科技部補助專題研究計畫成果報告
期末報告

懲罰效果的兩個實驗研究(第 2 年)

計 畫 類 別 ： 個別型計畫
計 畫 編 號 ： NSC 100-2410-H-004-065-MY2
執 行 期 間 ： 101 年 08 月 01 日至 103 年 07 月 31 日
執 行 單 位 ： 國立政治大學財政系

計 畫 主 持 人 ： 徐麗振

計畫參與人員： 碩士班研究生-兼任助理人員：林言翰
碩士班研究生-兼任助理人員：徐士閎
助教-兼任助理人員：巫美忻
大專生-兼任助理人員：郭姿辰
大專生-兼任助理人員：盧威涵
博士班研究生-兼任助理人員：林振陞

報 告 附 件 ： 出席國際會議研究心得報告及發表論文

處 理 方 式 ：
1.公開資訊：本計畫涉及專利或其他智慧財產權，2 年後可公開查詢
2.「本研究」是否已有嚴重損及公共利益之發現：否
3.「本報告」是否建議提供政府單位施政參考：否

中 華 民 國 103 年 10 月 31 日

中 文 摘 要： 我們以實驗方法探討當實驗參與者可以對稅率、稽核率、或懲罰率投票時的情況下，實驗參與者的租稅順從行為與民主效果。我們藉由增加一階段的電腦決策來控制選擇效果，此一方法也被用在 Dal Bó, Foster, and Putterman (2010) 的文章中。我們的實驗證據不支持民主效果。我們還發現當實驗參與者面對一較高的稽核率或懲罰率時，他們的租稅順從度較高。我們的發現建議，為提高租稅順從度，政府應該直接採用一較高的稽核率或懲罰率，不要讓選民經由民主過程決定。

我們以實驗方法探討在 weakest-link 和 best-shot 賽局中，獎勵和懲罰機制對自由捐獻公共財的效果。實驗參與者先自由捐獻他們稟賦的全部或一部份到公共財，然後每人捐獻量公共給所有組員知道，看到每位組員的捐獻後，每人決定要給其他每位組員多少獎勵或懲罰點數。我們的實驗證據顯示在 weakest-link 賽局中懲罰機制可以顯著提高捐獻額，但獎勵機制則無效果。反之，在 best-shot 賽局中獎勵機制可以顯著提高捐獻額，但懲罰機制則無效果。不論在 weakest-link 賽局還是 best-shot 賽局，都不需要同時用到兩種機制，因為額外使用另一機制並不會顯著提高捐獻。

中文關鍵詞： 租稅順從，民主效果，租稅稽核，懲罰，實驗，私人提供的公共財，獎勵

英 文 摘 要： Abstract. We experimentally investigate the compliance behavior and the democracy effect in an environment in which subjects are allowed to vote on the tax, audit, or fine rate. We control for the selection effect by adding a stage of computer decisions similar to that proposed by Dal Bo, Foster, and Putterman (2010). Our experimental evidence does not support the democracy effect. We also find that subjects behave more compliantly when a high audit rate or a high fine rate is applied than when a lower counterpart is applied. These findings suggest that, to improve compliance, the government should just impose a high audit or fine rate.

Abstract. We experimentally examine the effects of rewards and punishments on voluntary contributions in the weakest-link and best-shot game. Subjects voluntarily contribute some or all of their incomes to the public good and then after individual subjects' contributions are revealed, they decide to give reward or punishment points (if there are any) to others in their own groups. Experimental evidence from this paper shows that the instrument of punishments (but not rewards) is more effective in increasing voluntary contributions in the weakest-link game. On the contrary, the instrument of rewards (but not punishments) can significantly increase voluntary contributions in the best-shot game. In both games, using instruments of both rewards and punishments cannot improve voluntary contributions if either instrument has been adopted.

英文關鍵詞： tax compliance, democracy effect, tax auditing, punishment, experiment, privately provided public good, reward, weakest-link, best-shot

Experimental Study on the Effects of Democracy on Tax Compliance

Li-Chen Hsu*

Department of Public Finance
National Chengchi University
64, Section 2, Zhi-Nan Road
Wenshan, Taipei 11605, Taiwan
Email: lchsu@nccu.edu.tw
Phone: 886-2-29387313

Kamhon Kan

Institute of Economics
Academia Sinica
128, Section 2, Academia Road
Nankang, Taipei 11529, Taiwan
Email: kk@sinica.edu.tw
Phone: 886-2-27822791 ext 505

Abstract. We experimentally investigate the compliance behavior and the democracy effect in an environment in which subjects are allowed to vote on the tax, audit, or fine rate. We control for the selection effect by adding a stage of computer decisions similar to that proposed by Dal Bó, Foster, and Putterman (2010). Our experimental evidence does not support the democracy effect. We also find that subjects behave more compliantly when a high audit rate or a high fine rate is applied than when a lower counterpart is applied. These findings suggest that, to improve compliance, the government should just impose a high audit or fine rate.

JEL classification: H26, C91

Keywords: tax compliance; democracy effect; tax auditing; penalty; experiment

*Corresponding author.

1. INTRODUCTION

Because government revenues and thereby government spending are very much related to the amount of taxes that are reported honestly, improving tax compliance is a challenging task of tax authorities in many countries. To attain this goal, the central question concerns how individuals' tax compliance behavior is affected by various aspects of the tax system. To explore this question, appealing to empirical works using field data is a possibility. However, as pointed out by Andreoni, Erard, and Feinstein (1998), the difficulty of this approach is the lack of reliable information on taxpayers' reporting behavior. As a consequence, experimental methods are likely to be the most or even the only viable way.

Three aspects of the tax system are generally examined by experimental studies. They are the tax rate, the audit rate, and the penalty rate. In the experimental literature, some studies assume that these variables are fixed and examine the impacts of changes in these variables on compliance (Spicer and Becker, 1980; Spicer and Thomas, 1982; Becker, Büchner, and Sleeking, 1987; Alm, McKee, and Beck, 1990; Beck, Davis, and Jung, 1991; Collins and Plumlee, 1991; Alm, Jackson, and McKee, 1992b; Alm, McClelland, and Schulze, 1992; and Alm, Sanchez, and de Juan, 1995), while some other studies assume that some or all of these variables are non-fixed or endogenously determined (Alm, Jackson, and McKee, 1992a; Alm, Cronshaw, and Mckee, 1993; Alm and McKee, 2004; Clark, Friesen, and Muller, 2004; Gërxhani and Schram, 2006).

All aspects of the tax system in the above studies are set up by the government. Individuals have no chance to participate in the establishment of these institutions, nor can they express their preferences for the various aspects of the tax system. Under this framework, individuals who report income honestly can only rely on the institutions set up by the government or behave in the same way to punish tax dodgers. As a result, compliance may be kept at a low level. Although this outcome is pessimistic, it ignores the fact that in a political economy individuals often have direct or indirect influences on government policies.[1] In fact, some experimental studies have reached a conclusion that individual participation in the decision-making process can improve compliance or cooperation. In public goods experiments, for instance, Putterman, Tyran, and Kamei (2011) allow subjects to vote on whether "private account" or "public account" contributions are subject to penalties. They find that there is almost uniform support for penalizing non-contribution to the public account, and contributions to the public good are significantly higher when there are formal sanctions than when sanctions are absent.

In tax compliance experiments, Pommerehne, Hart and Frey (1994) suggest that a democratic process tends to raise tax morale and therefore tax compliance. Alm, Jackson, and McKee (1993) find that compliance is higher if subjects are allowed to select the public sector expenditure program themselves by majority voting. By designing an experiment in which the punishment is certain, that is, the audit probability is one, Feld and Tyran (2002) ask subjects to state their contributions for all possible voting outcomes. They find that the possibility of voting on fines significantly increases tax compliance since subjects who vote for the punishment scheme feel obliged to consistently comply with their decision by making larger contributions.

Alm, McClelland and Schulze (1999) obtain different results. They find that although

[1] An example of direct influences is the voting on tax increases to improve Atlanta's infrastructure. As is reported by The Economist (2011), Atlantans have the longest average rush-hour commute in America, and according to Georgia's government, the state spends less per head on transport than any other states with the exception of Tennessee. Since improving the infrastructure means raising taxes, in June 2010 Georgia's legislature decided to let citizens vote on whether to raise their own taxes. As for indirect influences, individuals may bring their influence to bear on or petition the legislative members to pass favorable laws or regulations.

the impact of voting on the tax rate is mixed, in all four fine sessions the majority votes for the low fine rate, and in the other four audit rate sessions the majority votes for the low audit rate. Furthermore, the average compliance rates in the vote stage are lower than the corresponding average compliance rates in the no-vote stage in all sessions. They appeal to the notion of the social norm that an individual will comply as long as he or she believes others will comply. The group decision regarding enforcement reveals the lack of a social norm of tax compliance and thus compliance with voting is lower than that without voting.

Despite the inconsistent results shown above, as pointed out by Dal Bó, Foster, and Putterman (2010), a central problem with the examination of the effects of democracy is that "one cannot rule out the possibility that there are unobserved factors that explain both responses to policies and either the degree of participation in policymaking or the particular policies selected." Briefly put, there is a selection problem. That is, the observed higher level of cooperation under voting may be attributed to individuals' inherent preferences for the chosen policy, and not simply because of their participation in the democratic process. To control for the selection effect, Dal Bó, Foster, and Putterman (2010) add a stage of computer decisions after voting. In their prisoner's dilemma experiment, individuals vote on two alternatives: modifying the payoff or not, and then the computer decides whether to consider the outcome of the majority voting or not. If the computer accepts the outcome of the majority voting, the final outcome is consistent with the result of the majority voting. If the computer rejects it, then the computer will decide whether to modify the payoff or not. The addition of computer decisions breaks the direct connection between the preference for the chosen alternative and the outcome of majority voting, and therefore the democracy effect can be properly measured.

This paper's identification strategy is inspired by that of Dal Bó, Foster, and Putterman (2010). We examine tax compliance in an environment that allows subjects to vote on the tax, audit, or fine rate. There are several major differences between Dal Bó, Foster, and Putterman's (2010) experimental design and ours. First, Dal Bó, Foster, and Putterman (2010) use a prisoner's dilemma game, while ours is a tax compliance game. Thus, their focus is very different from ours. Second, in Dal Bó, Foster, and Putterman's experiment subjects first play the game without voting for ten rounds and then majority voting is executed before the start of the eleventh round. The outcome of this one-time majority voting applies to the next ten rounds. In our experiment, the voting procedure occurs at the beginning of each of the ten rounds involving majority voting. Third, in Dal Bó, Foster, and Putterman's experiment subjects are informed of the outcome of majority voting before computer decisions, while in our experiment subjects are only informed of the tax, audit or fine rate to be applied, but not of the way in which it is determined (i.e., majority voting or random assignment by the computer). This is the most important difference between our experimental design and that of Dal Bó, Foster, and Putterman (2010).

With Dal Bó, Foster, and Putterman's experimental design, in which case whether the results of the majority voting are adopted or not is released to subjects, the signaling effect of the voting outcome is not controlled and this signaling effect may confound the identification of the causal effect of democracy. In the case where whether or not the majority voting results are adopted is announced (as in the case of Dal Bó, Foster, and Putterman, 2010), the signaling effect arises from the fact that a subject would be able to decipher other subjects' preferences from the voting outcomes and respond to this expectation of his or her peers' preferences (see a brief discussion in footnote 8 of Dal Bó, Foster, and Putterman, 2010). For example, in the tax compliance setting of the current study, if a low tax rate is determined by majority voting and this is conveyed to subjects, a subject may infer that most of his or her peers in the group may not comply because they prefer ~~to have~~ a low tax rate. This signaling effect is found in Alm, McClelland and Schulze (1999), who attribute the effect to social norms.

We control ~~for~~ the signaling effect by not revealing the outcomes of majority voting to the subjects. One may suspect that under this setting subjects may not perceive that they were involved in a democratic situation. However, to test the democracy effect, there must be a tradeoff between controlling for the signaling effect and sacrificing the subjects' perceptions of democratic participation. To fix the latter problem, we informed subjects that the computer would consider the outcome of majority voting based on a certain probability. We did not inform them of the exact magnitude of this probability, which we set it at 0.7, so that in most cases the computer's decisions would coincide with the outcomes of majority voting.

Compared with Dal Bó, Foster, and Putterman (2010), the drawback of our experimental design is that we do not observe the subjects' voting behavior in the no-vote regime of the experiment. This implies that we are not able to compare the levels of compliance between settings with and without democracy among individuals who voted for a particular tax, audit or fine rate. Instead, since a low tax, audit, or fine rate is used in the no-vote stage, we compare the level of compliance in the no-vote stage with that when a low tax, audit or fine rate is applied in the voting stage.

We circumvent this deficiency in terms of controlling for unobserved heterogeneity by using a fixed effects regression strategy. With repeated observations based on subjects' compliance behavior, we use a fixed effects regression specification to control for their time-invariant heterogeneity in tax compliance behavior. The unobserved factors that remain in determining a subject's compliance behavior will be time-varying (i.e., changes over each round) and these time-varying factors are unlikely to be correlated with his or her voting behavior. Furthermore, we examine the difference in a subject's compliance behavior when he or she voted for a high versus a low tax, audit, or fine rate. If our fixed effects specification is successful in controlling for the correlation between compliance behavior and voting behavior due to unobserved factors, we will not find a significant difference in compliance when he or she voted differently.

The main result of this paper does not support the democracy effect. That is, democratic participation does not have a significant and positive impact on compliance. Furthermore, our finding shows that subjects behave more compliantly when a high audit rate or a high fine rate is applied than when a lower counterpart is applied and their voting decisions have no significant effects on their compliance behavior.

The remainder of this paper is organized as follows. Section 2 describes the experimental design. Section 3 presents the theoretical predictions. Section 4 reports the results of the experiment, and Section 5 concludes.

2. EXPERIMENTAL DESIGN

The fundamental experimental design of this paper is similar to those of the experiments on VCM (the voluntary-contribution mechanism) and tax compliance. In the experiment, each subject receives an exogenous amount of income and he or she pays the tax according to the income that he or she declares. The tax is used to provide the public good that benefits only the members within the same group. After declaration, the subject's true income is audited based on some probability. The subjects who are audited and caught cheating will pay the evaded taxes and fines. A subject's original income net of the tax he or she has paid and the evaded tax and fines, if there are any, is his or her private good consumption. His or her payoff is the sum of the public good consumption and his or her private good consumption. To prevent any emotional responses, neutral terms are used in the experimental instructions. Furthermore, because the tax authority simply collects taxes and fines without making any decisions in the experiment, the role of the tax authority is not particularly mentioned.

[Table 1 about here]

Six treatments were conducted in this research. They are denoted as Tax-o1 (tax rate-order one), Tax-o2 (tax rate-order two), Audit-o1 (audit rate-order one), Audit-o2 (audit rate-order two), Fine-o1 (fine rate-order one), and Fine-o2 (fine rate-order two). The framework of the experiment and the magnitudes of various aspects of the tax system are provided in Table 1. Four sessions were conducted for each treatment and 12 subjects were recruited for each session, for a total of 288 subjects used in this study. To increase anonymity, two independent sessions under the same treatment were run at the same time, but subjects were unaware of this. All subjects were undergraduate students at National Chengchi University in Taiwan and none of them had previously participated in tax compliance or public goods experiments.

Each treatment consisted of two parts, and each part contained 10 rounds. Subjects were informed of the contents of the two parts at the beginning of the experiment. All the experimental settings in the first part (rounds 1–10) of the Tax-o1, Audit-o1, and Fine-o1 treatments were the same, while in the second part (rounds 11–20) a voting process was added and subjects in the three treatments voted on different variables of the tax system. The three variables to be voted on were the tax rate, audit rate, and fine rate. Since subjects' compliance behavior and their attitudes towards the three variables may have been affected by the timing of voting, to control for the order effect, three corresponding treatments with the opposite order of the two parts were also conducted. They are indicated as the Tax-o2 treatment, the Audit-o2 treatment, and the Fine-o2 treatment. Except for the order of voting, the three treatments were exactly the same as for their counterparts.

The experimental procedures of the Tax-o1, Audit-o1, and Fine-o1 treatments were as follows. Subjects made decisions in each of the 20 rounds. In each round, the 12 subjects in the same session were randomly and anonymously divided into three groups of size $n = 4$. To minimize the repeated-game effect, they were re-matched when a new round started. At the beginning of each round, four income levels (70, 90, 110 and 130 points) and four codes (A, B, C, and D) were randomly assigned to the four subjects in the same group. Call the income assigned to a subject his or her true income w_i. When a new round started, the four levels of income and the four codes were randomly reassigned. A subject knew his or her own code and income and the distribution of income, but not the income for each of the other three group members.

There were two stages in each round of the first part of the Tax-o1, Audit-o1, and Fine-o1 treatments. In stage one, the declaration stage, each subject was required to report a level of income R_i ($0 \le R_i \le w_i$), and the reported income was taxed at the rate $t = 0.2$. The tax was invested in the public account (the public good), and the rest of the income was maintained in the subject's private account (the private good). The marginal per capita return (MPCR) of the public good was set at $m = 0.5$. That is, each point invested in the public good yielded *every* group member a return of 0.5 points. Note that m had to satisfy the condition $1/n < m < 1$ so that each individual had the incentive to cooperate and to cheat. After all subjects had reported their incomes, they proceeded to the second stage, the auditing stage, in which each subject was audited by a probability $p = 0.1$. It is assumed that a subject's true income was revealed once he or she was audited. Any subject who was audited and caught cheating had to pay the evaded tax plus a penalty, which was twice the amount of the evaded tax. For simplicity, we state that the fine rate was 3 and denote it as f.

Given the above procedures, the expected monetary payoff for each subject i in each round of the first part of the Tax-o1, Audit-o1, and Fine-o1 sessions is given by

$$\pi_i = (1-p)(w_i - tR_i) + p[w_i - tR_i - ft(w_i - R_i)] + mt\sum_{j=1}^{n} R_j. \quad (1)$$

In equation (1), the sum of the first two terms is the subject's expected private good consumption and the third term is his or her public good consumption.

The tax rate t = 0.2, audit rate p = 0.1, and fine rate f = 3 serve as the benchmark. When subjects moved to the second-part of the experiment, an additional voting process was added at the beginning of each round, and one of these benchmark values was to be voted against another higher value. Specifically, in the Tax-o1 treatment the four members in the same group voted between two alternative levels of tax rates, 0.2 and 0.4; in the Audit-o1 treatment the four members voted between two alternative levels of audit rates, 0.1 and 0.4; and in the Fine-o1 treatment the four members voted between two alternative levels of fine rates, 3 and 6. The other two variables that were not to be voted on remained at the same levels as in the first-part of the experiment. As a consequence, there were three stages in each round of the second part of the Tax-o1, Audit-o1, and Fine-o1 treatments: a voting stage, a declaration stage, and an auditing stage. Except for the variable to be voted on, Tax-o1, Audit-o1, and Fine-o1 were exactly the same in every other aspect.

Let us explain the second part in more detail by taking the Tax-o1 treatment as an example. At the beginning of each round of rounds 11 to 20, subjects were required to vote between two tax rates, 0.2 and 0.4, for their own groups via majority voting. Subjects were informed that after all group members had made their own voting decisions, the computer would randomly determine whether to accept the outcome of majority voting or not. If the computer accepted the outcome of majority voting, the tax rate for the group was determined accordingly. If the computer rejected the outcome of majority voting or if a tie occurred, the computer would randomly assign either tax rate to the group.

We set the probability that the computer would randomly accept the outcome of majority voting to be 0.7. Once the computer rejected the outcome of majority voting or when a tie occurred, the probability that the computer would randomly assign either tax rate to the group was 0.5. Subjects were only informed of the above procedure and the final tax rate for their own group. They were unaware of the outcome of majority voting, the decision made by the computer, and the information regarding the probabilities for the computer's random choices. These settings aim to reduce speculation on the part of the subjects regarding other group members' voting decisions, and along with the setup where the size of each group was four persons, the selection problem can be kept to a minimum via the maximum possible intervention from the computer.

Following the voting stage, the second stage (the declaration stage) and the third stage (the auditing stage) of the second part of the Tax-o1, Audit-o1, and Fine-o1 treatments were exactly the same as the first and second stages in the first part of the three treatments. Given the above procedure, the expected monetary payoff for the subject in the second-part of the Tax-o1, Audit-o1, and Fine-o1 treatments was the same as in equation (1) except that the tax, audit, or fine rates were determined by majority voting and computer decisions.

At the end of each round of the experiment, each subject was informed of the result, which consisted mostly of the following information: the outcome of the voting stage (if there was one), the subject's declaration of income, his or her investment in the public account according to his or her declaration, the total income declared and the total investment in the public account excluding and including the subject's own investment, the code of the subject who was audited, the subject's payoff from his or her private account, the subject's payoff from the public account, the reduction in the subject's payoff if he or she was caught under-reporting, and the subject's payoff for this round.

In all sessions, subjects were given written instructions in Chinese. The experimenter read the instructions aloud and answered any questions raised by the subjects. After reading the instructions, subjects were required to answer four questions in relation to the calculation of payoffs and the experimental procedures.[2] The experiment would not start until everyone had answered all questions correctly. Each session lasted about 90 minutes. The average payoff (including a participation fee of NT$100) for all participants was NT$529.98 (with a standard deviation of NT$24.85, a maximum of NT$595, and a minimum of NT$445.7).[3]

3. THEORETICAL PREDICTIONS

To have a clear-cut theoretical prediction of subjects' behavior, it is assumed that all subjects were self-interested and maximized their own monetary payoffs, and that this feature was common knowledge to all subjects. Recall that when all the variables of the tax system were exogenously given in the no-vote part of the experiment, the subject's expected monetary payoff was characterized by equation (1). Differentiating equation (1) with respect to R_i yields

$$\partial\pi_i / \partial R_i = t(pf + m - 1). \tag{2}$$

Given the benchmark values of p = 0.1 and f = 3, and m = 0.5, equation (2) is certainly negative, implying that $R_i^* = 0$ for all i. That is, the dominant strategy for a self-interested and reward-maximizing subject is to report zero income and hence he or she earns an expected payoff of $\pi_i = w_i(1 - pft) = 0.94w_i$.

To find the equilibrium when a voting stage is involved, we can construct a two-stage game and solve the game by backward induction. The game proceeds as follows. In the first stage, all group members vote on two alternative levels of the tax, audit, or fine rate. Then, based on some probabilities, the computer randomly determines whether to accept the outcome of majority voting, and randomly assigns either level to the group if it rejects this outcome or if a tie occurs. To be consistent with our experimental design, it is assumed that subjects are only aware of the final tax, audit, or fine rates applied to their groups. In the second stage, given the final outcomes of the tax, audit, or fine rates, subjects declare income simultaneously and, after declaration, they are audited by some probability. When a subject makes his or her voting decision in the first stage, he or she assumes that all other group members have made their optimal voting decisions. When a subject makes his or her declaration decision in the second stage, he or she assumes that other group members have chosen their optimal levels of declarations, and takes other group members' voting decisions and the computer's random assignment into consideration.

It is evident that once the tax, audit, or fine rate has been determined in the first stage, the subject's expected payoff will be characterized by equation (1), and as a result the equilibrium in the second stage is still solved by equation (2). By considering the equilibrium strategy adopted in the second stage, the subject makes his or her best voting decision in the first stage.

Let us start with the case in which the two tax rates are to be voted on. When in the first stage subjects vote between two tax rates, 0.2 and 0.4, they are aware of the fact that either tax rate will be selected eventually. They also understand that their votes will to some extent affect the outcome of majority voting and that this outcome will be accepted by the computer according to some probability. Hence, the subject will vote for a tax rate that

[2] An English translation of the Subjects' Instructions and quiz questions for the Tax-o1 treatment is provided in the Appendix, which is not intended for publication.

[3] When these sessions were conducted, the exchange rate between the NT (New Taiwan) dollar and the US dollar was about 30:1. The part-time hourly wage rate for an undergraduate student in Taiwan is about NT$120.

yields him or her the higher expected payoff once the second stage arrives. In the second stage, since the sign of equation (2) is irrelevant to the tax rate and is negative given the benchmark values of *p* and *f*, the dominant strategy for the subject is still reporting zero income regardless of the outcome in the first stage. Given that zero income will be reported, $\pi_i = w_i(1-pft) = 0.94w_i$ if the tax rate is 0.2 and $\pi_i = w_i(1-pft) = 0.88w_i$ if the tax rate is 0.4. Hence, the subgame perfect equilibrium is that the subject votes for the low tax rate of 0.2 in the first stage and reports zero income in the second stage.

By applying similar analyses, the subgame perfect equilibrium for the case in which the two audit rates are to be voted on is that the subject votes for the high audit rate of 0.4 in the first stage and reports full income in the second stage. When the two fine rates are to be voted on, the subgame perfect equilibrium is that the subject votes for the high fine rate of 6 in the first stage and reports full income in the second stage.

Finally, how will the democracy effect affect the compliance behavior? The hypothesis of the democracy effect is that individuals will behave more cooperatively if they are provided with the opportunity to be involved in a political process than if they are not, and the more cooperative behavior is irrelevant to the outcome of the political process. We have employed a computer decision to control for the selection effect. To have an equal basis for comparison, we need to control further for the tax, audit, and fine rate so that these variables have the same value with and without voting. Because the tax, audit, and fine rates are low without voting, the prediction of the democracy effect is that those subjects for whom the low tax, audit, or fine rates in the voting rounds are applied will behave more cooperatively than when they are in the no-vote rounds.

In the following section, we will test the above equilibrium predictions, especially the democracy effect. In addition, we will investigate the subjects' compliance behavior as well as the compliance behavior conditional on their voting decisions and the final magnitudes of the tax, audit, and fine rates applied to them.

4. EXPERIMENTAL RESULTS

[Table 2 about here]

We conducted twenty-four sessions in April and May of 2012 in the computer lab of the Department of Public Finance at National Chengchi University in Taiwan. Table 2 reports that of the 288 subjects recruited, 74.31 percent of them were female, they had been in the university for an average of 2.11 years, the average age was 20.01 years, and 78.47 percent of them had taken economics course(s). The scale of the indicator "donation" ranged from one to six and the average was 2.09, meaning that, on average, subjects donated about NT$500 to NT$1,000 to charities during the year 2011. The scale of "risk-taking" ranged from 0 to 10, with 0 indicating not ready for taking any risks and 10 indicating fully prepared to take risks. The average level of risk-taking was 5.17, meaning that, on average, the subjects' attitude toward risks was modest.

