round).

Phase 2

Tax payer

In each round tax payers will decide how much to pay in taxes, but they will not roll the die anymore. Otherwise, the experiment will carry on as in Phase 1.

Tax agent

In this phase tax agents will not be in charge of carrying out the fiscal inspection by rolling the die. Tax agents will only be asked to estimate the total amount of taxes paid by tax payers in their group in each round. Inspection will be carried out by a computerized random mechanism which assigns the probability of 1/6 (16.67%) for an inspection

in each round. The tax agents will still receive information about the total of taxes paid in their group and the taking place of an inspection.
Contingent: "Their earning will be determined, as in Phase 1, by the total of tax paid multiplied by 1.5."

Final payment

Contingent: "The payment will be determined for all participants in a cumulative way, by summing up the earnings in all 30 rounds that compose the experiment."
Flat: "The payment will be determined for tax payers in a cumulative way, by summing up the earnings in all 30 rounds that compose the experiment. The payment of tax agents is equal for all and independent of the proceeding of the experiment. The exact amount will be revealed at the end of the experiment."
The final payment will occur in Euros at the end of the experiment. The exchange rate is set at 40 Cents every 100 tokens (100 tokens = 0.4 Euros).

Appendix B

Here we report the questionnaires participants filled in after condition Human *(Questionnaire 1) and condition* Machine *(Questionnaire 2) alongside with the mean and standard deviation for each of the items. The original text is translated from Italian.*

Questionnaire 1

- Tax payers
  - When I paid my taxes I did it in the group's interest. (mean=2.9, s.d.=1.29)
  - When I paid my taxes I did it in my personal interest. (mean=3.91, s.d.=1.08)
  - When I paid my taxes I did it to avoid sanctions. (mean=3.63, s.d.=1.23)

- I think the tax agent followed the rules in a transparent way. (mean=3.34, s.d.=1.48)
- I think the tax agent acted in his own interest. (mean=2.68, s.d.= 1.4)
- I think the tax agent acted in the group's interest. (mean=2.17, s.d.= 1.13)
- If I had been the tax agent, I would have acted the same way he or she did. (mean=3.16, s.d.= 1.51)

- Tax agents
  - I acted in the interest of the whole group. (mean=2.87, s.d.=1.33)
  - I acted in my personal interest. (mean=2.23, s.d.= 1.50)
  - I acted to enforce the rules. (mean=4.4, s.d.=0.97)
  - I think another participant in my role would have acted the same way I did. (mean=3.8, s.d.=1.21)

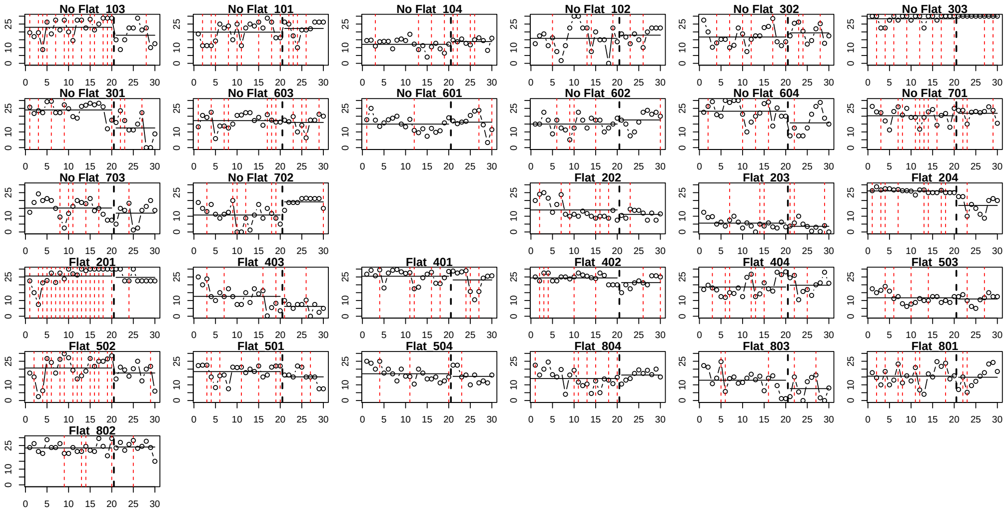
Questionnaire 2

- Tax payers
  - When I paid my taxes I did it in the group's interest. (mean=2.85, s.d.=1.29)
  - When I paid my taxes I did it in my personal interest. (mean=3.98, s.d.=1.03)
  - When I paid my taxes I did it to avoid sanctions. (mean=3.6, s.d.=1.15)
  - Since the tax agent was replaced by a random device, I behaved in a different way. (mean=2.98, s.d.=1.51)

Appendix C

Below we report audit and compliance patterns of all groups in the experiment.

Figure 4: Audit and compliance patterns



Note: Each box represents a group of 4 TPs and 1 TA. Rounds are listed on the x-axis, while average contributions are displayed on the y-axis. The red vertical lines represent rounds in which an inspection took place. The line connecting the circles traces average contribution lavels. The black horizontal lines show average contributions in condition *Human* and *Machine*, respectively.