4.1. A General Look at Compliance Behavior and Voting Decisions

[Tables 3.1, 3.2, and 3.3 about here]

Tables 3.1 through 3.3 summarize the data resulting from the first 10 rounds and the last 10 rounds in each treatment. Round averages and standard errors of compliance rates are depicted in Figures 1 and 2.[4] In addition, Figure 1 provides the information regarding the average compliance rates conditional on subjects' voting decisions, and Figure 2 provides the average compliance rates conditional on the magnitudes of the tax, audit, or fine rate applied to the subjects. The compliance rate for a subject is defined as his or her reported income divided by his or her true income.

[Figures 1 and 2 about here]

Several observations arise by looking at Tables 3.1 through 3.3. First, it is observed that average compliance rates for all six treatments lay in between 0.54 to 0.64 in the first ten rounds, and declined to an average of 0.36 to 0.55 in last ten rounds. Second, in the no-vote rounds, although on average 26.04 percent of the subjects in the Audit-o2 treatment and 21.46 percent of the subjects in the Fine-o2 treatment reported zero income, these magnitudes are far below a hundred percent as predicted by the theory. Even fewer than ten percent of the subjects reported zero income in each of the other four treatments.

Third, in rounds with voting, on average, less than half of the subjects voted for the stricter values of the various variables of the tax system. Specifically, on average, 49.58 percent of the subjects in Tax-o1 and 43.33 percent of the subjects in Tax-o2 voted for the high tax rate. The percentages of subjects voting for the high fine rate were a little bit lower (42.5 percent in Fine-o1 and 35.83 percent in Fine-o2), but, on average only 31.04 percent of the subjects in the Audit-o1 treatment and 22.29 percent of the subjects in the Audit-o2 treatment voted for the high audit rate. These observations suggest that, on average, subjects preferred a less strict auditing environment. In addition, these observations are inconsistent with the theoretical predictions that subjects will vote for the low tax rate and high audit and fine rates when voting is allowed.

Fourth, in rounds with voting, there were only, on average, 6.04 percent to 17.5 percent of the subjects declaring zero income, and 7.5 percent to 19.79 percent of the subjects complying fully. These observations also fail to meet the theoretical predictions when voting is allowed.

We summarize the above observations in Result 1 as follows:

Result 1: The theoretical prediction of zero compliance for the no-vote rounds in all six treatments and the vote rounds for the two tax treatments does not hold. The complete-compliance prediction for the vote rounds of the two audit treatments and the two fine treatments also fails. Furthermore, the experimental evidence does not support the theoretical predictions that subjects will vote for the low tax rate and report zero income, and that subjects will vote for the high audit rate and the high fine rate and comply fully.

4.2. Regression Analysis of the Effects of Democracy

We now examine whether democracy effects exist; that is, whether the level of compliance was higher if subjects were allowed to vote on the values of various aspects of the tax system than when they were not. We estimate the following fixed effects regression model of tax compliance.

$$\text{compliance rate}_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}, \tag{3}$$

[4] The standard error of the sample mean is calculated as $\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)}\Big/\sqrt{n}$, where x_i is the value of the observation, $\bar{x}$ is the sample mean, and n is the sample size.

where the i and t subscripts respectively represent subjects and rounds, x_{it} is a set of variables characterizing the experimental setting and subject behavior, β is a vector of coefficients to be estimated, μ_i is a fixed effects parameter, capturing subject i's time invariant heterogeneity in his or her compliance behavior, and ε_{it} is an error term, which is assumed to be normally distributed. It is noted that x_{it} does not include a constant term. This is to facilitate the interpretation of the coefficient estimates and is to be explained below. The fixed effects specification, by controlling for time-invariant heterogeneity, allows us to obtain within-subject variations in tax compliance in response to changes in the experimental setting and outcomes (e.g., the prevailing tax, audit, or fine rate).

[Tables 4, 5.1 and 5.2 about here]

In addition to using the whole sample (i.e., from rounds 1 to 20) to estimate equation (3), we also use a subsample consisting of observations from rounds 6–15. This is for the purpose of the robustness check. The rationale for this sample restriction is that there may be unobservable factors, e.g., learning and dynamic peer effects, affecting a subject's compliance behavior. The definitions of variables used in the regression are listed in Table 4 and will be explained in detail later. The results in Tables 5.1 and 5.2 are obtained using a quadratic function of the round number to control for these unobserved effects over the whole of the sample periods (rounds 1–20). Any misspecification of the unobserved factors over the rounds of the experiment will affect the results. This is because we rely on a change in the compliance rates surrounding the change in the tax, audit, or fine rate regime (whether or not voting is allowed) occurring in round 11, while allowing rounds of the experiment to have a smooth effect on compliance rates and thereby identify the effect of democracy. Any un-captured non-linearity in the effects of rounds of the experiment may bias the estimates of "No-vote," "High rate," and "Low rate" as these variables are more or less defined by the round number of the experiment.

By restricting the sample to rounds 6–15, we use a quadratic function of the round number to approximate these unobserved effects for a shorter duration surrounding the introduction or termination of voting in the transitional round (round 11). This makes the quadratic functional form less liable to misspecification. This is similar in spirit to the program evaluation literature's local polynomial approach for regression discontinuity design, where outcomes observed near the timing of a policy change are used (see Imbens and Lemieux, 2008).

To have a better understanding of compliance behavior, we also analyze the estimated fixed effects $\hat{\mu}_i$ by estimating a regression model as follows:

$$\hat{\mu}_i = \omega_i \gamma + e_i,$$

where ω_i is subject i's time invariant characteristics (e.g., gender, age, risk attitude, previous donation behavior), which were collected after the experiments, γ is a vector of coefficients to be estimated, and e_i is a normally distributed error term.

We use two different sets of explanatory variables x_{it} to explain the compliance rate. The first set, denoted as Specification I, consists of the variables "No-vote" (1 if voting was not allowed and 0 otherwise), "Round" (round number) and its square, "True income" (received income), the subjects' voting decision "Voted high" (1 if the subject voted for a high tax, audit, or fine rate, and 0 otherwise), "High rate" (1 if a high tax, audit, or fine rate was applied to the subject and 0 otherwise), and “Low rate” (1 if a low tax, audit, or fine rate was applied to the subject and 0 otherwise). Definitions of variables are listed in Table 4.

It is noted that we do not include a constant term in equation (3) such that we could include mutually exclusive dummy variables “High rate,” “Low rate,” and “No-vote.” The coefficients of these variables represent the average compliance rates associated with these mutually exclusive events holding other things constant. This innocuous specification makes interpretations of the results more straightforward. Moreover, the quadratic form of the round number is to control for unobserved heterogeneity, which changes with the rounds of the experiment that a subject has played. These unobserved factors include learning by subjects and dynamics generated by interactions among subjects. We assume that the effects of these unobserved factors are a smooth function of the round number. The regression results are reported in Table 5.1.

In Table 5.1 the coefficients of “No-vote,” “High rate,” and “Low rate” denote the average compliance rates in the mutually exclusive events. The coefficient estimates show that the average compliance rates for the rounds when voting was not allowed (“No-vote” = 1), the tax, audit, or fine rate applied was high with voting allowed (“High rate” = 1), and the tax, audit, or fine rate applied was low with voting allowed (“Low rate” = 1) are all positive and significantly different from zero at conventional levels in all six treatments. Given that the tax, audit, and fine rates were always low in the rounds when voting was now allowed, to test the effect of democracy on compliance, we examine whether or not the compliance rate when voting was not allowed is equal to the compliance rate when the rate applied was low in the rounds with voting allowed. That is, we test the equality of the coefficients “No-vote” and “Low rate.”

As suggested by the *p*-values reported in Table 5.1, it turns out that the null hypothesis of equality in compliance rates is accepted in all treatments except in the Audit-o2 and Fine-o1 treatments. However, in the Audit-o2 treatment, the average compliance rate was actually higher in the rounds with voting not allowed (86.58 percent vs. 72.64 percent). We also find higher compliance rates in the no-vote rounds for the Tax-o2, Audit-o1, and Fine-o2 treatments, even though the differences are statistically insignificant. For the Fine-o1 treatment, the compliance rate was higher when voting was allowed, as indicated by the *p*-value of the equality test of the “No-vote” and “Low rate” coefficient estimates. This is the only case indicating that democracy raises compliance. Thus, except for Fine-o1, our results suggest that democracy does not seem to improve compliance.

One may be concerned that the coefficient estimates are confounded by the effect of social norms (see Alm, McClelland and Schulze, 1999) as the actual tax, audit, or fine rates in the voting stage may reveal the preferences of a subject’s peers in the same group and the subject may react to expectations about his or her peers’ preferences. However, given that we have a randomization mechanism, where the computer decides the tax, audit, or fine rate, subjects are not able to infer his or her peers’ voting behavior or preferences.

[Tables 5.3 and 5.4 about here]

The results on the effect of democracy in Tables 5.3, using observations from rounds 6–15, suggest that the conclusion based on the results in Tables 5.1 still holds. The only exception is that the *p*-value of the equality test for the “No-vote” and “Low rate” coefficient estimates is now no longer statistically significant for the Fine-o1 treatment. As a consequence, our results do not support the existence of an effect of democracy on compliance.

A comparison of the coefficients for “High rate” and “Low rate” in Tables 5.1 and 5.3 suggests that when a high tax, audit, or fine rate prevailed, a subject complied more as the coefficients for "High rate" are significantly larger than those for "Low rate" in all treatments except Tax-o2. It is reasonable for subjects to comply more when faced with a higher audit or fine rate, but it is somewhat counter-intuitive to see that the compliance rate is also higher when the tax rate is higher. Since some randomization mechanism has been added to the determination of the tax, audit, and fine rates, this effect is unlikely to arise from a peer effect. This is because the tax, audit or fine rate that would prevail is not totally determined by majority voting, such that a subject may not perceive the tax, audit, or fine rate as a signal of other subjects' preferences.

It is also interesting to see that a subject's own voting decision did not affect his or her compliance behavior as suggested by the coefficient estimates for "Voted high." As shown by Tables 5.1 and 5.3, the coefficient estimates for "Voted high" in all treatments are statistically insignificant. This implies that after controlling for subjects' unobserved heterogeneity on compliance behavior, their voting behavior related to the tax, audit, or fine rate does not affect their compliance behavior.

Moreover, subjects receiving higher income had lower compliance rates, as indicated by the negative coefficients of "True income." However, this effect is insignificant in all six treatments when only observations from rounds 6–15 are used in the estimation. The above results are summarized in Result 2.

Result 2: Subjects behaved more compliantly when a high audit rate or a high fine rate was applied to them than when a lower counterpart was applied. Subjects' voting decisions had no significant effects on their compliance behavior. Income had no significant effect on compliance.

Tables 5.2 and 5.4 report the effects of subject characteristics on the fixed effects estimate, denoted by $\hat{\mu}_i$, which represents the subject's specific average compliance rate after controlling for experimental settings and outcomes. The results in Table 5.2 show that subjects who were more risk tolerant had lower compliance rates in all treatments as suggested by the negative coefficient estimates of "Risk-taking." However, the coefficient estimates are not statistically significant for the Tax-o2 and Fine-o1 treatments. The age of the subjects does not have a consistent effect across treatments. Its coefficient estimates are positive in some treatments (Tax-o1 and Audit-o1) and negative in others, and statistically significant only in the Audit-o2 treatment.

Having taken at least one economics course (i.e., "econ" = 1) has a negative effect on tax compliance, but only the estimates for the Audit-o1 and Fine-o1 treatments are statistically significant at conventional levels. Except for the Tax-o2 and Fine-o2 treatments, the coefficient estimates for "Donation" are all positive, but only the estimate for the Fine-o1 treatment is statistically significant. This seems to indicate that more charitable individuals are also more tax compliant. The results in Table 5.4 are similar to those in Table 5.2. We summarize these results in Result 3.

Result 3: Gender generally had no significant impacts on compliance. The attitude toward risks had a significant and negative impact on compliance, especially when it was the audit rate to be voted on. Age and having taken economics course(s) had significant and negative impacts on compliance in some treatments. The amount of money donated to charities had a significant and positive effect on compliance in only the Fine-o1 treatment.

[Tables 6.1 and 6.2 about here]

It is informative to compare these fixed effects estimates with the OLS estimates (i.e., unobserved heterogeneity not controlled for). In Table 6.1, the OLS results indicate that "Voted high" is statistically significant for the Audit-o1, Audit-o2, Fine-o1, and Fine-o2 treatments. As shown in Table 6.2, this finding holds in the first three treatments even when only observations from rounds 6–15 are used for the OLS estimation. This implies that there is time-invariant unobserved heterogeneity affecting subjects' compliance behavior and this unobserved heterogeneity also affects their preferences for the audit and fine rates.

One may be concerned that our fixed effects specification is not able to eliminate unobserved factors affecting both compliance behavior and voting behavior, i.e., there may be some round-specific shocks affecting both kinds of behavior. Even though the coefficient estimate of "Voted high" is statistically insignificant in Tables 5.1 and 5.3, the effect of voting behavior may be non-linear. To examine such a possibility we run estimate (3) again with a richer set of explanatory variables, which is denoted as Specification II and consists of interactions of "Voted high" and "Voted low" with "High rate" and "Low rate" (denoted as "Voted high × High rate," "Voted high × Low rate," "Voted low × High rate" and "Voted low × Low rate," respectively). Tables 7.1 through 7.4 report the results.

[Tables 7.1, 7.2, 7.3, and 7.4 about here]

The results in Table 7.1 suggest that, across all treatments, there are not many differences in a subject's compliance rate when he or she voted for a high or a low rate for tax, audit or fine, given that a low rate applied. The only exception is the difference for treatment Audit-o2, for which the compliance rate is higher by 7.70 percentage points (p-value = 0.04) when a subject voted for a high audit rate than when a subject voted for a low audit rate. The rest of the differences are statistically insignificant as indicated by the Wald test's p-values.

When a high rate applied, a subject's compliance rate was lower when he or she voted for a high rate than when he or she voted for a low rate. For the Tax-o1, Audit-o1 and Fine-o1 treatments, the differences are statistically significant. However, most of these differences disappear when we restrict observations to those for rounds 6–15 (see Table 7.3). Only the difference in coefficient estimates between "Voted high × High rate" and "Voted low × High rate" for Tax-o1 remains statistically significant (p-value = 0.04).

We next focus on the comparisons of tax compliance in the no-vote rounds versus the rounds when voting was allowed and a low rate applied that was conditional on subjects' voting decisions. As reported in Table 7.1, for subjects voting for a low tax, audit or fine rate, this difference is statistically significant for the Audit-o2 treatment (p-value = 0.00) and marginally significant for the Fine-o1 treatment (p-value = 0.08). For subjects voting for a high tax, audit or fine rate, there are no significant differences in the compliance rate during the no-vote rounds and the low-rate rounds in the voting stage as indicated by the p-values of the Wald test. The findings are similar when observations are confined to those for rounds 6–15.

Overall, the results suggest that there are not many differences in a subject's compliance rate when he or she voted for a high rate versus a low rate. This implies that our fixed effects specification is able to control for unobserved heterogeneity affecting tax compliance and voting behavior and our finding of no effect of democracy is not confounded by the subjects' unobserved heterogeneity. The above results are summarized in Result 4.

Result 4: The democracy effect exists only in the Fine-o1 treatment if we look at the entire twenty rounds of the experiment, and it does not hold in all six treatments if we look at only the middle ten rounds of the experiment.

5. CONCLUSION

This paper experimentally investigates individuals' compliance behavior when they are allowed to vote for the tax, audit, or fine rate. The democracy effect is also examined by this paper. Since individuals may behave more compliantly if the outcome of voting happens to be consistent with their inherent preferences, we apply Dal Bó, Foster, and Putterman's (2010) approach by adding a stage of computer decisions after voting to control for this selection problem. In addition, we also control for the order effect by switching the order of the rounds with voting and the rounds without voting.

The main findings of our paper are the following. First, subjects generally preferred a less severe auditing environment. On average more than half of the subjects voted for the low tax rate and the low fine rate, and even more than seventy percent of the subjects voted for the low audit rate. Second, income and gender generally had no significant impact on compliance. The attitude toward risks had a significant and negative impact on compliance, especially when the audit rate was to be voted on. Third, subjects for whom the high audit rate or high fine rate was applied behaved significantly more compliantly than subjects for whom the lower counterpart was applied. Fourth, our experimental finding did not support the democracy effect. That is, the institutions that allowed subjects to vote for the tax, audit, or fine rate did not have a positive and significant impact on compliance.

The results from our experiment have some policy implications. First, given the evidence that the democracy effect does not hold, democratic participation cannot improve compliance. Furthermore, because more than half of the subjects voted for a low tax rate and a low fine rate, and more than seventy percent of the subjects voted for a low audit rate, democratic participation may even have deteriorated compliance and government revenues. Hence, allowing individuals to vote on the magnitudes of the various aspects of the tax system may not be appropriate in a tax auditing environment. To improve compliance, a more effective way for the tax authority is to enforce the high audit rate and high fine rate directly.

Table 1. Framework of the experiment and parameters used for various aspects of the tax system

Treatment	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
Rounds 1–10						
Voting	no	yes: on *t*	no	yes: on *p*	no	yes: on *f*
Tax rate (*t*)	0.2	0.2 vs. 0.4	0.2	0.2	0.2	0.2
Audit probability (*p*)	0.1	0.1	0.1	0.1 vs. 0.4	0.1	0.1
Fine rate (*f*)	3	3	3	3	3	3 vs. 6
Rounds 11–20						
Voting	yes: on *t*	no	yes: on *p*	no	yes: on *f*	no
Tax rate (*t*)	0.2 vs. 0.4	0.2	0.2	0.2	0.2	0.2
Audit probability (*p*)	0.1	0.1	0.1 vs. 0.4	0.1	0.1	0.1
Fine rate (*f*)	3	3	3	3	3 vs. 6	3

Table 2. Individual characteristics

Treatment	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2	All
Female	0.7708 (0.4247)	0.6458 (0.4833)	0.8125 (0.3944)	0.7708 (0.4247)	0.7708 (0.4247)	0.6875 (0.4684)	0.7431 (0.4370)
Class	2.2917 (1.1291)	2.125 (1.0442)	1.875 (0.8903)	2.1458 (1.0516)	2.1458 (1.0717)	2.0833 (1.0883)	2.1111 (1.0449)
Age	20.1667 (1.4192)	19.9167 (1.3182)	19.75 (1.0417)	20.25 (1.2965)	20.0833 (1.2520)	19.8958 (1.1893)	20.0104 (1.2569)
Taken econ course(s)	0.8125 (0.3944)	0.7917 (0.4104)	0.7917 (0.4104)	0.75 (0.4376)	0.7083 (0.4593)	0.8542 (0.3567)	0.7847 (0.4110)
Donation	2 (0.9676)	1.9375 (0.7553)	2.0417 (0.9444)	1.9792 (0.8627)	1.9583 (0.8495)	2.625 (1.2820)	2.0903 (0.9783)
Risk-taking	5.4792 (2.0935)	5.1458 (2.1237)	5.2708 (2.2096)	5.6042 (2.3039)	4.6875 (2.3078)	4.8542 (2.3519)	5.1736 (2.2340)

Standard deviations are in parentheses.

Table 3.1. Descriptive statistics for the Tax-o1 and Tax-o2 treatments

	Tax-o1		Tax-o2	
	Rounds 1–10	Rounds 11–20	Rounds 1–10	Rounds 11–20
(1) Average compliance rate	0.5439 (0.2150)	0.4078 (0.2404)	0.6048 (0.2272)	0.4482 (0.2721)
(2) Percentage of subjects declaring zero income	9.58% (0.2042)	15.42% (0.2843)	6.04% (0.1943)	9.17% (0.2305)
(3) Percentage of subjects fully complying	8.96% (0.1716)	7.5% (0.1804)	10.63% (0.2453)	7.92% (0.2031)
(4) Average earnings (points)	107.8925 (6.4893)	106.9104 (12.5992)	115.0925 (8.4506)	105.155 (8.9406)
(5) Percentage of subjects voting for $t = 0.4$ in the voting stage	–	49.58% (0.4084)	43.33% (0.3652)	–
(6) Percentage of time computer adopts the group decision	–	73.33% (0.1342)	62.5% (0.1212)	–
(7) Percentage of subjects being applied $t = 0.4$ in the voting stage	–	47.5% (0.1792)	45.83% (0.1569)	–
(8) Average compliance rate conditional on voting for $t = 0.4$	–	0.4337 (0.2720)	0.6275 (0.2578)	–
(9) Average compliance rate conditional on voting for $t = 0.2$	–	0.4052 (0.2312)	0.5845 (0.2345)	–
(10) Average compliance rate conditional on being applied $t = 0.4$	–	0.4812 (0.2824)	0.6087 (0.2509)	–
(11) Average compliance rate conditional on being applied $t = 0.2$	–	0.3574 (0.2463)	0.6090 (0.2506)	–
(12) Average compliance rate for subjects voting for $t = 0.4$ and $t = 0.4$ is applied	–	0.4756 (0.2866)	0.6257 (0.2695)	–
(13) Average compliance rate for subjects voting for $t = 0.2$ and $t = 0.4$ is applied	–	0.4614 (0.2855)	0.6050 (0.2592)	–
(14) Average compliance rate for subjects voting for $t = 0.4$ and $t = 0.2$ is applied	–	0.3570 (0.2847)	0.6562 (0.2458)	–
(15) Average compliance rate for subjects voting for $t = 0.2$ and $t = 0.2$ is applied	–	0.3728 (0.2293)	0.5838 (0.2604)	–

The observations are the averages of all subjects' average choices over the period specified. The numbers in parentheses are the standard deviations of all subjects' average choices over that period.

Table 3.2. Descriptive statistics for the Audit-o1 and Audit-o2 treatments

	Audit-o1		Audit-o2	
	Rounds 1–10	Rounds 11–20	Rounds 1–10	Rounds 11–20
(1) Average compliance rate	0.6354 (0.2592)	0.5540 (0.2537)	0.5515 (0.2663)	0.3636 (0.2855)
(2) Percentage of subjects declaring zero income	5.63% (0.1749)	8.33% (0.2186)	17.5% (0.2892)	26.04% (0.3999)
(3) Percentage of subjects fully complying	14.58% (0.2939)	16.25% (0.2915)	14.79% (0.2449)	7.71% (0.1640)
(4) Average earnings (points)	110.99 (6.0136)	106.5279 (7.4026)	107.7475 (5.8442)	104.1296 (8.5336)
(5) Percentage of subjects voting for $p = 0.4$ in the voting stage	–	31.04% (0.3932)	22.29% (0.3197)	–
(6) Percentage of time computer adopts the group decision	–	67.5% (0.1756)	71.67% (0.1521)	–
(7) Percentage of subjects being applied $p = 0.4$ in the voting stage	–	30.83% (0.1761)	24.17% (0.1471)	–
(8) Average compliance rate conditional on voting for $p = 0.4$	–	0.6592 (0.2688)	0.6931 (0.2084)	–
(9) Average compliance rate conditional on voting p for = 0.1	–	0.5020 (0.2472)	0.5153 (0.2797)	–
(10) Average compliance rate conditional on being applied $p = 0.4$	–	0.787 (0.2262)	0.7705 (0.2479)	–
(11) Average compliance rate conditional on being applied $p = 0.1$	–	0.4531 (0.2825)	0.4835 (0.2929)	–
(12) Average compliance rate for subjects voting for $p = 0.4$ and $p = 0.4$ is applied	–	0.8336 (0.2075)	0.8346 (0.2194)	–
(13) Average compliance rate for subjects voting for $p = 0.1$ and $p = 0.4$ is applied	–	0.7494 (0.2488)	0.7647 (0.2526)	–
(14) Average compliance rate for subjects voting for $p = 0.4$ and $p = 0.1$ is applied	–	0.5548 (0.3101)	0.6153 (0.2623)	–
(15) Average compliance rate for subjects voting for $p = 0.1$ and $p = 0.1$ is applied	–	0.3958 (0.2583)	0.4450 (0.3067)	–

The observations are the averages of all subjects' average choices over the period specified. The numbers in parentheses are the standard deviations of all subjects' average choices over that period.

Table 3.3. Descriptive statistics for the Fine-o1 and Fine-o2 treatments

	Fine-o1		Fine-o2	
	Rounds 1–10	Rounds 11–20	Rounds 1–10	Rounds 11–20
(1) Average compliance rate	0.5633 (0.2415)	0.4468 (0.2560)	0.5681 (0.2569)	0.4047 (0.2626)
(2) Percentage of subjects declaring zero income	9.79% (0.2068)	16.67% (0.2846)	13.54% (0.2539)	21.46% (0.3003)
(3) Percentage of subjects fully complying	15.83% (0.2827)	14.58% (0.2601)	19.79% (0.2646)	11.04% (0.1949)
(4) Average earnings (points)	108.2571 (8.1316)	104.6079 (7.8425)	108.2342 (7.3459)	104.3917 (7.0479)
(5) Percentage of subjects voting for $f = 6$ in the voting stage	–	42.5% (0.4097)	35.83% (0.3847)	–
(6) Percentage of time computer adopts the group decision	–	75% (0.1473)	73.33% (0.0883)	–
(7) Percentage of subjects being applied $f = 6$ in the voting stage	–	40.83% (0.1724)	33.33% (0.1742)	–
(8) Average compliance rate conditional on voting for $f = 6$	–	0.4862 (0.3012)	0.6195 (0.2975)	–
(9) Average compliance rate conditional on voting for $f = 3$	–	0.4198 (0.2609)	0.5191 (0.2634)	–
(10) Average compliance rate conditional on being applied $f = 6$	–	0.5832 (0.2990)	0.7123 (0.2620)	–
(11) Average compliance rate conditional on being applied $f = 3$	–	0.3468 (0.2807)	0.5134 (0.2839)	–
(12) Average compliance rate for subjects voting for $f = 6$ and $f = 6$ is applied	–	0.5814 (0.2903)	0.7350 (0.2894)	–
(13) Average compliance rate for subjects voting for $f = 3$ and $f = 6$ is applied	–	0.5396 (0.3354)	0.6979 (0.2567)	–
(14) Average compliance rate for subjects voting for $f = 6$ and $f = 3$ is applied	–	0.3681 (0.3512)	0.5344 (0.3427)	–
(15) Average compliance rate for subjects voting for $f = 3$ and $f = 3$ is applied	–	0.3278 (0.2842)	0.4863 (0.2862)	–

The observations are the averages of all subjects' average choices over the period specified. The numbers in parentheses are the standard deviations of all subjects' average choices over that period.

Table 4. Variable Definitions

	Definition
Compliance rate	Percentage of income reported (i.e., $100 \times \frac{\text{reported income}}{\text{true income}}$).
No-vote	Dummy variable indicating that the current round is a no-vote round.
Round	Round number.
Round^2	Round number squared.
True income	Income received by the subject.
Voted high	Dummy variable indicating that the subject voted for a high tax, fine, or audit rate in the round when voting was allowed.
High rate	Dummy variable indicating that a high tax, fine, or audit rate applied to a subject in the round when voting was allowed.
Low rate	Dummy variable indicating that a low tax, fine, or audit rate applied to a subject in the round when voting was allowed.
Voted high × High rate	Dummy variable indicating that the subject voted for a high tax, fine, or audit rate and that a high tax, fine, or audit rate was applied.
Voted low × High rate	Dummy variable indicating that the subject voted for a low tax, fine, or audit rate and that a high tax, fine, or audit rate was applied.
Voted high × Low rate	Dummy variable indicating that the subject voted for a high tax, fine, or audit rate and that a low tax, fine, or audit rate was applied.
Voted low × Low rate	Dummy variable indicating that the subject voted for a low tax, fine, or audit rate and that a low tax, fine, or audit rate was applied.

Table 5.1. Specification I: Fixed effects regression results with observations from all rounds

Independent variables	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
No-vote	48.4900** (7.0403)	25.3432** (6.2502)	73.8207** (5.7704)	86.5757** (7.0381)	95.3144** (7.3447)	66.1644** (7.3358)
Round	–1.9049** (0.2795)	–1.5860** (0.2593)	–1.5829** (0.2334)	–2.4772** (0.2782)	–2.7390** (0.2987)	–1.4463** (0.3016)
$Round^2$	0.0739** (0.0271)	0.0258 (0.0250)	0.0725** (0.0226)	0.0779** (0.0269)	0.0237 (0.0288)	0.0844** (0.0291)
True income	–0.0726** (0.0365)	–0.0214 (0.0337)	–0.0578* (0.0307)	–0.0272 (0.0361)	–0.0124 (0.0389)	–0.0537 (0.0394)
Voted high	–3.4909 (2.8147)	–1.1126 (2.5574)	–1.8947 (2.5636)	3.8298 (3.2268)	–4.0614 (3.0671)	–1.0789 (3.1408)
High rate	61.0472** (7.4753)	26.8079** (6.6498)	104.0498** (6.2119)	105.2370** (7.2696)	126.6290** (7.8697)	80.1713** (7.8517)
Low rate	49.3858** (7.4847)	25.1043** (6.4776)	71.2627** (6.0567)	72.6438** (7.0026)	102.8341** (7.6517)	63.8291** (7.5605)
Within R^2	0.1431	0.1437	0.2656	0.2805	0.1920	0.1468
Observations	960	960	960	960	960	960
p-value of Wald test for equality of coefficients						
"No-vote" = "Low rate"	0.8052	0.9431	0.3675	0.0000	0.0426	0.5310
"Low rate" = "High rate"	0.0000	0.4437	0.0000	0.0000	0.0000	0.0000

Results from the estimation of the fixed effects model: $\text{compliance rate}_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$. Standard errors are in parentheses. The notation ** denotes the 5% significance level and * denotes the 10% significance level.

Table 5.2. Specification I: Explaining the fixed effects

Independent variables	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
Female	8.3942 (6.3310)	9.4742 (6.9558)	8.9371 (7.0516)	1.1266 (8.9839)	–2.5519 (6.5239)	6.9051 (7.7868)
Risk-taking	–3.6683** (1.3219)	–2.7024 (1.7376)	–5.1434** (1.3608)	–3.1788* (1.6272)	–1.6439 (1.3254)	–3.6339** (1.3823)
Age	1.4886 (2.2507)	–3.8828 (2.3144)	1.5966 (3.0489)	–6.5444** (2.8539)	–3.0448 (2.6943)	–6.0408 (3.9014)
Econ	–12.0047 (7.9900)	–7.0448 (8.2979)	–18.2287** (8.7328)	–6.9888 (7.2228)	–10.6680* (5.9731)	–0.3621 (8.9131)
Donation	4.1651 (3.3243)	–1.0367 (4.5235)	5.0151 (3.5367)	5.3237 (3.7964)	12.5602** (2.5280)	–0.3986 (3.1091)
Constant	–14.9685 (44.5650)	92.7054* (49.2715)	–7.4921 (59.3486)	144.1756** (58.2793)	53.7814 (54.4691)	134.4339* (77.6586)
R^2	0.2100	0.1701	0.3733	0.2226	0.3080	0.1763
Observations	48	48	48	48	48	48

Results from the regression model $\hat{\mu}_i = \omega_i\gamma + e_i$. Standard errors are in parentheses. The notation ** denotes the 5% significance level and * denotes the 10% significance level.

Table 5.3. Specification I: Robustness check using observations from rounds 6–15

Independent variables	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
No-vote	74.8693** (8.9205)	63.1395** (9.1038)	87.8736** (7.9574)	72.7349** (9.9448)	59.8587** (10.6645)	78.5821** (11.0670)
Round	–1.5379** (0.7791)	–1.2680* (0.7226)	–2.5832** (0.6834)	–1.9594** (0.8070)	–1.8611** (0.8738)	–0.7963 (0.8745)
$Round^2$	0.0016 (0.1488)	–0.0459 (0.1377)	0.2513* (0.1312)	0.0089 (0.1535)	0.1805 (0.1670)	0.1773 (0.1670)
True income	–0.0653 (0.0500)	–0.0314 (0.0466)	–0.0555 (0.0452)	–0.0324 (0.0514)	–0.0585 (0.0560)	–0.0617 (0.0573)
Voted high	–5.3983 (3.8442)	–2.8893 (3.5251)	–3.5117 (3.6219)	2.2503 (4.7688)	–0.9765 (4.4304)	0.2437 (4.7121)
High rate	84.2109** (9.5322)	66.5032** (9.1794)	118.1157** (8.6131)	95.6164** (10.8841)	85.4649** (11.1287)	97.3472** (11.3343)
Low rate	75.9829** (9.1353)	67.0194** (9.0482)	90.4694** (8.4342)	61.0880** (10.1012)	62.2287** (11.2152)	79.3970** (10.9516)
Within R^2	0.0382	0.0557	0.1838	0.2062	0.0942	0.0864
Observations	480	480	480	480	480	480
p-value of Wald test for equality of coefficients						
"Low rate" = "No-vote"	0.8136	0.4085	0.5303	0.0141	0.6527	0.8775
"Low rate" = "High rate"	0.0164	0.8678	0.0000	0.0000	0.0000	0.0000

Results from the estimation of the fixed effects model: $\text{compliance rate}_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$. Standard errors are in parentheses. The notation ** denotes the 5% significance level and * denotes the 10% significance level.

Table 5.4. Specification I: Explaining the fixed effects

Independent variables	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
Female	7.8282 (7.2601)	8.0260 (7.6919)	8.2303 (6.5964)	1.4870 (9.8615)	–4.4206 (7.2062)	7.5537 (8.3211)
Risk-taking	–4.1316** (1.4395)	–2.9415 (1.8792)	–4.7057** (1.3907)	–3.6561** (1.6570)	–1.6361 (1.4602)	–3.4833** (1.4986)
Age	1.8084 (2.6804)	–4.5421* (2.6688)	0.5896 (3.3280)	–6.6866** (2.9230)	–2.0546 (2.7449)	–5.6249 (4.1605)
Econ	–13.9015 (8.4265)	–8.1938 (9.4543)	–19.1466* (10.0040)	–6.0740 (7.4275)	–17.3939** (6.1024)	0.2002 (9.8435)
Donation	5.3216 (3.8321)	–0.5689 (4.9564)	5.3852 (4.0946)	5.5889 (4.0490)	13.9321** (2.6216)	–0.7593 (3.2917)
Constant	–19.2142 (53.3211)	108.0061* (56.9401)	10.6337 (64.6409)	148.2406** (59.8066)	37.3777 (54.4439)	125.4494 (82.2772)
R^2	0.2072	0.1572	0.3246	0.2172	0.3525	0.1492
Observations	48	48	48	48	48	48

Results from the regression model $\hat{\mu}_i = \omega_i \gamma + e_i$. Standard errors are in parentheses. The notation ** denotes the 5% significance level and * denotes the 10% significance level.

Table 6.1. Specification I: OLS results with observations from all rounds

Independent variables	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
No-vote	46.2170** (8.2417)	48.6928** (8.1522)	56.6799** (6.8585)	46.7052** (8.5355)	54.8843** (5.9348)	51.9108** (7.5367)
Round	46.2170** (8.2417)	–1.6683** (0.3295)	–2.7895** (0.3518)	–1.4870** (0.3431)	–1.5493** (0.3046)	–2.5254** (0.3688)
$Round^2$	0.0723** (0.0322)	0.0258 (0.0250)	0.0198 (0.0355)	0.0884** (0.0294)	0.0776** (0.0283)	0.0819** (0.0269)
True income	–0.0911 (0.0649)	–0.0214 (0.0337)	–0.0150 (0.0547)	–0.0094 (0.0640)	–0.0327 (0.0493)	–0.0103 (0.0448)
Voted high	–1.3401 (5.6237)	6.3796 (4.6751)	9.8700* (5.6965)	11.6060* (6.0590)	20.0481** (5.8897)	14.1401** (6.0096)
High rate	56.7847** (9.1419)	46.2458** (8.9092)	83.8310** (6.7666)	56.4558** (9.1109)	76.5302** (6.8849)	68.0295** (7.2324)
Low rate	46.9842** (8.1786)	44.3851** (9.0150)	57.9843** (6.8215)	39.1290** (8.8622)	45.7452** (6.3596)	35.1752** (7.3079)
R^2	0.7113	0.7613	0.8061	0.6691	0.7397	0.6673
Observations	960	960	960	960	960	960
p-value of Wald test for equality of coefficients						
"Low rate" = "No-vote"	0.8584	0.3329	0.7181	0.0984	0.0079	0.0002
"Low rate" = "High rate"	0.0118	0.6167	0.0000	0.0002	0.0000	0.0000

Results from the estimation of the fixed effects model: $\text{compliance rate}_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$. Standard errors are in parentheses. The notation ** denotes the 5% significance level and * denotes the 10% significance level.

Table 6.2. Specification I: OLS results with observations from rounds 6–15

Independent variables	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
No-vote	49.9407** (7.8295)	47.3791** (10.0278)	61.7721** (8.9078)	44.7915** (9.3077)	52.0048** (8.9028)	48.7820** (10.6640)
Round	–1.4801** (0.6850)	–1.5177* (0.8196)	–1.9833* (1.0412)	–0.8229 (1.0795)	–2.6412** (0.8067)	–2.0837** (0.9607)
$Round^2$	0.0123 (0.1645)	–0.0014 (0.1394)	0.1617 (0.1648)	0.1818 (0.2059)	0.2333 (0.1415)	0.0261 (0.1537)
True income	–0.1112 (0.0698)	0.0034 (0.0825)	–0.0150 (0.0547)	–0.0141 (0.0787)	–0.0413 (0.0739)	0.0076 (0.0764)
Voted high	–0.5408 (6.2500)	4.9625 (5.3538)	11.0601* (5.6858)	15.1154* (7.9685)	14.9244** (6.1402)	12.2087 (7.7232)
High rate	55.3716** (11.2469)	47.2902** (11.3243)	87.3359** (9.0478)	56.8732** (11.9315)	75.9974** (9.8909)	70.5521** (9.2375)
Low rate	49.2713** (9.4871)	45.4855** (11.2446)	57.3416** (10.0141)	40.5099** (10.8469)	49.4555** (8.9197)	33.6969** (9.7859)
R^2	0.6874	0.7507	0.7837	0.6382	0.7171	0.6450
Observations	480	480	480	480	480	480
p-value of Wald test for equality of coefficients						
“Low rate” = “No-vote”	0.8796	0.7352	0.4242	0.5346	0.5873	0.0111
“Low rate” = “High rate”	0.1850	0.7162	0.0000	0.0085	0.0000	0.0000

Results from the estimation of the fixed effects model: $\text{compliance rate}_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$. Standard errors are in parentheses. The notation ** denotes the 5% significance level and * denotes the 10% significance level.

Table 7.1. Specification II: Fixed effects regression results with observations from all rounds

Independent variables	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
No-vote	48.7943** (7.0349)	25.2179** (6.2625)	74.4603** (5.7591)	87.0896** (7.0307)	78.6984** (7.4092)	66.1462** (7.3389)
Round	–1.9091** (0.2792)	–1.5862** (0.2594)	–1.5677** (0.2328)	–2.4774** (0.2778)	–2.7511** (0.2987)	–1.4411** (0.3019)
Round^2	0.0730** (0.0270)	0.0257 (0.0250)	0.0743** (0.0225)	0.0780** (0.0269)	0.0227 (0.0288)	0.0840** (0.0291)
True income	–0.0714* (0.0365)	–0.0210 (0.0338)	–0.0581* (0.0306)	–0.0304 (0.0360)	–0.0108 (0.0389)	–0.0544 (0.0395)
Voted high × High rate	56.2613** (7.1228)	25.2134** (6.4092)	98.5984** (6.0166)	104.4199** (8.0389)	104.3767** (8.0205)	78.2969** (7.5658)
Voted low × High rate	64.1375** (7.6806)	27.1639** (6.7233)	107.5370** (6.3472)	108.6910** (7.4548)	112.4620** (7.6626)	81.1167** (8.0693)
Voted high × Low rate	48.5925** (7.4526)	24.4185** (6.4255)	72.9005** (5.9725)	80.1643** (7.8192)	84.4265** (8.1158)	63.5668** (7.5077)
Voted low × Low rate	48.2400** (7.5063)	24.7237** (6.5631)	70.6964** (6.0432)	72.4612** (6.9912)	85.3423** (7.6126)	63.5590** (7.5820)
Within R^2	0.1458	0.1438	0.2707	0.2838	0.1937	0.1470
Observations	960	960	960	960	960	960
	p-value of Wald test for equality of coefficients					
"Voted high × High rate" = "Voted low × High rate"	0.0383	0.5696	0.0187	0.4058	0.0563	0.5427
"Voted high × Low rate" = "Voted low × Low rate"	0.9218	0.9279	0.4675	0.0400	0.8106	0.9984
"Voted low × Low rate" = "No-vote"	0.8817	0.8852	0.1899	0.0000	0.0773	0.4915
"Voted high × Low rate" = "No-vote"	0.9602	0.8343	0.6525	0.1165	0.1988	0.5644

Results from the estimation of the fixed effects model: $\text{compliance rate}_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$. Standard errors are in parentheses. The notation ** denotes the 5% significance level and * denotes the 10% significance level.

Table 7.2. Specification II: Explaining the fixed effects

Independent variables	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
Female	8.3639 (6.3056)	9.4822 (6.9573)	9.0035 (7.0568)	1.0146 (9.0424)	–2.5492 (6.5287)	6.8734 (7.7846)
Risk-taking	–3.6490** (1.3097)	–2.6982 (1.7391)	–5.0947** (1.3646)	–3.1759* (1.6326)	–1.6045 (1.3287)	–3.6317** (1.3824)
Age	1.4779 (2.2420)	–3.8845 (2.3125)	1.5865 (3.0427)	–6.4532** (2.8647)	–3.0534 (2.6938)	–6.0367 (3.9025)
Econ	–12.0665 (7.9450)	–7.0663 (8.3026)	–18.4542** (8.7210)	–6.9603 (7.2656)	–10.7037* (5.9716)	–0.3103 (8.9036)
Donation	4.1571 (3.3175)	–1.0278 (4.5263)	5.0191 (3.5286)	5.2785 (3.8099)	12.5102** (2.5313)	–0.4064 (3.1086)
Constant	–14.7684 (44.3925)	92.7122* (49.2523)	–7.4335 (59.3146)	142.4662** (58.4443)	53.8909 (54.4632)	134.3394* (77.6668)
R^2	0.2105	0.1701	0.3719	0.2181	0.3057	0.1761
Observations	48	48	48	48	48	48

Results from the regression model $\hat{\mu}_i = \omega_i\gamma + e_i$. Standard errors are in parentheses. The notation ** denotes the 5% significance level and * denotes the 10% significance level.

Table 7.3. Specification II: Fixed effects regression results with observations from rounds 6–15

Independent variables	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
No-vote	74.7112** (8.9083)	22.6162** (8.4120)	3.9181 (8.4090)	71.9083** (9.9987)	62.7907** (10.1861)	51.0277** (10.2742)
Round	–1.4657* (0.7795)	–1.2826* (0.7231)	–2.5573** (0.6831)	–1.9557** (0.8073)	–1.8683** (0.8753)	–0.7899 (0.8742)
Round^2	0.0206 (0.1491)	–0.0438 (0.1378)	0.2536* (0.1310)	0.0075 (0.1536)	0.1791 (0.1673)	0.1754 (0.1670)
True income	–0.0655 (0.0499)	–0.0310 (0.0466)	–0.0519 (0.0453)	–0.0345 (0.0514)	–0.0584 (0.0561)	–0.0646 (0.0574)
Vote high × High rate	76.2453** (9.7121)	21.9978** (8.5085)	27.3720** (9.5312)	94.4057** (11.1669)	87.0091** (10.9874)	67.7520** (10.7343)
Voted low × High rate	87.1136** (9.7146)	27.5008** (9.0173)	36.5426** (9.1129)	96.3604** (10.9253)	89.0562** (11.2565)	73.1142** (11.4498)
Voted high × Low rate	72.9388** (9.6830)	25.1127** (8.5727)	5.0919 (9.1898)	64.6172** (9.9439)	64.7639** (10.7200)	54.7727** (10.2857)
Vote low × Low rate	73.7174** (9.2473)	25.5563** (8.8532)	5.5447 (8.6202)	59.7863** (10.2271)	64.9876** (10.7012)	50.8343** (10.5289)
Within R^2	0.0432	0.0572	0.1872	0.2075	0.0943	0.0892
Observations	480	480	480	480	480	480
	p-value of Wald test for equality of coefficients					
"Voted high × High rate" = "Voted low × High rate"	0.0411	0.2491	0.1016	0.7794	0.7479	0.4307
"Voted high × Low rate" = "Voted low × Low rate"	0.8746	0.9237	0.9160	0.3974	0.9674	0.4909
"Voted low × Low rate" = "No-vote"	0.8400	0.5432	0.6983	0.0113	0.6798	0.9712
"Voted high × Low rate" = "No-vote"	0.7352	0.6286	0.8170	0.2409	0.7619	0.5640

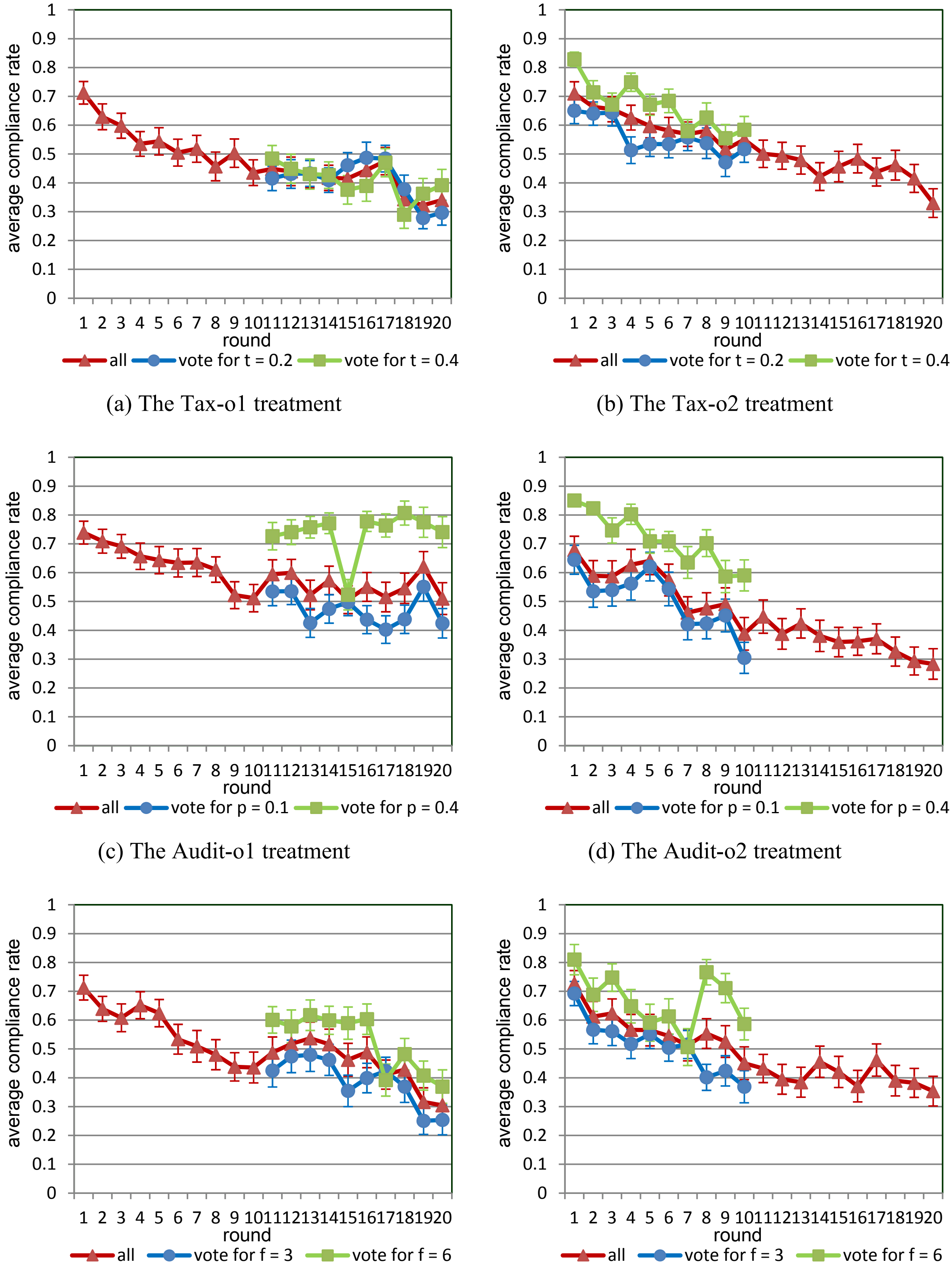
Results from the estimation of the fixed effects model: $\text{compliance rate}_{it} = x_{it}\beta + \mu_i + \varepsilon_{it}$. Standard errors are in parentheses. The notation ** denotes the 5% significance level and * denotes the 10% significance level.

Table 7.4. Specification II: Explaining the fixed effects

Independent variables	Tax-o1	Tax-o2	Audit-o1	Audit-o2	Fine-o1	Fine-o2
Female	7.7901 (7.2812)	8.0537 (7.7089)	8.2590 (6.5986)	1.4941 (9.8810)	–4.4413 (7.2149)	7.3322 (8.2876)
Risk-taking	–4.0927** (1.4258)	–2.9184 (1.8893)	–4.6532** (1.3870)	–3.6504** (1.6583)	–1.6259 (1.4608)	–3.4780** (1.4949)
Age	1.7750 (2.6774)	–4.5520* (2.6575)	0.5030 (3.3149)	–6.5995** (2.9240)	–2.0628 (2.7446)	–5.6270 (4.1519)
Econ	–13.9489 (8.3869)	–8.1020 (9.4512)	–19.1291* (9.9774)	–6.0455 (7.4512)	–17.3898** (6.1020)	0.7433 (9.8124)
Donation	5.3429 (3.8596)	–0.4906 (4.9596)	5.4025 (4.0889)	5.4934 (4.0529)	13.9405** (2.6216)	–0.8120 (3.2936)
Constant	–18.7289 (53.2364)	107.8417* (56.7977)	11.9952 (64.5199)	146.6068** (59.8118)	37.4910 (54.4349)	125.2916 (82.1052)
R^2	0.2052	0.1565	0.3228	0.2145	0.3524	0.1479
Observations	48	48	48	48	48	48

Results from the regression model $\hat{\mu}_i = \omega_i \gamma + e_i$. Standard errors are in parentheses. The notation ** denotes the 5% significance level and * denotes the 10% significance level.

Figure 1. Average Compliance Rates Conditional on Subjects' Voting Decisions



(a) The Tax-o1 treatment

(b) The Tax-o2 treatment

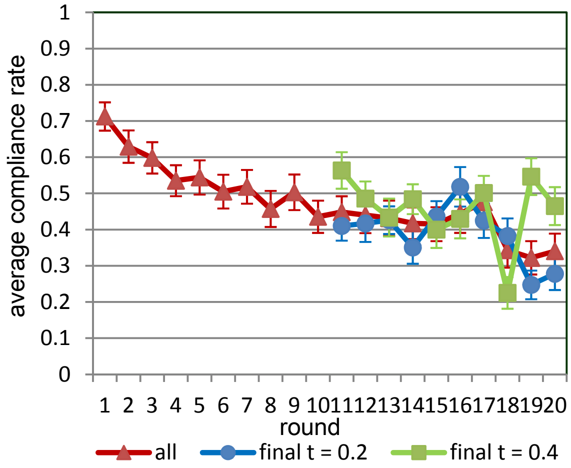
(c) The Audit-o1 treatment

(d) The Audit-o2 treatment

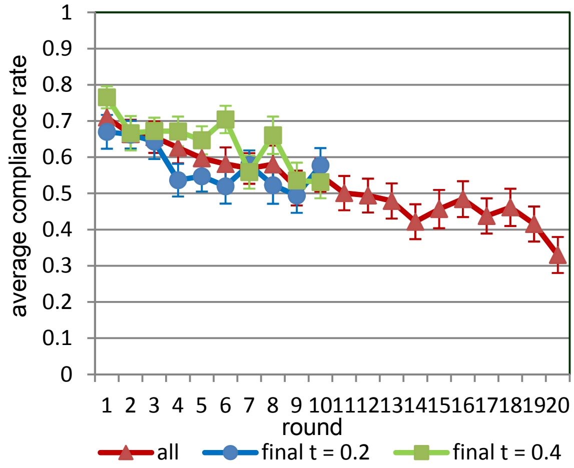
(e) The Fine-o1 treatment

(f) The Fine-o2 treatment

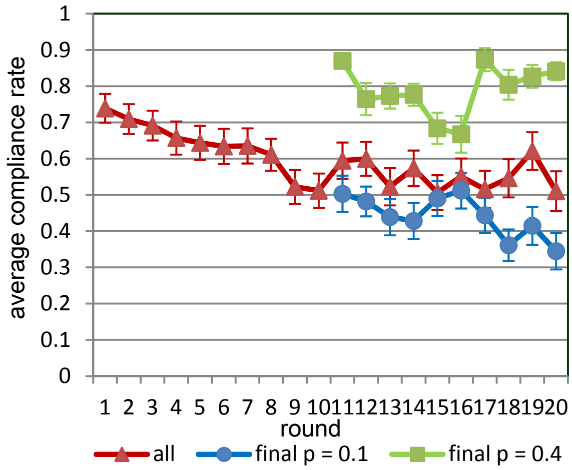
Figure 2. Average Compliance Rates Conditional on the Outcome Being Applied



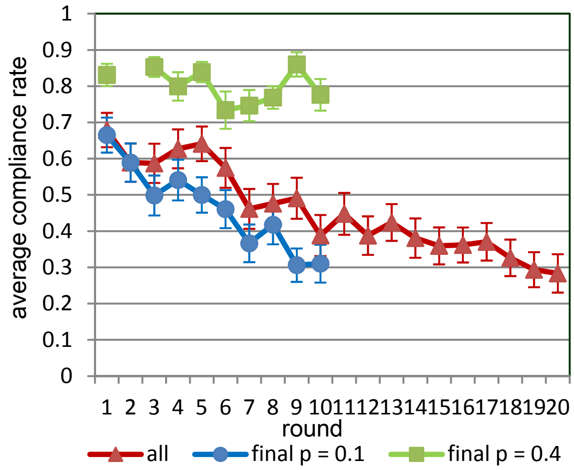
(a) The Tax-o1 treatment



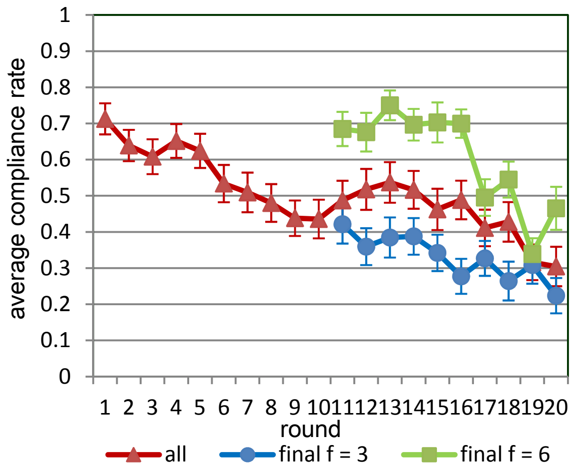
(b) The Tax-o2 treatment



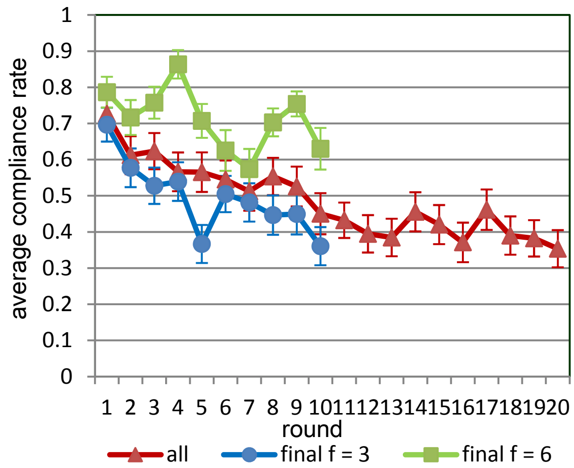
(c) The Audit-o1 treatment



(d) The Audit-o2 treatment



(e) The Fine-o1 treatment



(f) The Fine-o2 treatment

APPENDIX—Not to be published

Subjects' Instructions for the Tax-o1 Treatment

Subject ID number: ______

Welcome to the experiment. This is an experiment about individual economic decisions. Besides receiving a participation fee of NT$100 for appearing on time, if you follow the instructions closely and make your decisions carefully, you may earn an additional sufficient amount of money. All participants will be paid in cash at the end of the experiment. This study is funded by a grant from the National Science Council.

The experiment will last about 80 minutes.

The experiment includes two parts and each part consists of 10 rounds, for a total of 20 rounds. In the experiment your payoff will be represented by "points." At the end of the experiment, every 5 points can be exchanged for NT$1. In each round you and the other participants will be randomly assigned to groups of four, and each of the group members will have a code A, B, C, and D, respectively. You will only know about your own code, but not the codes of other people, and neither will you know who the other three members in your group are. When a new round starts, you will be randomly re-matched and reassigned the code.

At the beginning of each round, we will deposit randomly 70, 90, 110, and 130 points in the four group members' private accounts, respectively. When a new round begins, the four different amounts of points will be re-deposited randomly in the four group members' private accounts. You will only know the number of points deposited in your own private account, and this number of points may be different across rounds. Notice that the total points that we deposited into the four group members' private accounts is 400 points, with an average of 100 points.

We now clearly describe the two parts of the experiment.

Part I:

Every round of the Part I experiment contains two stages: Declaration and Check.

Stage One: Declaration. Everyone reports the number of points that is originally deposited in his or her private account. Twenty percent of the points that you report will be invested in the public account and the rest of the points will be retained in your private account. When you enter your decision in column (2), the computer will automatically calculate the numbers in columns (3) and (4).

Declaration Form
Round 1

Subject ID number: 27
Code: B

(1) The number of points deposited in your private account: 110
(2) The number of points you report: [ ]
[(2) ≤ (1) and must be zero or a positive integer.]
(3) The number of points you invest in the public account: [ ]
[= 20%×(2)]
(4) The number of points retained in your private account: [ ]
[= (1) – (3)]

Submit

The private account is owned by you, while the public account is owned together by the four members in the same group. The payoff you earn from your private account is the number of points left in your private account. Every point invested in the public account will earn each group member 0.5 points. The payoff you earn from the public account equals 0.5 times the total points invested by you and the other three members in your group.

Stage Two: Check. After everyone reports the number of points in his or her private account, the computer will randomly select a letter from A to J. If the letter that the computer selects is A, then the computer will check whether the number of points in his or her private account that A reported coincides with the number of points that we deposited into A's private account. If the letter the computer selects is B, then B will be checked; if the letter the computer selects is C, then C will be checked, and so on. Since there are only A, B, C, and D in the group, nobody will be checked if the letter that the computers selects is between E and J. The result of checking, which is similar to the following, will appear on your computer screen. The subject whose code is framed by red lines will be checked.

F C D I A E H J $\boxed{\text{B}}$ G

If the reported points of the person selected by the computer are fewer than the points we deposited in his or her private account, then we will deduct twenty percent of the difference from his or her payoff for this round. In addition, twice this amount will also be deducted. Hence, the total amount of the deduction is:

Deduction = 3×20%×(the amount of points deposited in the subject's private account – the amount of points declared by the subject)

The amount of the deduction in excess of his or her payoff this round will be subtracted from his or her payoff in previous rounds or later rounds if not enough.

Your Payoff in Each Round of Part I:

Your payoff in each round of Part I equals the amount of points remaining in your private account, plus the payoff you earned from the public account minus the deduction if there is any.

We provide several examples below to explain how your pay payoff is determined.

Example 1. Suppose that 90 points are originally deposited in your private account and you report 67 points. Therefore, you invest 13.4 points (=20% × 67) in the public account and the number of points retained in your private account is 76.6 points (= 90 – 13.4). Suppose that the other three group members report a total of 285 points, so that the total number of points reported by your group is 67+285 = 352 points and the total points invested in the public account by your group amount to 70.4 points (= 20% × 352). Suppose that you are not checked by the computer. Your payoff is 76.6 + 0.5×70.4 = 111.8 points.

Example 2. Suppose that 110 points are originally deposited in your private account and you report 83 points. Therefore, you invest 16.6 points (=20% × 83) in the public account and the number of points retained in your private account is 93.4 points (= 110 – 16.6). Suppose that the other three group members report a total of 236 points, so that the total points reported by your group is 83+236 = 319 points and the total points invested in the public account by your group are 63.8 points (= 20% × 319). Suppose that you are checked by the computer so your deduction is 3×20%×(110 – 83) = 16.2 points. Your payoff is 93.4 + 0.5×63.8–16.2 = 109.1 points.

After all participants have submitted their reporting decisions and the computer has finished the work of checking, a result report, which is similar to the following, will appear on your computer screen:

Result Report
Round 1

Subject ID number: 27
Code: B

The number of points originally deposited in your private account: 110
The total points deposited in the private accounts of the four members in your group: 400

The result of declaration:

	You	The total of the other three members	The total of your group
(1) The number of points reported by every group member:	83	236	319
(2) The number of points invested in the X account by every group member: [=20%×(1)]	16.6	47.2	63.8

The result of computer check: B is checked
Your payoff:

(3) The number of points retained in your private account:	93.4
(4) The number of points you earned from the public account: [=0.5× the total points invested in the public account by your group]	31.9
(5) Your deduction	16.2
(6) Your payoff this round = (3)+(4)–(5)	109.1

Part II

Each round of Part II contains three stages: Majority Voting, Declaration, and Check. The second stage, Declaration, and the third stage, Check, are the same as the first and the second stages in Part I. We now clearly explain the first stage, Majority Voting.

Stage One: Majority Voting. At the beginning of each round of the Part II experiment the four members in the same group vote on two proportions, 20 percent and 40 percent. The proportion that gains three or more votes wins. Then the computer will randomly determine whether to accept this outcome of the majority voting. If the computer accepts the outcome of the majority voting, then the winning proportion times the number of the points the subject declares will be deposited into the public account. If the computer rejects the outcome of the majority voting, then the computer will randomly select between the two proportions. If a tie occurs in the majority voting, the computer will also randomly select between the two proportions.

A table similar to the following will appear on everyone's computer screen:

Voting Sheet
Round 11
Subject ID number: 27
Code: C

Which proportion do you agree to be applied to all four members in your group? This proportion times the number of points a subject declares will be the amount of points that he or she deposits into the public good.

◯ 20% ◯ 40%

Notice that we will not inform you of the result of the majority voting, nor will we let you know whether the computer accepts the result of the majority voting or not. We will only notify you about the proportion that is eventually determined. We assure you that this proportion is definitely determined by the above process.

If 20 percent is eventually applied to your group, then your screen will display:

Round 11
20% of the amount you declare will be deposited in the public account.

If 40 percent is eventually applied to your group, then your screen will display:

Round 11
40% of the amount you declare will be deposited in the public account.

Stage Two and Stage Three: The second stage (declaration) and the third stage (check) are exactly the same as the first and the second stage in the Part I experiment, except that the deduction becomes:

Deduction= 3 × the proportion determined in Stage One×(the amount of points deposited in the subject's private account – the amount of points declared by the subject)

At the end of each round of Part II a result report, which is similar to that in Part I, will appear on your computer screen. Your total earnings from this experiment will be the sum of the earnings that you earn in each of the 20 rounds plus a participation fee of NT$100. Please do not talk to each other during the experiment. Your decisions and payoffs will be kept secret both during and after the experiment. There will be no link between your personal identity and the experimental data.

Quiz Questions

1. Suppose that in a round of Part I 110 points are deposited in your private account and you declare 76 points. How many points are left in your private account?
2. Suppose that in a round of Part I the four members in your group declare a total of 303 points. How much do you earn from the public account?
3. Suppose that in a round of Part I 90 points are deposited in your private account and you declare 82 points. Suppose also that in the Check stage you are selected by the computer. How many points will be deducted from your payoff?
4. Suppose that in the stage of Majority Voting of a round in Part II three members in your group vote for 20 percent and the other one member votes for 40 percent. Which of the following is correct?
    a. The amount of points that each member deposits in the public account are determined by the winning proportion through majority voting. Therefore, it is 20 percent in this example.
    b. The computer will step in only if a tie occurs. Therefore, in this example the computer will not intervene in the outcome of majority.
    c. The computer will randomly determine whether to accept the outcome of the majority voting or not. If the computer rejects the outcome of the majority voting, then it will randomly assign either 20 percent or 40 percent to this group.
    d. The computer will randomly determine whether to accept the outcome of the majority voting or not. If the computer rejects the outcome of the majority voting, then the computer will assign the proportion that gains fewer votes to the group. In this example it is 40 percent.

REFERENCES

Alm, James, Mark B. Cronshaw, and Michael McKee. 1993. "Tax Compliance with Endogenous Audit Selection Rules," *Kyklos,* 46(1), 27–45.

Alm, James, Betty R. Jackson, and Michael McKee. 1992a. "Institutional Uncertainty and Taxpayer Compliance," *American Economic Review,* 82(4), 1018–1026.

Alm, James, Betty R. Jackson, and Michael McKee. 1992b. "Estimating the Determinants of Taxpayer Compliance with Experimental Data," *National Tax Journal,* 45(1), 107–114.

Alm, James, Betty R. Jackson, and Michael McKee. 1993. "Tax Exchange, Collective Decision Institutions, and Tax Compliance," *Journal of Economic Behavior and Organization,* 22(3), 285–303.

Alm, James, Gary H. McClelland, and William D. Schulze. 1992. "Why Do People Pay Taxes?" *Journal of Public Economics,* 48(1), 21–38.

Alm, James, Gary H. McClelland, and William D. Schulze. 1999. "Changing the Social Norm of Tax Compliance by Voting," *Kyklos*, 52(2), 141–171.

Alm, James and Michael McKee. 2004. "Tax Compliance as a Coordination Game," *Journal of Economic Behavior and Organization,* 54(3), 297–312.

Alm, James, Michael McKee, and William Beck. 1990. "Amazing Grace: Tax Amnesties and Compliance," *National Tax Journal,* 43(1), 23–37.

Alm, James, Isabel Sanchez, and Ana de Juan. 1995. "Economic and Noneconomic Factors in Tax Compliance," *Kyklos,* 48(1), 3–18.

Andreoni, James, Brian Erard, and Jonathan Feinstein. 1998. "Tax Compliance," *Journal of Economic Literature* 36(2), 818–860.

Beck, Paul J., Jon S. Davis, and Woon-Oh Jung (1991). "Experimental Evidence on Taxpayer Reporting under Uncertainty," *Accounting Review*, 66, 535–558.

Becker, W., H. Büchner, and S. Sleeking. 1987. "The Impact of Public Transfer Expenditures on Tax Evasion: An Experimental Approach," *Journal of Public Economics,* 34(2), 243–252.

Clark, Jeremy, Lana Friesen, and Andrew Muller. 2004. "The Good, the Bad, and the Regulator: An Experimental Test of Two Conditional Audit Schemes," *Economic Inquiry,* 42(1), 69–87.

Collins, Julie H. and R. David Plumlee. 1991. "The Taxpayer's Labor and Reporting Decision: The Effects of Audit Schemes," *The Accounting Review,* 66(3), 559–576.

Dal Bó, Pedro, Andrew Foster, and Louis Putterman (2010), Institutions and Behavior: Experimental Evidence on the Effects of Democracy, *American Economic Review,* 100(5), December, 2205–2229.

Feld, Lars P. and Jean-Robert Tyran (2002). "Tax Evasion and Voting: An Experimental Analysis," *Kyklos*, 55(2), 197–222.

Gërxhani, Klarita and Arthur Schram (2006). "Tax Evasion and Income Source: A Comparative Experimental Study," *Journal of Economy Psychology*, 27, 402–422.

Imbens, Guido W. and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615–635.

Pommerehne, Werner W., Albert Hart and Bruno S. Frey. 1994. "Tax Morale, Tax Evasion, and the Choice of Tax Policy Instruments in Different Political Systems," *Public Finance/Finances Publiques*, 49 (Supplement), 52–69.

Putterman, Louis, Jean-Robert Tyran, and Kenju Kamei. 2011. "Public Goods and Voting on Formal Sanction Schemes," *Journal of Public Economics*, 95(9–10), 1213–1222.

Spicer, Michael W. and Lee A. Becker. 1980. "Tax Inequity and Tax Compliance: An Experimental Approach," *National Tax Journal,* 33(2), 171–175.

Spicer, Michael W. and J. Everett Thomas (1982). "Audit Probabilities and the Tax Evasion Decision: An Experimental Approach," *Journal of Economic Psychology*, 2, 241–245.

The Economist (2011). “Transport in Georgia: Spare a Penny? Taxes for Infrastructure,” *The Economist*, 401(8763), December 10, 2011, 32–34.

Reward and Punishment in Weakest-Link and Best-Shot Games: An Experimental Study

Li-Chen Hsu*

Department of Public Finance
National Chengchi University
64, Section 2, Zhi-Nan Road
Wenshan, Taipei 11605, Taiwan
Email: lchsu@nccu.edu.tw
Phone: 886-2-29387313

Abstract. We experimentally examine the effects of rewards and punishments on voluntary contributions in the weakest-link and best-shot game. Subjects voluntarily contribute some or all of their incomes to the public good and then after individual subjects' contributions are revealed, they decide to give reward or punishment points (if there are any) to others in their own groups. Experimental evidence from this paper shows that the instrument of punishments (but not rewards) is more effective in increasing voluntary contributions in the weakest-link game. On the contrary, the instrument of rewards (but not punishments) can significantly increase voluntary contributions in the best-shot game. In both games, using instruments of both rewards and punishments cannot improve voluntary contributions if either instrument has been adopted.

JEL classification: H41, C91

Keywords: privately provided public goods; punishment; reward; weakest-link; best-shot

* Financial support from the Ministry of Science and Technology of Taiwan (grant number: NSC 100–2410–H–004–065–MY2) is gratefully acknowledged.

1. INTRODUCTION

Attributing to its inherent characteristics of non-rivalry and non-excludability, free-riding is a common phenomenon in public good provision. This phenomenon was first introduced by Samuelson (1954) in his pioneered paper and was also examined and verified by many experimental studies (Andreoni, 1988, 1993, 1995; Isaac and Walker, 1988a, 1988b; Isaac, Walker, and Thomas, 1984; Isaac, Walker, and Williams, 1994). Although relying on public sector provision (and funded the public goods by taxes) is a method to resolve the free-riding problem, still enormous public goods are provided by the private sector alone and many of them are surrounded by in our daily life.

When encountering free-riding behavior of others, defection is a possible strategy to respond to those members who free ride. However, defection is a passive instrument of punishments and a pessimistic outcome of defection is an even lower level of public goods. To solve the inefficient problem in voluntary public good provision, the establishment of punishment mechanisms is a possible way, and this method has been studied by many experimental researchers. In the literature of punishments, punishment mechanisms are not only used in the discussion of how one can punish free riders, but are also used on issues of employers or hard-working workers dealing with co-workers who shirk (Carpenter, 2007; Fehr and Gächter, 2000; Fehr and Gächter, 2002; Fehr, Gächter, and Kirchsteiger, 1997; Fehr and Schmidt, 2007; Gächter, Renner, and Sefton, 2008; Gürerk, Irlenbusch, and Rockenbach, 2006; Herrmann, Thöni and Gächter, 2008; Nikiforakis and Normann, 2008; Reuben and Riedl, 2009; Sefton, Shupp, and Walker, 2007). Among these studies, the punishment mechanism proposed by Fehr and Gächter (2000) has been broadly discussed by subsequent examinations.

There are two treatments in Fehr and Gächter's (2000) experiment: the partner treatment and the stranger treatment. Each treatment consists of two conditions: the punishment condition and the no-punishment condition. In the no-punishment condition, each subject's earnings from the public good depend on the sum of the contributions made by all group members. The earnings from the public good plus the earnings from the subject's private good (that is, the rest of her endowment not contributed to the public good) are the subject's payoff. In the punishment condition, a second stage is added and the amount of the contribution made by each group member is declared at the end of stage one. After the declaration, everyone decides on the number of punishment points that she gives to each of the other group members. The punishment points that she gives out will reduce her own payoff in an increasing manner, that is, the marginal cost of punishments is increasing. Each punishment point she receives from other group members will reduce her first-stage payoff by one tenth. Because punishments are costly and any punishment points giving out are no longer change the outcome in the first stage, the theory predicts that everyone will not use the stick of punishments and hence the punishment mechanism has no effect in public good provision.

Fehr and Gächter (2000) found that in both partner and stranger treatments the levels of contributions in the punishment condition are significantly higher than those in the no-punishment condition, apparently contradicting the above theoretical prediction. Furthermore, when the opportunity of punishments is available, average contributions in the partner treatment approach the fully-cooperative level across rounds, and although average contributions in the stranger treatment is lower, they still increase across rounds and remain at a high level. On the contrary, when punishing others is not allowed, average contributions are low and move towards the fully non-cooperative level in both treatments.[1]

[1] Fehr and Gächter's (2000) punishment mechanism has two problems. First, the cost of punishments may exceed the gain from enforcing punishments, and thus the average may be lower with the punishment mechanism than without it. However, if subjects play more rounds, this inefficient problem could vanish.

In addition to the punishment mechanism, Andreoni, Harbaugh, and Vesterlund (2003) also investigate the effects of rewards on cooperation. They conducted treatments with punishments, rewards, and both punishments and rewards. In their experiment, every cent of punishments the subject gives to others will cost her one cent, and every cent of punishments she receives from others will reduce her payoff by five cents. Similar, every cent of rewards giving out will cost her one cent and every cent of rewards receiving from others will increase her payoff by five cents. They found that using the reward mechanism alone yields the lower average contributions than using the punishment mechanism alone. Using both rewards and punishments produces the highest average contributions.

Because Fehr and Gächter (2000) use a VCM (voluntary contribution mechanism) framework with a linear payoff function, most of the subsequent studies follow their setting. With a linear payoff function, a subject's earnings from the public good provision depend on the sum of individual group members' contributions. Hence, individual group members' contributions are perfect substitutions in public good provision. However, in reality many public goods are not in an additive form. The weakest-link and best-shot games proposed by Hirshleifer (1983) and Harrison and Hirshleifer (1989) are two examples. In the weakest-link game the amount of the public good is determined by the minimum of the contributions made by all group members. On the contrary, in the best-shot game the amount of the public good only depends on the member who makes the maximum contribution. The Nash equilibria of the weakest-link game are that all group members contribute the same amount, regardless of the size of the group. Brandts and Cooper's (2006) experimental finding suggests that the average contribution converges quickly to zero as long as the size of the group is four or above.

This paper examines the effects of rewards and punishments on voluntary contributions to public goods in the weakest-link and best-shot games. To our knowledge only Hamman, Rick, and Weber (2007) consider the punishment and reward mechanisms in the two games, and the reward and punishment increases or reduces every group member's earnings equally. We conducted four treatments for each game. In one treatment the instruments of both rewards and punishments are not available, in one treatment subjects are allowed to use both instruments, and in the other two treatments either rewards or punishments are available. Experimental evidence shows find that the instrument of punishments can increase contributions to the public good, but the instrument of rewards cannot. On the contrary, the instrument of rewards is more efficient in increasing contributions to the public good, but the instrument of punishments does not. In both games using both instruments is not necessary, since the level of contributions do not differ from that when only either reward mechanism or punishment mechanism is available.

The remainder of this paper is organized as follows. Section 2 describes the experimental design. Section 3 presents the theoretical predictions. Section 4 reports the results of the experiment, and Section 5 concludes.

2. EXPERIMENTAL DESIGN

Two types of public good provision games are examined in this study: a weakest-link game and a best-shot game. In the weakest-link game the amount of the public good is the

---

For instance, Gächter, Renner, and Sefton (2008) found that if the experiment persists for fifty rounds, then the punishment mechanism can increase average payoff. Sefton, Shupp, and Walker (2007) also found that the lower payoff phenomenon disappears in latter rounds of the experiment. Second, Herrmann, Thöni, and Gächter (2008, p. 1362) pointed out that Fehr and Gächter's (2000) punishment may produce an antisocial punishment outcome in which the subjects who free ride more punish their group members who behave more cooperatively.

minimum of the voluntary contributions by all group members, i.e., $G = \min_i(g_i)$, where g_i is the voluntary contribution by group member *i* and *G* is the amount of the public good for the group. By contrast, in the best-shot game the amount of the public good is the maximum of the voluntary contributions by all group members, i.e., $G = \max_i(g_i)$.

Four treatments were conducted for the weakest-link game. They were the weakest-none (weakest-link without rewards and punishments) treatment, the weakest-reward (weakest-link with rewards) treatment, the weakest-punish (weakest-link with punishments) treatment, and the weakest-reward-punish (weakest-link with both rewards and punishments) treatment. Symmetrically, four other treatments were also conducted for the best-shot game: the best-none (best-shot without rewards and punishments) treatment, the best-reward (best-shot with rewards) treatment, the best-punish (best-shot with punishments) treatment, and the best-reward-punish (best-shot with both rewards and punishments) treatment. Table 1 illustrates the framework of the experiment.

[Table 1 about here]

Four sessions were conducted for each treatment and nine subjects were recruited for each session, for a total of 288 subjects used in this study. All subjects were undergraduate students at National Chengchi University in Taiwan and none of them had previously participated in public goods experiments. All the experiments were conducted in the computer lab of the Department of Public Finance at the National Chengchi University. Subjects made decisions in twenty rounds. In each rounds the nine subjects in the same session were randomly and anonymously divided into three groups of n = 3. When each new round began, the nine subjects were randomly re-matched to prevent the reputation effect.

Each subject received an exogenous amount of income of ten points in each round, and he or she determined how much of the income to be allocated to a public account (a public good). The rest of the income was remained in the subject's private account (a private good). The return from the public good to *each* subject was two times the amount of the public good. The return from the private good to a subject was the amount of income remaining in his or her private account. When the instruments of rewards and punishments were absent, a subject's payoff per round was the sum of the returns from the public good and the private good. Hence, a subject's payoff per round was

$$\pi_i^{weakest-none} = y - g_i + k\min\{g_1, ..., g_n\} \tag{1}$$

in the weakest-none treatment and was

$$\pi_i^{best-none} = y - g_i + k\max\{g_1, ..., g_n\} \tag{2}$$

in the best-none treatment, in which y = 10 was the income per round and k = 2 was the marginal return of the public good to each group member. Notice that k must be greater than one to ensure that everyone has an incentive to contribute to the public good. At the end of each round, each subject was informed of a result report, which contained the information regarding the amount of points that each of the three members allocated to the public account, the minimum (in the weakest-none treatment) or maximum (in the best-none treatment) of them, the return from the public account, the return from the subject's private account, and the subject's payoff for this round.

In the remaining six treatments with either rewards or punishments or both, each round of the experiment consisted of two stages. Subjects were informed of the contents of the two stages at the beginning of the experiment. The first stage of the six treatments was the same as the weakest-none and best-none treatments, but the second stage differed. More specifically, in the weakest-reward and best-reward treatments subjects were provided with the opportunity of rewarding other group members. They decided on whether or not to reward the other group members and how many reward points to give out to each of them. For every reward point a subject gave out, his or her payoff was reduced by one point, and for each reward point a subject received, his or her payoff increased by two points. Similarly, in the weakest-punish and best-punish treatments subjects were provided with the opportunity of punishing other group members. Every punishment point giving out to other group members reduced a subject's payoff by one point, and every punishment point receiving from other group members reduced a subject's payoff by two points. Finally, in the weakest-reward-punish and best-reward-punish treatments the opportunities of both rewarding and punishing other group members were available to each subject. The impacts of the reward/punishment points giving out to and receiving from other group members on subjects' payoffs were the same as above.

In sum, a subject's payoff at the end of stage two in each round was

$$\pi_i^w = y - g_i + k\min\{g_1, ..., g_n\} - \sum_{j\neq i} r_i^j - \sum_{j\neq i} p_i^j + c\sum_{j\neq i} r_j^i - d\sum_{j\neq i} p_j^i \tag{3}$$

in the three weakest-link treatments, and was

$$\pi_i^b = y - g_i + k\max\{g_1, ..., g_n\} - \sum_{j\neq i} r_i^j - \sum_{j\neq i} p_i^j + c\sum_{j\neq i} r_j^i - d\sum_{j\neq i} p_j^i \tag{4}$$

in the three best-shot treatments. In equations (3) and (4) r_i^j was the number of reward points that *i* gave out to *j*, p_i^j was the number of punishment points that *i* gave out to *j*, r_j^i was the number of reward points that *j* gave out to *i*, p_j^i was the number of punishment points that *j* gave out to *i*, and $c = d = 2$. Notice that $r_i^j = r_j^i = 0$ for all *i*, *j* in the weakest-punish and best-punish treatments and $p_i^j = p_j^i = 0$ for all *i*, *j* in the weakest-reward and best-reward treatments.

At the end of each round of the experiment, each subject received a result report, in which the result from the first stage was replicated and the result from the second stage was declared. The latter information includes the following information: the total numbers of reward and punishment points (if there were any) that each group member gave out, the total numbers of reward and punishment points (if there were any) that each group member received, and the subject's payoff for this round. The subject's payoff from the experiment was the sum of his or her payoff from all twenty rounds plus a participation fee of NT$100.

In all sessions, subjects were given written instructions in Chinese. The experimenter read the instructions aloud and answered any questions raised by the subjects. After reading the instructions, subjects were required to answer three (in the weakest-none and best-none treatments) to four (in all other treatments) questions in relation to the calculation of payoffs. The experiment would not start until everyone had answered all questions correctly. Each session lasted about 70 minutes. The average payoff (including a participation fee of NT$100) for all participants was NT$442.15 (with a standard deviation of NT$84.84, a maximum of NT$673, and a minimum of NT$205).[2]

3. THEORETICAL PREDICTIONS

To have a clear-cut theoretical prediction of subjects' decisions, it is assumed that all subjects were self-interested and maximized their own monetary payoffs, and that this feature was common knowledge to all subjects. Let us start with the weakest-none treatment. By equation (1), because the level of the public good is the minimum of the contributions by members in the group, any group member unilaterally increases his or her contribution to the public good will only reduce his or her own payoff and others' payoffs will remain unchanged. Hence, any sets of contributions in which every group member contributes the same amount of income to the public good are Nash equilibria, i.e., $g_i = g_j$ for all $j \neq i$ and $0 \le g_i \le y$. Apparently, among these Nash equilibria $g_i = g_j = y$ is the only Pareto efficient allocation. In this equilibrium the payoff for every group member is $\pi_i = 2y = 20$, which is also the maximum payoff that anyone can earn from the weakest-none treatment. Apparently the minimum payoff is zero, which is resulted from the outcome that the subject contributes all of the income *y* to the public good, but at least one other group members contributes nothing.

Next, by equation (2), because the level of the public good in the best-none treatment is the maximum of the contributions by members in the group, if any member *i* contributes more income to the public good than do others, then other members reducing their own contributions to zero can increase their own payoffs. Hence, in the equilibrium only one member contributes to the public good. Furthermore, because $k > 1$, for person *i* who makes a positive contribution, his or her payoff is maximized if he or she contributes all of the income to the public good. Because in the experiment every subject is endowed with the same amount of income, there are multiple Nash equilibria in the best-none treatment, and in each Nash equilibrium only one group member contributes all of his or her income to the public good, i.e., $g_i = y$ and $g_j = 0$ for all $j \neq i$. Obviously all of the Nash equilibria are Pareto efficient. In equilibrium the payoff for the member who contributes all of the income is $\pi_i = 2y = 20$ and each of the other group members earns $\pi_j = y + 2y = 30$.

Now let us find out the subgame perfect Nash equilibria for the treatments with either rewards or punishments or both. We can solve the game backwards by starting with the second stage of the game. When the second stage arrives, in spite of the outcome in the first stage, the best strategy for a subject is to give out zero reward/punishment points, because punishing or rewarding other group members will only reduce the subject's own payoff. This solution can also be seen from the differentiation of equations (3) and (4) with respect to r_i^j and p_i^j respectively. That is, $\partial\pi_i^w / \partial r_i^j = \partial\pi_i^w / \partial p_i^j = \partial\pi_i^b / \partial r_i^j = \partial\pi_i^b / \partial p_i^j = -1$, which implies that in the second stage the best strategy for each subject *i* is to choose $r_i^j = p_i^j = 0$. Given that in the second stage no one will punish or reward anyone else, in the first stage subjects simply maximize the payoffs earned from the public good and the private good, and it is evident that the equilibria are the same as those in the weakest-none and best-none treatments. That is, in the weakest-reward, weakest-punish, and weakest-reward-punish treatments all subjects in the same group contribute the same amount of the income to the public good, and in the best-reward, best-punish, and best-reward-punish treatments one member contributes all of his or her income and the other two members

[2] When these sessions were conducted, the exchange rate between the NT (New Taiwan) dollar and the US dollar was about 30:1. The part-time hourly wage rate for an undergraduate student in Taiwan is about NT$120.

contribute nothing to the public good.

In the following section, the above equilibrium predictions will be tested. In addition, we will investigate the subjects' reward and punishment decisions conditional on their contribution behavior.

4. EXPERIMENTAL RESULTS

We conducted thirty-two sessions in October and November of 2013 in the computer lab of the Department of Public Finance at National Chengchi University in Taiwan. Table 2 reports that of the 288 subjects recruited, 70.83 percent of them were female (standard deviation = 0.4553), they had been in the university for an average of 2.52 years (standard deviation = 1.0850), the average age was 19.93 years (standard deviation = 1.2574), and 84.38 percent of them had taken economics course(s) (standard deviation = 0.3637). There are no significant differences in these characteristics between treatments. Hence, any differences in subjects' contribution or punishment/reward behavior are not result from these inherent characteristics of subjects.

[Table 2 about here]

4.1. A General Look at Contribution and Reward/Punishment Behavior

[Tables 3, 4, 5, and 6 about here]

Tables 3 through 6 summarize the data resulting from the whole 20 rounds and the last 10 rounds in each treatment. Round averages and standard errors of contributions to the public good are depicted in Figures 1 and 2.[3]

[Figures 1 and 2 about here]

Several observations arise from the preliminary statistics. First, by looking at Figures 1 and 2, it is observed that average contributions to the public good reveal no downward trends across rounds in all four weakest-link treatments. The average contributions to public good are about the same across rounds in the weakest-none and weakest-reward treatments, and they steadily increase in the weakest-punish and weakest-reward-punish treatments as subjects play more rounds. By contrast, average contributions to the public good fluctuate across rounds and reveal slight downward trends in the four best-shot treatments.

Second, by looking at Tables 4 and 6, it shows that average contributions to the public good in the last ten rounds are higher in the weakest-link treatments than in the best-shot treatments. Average contributions to the public good in the last ten rounds is 5.68 points in the weakest-none treatment, 6.14 points in the weakest-reward treatment, 7.16 points in the weakest-punish treatment, and 7.15 points in the weakest-reward-punish treatments, which are correspondently higher than the 3.70 points in the best-none treatment, 5.01 points in the best-reward treatment, 4.24 points in the best-punish treatment, and 4.67 points in the best-reward-punish treatment.

Third, there are significantly fewer subjects making zero contributions to the public good in the weakest-link treatments than in the best-shot treatments. In the last ten rounds, an average of 12.5 percent of the subjects in the weakest-none treatment, 5.83 percent of the

---

[3] The standard error of the sample mean is calculated as $\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1)}\Big/\sqrt{n}$, where x_i is the value of the observation, $\bar{x}$ is the sample mean, and n is the sample size.

subjects in the weakest-reward treatment, and zero percent of the subjects in both the weakest-punish and weakest-reward-punish treatments contribute nothing to the public good, while these magnitudes are 41.11 percent in the best-none treatment, 28.06 percent in the best-reward, 30 percent in the best-punish, and 37.5 percent in the best-punish-reward treatment.

Fourth, by contrast, more subjects contribute the entire income to the public good in the best-shot treatments than in the weakest-link treatments. In the last ten rounds, an average of 20.83 percent of the subjects in the best-none treatment, 29.44 percent of the subjects in the best-reward treatment, 18.33 percent of the subjects in the best-punish treatment, and 28.06 percent of the subjects in the weakest-reward-punish treatment contribute ten points to the public good, while these magnitudes are only 4.72 percent in the weakest-none treatment, 26.39 percent in the weakest-reward, 9.17 percent in the weakest-punish, and 12.78 percent in the weakest-punish-reward treatment.

Fifth, in both weakest-link and best-shot treatments, when subjects are only provided with either the opportunity of rewarding others or the opportunity of punishing others, more subjects are willing to reward others than to punish others. On the contrary, when subjects have both the opportunities of rewarding and punishing other, they use punishments more often. For instance, in the last ten rounds an average of 28.33 percent of the subjects in the weakest-reward treatment and 38.06 percent of the subjects in the best-reward treatment give out reward points to others, while an average of 20.28 percent of the subjects in the weakest-punish and 17.5 percent of the subjects in the best-punish treatment ever punished others. However, in the last ten rounds of the weakest-reward-punish treatment, on average 22.22 percent of the subjects have ever punished others and 15.28 percent of the subjects have ever rewarded others. The two corresponding numbers are 19.44 percent and 13.89 percent in the best-reward-punish treatment.

4.2. Verifying the Theoretical Predictions

We now verify the theoretical predictions for the weakest-link and the best-shot games. Recall that in the stage of contributions of the four weakest-link treatments any combinations of group members' contributions in which all group members contribute the same amount of the income to the public good are Nash equilibria. Among these Nash equilibria, the equilibrium that every group member contributes his or her whole income is also Pareto efficient. In the four best-shot treatments, the only Nash equilibrium is that only one group member contributes and contributes all of his or her income to the public good. In the stage when the opportunities of punishments and rewards are available, if there are any, the Nash equilibrium for both weakest-link and best-shot treatments is that every subjects should give out zero punishment and reward points.

The experimental evidence does not support the above Nash equilibria. Let us start with the stage of contributions. In all 20 rounds of the weakest-none treatment, on average only 15 percent of the groups satisfy the Nash equilibrium prediction that all three group members contribute the same amount to the public good. By looking at only the last 10 rounds, this magnitude increases, but still is about 20.8 percent. More specifically, almost no groups meet the Nash equilibrium prediction of equal contributions in the first 10 rounds, and although more groups meet this prediction in latter rounds, in the last round there are still only 41.7 percent of groups satisfying the Nash equilibria, far below 100 percent predicted by the theory. As for the Pareto efficient allocation that all group members contribute all of their incomes to the public good, only one group in the last round meets this standard.

The opportunity of rewarding others only slightly improves the achievement of Nash equilibria. Of all the 20 rounds in the weakest-reward treatment, on average 22.08 percent of the group meet the Nash equilibrium prediction of equal contributions. By looking at only the last 10 rounds, this number increases to 30 percent. Furthermore, starting round 7 two to three out of nine groups meet the Pareto efficient criterion and across all 20 rounds on average 16.7 percent of the groups behave efficiently.

By contrast, the Nash equilibria barely occur in the weakest-punish treatment. By looking at the whole 20 rounds and the last 10 rounds, only on average only 9.2 percent of the groups meet the equal contributions standard. The Pareto efficient outcome appears in only one group in round 17.

In the weakest-reward-punish treatment, on average 15 percent of the groups make equal contributions in all 20 rounds, and this number increases to 19.2 percent by looking at only the last 10 rounds. Over the whole 20 rounds, a total of only four groups (one group in round 18, one group in round 19, and one group in round 20), or 1.7 percent of the groups, make the efficient contribution.

We now go to the Nash equilibrium prediction for the four best-shot treatments. In the best-none treatment, there are 24.17 percent of the groups in all 20 rounds and 27.5 percent of the groups in the last 10 rounds with only one group member contributing to the public good. However, these numbers drop to 10.42 percent and 12.5 percent if we confine to the Nash equilibrium that this one member contributes exactly 10 points to the public good.

In the best-reward treatment, on average only 10.83 percent of the groups in the 20 rounds and 16.67 percent of groups in the last 10 rounds have only group member contributes to the public good. The number of groups meets the Nash equilibrium is even lower: on average there are only 7.5 percent of the groups in the 20 rounds and 8.33 percent of the groups in the last 10 rounds with one group member contributing exactly 10 points to the public good.

As compared with the best-reward treatment, in the best-punish treatment there are more groups with only one member giving to the public good, but still quite few groups satisfying the Nash equilibrium. On average 20.42 percent of the groups in the 20 rounds and 20.83 percent of the groups in the last 10 rounds have only one member giving to the public goods. However, these numbers are only 8.75 percent and 7.5 percent if we stick to the Nash equilibrium prediction of giving 10 points.

The best-reward-punish treatment does not yield more Nash equilibrium cases. On average there are 17.08 percent of groups in the 20 rounds and 21.67 percent of the groups in the last 10 rounds with only one member giving to the public good. Only 9.58 percent of the groups in the 20 rounds and 14.17 percent of the groups in the last 10 rounds satisfy the Nash equilibrium prediction.

[Figures 3 and 4 about here]

Figures 3 and 4 depict the ratios of subjects giving out reward or punishment points across rounds in various treatments. Many subjects are willing to reward other group members in the beginning of the experiment. Figure 3 shows that there are as high as 63.89 percent of the subjects in the weakest-reward treatment and 69.44 percent of the subjects in the best-reward treatment giving out reward points to others in round 1. The levels of reward points giving out drop soon in early stage of the two treatments, and down to a low level in the final several rounds. However, in round 20 still 25 percent of the subjects in the best-reward treatment and 16.67 percent of the subjects in the weakest-reward-punish treatment give out reward points. The ratios of giving out reward points are lower in both the weak-reward and best-reward-punish treatments, but there are still respectively 13.89 percent of subjects in the two treatments rewarding others.

Figure 4 shows that the ratios of subjects punishing other group members start high in the four treatments with the opportunity of punishments. Though it reveals a downward trend in all four treatments, and even in round 18 only 5.56 percent of the subjects in the best-reward-punish treatment use the tool of punishments, this percentage rises soon in round 19, and ends up of 22.22 percent in round 20. According to Tables 3 and 5, over the entire 20 rounds on average 26.94 percent of the subjects in the weakest-punish treatment, 24.72 percent of the subjects in the weakest-reward-punish treatment, 22.64 percent of the subjects in the best-punish treatment, and 21.67 percent of the subjects in the best-reward-punish treatment have ever punished other group members. According to Tables 4 and 5, these magnitudes drop in rounds 11–20, but still way above 0 percent. More specifically, on average 20.28 percent of the subjects in the weakest-punish treatment, 22.22 percent of the subjects in the weakest-reward-punish treatment, 17.5 percent of the subjects in the best-punish treatment, and 19.44 percent of the subjects in the best-reward-punish treatment have ever punished others.

We summarize the above observations in Result 1 as follows:

Result 1: In the four weakest-link treatments, the subgame perfect Nash equilibrium prediction of equal contributions and zero punishment/reward points does not hold. In the four best-shot treatments, the subgame perfect Nash equilibrium prediction of only one member giving 10 points and zero punishment/reward points does not hold, either.

4.3. The Effectiveness of Rewards and Punishments

The feature of the weakest-link game is that everyone's payoff is determined by the minimum of the contributions to the public good made by all group members. As a consequence, everyone must make the same contribution to the public good, and the one who contributes fewer than others will drag down all other group members' payoff. On the contrary, the feature of the best-shot game is that everyone's payoff is determined only by the maximum of the contributions to the public good made by all group members. Hence, exactly only one member making contributions is enough and this member should make the maximum possible contribution, that is, all of her income. Any contributions of an equal or lower amount made by other group members are simply a waste. How effective in raising contributions are the instruments of punishments and rewards in the two games with such divergent features?

Let us start with the four weakest-link treatments. Figure 1 shows that average contributions in the weakest-punish and weakest-reward-punish treatments increase steadily across rounds. While average contributions in the weakest-reward treatment remain in the range of 5 to 6.5 points across rounds, they are lower than the average contributions in the weakest-punish and weakest-reward-punish treatments in every round of rounds 10–20, and even lower than the average contributions in the weakest-none treatment in the final two rounds.

By using the session-level data, two-sided Mann-Whitney U tests show that there are no significant differences in contributions between the weakest-reward and weakest-none treatments, between the weakest-punish and weakest-none treatments, between the weakest-reward and weakest-reward-punish treatments, and between the weakest-punish and weakest-reward-punish treatments in any round of the experiment, seemingly suggesting that neither rewards no punishments have significant effects in improving contributions. However, by using the subject-level data, two-sided Mann-Whitney U tests show that contributions are significantly higher in the weakest-punish treatment than in the weakest-none treatment in every round of rounds 7–18 and by looking at the average of rounds 6–10, 11–15, 1–10, 11–20, and 1–20 (all $p \leq 0.05$). Contributions are also significantly higher in the weakest-reward-punish treatment than in the weakest-none treatment in every round of rounds 8, 12, 13, 17, 18, and 20 and by looking at the average of rounds 11–15, 16–20, and 1–20 (all $p \leq 0.05$). A comparison between the weakest-punish treatment and the weakest-reward-punish treatment shows that contributions in every round of the two treatments do not differ significantly (even by using the significance level 0.1). These findings suggest that the instrument of punishments raises contributions to the public good significantly, and using the additional instrument of rewards does not help improve contributions further. The latter result is also verified by the comparison between the weakest-reward treatment and the weakest-none treatment: contributions in every round of the weakest-reward treatment do not differ significantly from those in the weakest-none treatment (even by using $p = 0.1$).

We now look at the four best-shot treatments. By using the group-level data, two-sided Mann-Whitney U tests show that contributions are higher in the best-reward treatment than in the best-none treatment in rounds 1–5, 8–10, 13, and 19, and by looking the average of rounds 1–5, 6–10, 11–15, 16–20, 1–10, 11–20, and 1–20 ($p < 0.05$). However, contributions in the best-punish treatment are significantly higher than average contributions in the best-none treatment in only rounds 1, 2, and 13. Furthermore, contributions in the best-reward-punish treatment do not differ significantly from contributions in both the best-reward and best-punish treatments. These findings suggest that the instrument of rewards significantly improves contributions, but the instrument of punishments does not. Using instruments of both rewards and punishments does not improve contributions over using either instrument. Results by using the subject-level data are similar to those by using the group-level data.

We summarize the above observations in Result 2 in the following:

Result 2: In the weakest-link game, the instrument of punishments significantly raises contributions, but the instrument of rewards does not. On the contrary, the instrument of rewards significantly improves contributions, but the instrument of punishments does not. Employing both instruments cannot yield more contributions than employing only the instrument of punishments (in the weakest-ling game) and the instrument of rewards (in the best-shot game).

5. CONCLUSION

This paper experimentally examines the effectiveness of rewards and punishments on voluntary public good provisions in the weakest-link and best-shot games. In the weakest-link game any group member's payoff is determined by the minimum of the contributions made by all group members, while in the best-shot game any group member's payoff only depends on the maximum of the contributions made by the group. The subgame perfect Nash equilibria for the weakest-link game is that all group members should make the same amount of contributions. On the contrary, in the best-shot game only one group member making contributions is enough and she should contributes all of her income to the public good. Furthermore, when the instruments of rewards or punishments or both are available, in both the weakest-link and best-shot games any player should not use them, because using them will simply lower her own payoff.

Experimental evidence of this paper does not support the above theoretical predictions. In the weakest-link treatments, regardless of the availability of rewards or punishments or both, few groups satisfy the Nash equilibrium prediction that all group members make the same contribution. Similarly, quite few groups in the best-shot treatments have only one group member contributing her whole income to the public good. In both the weakest-link and best-shot treatments, there are a fair amount of subjects punish or reward other group members, even in the end round of the experiment. This paper also finds that the instrument of punishments significantly increases contributions to the public good in the weakest-link game and the instrument of rewards is not effective in public good provision. On the contrary, the instrument of rewards significantly increases contributions, while the instrument of punishments does not.

Table 1. Framework of the experiment

Treatment	Multiple of public good (k)	Multiple of reward (m)	Multiple of punishment (n)
Weakest-none	2	–	–
Weakest-reward	2	2	–
Weakest-punish	2	–	2
Weakest-reward-punish	2	2	2
Best-none	2	–	–
Best-reward	2	2	–
Best-punish	2	–	2
Best-reward-punish	2	2	2

Table 2. Individual characteristics

Treatment	Female	Class	Age	Taken economics course(s)
Weakest-none	0.8056 (0.4014)	3.1667 (1)	20.4722 (1.1585)	0.8056 (0.4014)
Weakest-reward	0.6667 (0.4781)	2.7778 (0.9292)	20.2222 (1.1492)	0.9167 (0.2803)
Weakest-punish	0.75 (0.4392)	2.5556 (1.0809)	19.9722 (1.1081)	0.8056 (0.4014)
Weakest-reward-punish	0.6389 (0.4871)	2.5556 (1.2293)	19.9722 (1.0820)	0.9167 (0.2803)
Best-none	0.6389 (0.4871)	2.3056 (1.0370)	19.7222 (1.1616)	0.8611 (0.3507)
Best-reward	0.75 (0.4392)	2.2778 (1.0032)	19.6389 (1.3342)	0.8056 (0.4014)
Best-punish	0.6944 (0.4672)	2.1667 (1.0556)	19.5833 (1.1307)	0.8611 (0.3507)
Best-reward-punish	0.7222 (0.4543)	2.3611 (1.0731)	19.8889 (1.6865)	0.8611 (0.3507)
All	0.7083 (0.4553)	2.5208 (1.0850)	19.934 (1.2574)	0.8542 (0.3536)

Standard deviations are in parentheses.

Table 3. Descriptive statistics for the weakest-link treatments from rounds 1–20

	Weakest-none	Weakest-reward	Weakest-punish	Weakest-reward-punish
(1) Average of points invested in the public good	5.5722 (1.9217)	6.1514 (2.2284)	6.55 (1.3735)	6.4583 (1.4438)
(2) Average of the minimum of the points invested in the public good by group members	4.3625 (2.1961)	5.0333 (2.3599)	5.5292 (1.1739)	5.675 (1.0864)
(3) Percentage of subjects investing zero points in the public good	9.03% (0.2326)	3.33% (0.0548)	0.28% (0.0167)	0% (0)
(4) Percentage of subjects investing ten points in the public good	5.83% (0.1137)	24.31% (0.3429)	7.78% (0.2051)	7.78% (0.1873)
(5) Average earnings (points)	13.1528 (2.9567)	14.8736 (3.9610)	13.0417 (1.8547)	14.0292 (1.0816)
(6) Percentage of subjects giving out reward points to others	–	34.31% (0.2510)	–	19.86% (0.2562)
(7) Percentage of subjects receiving reward points from others	–	36.67% (0.2670)	–	23.19% (0.2159)
(8) Percentage of subjects giving out punishment points to others	–	–	26.94% (0.2132)	24.72% (0.2330)
(9) Percentage of subjects receiving punishment points from others	–	–	25.83% (0.1899)	25.83% (0.1528)
(10) Average of the reward points giving out to others	–	2.0534 (1.6234)	–	1.2390 (0.3173)
(11) Average of the reward points receiving from others	–	2.0108 (1.2246)	–	1.1203 (0.1915)
(12) Average of the punishment points giving out to others	–	–	1.6608 (1.3764)	1.3715 (0.3935)
(13) Average of the punishment points receiving from others	–	–	1.9108 (1.3512)	1.4189 (0.4070)
(14) Average points invested in the public good conditional on having given out reward points to others	–	6.2815 (2.1118)	–	6.2631 (1.6764)
(15) Average points invested in the public good conditional on having received reward points from others	–	6.8634 (2.2066)	–	6.4824 (1.5278)
(16) Average of points invested in the public good conditional on having given out punishment points to others	–	–	6.6410 (1.3619)	6.6576 (1.4559)
(17) Average of points invested in the public good conditional on having received punishment points from others	–	–	5.6674 (1.8389)	5.7750 (1.3295)

The observations are the averages of all subjects' average choices over the period specified. The numbers in parentheses are the standard deviations of all subjects' average choices over that period.

Table 4. Descriptive statistics for the weakest-link treatments from rounds 11–20

	Weakest-none	Weakest-reward	Weakest-punish	Weakest-reward-punish
(1) Average of points invested in the public good	5.6833 (2.6911)	6.1417 (2.6511)	7.1611 (1.5540)	7.1528 (1.6811)
(2) Average of the minimum of the points invested in the public good by group members	4.7583 (2.8956)	5.225 (2.8498)	6.3333 (1.4456)	6.525 (1.4226)
(3) Percentage of subjects investing zero points in the public good	12.5% (0.3018)	5.83% (0.1105)	0% (0)	0% (0)
(4) Percentage of subjects investing ten points in the public good	4.72% (0.1444)	26.39% (0.4311)	9.17% (0.2322)	12.78% (0.2503)
(5) Average earnings (points)	13.8333 (3.5588)	15.2472 (4.6445)	14.2972 (2.4918)	15.0722 (1.6438)
(6) Percentage of subjects giving out reward points to others	–	28.33% (0.2893)	–	15.28% (0.2613)
(7) Percentage of subjects receiving reward points from others	–	30.56% (0.3180)	–	18.06% (0.2573)
(8) Percentage of subjects giving out punishment points to others	–	–	20.28% (0.2249)	22.22% (0.2642)
(9) Percentage of subjects receiving punishment points from others	–	–	19.17% (0.1857)	23.33% (0.2042)
(10) Average of the reward points giving out to others	–	2.4693 (2.6986)	–	1.2773 (0.3784)
(11) Average of the reward points receiving from others	–	2.5462 (1.8495)	–	1.1405 (0.2229)
(12) Average of the punishment points giving out to others	–	–	1.645 (1.6213)	1.3553 (0.4506)
(13) Average of the punishment points receiving from others	–	–	2.1008 (2.3153)	1.3747 (0.5002)
(14) Average points invested in the public good conditional on having given out reward points to others	–	6.7796 (2.7324)	–	7.9062 (1.3913)
(15) Average points invested in the public good conditional on having received reward points from others	–	7.7727 (2.3715)	–	8.1191 (1.4752)
(16) Average of points invested in the public good conditional on having given out punishment points to others	–	–	7.8167 (1.2378)	7.7382 (1.6909)
(17) Average of points invested in the public good conditional on having received punishment points from others	–	–	6.6569 (1.8221)	6.5706 (1.6307)

The observations are the averages of all subjects' average choices over the period specified. The numbers in parentheses are the standard deviations of all subjects' average choices over that period.

Table 5. Descriptive statistics for the best-shot treatments from rounds 1–20

	Best-none	Best-reward	Best-punish	Best-reward-punish
(1) Average of points invested in the public good	3.7292 (2.0402)	5.4514 (2.2368)	4.3389 (2.3653)	5.0486 (2.4601)
(2) Average of the maximum of the points invested in the public good by group members	6.9958 (0.9582)	8.6583 (0.6611)	7.2792 (1.0944)	8.3667 (1.0414)
(3) Percentage of subjects investing zero points in the public good	35.83% (0.2982)	20.56% (0.2137)	28.89% (0.3064)	30% (0.2888)
(4) Percentage of subjects investing ten points in the public good	17.08% (0.1730)	26.81% (0.2525)	16.25% (0.2288)	26.11% (0.2824)
(5) Average earnings (points)	20.2625 (1.5042)	23.0972 (2.0494)	18.3569 (1.5082)	20.0444 (2.0474)
(6) Percentage of subjects giving out reward points to others	–	47.64% (0.2579)	–	18.19% (0.2303)
(7) Percentage of subjects receiving reward points from others	–	49.44% (0.2239)	–	18.75% (0.1903)
(8) Percentage of subjects giving out punishment points to others	–	–	22.64% (0.2609)	21.67% (0.2290)
(9) Percentage of subjects receiving punishment points from others	–	–	26.94% (0.1338)	22.5% (0.1417)
(10) Average of the reward points giving out to others	–	2.1930 (1.3434)	–	1.6841 (0.9512)
(11) Average of the reward points receiving from others	–	2.3022 (0.9293)	–	1.6707 (0.5299)
(12) Average of the punishment points giving out to others	–	–	2.3050 (1.4698)	2.5424 (2.0991)
(13) Average of the punishment points receiving from others	–	–	2.2662 (1.0169)	2.7716 (2.0474)
(14) Average points invested in the public good conditional on having given out reward points to others	–	5.4505 (2.6134)	–	5.8176 (2.9922)
(15) Average points invested in the public good conditional on having received reward points from others	–	7.6655 (2.0605)	–	8.0710 (2.0000)
(16) Average of points invested in the public good conditional on having given out punishment points to others	–	–	5.4262 (2.2979)	6.3605 (1.9036)
(17) Average of points invested in the public good conditional on having received punishment points from others	–	–	2.8556 (2.1625)	2.4240 (2.4818)

The observations are the averages of all subjects' average choices over the period specified. The numbers in parentheses are the standard deviations of all subjects' average choices over that period.

Table 6. Descriptive statistics for the best-shot treatments from rounds 11–20

	Best-none	Best-reward	Best-punish	Best-reward-punish
(1) Average of points invested in the public good	3.7028 (2.3112)	5.0056 (2.7406)	4.2444 (2.5634)	4.6722 (2.8460)
(2) Average of the maximum of the points invested in the public good by group members	7.025 (1.1480)	8.55 (0.9858)	7.0083 (1.3861)	8.0917 (1.4155)
(3) Percentage of subjects investing zero points in the public good	41.11% (0.3214)	28.06% (0.2617)	30% (0.3355)	37.5% (0.3384)
(4) Percentage of subjects investing ten points in the public good	20.83% (0.2103)	29.44% (0.3242)	18.33% (0.2408)	28.06% (0.3078)
(5) Average earnings (points)	20.3472 (1.8101)	22.9889 (2.4650)	18.3556 (2.2733)	19.8806 (2.6730)
(6) Percentage of subjects giving out reward points to others	–	38.06% (0.2817)	–	13.89% (0.2346)
(7) Percentage of subjects receiving reward points from others	–	40.28% (0.2699)	–	15% (0.1558)
(8) Percentage of subjects giving out punishment points to others	–	–	17.5% (0.2489)	19.44% (0.2437)
(9) Percentage of subjects receiving punishment points from others	–	–	22.5% (0.1680)	20.28% (0.1699)
(10) Average of the reward points giving out to others	–	2.1907 (1.9677)	–	1.5394 (0.7493)
(11) Average of the reward points receiving from others	–	2.0983 (1.1075)	–	1.7106 (0.6727)
(12) Average of the punishment points giving out to others	–	–	2.4093 (1.5923)	2.6774 (2.5522)
(13) Average of the punishment points receiving from others	–	–	1.9683 (1.0030)	2.9143 (2.1614)
(14) Average points invested in the public good conditional on having given out reward points to others	–	4.7192 (3.2752)	–	5.6667 (3.7211)
(15) Average points invested in the public good conditional on having received reward points from others	–	7.5893 (2.4610)	–	8.4394 (2.6440)
(16) Average of points invested in the public good conditional on having given out punishment points to others	–	–	5.2074 (3.0737)	6.2922 (2.7679)
(17) Average of points invested in the public good conditional on having received punishment points from others	–	–	2.6865 (2.4161)	2.2708 (2.9749)

The observations are the averages of all subjects' average choices over the period specified. The numbers in parentheses are the standard deviations of all subjects' average choices over that period.

Figure 1. Average contributions to the public good in the weakest-link treatments by round

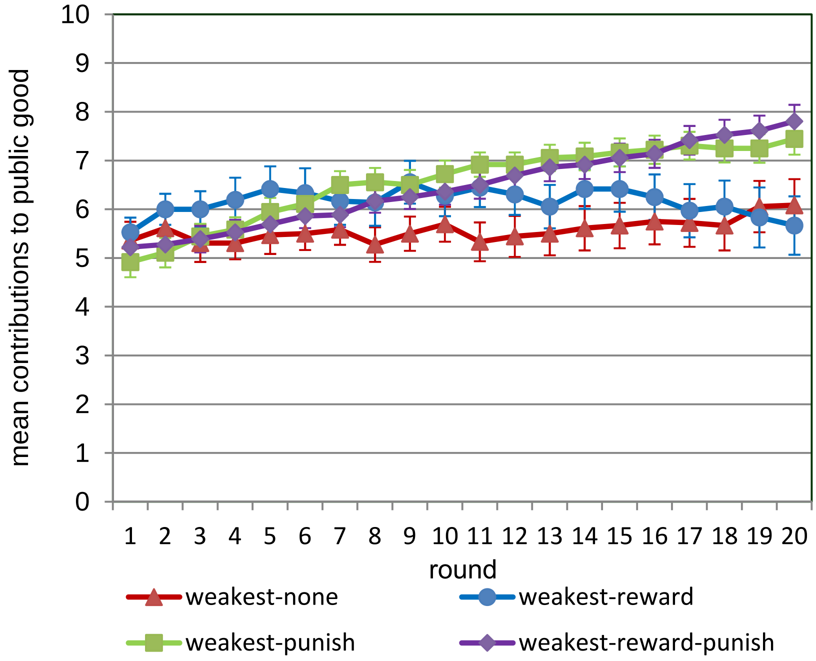


Figure 2. Average contributions to the public good in the best-shot treatments by round

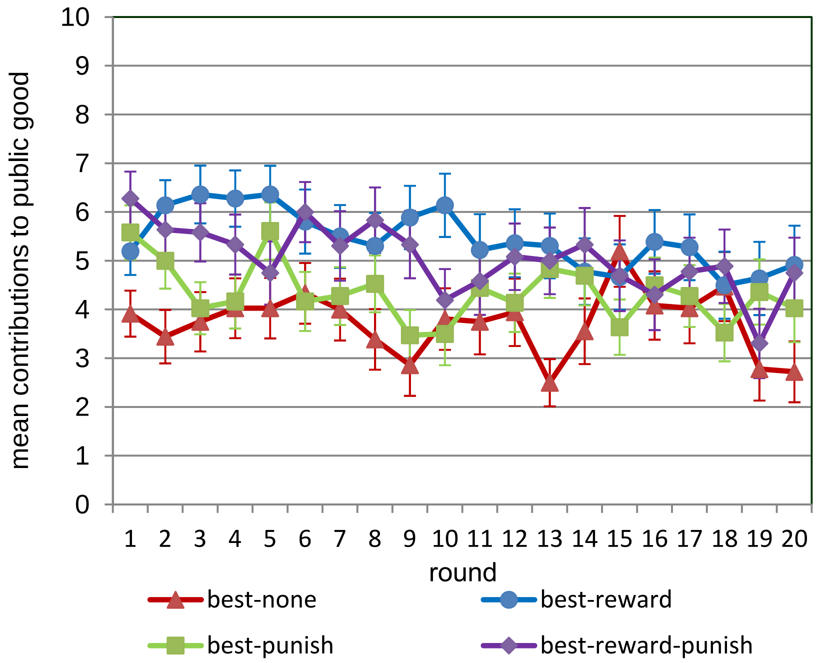


Figure 3. Ratios of subjects giving out reward points to other group members by round

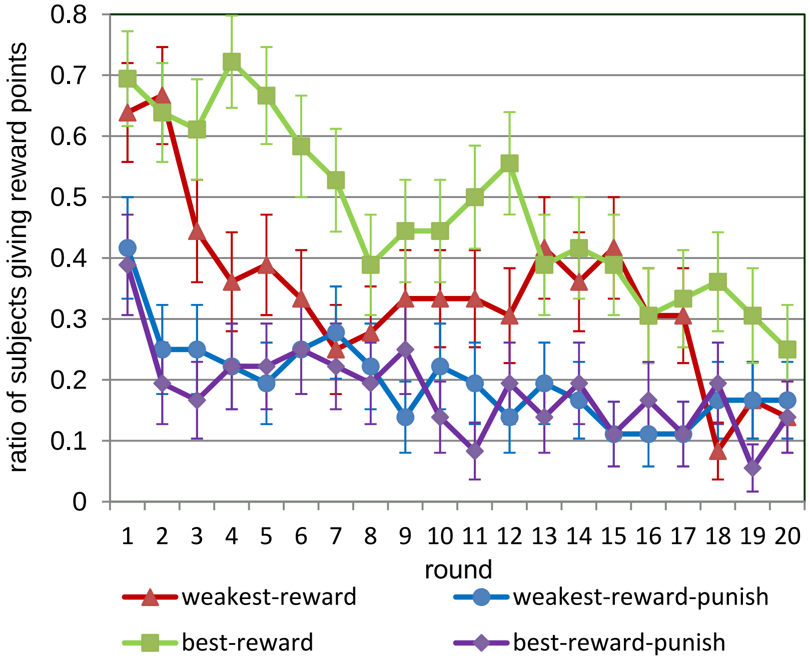
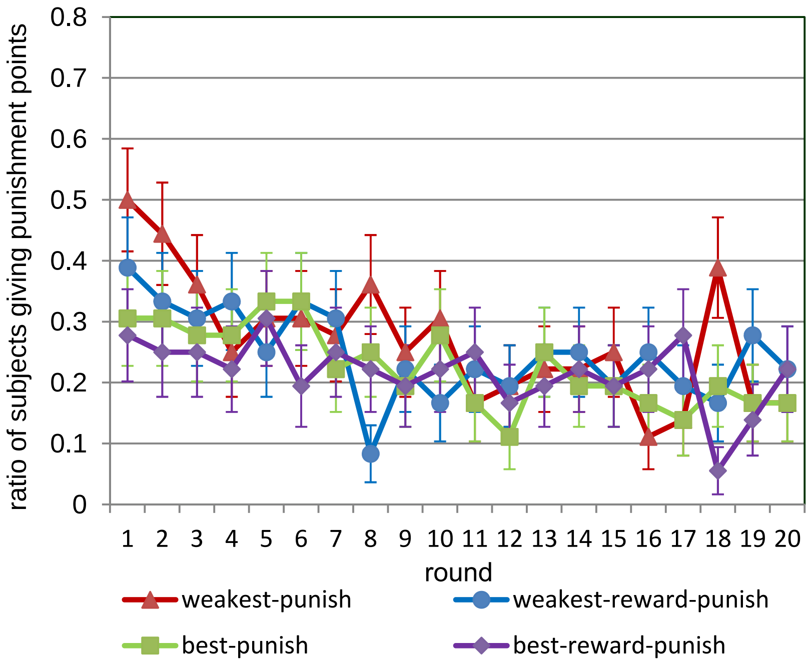


Figure 4. Ratios of subjects giving out punishment points to other group members by round



REFERENCES

Andreoni, James (1988), "Why Free Ride: Strategies and Learning in Public Goods Experiments," *Journal of Public Economics,* 37(3), 291–304.

Andreoni, James (1993), "An Experimental Test of the Public-Goods Crowding-Out Hypothesis," *American Economic Review,* 83(5), 1317–1327.

Andreoni, James (1995), "Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation," *Quarterly Journal of Economics,* 110(1), 1–21.

Andreoni, James, William Harbaugh, and Lise Vesterlund (2003), "The Carrot or the Stick: Rewards, Punishments, and Cooperation," *American Economic Review,* 93(3), 893–902.

Brandts, J. and D. J. Cooper (2006), "A Change Would Do You Good…: An Experimental Study of on How to Overcome Coordination Failure in Organization," *American Economic Review*, 96, 669–693.

Carpenter, Jeffrey P. (2007), "Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods," *Games and Economic Behavior,* 60(1), 31–51.

Fehr, Ernst, and Simon Gächter (2000), "Cooperation and Punishment in Public Goods Experiments," *American Economic Review,* 90(4), 980-994.

Fehr, Ernst and Simon Gächter (2002), "Altruistic Punishment in Humans," *Nature,* 415, 137-140.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger (1997), "Reciprocity as a Contract Enforcement Device: Experimental Evidence," *Econometrica,* 65(4), 833-860.

Fehr, Ernst, and Klaus M. Schmidt (2007), "Contracts and Fairness: Adding a Stick to the Carrot? The Interaction of Bonuses and Fines," *American Economic Review,* 97(2), 177-181.

Gächter, Simon, Elke Renner, and Martin Sefton (2008), "The Long-Run Benefits of Punishment," *Science,* 322(5907), 1510.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach (2006), "The Competitive Advantage of Sanctioning Institutions," *Science,* 312(5770), 108-111.

Hamman, John, Scott Rick, and Roberto A. Weber (2007), "Solving Coordination Failure with "All-or-None" Group-Level Incentives," *Experimental Economics*, 10, 285-303.

Harrison, Gleen and Jack Hirshleifer (1989), "An Experimental Evaluation of Weakest-Link/ Best-Shot Models of Public Goods," *Journal of Political Economy*, 97(1), 201-225.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter (2008), "Antisocial Punishment across Societies," *Science,* 319(5868), 1362-1367.

Hirshleifer (1983), "From Weakest-Link to Best-Shot: The Voluntary Provision of Public Goods," *Public Choice*, 41, 371-386.

Isaac, R. Mark and James M. Walker (1988a), "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism," *Quarterly Journal of Economics,* 103(1), 179-200.

Isaac, R. Mark and James M. Walker (1988b), "Communication and Free-Riding Behavior: The Voluntary Contributions Mechanism," *Economic Inquiry,* 26(4), 585-608.

Isaac, R. Mark, James M. Walker, and Susan H. Thomas (1984), "Divergent Evidence on Free-Riding: An Experimental Examination of Possible Explanations," *Public Choice,* 43(2), 113-150.

Isaac, R. Mark, James M. Walker, and Arlington W. Williams (1994), "Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups," *Journal of Public Economics,* 54(1), 1-36.

Nikiforakis, Nikos and Hans-Theo Normann (2008), "A Comparative Statics Analysis of Punishment in Public-Good Experiments," *Experimental Economics,* 11(4), 358-369.

Reuben, Ernesto and Arno Riedl (2009), "Public Goods Provision and Sanctioning in

Privileged Groups," *Journal of Conflict Resolution,* 53(1), 72-93.
Samuelson, Paul A (1954), "The Pure Theory of Public Expenditure," *Review of Economics and Statistics*, 36(4), 387-389.
Sefton, Martin, Robert Shupp, and James M. Walker (2007), "The Effect of Rewards and Sanctions in Provision of Public Goods," *Economic Inquiry,* 45(4), 671-690.

國科會補助專題研究計畫出席國際學術會議心得報告

日期： 102 年 7 月 31 日

計畫編號	NSC 100–2410–H–004–065–MY2		
計畫名稱	懲罰效果的實驗研究（第 2 年）		
出國人員姓名	徐麗振	服務機構及職稱	政治大學財政系教授
會議時間	102 年 6 月 28 日至 102 年 7 月 2 日	會議地點	Seattle, Washington, USA
會議名稱	(中文) 西部國際經濟學會第 88 屆年會 (英文) The 88th Annual Conference of Western Economic Association International		
發表題目	(中文) 政治經濟體系中租稅稽核的實驗研究 (英文) Tax Auditing in a Political Economy: An Experimental Study		

一、參加會議經過

WEAI 雖是北美西部經濟學會所舉辦的學術會議，但與會者來自全球各地，每年多達千人，今年也不例外。本年度會議在美國華盛頓州西雅圖市的 Grand Hyatt Seattle 飯店舉行，第一天 (6 月 28 日) 下午會議即開始，一直進行至 7 月 2 日中午。我今年投稿到 CEANA (Chinese Economic Association in North America) 所安排的場次，被大會排在 6 月 29 日 12:30 pm~2:15 pm 第[76]場。本場次的文章涵蓋範圍較廣，除了我的文章與租稅有關外，亦包含與管制與總體相關的文章。儘管如此，我仍獲得許多有用的建議。

二、與會心得

參與本會議讓我得以接觸來自世界各地的學者與經濟學不同領域的研究發展，與會學者們對我的文章所提出的建議對我修改文章與未來的研究皆有相當助益。

三、發表論文全文或摘要

Abstract – This paper examines the issue of tax compliance in an environment that the tax rate, the audit rate, and the fine rate are determined by majority voting. In the experiments, subjects play the tax compliance game in the first ten rounds, and then they vote on the tax parameters as indicated above in the latter ten rounds. To control for the order effect, the treatments with the opposite order are also conducted. We also control for the self-selection effect by adding a stage of computer decisions after

voting. The experimental evidence of this paper shows that more than half of the subjects vote for the low tax, audit, and fine rate, indicating that they prefer a less sever auditing environment. Furthermore, our experimental evidence does not support the democracy effect, that is, democracy does not have a positive and significant impact on compliance.

(全文請見附件)

四、建議：

國內學者可藉由參與國際會議和國外學者交流，並快速瞭解國際間相關研究的走向，對拓展國內學者在學術上的人際關係與視野皆有相當助益。故建議國科會可持續補助學者專家出席國際學術會議。

五、攜回資料名稱及內容：

會議議程，與會者通訊錄與近期出版之經濟相關書籍目錄。大部分資料皆可至 WEAI 網站取得。

六、其他：無。

Tax Auditing in a Political Economy: An Experimental Study

Li-Chen Hsu*
Department of Public Finance
National Chengchi University

Kamhon Kan
Institute of Economics
Academia Sinica

Abstract – This paper examines the issue of tax compliance in an environment that the tax rate, the audit rate, and the fine rate are determined by majority voting. In the experiments, subjects play the tax compliance game in the first ten rounds, and then they vote on the tax parameters as indicated above in the latter ten rounds. To control for the order effect, the treatments with the opposite order are also conducted. We also control for the self-selection effect by adding a stage of computer decisions after voting. The experimental evidence of this paper shows that more than half of the subjects vote for the low tax, audit, and fine rate, indicating that they prefer a less sever auditing environment. Furthermore, our experimental evidence does not support the democracy effect, that is, democracy does not have a positive and significant impact on compliance.

Keywords: tax compliance; democracy effect; penalty; tax audit; experiments

*Corresponding author. Financial support from the National Science Council of Taiwan (grant number: NSC 100-2410-H-004-065-MY2) is gratefully acknowledged. We also thank Professor C.C. Yang, the Institute of Economics at Academia Sinica, and the Department of Public Finance at National Chengchi University for providing us with all the facilities.

1. INTRODUCTION

Since the publication of the prominent paper by Allingham and Sandmo (1972), the issue of tax compliance has attracted an abundance of studies. Allingham and Sandmo's model and the subsequent examinations generally mimic the real tax-auditing environment. Specifically, individuals are endowed with income, and they are required to report income and pay the tax on the income declared. After declaration, the tax authority executes an auditing process based on some audit rules and the individuals who are caught cheating are fined.

An odd phenomenon regarding tax auditing is that while audit rates are often low in most countries, zero or nearly zero compliance was never found. As reported by Andreoni, Erard, and Feinstein (1998), the audit rate in the United States has fallen sharply from roughly 6% in 1960s to only 1% in the 1990s. Despite non-zero compliance, vast evasions are still present. Hence, the purpose of tax auditing is to deter evasion and therefore raise government revenues and government spending. To attain these goals, the central questions are which aspects of the tax system are more effective to achieve this goal and how tax compliance can be further improved. To explore these questions, appealing to empirical works using field data is a possibility. However, as pointed out by Andreoni, Erard, and Feinstein (1998), the difficulty is the lack of reliable information on individual reporting behavior. As a consequence, experimental methods may be the most and even the only useful way.

Three aspects of the tax system are generally examined by experimental studies. They are the tax rate, the audit rate, and the penalty rate. Some experimental studies assume these fiscal variables are fixed and vary one at a time to observe the comparative static impacts on compliance (Spicer and Becker, 1980; Becker, Büchner, and Sleeking, 1987; Alm, McKee, and Beck, 1990; Collins and Plumlee, 1991; Alm, Jackson, and McKee, 1992b; Alm, McClelland,

and Schulze, 1992; Alm, Jackson, and McKee, 1993; Alm and McKee, 2006), while some other studies assume non-fixed or endogenously determined audit rates. For instance, Alm, Jackson, and McKee (1992a, 1992b, 1993), Alm, McClelland, and Schulze (1992), and Alm, Sanchez, and de Juan (1995) investigate the effects of uncertain audit probabilities on compliance. Alm, Cronshaw, and Mckee (1993) and Clark, Friesen, and Muller (2004) consider an endogenous audit rule conditional on individuals' own declarations in current or prior periods. Alm and McKee's (2004) examine an endogenous audit rule that considers interactions between subjects.

All aspects of the tax system in the above studies are set up by the tax authority, and individuals simply follow the institutions that have already been set. Under this framework, individuals who report income honestly can only punish tax dodgers through an informal sanction, that is, by behaving in the same way. As a result, compliance will be kept at a low level. Though this outcome is pessimistic, it ignores a fact that in a political economy voters/citizens often have direct or indirect influences on government policies.[1] Some experimental studies have reached a conclusion that individual participation in the decision process can improve cooperation. In public goods experiments, for instance, Andreoni, Harbaugh, and Vesterlund (2003), Carpenter (2007), Fehr and Gächter (2000, 2002), Gächter, Renner, and Sefton (2008), Gürerk, Irlenbusch, and Rockenbach (2006), and Sefton, Shupp, and Walker (2007) have found that formal sanctions on free-riding significantly raise cooperation as compared with informal sanctions. Pommerehne, Hart and Frey (1994) also suggest that

---

[1] An example of direct influences is the voting on tax increases for improving Atlanta's infrastructure. As is reported by *The Economist* (Dec. 10th 2011), Atlantans have the longest average rush-hour commute in America, and according to Georgia's government, the state spends less per head on transport than any other states except Tennessee. Since improving the infrastructure means raising taxes, in June 2010 Georgia's legislature decided to let citizens vote on whether to raise their own taxes. As for indirect influences, individuals may affect or petition the legislative members to pass favorable laws or regulations.

democratic process tends to raise tax morale and therefore tax compliance.

There are several experimental examinations on the effect of democracy on tax compliance. Among these, Alm, Jackson, and McKee (1993) find that compliance is higher if subjects are allowed to select the public sector expenditure program themselves by majority voting. Furthermore, compliance is higher when the selected program is known to have a widespread support than to have narrow support since individuals believe that others will comply more fully in the former. Alm, McClelland and Schulze (1999) investigate the effects of voting the tax rate, fine rate, and audit rate on compliance. They find that in two of the tax sessions the majority votes for the low tax rate and in the other two tax sessions the majority votes for the high tax rate. In all four fine sessions the majority votes for the low fine rate, and in the other four audit rate sessions the majority votes for the low audit rate. Furthermore, the average compliance rates in the vote stage are lower than the corresponding average compliance rates in the no-vote stage in all sessions. They appeal to the notion of social norm that an individual will comply as long as he (she) believes others will comply. The group decision on enforcement reveals a lack of social norm of tax compliance and thus compliance with voting is lower than that without voting.

By designing an experiment in which the punishment is certain, that is, the audit probability is one, Feld and Tyran (2002) ask subjects to state their contributions for all possible voting outcomes. They find that the possibility of voting on fine significantly increases tax compliance since subjects who vote for the punishment scheme feel obliged to consistently comply with their decision by making higher contributions. Similar results are found in Putterman, Tyran, and Kamei (2011), in which subjects vote on whether “private account” or “public account” contributions are subject to penalty. They find that there is almost uniform support for penalizing noncontribution to the public account, and contributions to the public good are significantly higher when there are formal sanctions than when sanctions are absent.

Despite the above inconsistency in the findings of democracy on tax compliance, as pointed out by Bó, Foster, and Putterman (2010), a central problem with these kinds of experiments is that "one cannot rule out the possibility that there are unobserved factors that explain both responses to policies and either the degree of participation in policymaking or the particular policies selected." Briefly speaking, there is a self-selection problem left behind. That is, the observed higher level of cooperation under voting may be attributed to individuals' inherent preference for the chosen policy, not because of the participation in the democratic process. To control for the self-selection problem, in Bó, Foster, and Putterman's prisoner's dilemma experiment a stage of computer decision is added after voting. Specifically, individuals first vote on two alternatives: modifying the payoff or not. Then the computer decides to consider the outcome of majority voting or not. If the former situation occurs, the final outcome is consistent with the result of majority voting. If the latter situation occurs, then the computer decides whether to modify the payoff or not. The addition of computer decisions breaks the direct connection between the preference for the chosen alternative and the outcome of majority voting, and therefore the democracy effect can be properly measured.

This paper applies Bó, Foster, and Putterman's (2010) approach to examine tax compliance in an environment that tax, audit rate, and fine rates are determined by majority voting. There are several major differences between Bó et al.'s experimental design and ours. First, Bó et al. use a prisoner's dilemma game, but ours is a tax compliance game. Second, in Bó et al.'s experiment subjects first play the game without voting for ten rounds. Then a majority voting is executed before the start of the eleventh round and the outcome of this one-time majority voting applies to the next ten rounds. In our experiment, the voting procedure occurs in the beginning of each of the ten rounds being applied majority voting. In addition, we control for the order effect by switching the order of the rounds with voting and the rounds without voting. That is, in some sessions the ten rounds without voting are run first then the other ten rounds with voting are run afterward. In some other sessions, the ten rounds with voting are run first, followed by the ten rounds without voting. Third, in Bó et al.'s experiment subjects are informed of the outcome of majority voting before computer making decisions, while in our experiment subjects are only informed of the final outcomes applied to their groups. They are unaware of the outcome of majority voting, nor the decision made by the computer.

The experimental evidence of this paper shows that more than half of subjects vote for the lower tax rate or the lower fine rate, and there are two thirds to three fourths of the subjects vote for the lower audit rate. These findings suggest that most subjects prefer a low-tax policy and a less severe auditing environment. Furthermore, our experimental results do not support the democracy effect. That is, democratic participation does not have positive impacts on compliance.

The remainder of this paper is organized as follows. Section 2 describes the experimental design. Section 3 presents the theoretical predictions. Section 4 reports the results of the experiment, and Section 5 concludes.

2. EXPERIMENTAL DESIGN

The fundamental experimental design is similar to those of the experiments on VCM (the voluntary-contribution mechanism) and tax compliance. In the experiment, each subject receives an exogenous amount of income and he (she) pays the tax according to the income that he (she) declares. The tax is used to provide the public good that benefits only the members in the same group. After declaration, the subject's true income is audited based on some probability. The subjects who are audited and caught cheating will pay the evaded taxes and fines. A subject's original income net of the tax he (she) has paid and the evaded tax and fines, if there are any, is his (her) private good consumption. His (her) payoff is the sum of the public good consumption and his (her) private good consumption. To prevent any emotional responses, neutral terms are used in the experimental instructions. For instance, in the instructions the public good is written as "public account," the tax paid by a subject is instructed as the subject's investment in the public account, the term "audited" is phrased as "checked," and the payback of the evaded tax and fines is denoted as a "reduction" in the subject's payoff. Furthermore, since the tax authority simply collects taxes and fines without making any decisions in the experiment, the role of the tax authority is not particularly mentioned.

[Table 1 about here]

Six treatments were conducted in this research. They are denoted as the tax-o1 (tax rate-order one) treatment, the tax-o2 (tax rate-order two) treatment, the aud-o1 (audit rate-order one) treatment, the aud-o2 (audit rate-order two) treatment, the fin-o1 (fine rate-order one) treatment, and the fin-o2 (fine rate-order two) treatment. The framework of the experiment and the tax parameters used in each treatment is provided in Table 1. Four sessions were conducted in each treatment and twelve subjects were recruited for each session, for a total of 288 subjects used for this study. Two independent sessions under the same treatment were run at the same time, but subjects were unaware of this. This design prevents subjects from inferring the components of their groups so that the reputation effect can be kept at a minimally possible level. All subjects were undergraduate students at National Chengchi University and none of them had ever participated in any public goods or tax compliance experiments.

Each treatment consisted of two parts, and each part contained ten rounds. Subjects were informed of the contents of the two parts in the beginning of the experiment. All the experimental settings in the first part (rounds 1–10) of the tax-o1, aud-o1, and fin-o1 treatments were the same, while in the second part (rounds 11–20) a voting process was added and subjects in the three treatments voted on different tax variables. Since subjects' compliant behavior and their attitudes towards the tax variables may be affected by the timing of voting, to control for the order effect, three corresponding treatments with the opposite order of the two parts were also conducted. They are indicated as the tax-o2 treatment, the aud-o2 treatment, and the fin-o2 treatment. The three treatments were exactly the same as the tax-o1, aud-o1, and fin-o1 treatments, respectively, except that the voting process appeared in the first ten rounds instead of the latter ten rounds of the experiment.

The experimental procedures of the tax-o1, aud-o1, and fin-o1 treatments were run as follows. Subjects made decisions in each of the twenty rounds. In each round, the twelve subjects in the same session were randomly and anonymously divided into three groups of size $n = 4$, and they were re-matched when a new round started. Under this setting, one independent observation is obtained from each session. At the beginning of each round, four income levels (70, 90, 110 and 130 points) and four codes (A, B, C, and D) were randomly assigned to the four subjects in the same group. Call the income assigned to a subject his or her true income w_i. When a new round started, the four levels of income and the four codes were randomly reassigned. A subject knew his (her) own code and income and the distribution of income, but not the income for each of the other three group members.

There were two stages in each round of the first part of the tax-o1, aud-o1, and fin-o1 treatments. In stage one, the declaration stage, each subject was required to report a level of income R_i ($0 \le R_i \le w_i$), and the reported income was taxed at the rate $t = 0.2$. The tax was invested in the public account (the public good), and the rest of the income was maintained in the subject's private account (the private good). The marginal per capita return (MPCR) of the public good was set at $m = 0.5$. Notice the selection of m must satisfy the condition $1/n < m < 1$ so that each individual has the incentive to cooperate and to cheat. That is, every point invested in the public good yielded *every* group member a return of 0.5 points. After all subjects had reported their incomes, they proceeded to the second stage, the auditing stage, in which each subject was audited by a probability $p = 0.1$. It is assumed that a subject's true income was revealed once he (she) was audited. Any subject that was audited and caught cheating had to pay the evaded tax and a fine. The fine was twice the amount of the evaded tax. Hence, the subject who was caught cheating incurred a penalty which was three times the amount of the tax that he (she) had evaded. The fine rate is denoted as f and $f = 3$.

Given the above procedures, the expected monetary payoff for each subject i in each round of the first part of the tax-o1, aud-o1, and fin-o1 sessions is given by

$$\pi_i = (1-p)(w_i - tR_i) + p[w_i - tR_i - ft(w_i - R_i)] + mt\textstyle\sum_{j=1}^{n} R_j. \qquad (1)$$

In (1), the sum of the first two terms is the subject's expected private good consumption and the third term is his (her) public good consumption. As is indicated by (1), the evaded taxes and fines retrieved were simply discarded and were not included in the public good. This setting aims to exclude any benefits resulting from the subject's tax-dodging behavior, and in turn to eliminate any incentives that may deteriorate the subject's compliance decisions attributed to this kind of benefits.

The tax rate $t = 0.2$, audit rate $p = 0.1$, and fine rate $f = 3$ serve as the benchmark of the three fiscal variables. When subjects moved to the second-part of the experiment, an additional voting process was added in the beginning of each round, and one of these benchmark values was to be voted against another higher value. Specifically, in the tax-o1 treatment the four members in the same group voted between two alternative levels of tax rates, 0.2 and 0.4; in the aud-o1 treatment the four members voted between two alternative levels of audit rates, 0.1 and 0.4; and in the fin-o1 treatment the four members voted between two alternative levels of fine rates, 3 and 6. In each treatment, the other two fiscal variables that were not determined via voting remained at the same levels as in the first-part of the experiment. As a consequence, there were three stages in the second part of the tax-o1, aud-o1, and fin-o1 treatments: a voting stage, a declaration stage, and an auditing stage.

Let us explain the experimental procedure in more detail by using the tax-o1 treatment as an example. In the beginning of each round of rounds 11 to 20 of the tax-o1 treatment, subjects were required to vote between two tax rates, 0.2 and 0.4, for their own groups via a majority voting. Subjects were informed that after all group members had made their own voting decisions, the computer would randomly determine whether to accept the outcome of majority voting or not. If the computer accepted the outcome of majority voting, the tax rate for the group was determined accordingly. If the computer rejected the outcome of majority voting, then the computer would randomly assign one of the two tax rates for the group. If a tie occurred in majority voting, the computer would also randomly assign either tax rate for the group.

In the computer program, we set the probability that the computer randomly accepts the outcome of majority voting to be 0.7 (so the probability of rejection is 0.3). Once the computer rejects the outcome of majority voting or if a tie occurs in majority voting, the probability that the computer randomly assigns either tax rate to the group is 0.5. Subjects were only informed of the above process and the final outcome of the tax rates for their own groups. They were unaware of the outcome of majority voting, the decision made by the computer, nor the information regarding the setup of the probabilities for the computer's random choices.[2] This setting along with the set up that the size of groups is four aim to make the greatest effort to eliminate any self-selection problems by allowing the maximum possible computer interventions.

The same procedure applied to the aud-o1 and fin-o1 treatments, except that in the aud-o1 treatment subjects voted between two alternative audit rates, 0.1 and 0.4, and in the fin-o1 treatment subjects voted between two alternative fine rates, 3 and 6. Followed by the voting stage, the second stage (the declaration stage) and the third stage (the auditing stage) of the second part of the tax-o1, aud-o1, and fin-o1 treatments are exactly the same as the first and second stages in the first part of the three treatments. Given the above procedure, the expected monetary payoff for the subject in the second-part of the tax-o1, aud-o1, and fin-o1 treatments is the same as in (1) except that the tax, audit, or fine rate was determined by majority voting and computer decisions.

The tax-o2, aud-o2, and fin-o2 treatments were exactly the same as the tax-o1, aud-o1, and fin-o1 treatments, respectively, except that the order of the first part and the second part was reversed. At the end of each round of the experiment, each subject was informed of a result report, which contained mainly the following information: the outcome of the voting stage (when there was one), the subject's declaration of income, his (her) investment in the public account according to his (her) declaration, the total income declared and the total investment in the public account by the other three group members and by the entire group, the code of the subject who

[2] This design differs from that in Bó, Foster, and Putterman (2010), in which subjects were informed of the outcomes of majority voting and whether the computer randomly chose to consider the vote.

was audited, the subject's payoff from his (her) private account, the subject's payoff from the public account, the reduction in the subject's payoff if he (she) was caught under-reporting, and the subject's payoff for this round.

In all sessions, subjects were given written instructions in Chinese. The experimenter read the instructions aloud and answered any questions raised by the subjects. After reading the instructions, subjects were required to answer four questions in relation to the calculation of payoffs and the experimental procedures. The experiment would not start until everyone answered all questions correctly. Each session lasted about 90 minutes. The average payoff (including a participation fee of NT$100) for all participants is NT$529.98 (with a standard deviation of NT$24.85, a maximum of NT$595, and a minimum of NT$445.7).[3]

3. THEORETICAL PREDICTIONS

To have a clear-cut theoretical prediction on subjects' behavior, it is assumed that all subjects are self-interested and maximize their own monetary payoffs, and that this characteristic is common knowledge among all subjects. Recall that when fiscal variables were exogenously given in the first part of the tax-o1, aud-o1, and fin-o1 treatments and the second part of the tax-o2, aud-o2, and fin-o2 treatments, the subject's expected monetary payoff was characterized by (1). Differentiating (1) with respect to R_i yields $\partial \pi_i / \partial R_i = t(pf + m - 1)$. Since t = 0.2 > 0, to have an inner solution the condition $pf + m = 1$ must be satisfied, and it is obvious that the chance is slim. Given p = 0.1, f = 3, and m = 0.5, we have $pf + m < 1$, implying that $R_i^* = 0$, for all i. That is, the dominant strategy for a self-interested and reward-maximizing

[3] When these sessions were conducted, the exchange rate between the NT (New Taiwan) dollar and the US dollar was about 30:1. The part-time hourly wage rate for an undergraduate student in Taiwan is about NT$120.

subject is reporting zero income. As a result, the subject's expected payoff is $\pi_i = w_i(1 - pft) = 0.94w_i$. However, if $R_i = w_i$ for all i, then the subject's expected payoff becomes $\pi_i = w_i(1-t) + mt\sum_{j=1}^{n} R_j = 0.8w_i + 40$, which is certainly greater than the expected payoff that everyone reports zero income, regardless of subjects being assigned an income of 70, 90, 110, or 130 points. Hence, this is a typical social dilemmas problem (Dawes, 1980): It is in everyone's self interest to report zero income and to pay zero taxes, although complete honesty is indeed a socially optimal decision.

To find the equilibrium when a voting stage is involved in the second part of the tax-o1, aud-o1, and fin-o1 treatments and the first part of the tax-o2, aud-o2, and fin-o2 treatments, we can construct a two-stage game and solve the game by backward induction. The game proceeds as follows. In the first stage, all group members vote on two alternative levels of the tax, audit, or fine rate. The computer then randomly determines whether to accept the outcome of majority voting based on some probability. In the second stage, given the outcome of majority voting and the computer decision, subjects declare income simultaneously and then they are audited according to some probability. When a subject makes his (her) voting decision, he (she) assumes all other group members have made their optimal voting decisions. When a subject makes his (her) declaration decision in the second stage, he (she) assumes that all other group members have chosen their optimal levels of declarations, and takes all group members' voting decisions into consideration. It is evident that once the tax, audit, or fine rate has been determined in the first stage, subjects' expected payoff will be characterized by (1), and as a result the dominant strategy in the second stage is still solved by the first-order condition $\partial\pi_i / \partial R_i = t(pf + m - 1)$. By considering the dominant strategy adopted in the second stage, subjects make their best voting decisions in the first stage.

Let us start with the situation in which the tax rate is determined by majority voting in the first stage and the audit rate is given at 0.1 and the fine rate is given at 3. When in the first stage subjects vote on the two tax rates, 0.2 and 0.4, they are aware of the fact that either tax rate will be selected eventually. They also understand that their votes will to some extent affect the outcome of majority voting and that this outcome will be accepted by the computer according to some probability. Hence, the subject will vote for a tax rate that is more advantageous to his (her) expected payoff once the second stage arrives. In the second stage, since the sign of the first-order condition $\partial \pi_i / \partial R_i = t(pf + m - 1)$ is negative and is irrelevant to the tax rate, the dominant strategy for the subject is still reporting zero income regardless of the outcome of the first stage. Given that zero income will be reported, the subject's expected payoff becomes $\pi_i = w_i(1 - pft) = 0.94w_i$ if the tax rate is 0.2, and is $\pi_i = w_i(1 - pft) = 0.88w_i$ if the tax rate is 0.4. Since the former is higher, subjects will vote for the low tax rate 0.2. Hence, the subgame perfect equilibrium is that subjects vote for the low tax rate 0.2 in the first stage and report zero income in the second stage.

Similar analysis applies to the situation that the audit rate is determined via majority voting and the other two fiscal variables remain exogenously given. Simply put, since the first-order condition for the maximization of the expected payoff is $\partial \pi_i / \partial R_i = t(pf + m - 1) = -0.04$ if the audit rate is 0.1 and is $\partial \pi_i / \partial R_i = t(pf + m - 1) = 0.14$ if the audit rate is 0.4, the subject will have extremely opposite decisions under each audit rate. Specifically, if in the first stage the audit rate turns out to be 0.1, then the subject will report zero income in the second stage and earns an expected payoff $\pi_i = w_i(1 - pft) = 0.94w_i$. If instead in the first stage the audit rate turns out to be 0.4, then the subject will report full income and earns an expected payoff

$\pi_i = w_i(1-t) + mt\sum_{j=1}^{n} w_j = 0.8w_i + 40$, which is certainly greater than $0.94w_i$ regardless of the value of w_i being assigned in the experiment. Given the result in the second stage, subjects will vote for the audit rate 0.4. Therefore, the subgame perfect equilibrium is that subjects vote for the high audit rate 0.4 in the first stage and report full income in the second stage.

Lastly, when the fine rate is determined by majority voting and the other two fiscal variables remain exogenously given, the first-order condition for the maximization of the expected payoff is $\partial\pi_i / \partial R_i = t(pf + m - 1) = -0.04$ if the fine rate is 3, and is $\partial\pi_i / \partial R_i = t(pf + m - 1) = 0.02$ if the fine rate is 6. Therefore, if the fine rate turns out to be 3 in the first stage, the subject will report zero income and earns an expected payoff $\pi_i = 0.94w_i$. If the fine rate turns out to be 6, then the subject will be completely honest and earns an expected payoff $\pi_i = 0.8w_i + 40$, which again is certainly greater than $0.94w_i$ given all the possible values of w_i used in the experiment. Therefore, the subgame perfect equilibrium is that subjects vote for the high fine rate 6 in the first stage and report full income in the second stage.

In sum, the dominant strategy Nash equilibrium for the no-voting rounds of all the six treatments is reporting zero income. For the voting-rounds, the subgame perfect equilibria are the following: In the two tax-rate treatments (tax-o1 and tax-o2), each subject will vote for the low tax rate 0.2 and report zero income; in the two audit-rate treatments (aud-o1 and aud-o2) and the two fine-rate treatments (fin-o1 and fin-o2), subjects will vote respectively for the high audit rate 0.4 and the high fine rate 6 and they will comply fully.

A prevalent finding in the VCM experiments is that the dominant strategy Nash equilibrium, i.e., zero contributions to the public good, almost never occurred (Andreoni 1988, 1993, 1995; Isaac and Walker 1988a, 1988b; Isaac, Walker, and Thomas 1984; Isaac, Walker, and Williams

1994). Since the framework of the tax compliance experiment in this study is similar to VCM, it is reasonable to anticipate that the equilibrium of zero compliance is unlikely to occur in the no-voting rounds and in the voting rounds of the tax-o1 and tax-o2 treatments. Furthermore, in spite of the previous findings in public goods experiments, the auditing procedure should also drive subjects' income-declaration decisions away from zero compliance. When the audit rate or the fine rate is to be voted, the subgame perfect equilibrium prediction that subjects will fully comply also seems unlikely to realize. Therefore, besides testing the equilibrium predictions, we will emphasize in more detail on the relationship between subjects' voting decisions and their compliant behavior. Specifically, we will particular test the following issues. First, will subjects behave more compliantly when they are involved in the political process than when they are not? Second, will subjects who vote for the stricter fiscal policies also behave more compliantly? Third, will subjects' compliant behavior differs across regimes in which different aspects of the tax system are to be voted?

4. EXPERIMENTAL RESULTS

[Table 2 about here]

We conducted twenty-four sessions in April and May of 2012 in the computer lab of the Department of Public Finance at National Chengchi University in Taiwan. Table 2 reports that of the 288 subjects recruited, 74.31 percent of them were female, on average they had been in the university for 2.11 years, the average age was 20.01 years, and 78.47 percent of them had taken economics course(s). The scale of the indicator "donation" ranged from one to six and the average was 2.09, meaning that on average subjects donated about NT$500 to NT$1,000 to charities during the year 2011. The scale of "risk-taking" ranged from 0 to 10, with 0 indicating not ready for taking any risks and 10 indicating fully prepared to take risks. The average level of risk-taking was 5.17, meaning that on average subjects' attitude toward risks was modest.

4.1. A General Look at the Compliant Behavior

[Tables 3.1, 3.2, and 3.3 about here]

Tables 3.1 through 3.3 summarize the data resulting from the first 10 rounds and the latter 10 rounds in each treatment. The compliance rate for a subject is defined as his (her) reported income divided by his (her) true income. Several observations arise by looking at subjects' compliance decisions. First, it is observed that average compliance rates for all six treatments lie in between 0.54 to 0.64 in the first ten rounds, then fall to the average of 0.36 to 0.55 in last ten rounds. Except in the aud-o1 treatment, average compliance rates generally decay across rounds. Second, in the no-voting rounds, though on average 26.04 percent of the subjects in the aud-o2 treatment and 21.46 percent of the subjects in the fin-o2 treatment reported zero income, these magnitudes are far below one as predicted by the theory. In each of the other four treatments, even less than ten percent of the subjects reported zero income. Therefore, the theoretical prediction of zero compliance for the no-voting rounds does not hold. Third, in rounds with voting, the percentages of subjects declaring zero income were still far below one in the two tax treatments (15.42 percent in tax-o1 and 6.04 percent in tax-o2), and less than twenty percent of the subjects full complied in the two audit and the two fine treatments. Hence, the theoretical predictions still fail to explain subjects' compliant decisions. We summarize the above observations in Result 1.

Result 1: Except in the aud-o1 treatment, average compliance rates generally decay across rounds in all other treatments. The theoretical prediction of zero compliance for the no-voting rounds in all six treatments and the voting-rounds for the two tax treatments does not hold. The theoretical prediction of fully comply for the voting rounds of the two audit treatments and the two fine treatments also fails.

[Figures 1 and 2 about here]

The round averages and the standard errors of compliance rates are depicted in Figures 1 and 2.[4] In addition, Figure 1 provides the information regarding the average compliance rates conditional on subjects' voting decisions, and Figure 2 provides similar information conditional on the magnitudes of the fiscal variables eventually applied to the subjects. Let us look simply at the average compliance rates across rounds first. Comparing the six figures in Figure 1 shows that within the three treatments that the voting stage occurs in the latter ten rounds (i.e., the tax-o1, aud-o1, and fin-o1 treatments), the tax-o1 treatment generally has the lowest average compliance rates across rounds, and the aud-o1 treatment has the highest average compliance rate in almost every round. The opposite trend is observed in the three treatments in which the voting stage occurs in the first ten rounds (i.e., the tax-o2, aud-o2, and fin-o1 treatments), but the differences in average compliance rates between the three treatments are smaller.

A two-sided Mann-Whitney U test confirms the above observations. Compliance rates are significantly higher in the aud-o1 treatment than in the tax-o1 treatment by looking at the average of rounds 1–5 (p = 0.0833), 6–10 (p = 0.0433), 1–10 (p = 0.0833), and 1–20 (p = 0.0833). However, compliance rates are significantly higher in the aud-o1 treatment than in the fin-o1 treatment in only round 19 (p = 0.0209). There exist no significant differences between

[4] The standard error of the sample mean is calculated as $\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)}\Big/\sqrt{n}$, where x_i is the value of the observation, $\bar{x}$ is the sample mean, and n is the sample size.

the tax-o1 and fin-o1 treatments and between any two of the tax-o2, aud-o2, and fin-o2 treatments in each round of the experiment.[5] We have the following result.

Result 2: Allowing subjects to vote on the audit probability is a more effective way to improve compliance than allowing subjects to vote on the tax rate, but only if the voting process occurs in the latter ten rounds of the experiment.

Experimental studies usually find that subjects need time to learn the game and other subjects' responses. The intuition for Result 2 is that if subjects are allowed to vote in the first ten rounds of the experiment, the effect of democracy will be outweighed by the downward trend of cooperation attributed to the free-riding incentive.

4.2. Voting Behavior and Compliance Conditional on Voting Decisions and Final Outcomes

Now let us look at the voting behavior. A first glance of row (5) in Tables 3.1, 3.2, and 3.3 tells us that on average less than half of the subjects voted for the stricter values of the fiscal variables. Specifically, on average 49.58 percent of the subjects in tax-o1 and 43.33 percent of the subjects in tax-o2 voted for the high tax rate. The percentages of subjects voting for the high fine rate were a little bit lower (42.5 percent in fin-o1 and 35.83 percent in fin-o2), but there were on average only 31.04 percent of the subjects in the aud-o1 treatment and 22.29 percent of the subjects in the aud-o2 treatment voted for the high audit rate. These observations suggest

[5] As mentioned in the Experimental Design, random re-matching was managed between every twelve subjects. Therefore, in this and other Mann-Whitney U tests, the average of the twelve subjects' choices in each round is used as the round observation and the average of the round observations over a certain period is used as the observation for that period.

that on average subjects preferred a less strict auditing environment.

We now look particularly at subjects' compliant behavior conditional on their voting decisions and the final outcomes applied to them. Row (8) through row (11) of Tables 3.1, 3.2, and 3.3 report related statistics. A quick glance at Tables 3.1, 3.2, and 3.3 tells us that on average subjects who voted for the higher tax, audit, or fine rate behaved more compliantly than subjects who voted for the lower counterpart. Similarly, except for the tax-o2 treatment, the average compliance rates for subjects being applied the higher tax, audit, or fine rate were also higher than the average compliance rates for subjects being applied the lower counterpart. The differences are much more substantial when audit rates and fine rates were to be voted.

The round averages depicted in Figures 1 and 2 not only echo the above observations but also reveal more information. When the voting stage appears in the first part (rounds 1–10) of the experiment, that is, of the tax-o2, aud-o2, and fin-o2 treatments, the average compliance rate in each of the three treatments started high, at about seventy percent, then generally declined across rounds and reached the lowest level in the final round of the experiment. This observation is similar to those found in many experiments on VCM, suggesting that adding a stimulus (voting) in the beginning of the tax compliance game does not help preventing the downward trend of cooperation. However, when the voting stage appeared in the second part (rounds 11–20) of the experiment, that is, of the tax-o1, aud-o1, and fin-o1 treatments, an obvious restart effect (Andreoni, 1988; Andreoni and Miller, 1993) occurred in round 11 and remained in several later rounds of the aud-o1 and fin-o1 treatments. In the aud-o1 treatment, the restart effect almost never faded so that the average compliance rate still maintained at 0.5102 in the final round.

Differences also exhibit between subjects who voted for the higher values of the fiscal variables and subjects who voted for the lower counterparts. As is observed from Figure 1, except in round 15 of the aud-o1 treatment, the average compliance rates were substantially higher for subjects who voted for the higher audit rate (0.4) than for subjects who voted for the lower counterpart (0.1). Similar trends appeared when the fiscal variable to be voted was the fine rate, though the differences were smaller.

Figure 2 shows that average compliance rates in both the aud-o1 and aud-o2 treatments were at least 80 percent in five out of ten rounds for the subjects who were applied the higher audit rate, whereas average compliance rates were only about 30 percent for the subjects who were applied the lower audit rate. Similar patterns are found in the fin-o1 and fin-o2 treatments, though the differences in average compliance rates between the subjects being applied the high fine rate and the subjects being applied the low fine rate are smaller than those found between aud-o1 and aud-o2.

Lastly, recall that the theoretical predictions for the voting rounds are that subjects will vote for the low tax rate 0.2 and report zero income; and subjects will vote for the high audit rate 0.4 and the high fine rate 6 and comply fully. The above observations do not support these predictions. We summarize the above observations in Result 3.

Result 3: The experimental evidence does not support the theoretical predictions that subjects will vote for the low tax rate 0.2 and report zero income; and subjects will vote for the high audit rate 0.4 and the high fine rate 6 and comply fully. Furthermore, subjects who voted for or were being applied to the high tax, audit, or fine rate behaved more compliantly than subjects who voted for or were being applied to the lower counterpart.

4.3. Effects of Democracy

[Tables 4.1, 4.2, 5.1, and 5.2 about here]

We now examine whether democracy effects exist, that is, whether the level of compliance is higher if subjects were allowed to vote than when they were not. Table 4 and Table 5 report the regression results from the fixed-effect ordinary least squares (OLS) by using two different regression equations, which are called respectively Specification I and Specification II. In both specifications, two different periods of observations are used. In Table 4.1 and Table 5.1, observations from all rounds are included. Since sharper fluctuations or downtrends usually occur in the beginning and final rounds of the game, in Table 4.2 and Table 5.2 only the observations from round 6 to round 15 are used.

In Table 4.1.1 the dependent variable is the compliance rate and the independent variables include no-voting stage (1 if yes and 0 otherwise), round number, the square of round number, subjects' true income, subjects' voting decision (1 if vote for the high tax, audit, or fine rate and 0 otherwise), whether the final rate is high (1 if the high rate applied and 0 otherwise), and whether the final rate is low (1 if the low rate applied and 0 otherwise). The estimated coefficients of the independent variables that are binary (0 or 1) indicate the estimated average compliance rates as are defined by the variables, and the range of these estimated coefficients is from 0 to 100 (percent). For instance, in the tax-o1 treatment the estimated coefficient of "no-voting stage" is 48.4900, saying that the estimated average compliance rate in the no-voting stage is 48.49 percent.

The estimated results reported by Table 4.1.1 show that the average compliance rates for subjects characterized by the variables "no-voting stage," "high rate applied," and "low rate applied" are all positive and significantly different from zero at five percent significance level in all six treatments. Furthermore, the variable “round” has a negative and significant effect on compliance in each treatment, indicating that compliance rates decrease across rounds. True income also has negative effects on compliance rates, but is significant only in the tax-o1 and aud-o1 treatments. Finally, the estimated coefficient of “vote for high rate” are negative in all six treatments, showing that the estimated average compliance rates for the subjects who vote for the high tax rate are negative. Though negative is unreasonable, they do not differ significantly from zero.

Since the tax, audit, and fine rates are given at the lower levels in the no-voting stage, to investigate the democracy effect, in the voting stage the tax, audit, or fine rate must be controlled at the same low levels. Whether democracy has an impact on compliance is inferred from the *p*-value of S1p1, which is the test result of the equality between the compliance rate of the no-voting stage and the compliance rate of the voting stage when the rate finally being applied is low. Only in aud-o2 and fin-o1 treatments the *p*-values of S1p1 are significant, indicating that only in these two treatments voting affects compliance. However, only the result of voting on the fine rate supports the democracy effect. In the aud-o2 treatment the estimated average compliance rate in the no-voting stage (86.58%) is actually higher than the estimated average compliance rate in the voting stage when the low audit rate is applied (72.64%). This means that that allowing voting on the audit rate indeed deteriorates compliance, thus contradicting the democracy effect.

Table 4.1.2 reports the effects of individual characteristics on compliance. In Table 4.1.2 the dependent variable μ_i is the average compliance rate per subject, and the between-subjects variations within the same treatment have been eliminated. It is shown that subjects who are able to bear more risks behave less compliantly in the tax-o1, aud-o1, aud-o2, and fin-o2 treatments. Only in the aud-o2 treatment age has a significant effect on compliance, and this effect is negative. Having taken an economics course (or courses) has a significant and negative effect on compliance in the aud-o1 and fin-o1 treatments. Subjects who donate more money to charities in the past year are significantly more compliant in the fin-o1 treatment.

By confining the sample periods to round 6 through round 15, Table 4.2.1 and Table 4.2.2 show similar estimated results as those found in Table 4.1.1 and Table 4.1.2. The only prominent exception is that S1p1 is now no longer significant in the fin-o1 treatment. As a consequence, the democracy effect does not exist in all six treatments if we forsake the relatively more fluctuant beginning and final five rounds of the game.

Table 5 reports the results from an alternative regression equation, the Specification 2. Table 5.1.1 shows that the estimated coefficients and significance levels of the dependent variables "no-voting stage," "round," "round^2," and "true income" are all similar to those in Table 4.1.1. The estimated coefficients of other independent variables, including "vote for high rate and high rate applied" (1 if yes and 0 otherwise), "vote for low rate and high rate applied" (1 if yes and 0 otherwise), "vote for high rate and low rate applied" (1 if yes and 0 otherwise), and "vote for low rate and low rate applied" (1 if yes and 0 otherwise), are all positive and significantly different from zero at the 5% significance level.

The *p*-values of S2p1 indicate the test results of the equality between the compliance rates of subjects who vote for the high rate and are actually applied the high rate and the compliance rates of subjects who vote for the low rate but are actually applied the high rate during the voting stage. It is shown that there are significant differences between these two groups of subjects in the tax-o1, aud-o1, and fin-o1 treatments. Similarly, the *p*-values of S2p2 indicate the test results of the equality between the compliance rates of subjects who vote for the high rate but are actually applied the low rate and the compliance rates subjects who vote for the low rate and are actually applied the low rate during the voting stage. Only in the aud-o2 treatment the *p*-value of S2p2 reveals a significant difference between these two groups. Notice that the significant differences derived from both S2p1 and s2p2 cannot be inferred as the effects of democracy, but they are rather the outcomes of the mixtures of the tastes for the levels of the fiscal variables and the actual levels being applied to them.

The *p*-values of S2p3 indicate the test results of the equality between the compliance rates of subjects who vote for the low rate and are actually applied the low rate and the compliance rates of subjects in the no-voting stage. By controlling the final outcome of the fiscal variables to be the low rate in both the no-voting stage and the voting stage, the test results of S2p3 explain the existence of the democracy effect. It is shown that the difference is substantially significant in the aud-o2 treatment (p = 0.0000), and marginally significant in the fin-o1 treatment (p = 0.0773). However, in the aud-o2 treatment the estimated average compliance rate is indeed higher in the no-voting stage than in the voting stage (87.09% vs. 72.46%), indicating that voting has a negative impact on compliance. Hence, the democracy effect exists only in the fin-o1 treatment if we look at the observations from all rounds.

Finally, the *p*-values of S2p4 indicate the test results of the equality between the compliance rates of subjects who vote for the high rate and are actually applied the low rate and the compliance rates of subjects in the no-voting stage. The differences are all insignificant in all six treatments.

Table 5.2.1 reports that the democracy effect originally found in the fin-o1 treatment vanishes if we look at the central ten rounds of the game, i.e., round 6 through round 15 (p = 0.6798 for S2p3). Though S2p3 remains highly significant for athe aud-o2 treatment (p =

0.0113), as indicated previously, this does not support the democracy effect since the estimated average compliance rate of subjects who vote for a low audit rate and are actually applied the low audit rate in the voting stage is instead lower than the estimated average compliance rate of subjects in the no-voting stage (59.79% vs. 71.91%). All other estimated results reported in Table 5.2.1 are similar to those in Table 5.1.1. Furthermore, the estimated results shown in both Table 5.1.2 and 5.2.2 are similar to those in 4.2.2. The above results associated with the regression estimations are summarized in Results 4 and 5.

Result 4: The compliance rate decay significantly across rounds in all six treatments. Income and gender generally have no significant impact on compliance. The attitude toward risks has a significant and negative impact on compliance, especially when it is the audit rate to be voted. Age and having taken economics course(s) have significant and negative impacts on compliance in some treatments. The amount of money donated to charities has a significant and positive effect on compliance in only the fin-o1 treatment.

Result 5: The democracy effect exists only in the fin-o1 treatments if we look at the entire twenty rounds of the experiment. The experimental evidence does not support the democracy effect if we look at only the central ten rounds of the experiment. The latter result holds regardless of which fiscal variable to be voted and whether the voting process is placed in the first or the latter ten rounds.

5. CONCLUSION

This paper investigates experimentally individuals' compliant behavior when the tax, audit, and fine rates are determined by majority voting. We apply Bó, Foster, and Putterman's (2010) approach by adding a stage of computer decision after voting to control for the self-selection problem. In addition, we control for the order effect by switching the order of the rounds with voting and the rounds without voting. The main findings of our paper are the following. First, the compliance rate decay significantly across rounds in all six treatments. Income and gender generally have no significant impact on compliance. The attitude toward risks has a significant and negative impact on compliance, especially when it is the audit rate to be voted. Second, subjects generally prefer a less severe auditing environment. They prefer a low tax, audit, or fine rate to a high counterpart. Third and most importantly, our major finding does not support the democracy effect. That is, allowing subjects to vote for the tax, audit, and fine rates do not have a positive and significant impact on compliance. This is because although the level of compliance may increase temporarily right after the addition of voting, the decay in compliance eventually outweighs this added stimulus of voting. Hence, on policy implications, our experimental evidence do not suggest using a democratic process to determine the magnitudes of these fiscal variables, at least under the framework our experimental design.

TABLE 1—FRAMEWORK OF THE EXPERIMENT AND TAX PARAMETERS

Treatment	Tax-o1	Tax-o2	Aud-o1	Aud-o2	Fin-o1	Fin-o2
Rounds 1–10						
Voting	no	yes: on *t*	no	yes: on *p*	no	yes: on *f*
Tax rate (*t*)	0.2	0.2 vs. 0.4	0.2	0.2	0.2	0.2
Audit probability (*p*)	0.1	0.1	0.1	0.1 vs. 0.4	0.1	0.1
Fine rate (*f*)	3	3	3	3	3	3 vs. 6
Rounds 11–20						
Voting	yes: on *t*	no	yes: on *p*	no	yes: on *f*	no
Tax rate (*t*)	0.2 vs. 0.4	0.2	0.2	0.2	0.2	0.2
Audit probability (*p*)	0.1	0.1	0.1 vs. 0.4	0.1	0.1	0.1
Fine rate (*f*)	3	3	3	3	3 vs. 6	3

TABLE 2—INDIVIDUAL CHARACTERISTICS

Treatment	Tax-o1	Tax-o2	Aud-o1	Aud-o2	Fin-o1	Fin-o2	All
Female	0.7708	0.6458	0.8125	0.7708	0.7708	0.6875	0.7431
	(0.4247)	(0.4833)	(0.3944)	(0.4247)	(0.4247)	(0.4684)	(0.4370)
Class	2.2917	2.125	1.875	2.1458	2.1458	2.0833	2.1111
	(1.1291)	(1.0442)	(0.8903)	(1.0516)	(1.0717)	(1.0883)	(1.0449)
Age	20.1667	19.9167	19.75	20.25	20.0833	19.8958	20.0104
	(1.4192)	(1.3182)	(1.0417)	(1.2965)	(1.2520)	(1.1893)	(1.2569)
Taken econ course(s)	0.8125	0.7917	0.7917	0.75	0.7083	0.8542	0.7847
	(0.3944)	(0.4104)	(0.4104)	(0.4376)	(0.4593)	(0.3567)	(0.4110)
Donation	2	1.9375	2.0417	1.9792	1.9583	2.625	2.0903
	(0.9676)	(0.7553)	(0.9444)	(0.8627)	(0.8495)	(1.2820)	(0.9783)
Risk-taking	5.4792	5.1458	5.2708	5.6042	4.6875	4.8542	5.1736
	(2.0935)	(2.1237)	(2.2096)	(2.3039)	(2.3078)	(2.3519)	(2.2340)

Standard deviations are in parentheses.

TABLE 3—DESCRIPTIVE STATISTICS

3.1 The tax-o1 and tax-o2 treatments

	Tax-o1		Tax-o2	
	Rounds 1–10	Rounds 11–20	Rounds 1–10	Rounds 11–20
(1) Average compliance rate	0.5439 (0.2150)	0.4078 (0.2404)	0.6048 (0.2272)	0.4482 (0.2721)
(2) Percentage of subjects declaring zero income	9.58% (0.2042)	15.42% (0.2843)	6.04% (0.1943)	9.17% (0.2305)
(3) Percentage of subjects fully complying	8.96% (0.1716)	7.5% (0.1804)	10.63% (0.2453)	7.92% (0.2031)
(4) Average earnings (points)	107.8925 (6.4893)	106.9104 (12.5992)	115.0925 (8.4506)	105.155 (8.9406)
(5) Percentage of subjects voting for t = 0.4 in the voting stage	–	49.58% (0.4084)	43.33% (0.3652)	–
(6) Percentage of time computer adopts the group decision	–	73.33% (0.1342)	62.5% (0.1212)	–
(7) Percentage of subjects being applied t = 0.4 in the voting stage	–	47.5% (0.1792)	45.83% (0.1569)	–
(8) Average compliance rate conditional on voting for t = 0.4	–	0.4337 (0.2720)	0.6275 (0.2578)	–
(9) Average compliance rate conditional on voting for t = 0.2	–	0.4052 (0.2312)	0.5845 (0.2345)	–
(10) Average compliance rate conditional on being applied t = 0.4	–	0.4812 (0.2824)	0.6087 (0.2509)	–
(11) Average compliance rate conditional on being applied t = 0.2	–	0.3574 (0.2463)	0.6090 (0.2506)	–
(12) Average compliance rate for subjects voting for t = 0.4 and t = 0.4 is applied	–	0.4756 (0.2866)	0.6257 (0.2695)	–
(13) Average compliance rate for subjects voting for t = 0.2 and t = 0.4 is applied	–	0.4614 (0.2855)	0.6050 (0.2592)	–
(14) Average compliance rate for subjects voting for t = 0.4 and t = 0.2 is applied	–	0.3570 (0.2847)	0.6562 (0.2458)	–
(15) Average compliance rate for subjects voting for t = 0.2 and t = 0.2 is applied	–	0.3728 (0.2293)	0.5838 (0.2604)	–

Note: The observations are the averages of all subjects' average choices over the period specified. The numbers in parentheses are the standard deviations of all subjects' average choices over that period.

3.2 The aud-o1 and aud-o2 treatments

	Aud-o1		Aud-o2	
	Rounds 1–10	Rounds 11–20	Rounds 1–10	Rounds 11–20
(1) Average compliance rate	0.6354 (0.2592)	0.5540 (0.2537)	0.5515 (0.2663)	0.3636 (0.2855)
(2) Percentage of subjects declaring zero income	5.63% (0.1749)	8.33% (0.2186)	17.5% (0.2892)	26.04% (0.3999)
(3) Percentage of subjects fully complying	14.58% (0.2939)	16.25% (0.2915)	14.79% (0.2449)	7.71% (0.1640)
(4) Average earnings (points)	110.99 (6.0136)	106.5279 (7.4026)	107.7475 (5.8442)	104.1296 (8.5336)
(5) Percentage of subjects voting for *p* = 0.4 in the voting stage	–	31.04% (0.3932)	22.29% (0.3197)	–
(6) Percentage of time computer adopts the group decision	–	67.5% (0.1756)	71.67% (0.1521)	–
(7) Percentage of subjects being applied *p* = 0.4 in the voting stage	–	30.83% (0.1761)	24.17% (0.1471)	–
(8) Average compliance rate conditional on voting for *p* = 0.4	–	0.6592 (0.2688)	0.6931 (0.2084)	–
(9) Average compliance rate conditional on voting *p* for = 0.1	–	0.5020 (0.2472)	0.5153 (0.2797)	–
(10) Average compliance rate conditional on being applied *p* = 0.4	–	0.787 (0.2262)	0.7705 (0.2479)	–
(11) Average compliance rate conditional on being applied *p* = 0.1	–	0.4531 (0.2825)	0.4835 (0.2929)	–
(12) Average compliance rate for subjects voting for *p* = 0.4 and *p* = 0.4 is applied	–	0.8336 (0.2075)	0.8346 (0.2194)	–
(13) Average compliance rate for subjects voting for *p* = 0.1 and *p* = 0.4 is applied	–	0.7494 (0.2488)	0.7647 (0.2526)	–
(14) Average compliance rate for subjects voting for *p* = 0.4 and *p* = 0.1 is applied	–	0.5548 (0.3101)	0.6153 (0.2623)	–
(15) Average compliance rate for subjects voting for *p* = 0.1 and *p* = 0.1 is applied	–	0.3958 (0.2583)	0.4450 (0.3067)	–

Note: The observations are the averages of all subjects' average choices over the period specified. The numbers in parentheses are the standard deviations of all subjects' average choices over that period.

3.3 The fin-o1 and fin-o2 treatments

	Fin-o1		Fin-o2	
	Rounds 1–10	Rounds 11–20	Rounds 1–10	Rounds 11–20
(1) Average compliance rate	0.5633 (0.2415)	0.4468 (0.2560)	0.5681 (0.2569)	0.4047 (0.2626)
(2) Percentage of subjects declaring zero income	9.79% (0.2068)	16.67% (0.2846)	13.54% (0.2539)	21.46% (0.3003)
(3) Percentage of subjects fully complying	15.83% (0.2827)	14.58% (0.2601)	19.79% (0.2646)	11.04% (0.1949)
(4) Average earnings (points)	108.2571 (8.1316)	104.6079 (7.8425)	108.2342 (7.3459)	104.3917 (7.0479)
(5) Percentage of subjects voting for $f = 6$ in the voting stage	–	42.5% (0.4097)	35.83% (0.3847)	–
(6) Percentage of time computer adopts the group decision	–	75% (0.1473)	73.33% (0.0883)	–
(7) Percentage of subjects being applied $f = 6$ in the voting stage	–	40.83% (0.1724)	33.33% (0.1742)	–
(8) Average compliance rate conditional on voting for $f = 6$	–	0.4862 (0.3012)	0.6195 (0.2975)	–
(9) Average compliance rate conditional on voting for $f = 3$	–	0.4198 (0.2609)	0.5191 (0.2634)	–
(10) Average compliance rate conditional on being applied $f = 6$	–	0.5832 (0.2990)	0.7123 (0.2620)	–
(11) Average compliance rate conditional on being applied $f = 3$	–	0.3468 (0.2807)	0.5134 (0.2839)	–
(12) Average compliance rate for subjects voting for $f = 6$ and $f = 6$ is applied	–	0.5814 (0.2903)	0.7350 (0.2894)	–
(13) Average compliance rate for subjects voting for $f = 3$ and $f = 6$ is applied	–	0.5396 (0.3354)	0.6979 (0.2567)	–
(14) Average compliance rate for subjects voting for $f = 6$ and $f = 3$ is applied	–	0.3681 (0.3512)	0.5344 (0.3427)	–
(15) Average compliance rate for subjects voting for $f = 3$ and $f = 3$ is applied	–	0.3278 (0.2842)	0.4863 (0.2862)	–

Note: The observations are the averages of all subjects' average choices over the period specified. The numbers in parentheses are the standard deviations of all subjects' average choices over that period.

TABLE 4—THE RESULTS OF FIXED-EFFECT OLS: SPECIFICATION 1

4.1 WITH OBSERVATIONS FROM ALL ROUNDS

4.1.1 compliance rate = $x_{it}\beta + \mu_i + \varepsilon_{it}$

	tax-o1	tax-o2	aud-o1	aud-o2	fin-o1	fin-o2
No-voting stage	48.4900** (7.0403)	25.3432** (6.2502)	73.8207** (5.7704)	86.5757** (7.0381)	95.3144** (7.3447)	66.1644** (7.3358)
Round	–1.9049** (0.2795)	–1.5860** (0.2593)	–1.5829** (0.2334)	–2.4772** (0.2782)	–2.7390** (0.2987)	–1.4463** (0.3016)
Round^2	0.0739** (0.0271)	0.0258 (0.0250)	0.0725** (0.0226)	0.0779** (0.0269)	0.0237 (0.0288)	0.0844** (0.0291)
True income	–0.0726** (0.0365)	–0.0214 (0.0337)	–0.0578* (0.0307)	–0.0272 (0.0361)	–0.0124 (0.0389)	–0.0537 (0.0394)
Vote for high rate	–3.4909 (2.8147)	–1.1126 (2.5574)	–1.8947 (2.5636)	3.8298 (3.2268)	–4.0614 (3.0671)	–1.0789 (3.1408)
High rate applied	61.0472** (7.4753)	26.8079** (6.6498)	104.0498** (6.2119)	105.2370** (7.2696)	126.6290** (7.8697)	80.1713** (7.8517)
Low rate applied	49.3858** (7.4847)	25.1043** (6.4776)	71.2627** (6.0567)	72.6438** (7.0026)	102.8341** (7.6517)	63.8291** (7.5605)
Observations	960	960	960	960	960	960
S1p1	0.8052	0.9431	0.3675	0.0000	0.0426	0.5310

Notes: The dependent variable is compliance rate. Standard errors are in parentheses. The notation ** denotes 5% significance level and * denotes 10% significance level. S1p1 is the *p*-value of the test for the equality between the compliance rates of the no-voting stage and the compliance rates of the voting stage when the rate actually applied is low.

4.1.2 EXPLAINING THE FIXED EFFECTS— $\mu_i = \omega_i \gamma + e_i$

	tax-o1	tax-o2	aud-o1	aud-o2	fin-o1	fin-o2
Female	8.3942 (6.3310)	9.4742 (6.9558)	8.9371 (7.0516)	1.1266 (8.9839)	–2.5519 (6.5239)	6.9051 (7.7868)
Risk-taking	–3.6683** (1.3219)	–2.7024 (1.7376)	–5.1434** (1.3608)	–3.1788* (1.6272)	–1.6439 (1.3254)	–3.6339** (1.3823)
Age	1.4886 (2.2507)	–3.8828 (2.3144)	1.5966 (3.0489)	–6.5444** (2.8539)	–3.0448 (2.6943)	–6.0408 (3.9014)
Econ	–12.0047 (7.9900)	–7.0448 (8.2979)	–18.2287** (8.7328)	–6.9888 (7.2228)	–10.6680* (5.9731)	–0.3621 (8.9131)
Donation	4.1651 (3.3243)	–1.0367 (4.5235)	5.0151 (3.5367)	5.3237 (3.7964)	12.5602** (2.5280)	–0.3986 (3.1091)
Constant	–14.9685 (44.5650)	92.7054* (49.2715)	–7.4921 (59.3486)	144.1756** (58.2793)	53.7814 (54.4691)	134.4339* (77.6586)
R^2	0.2100	0.1701	0.3733	0.2226	0.3080	0.1763
Observations	48	48	48	48	48	48

Notes: Standard errors are in parentheses. The notation ** denotes 5% significance level and * denotes 10% significance level.

4.2 WITH OBSERVATIONS FROM ROUNDS 6–15

4.2.1 compliance rate = $x_{it}\beta + \mu_i + \varepsilon_{it}$

	tax-o1	tax-o2	aud-o1	aud-o2	fin-o1	fin-o2
No-voting stage	74.8693** (8.9205)	63.1395** (9.1038)	87.8736** (7.9574)	72.7349** (9.9448)	59.8587** (10.6645)	78.5821** (11.0670)
Round	–1.5379** (0.7791)	–1.2680* (0.7226)	–2.5832** (0.6834)	–1.9594** (0.8070)	–1.8611** (0.8738)	–0.7963 (0.8745)
Round^2	0.0016 (0.1488)	–0.0459 (0.1377)	0.2513* (0.1312)	0.0089 (0.1535)	0.1805 (0.1670)	0.1773 (0.1670)
True income	–0.0653 (0.0500)	–0.0314 (0.0466)	–0.0555 (0.0452)	–0.0324 (0.0514)	–0.0585 (0.0560)	–0.0617 (0.0573)
Vote for high rate	–5.3983 (3.8442)	–2.8893 (3.5251)	–3.5117 (3.6219)	2.2503 (4.7688)	–0.9765 (4.4304)	0.2437 (4.7121)
High rate applied	84.2109** (9.5322)	66.5032** (9.1794)	118.1157** (8.6131)	95.6164** (10.8841)	85.4649** (11.1287)	97.3472** (11.3343)
Low rate applied	75.9829** (9.1353)	67.0194** (9.0482)	90.4694** (8.4342)	61.0880** (10.1012)	62.2287** (11.2152)	79.3970** (10.9516)
Observations	480	480	480	480	480	480
S1p1	0.8136	0.4085	0.5303	0.0141	0.6527	0.8775

Notes: The dependent variable is compliance rate. Standard errors are in parentheses. The notation ** denotes 5% significance level and * denotes 10% significance level. S1p1 is the *p*-value of the test for the equality between the compliance rates of the no-voting stage and the compliance rates of the voting stage when the rate actually applied is low.

4.2.2 EXPLAINING THE FIXED EFFECTS— $\mu_i = \omega_i \gamma + e_i$

	tax-o1	tax-o2	aud-o1	aud-o2	fin-o1	fin-o2
Female	7.8282 (7.2601)	8.0260 (7.6919)	8.2303 (6.5964)	1.4870 (9.8615)	–4.4206 (7.2062)	7.5537 (8.3211)
Risk-taking	–4.1316** (1.4395)	–2.9415 (1.8792)	–4.7057** (1.3907)	–3.6561** (1.6570)	–1.6361 (1.4602)	–3.4833** (1.4986)
Age	1.8084 (2.6804)	–4.5421* (2.6688)	0.5896 (3.3280)	–6.6866** (2.9230)	–2.0546 (2.7449)	–5.6249 (4.1605)
Econ	–13.9015 (8.4265)	–8.1938 (9.4543)	–19.1466* (10.0040)	–6.0740 (7.4275)	–17.3939** (6.1024)	0.2002 (9.8435)
Donation	5.3216 (3.8321)	–0.5689 (4.9564)	5.3852 (4.0946)	5.5889 (4.0490)	13.9321** (2.6216)	–0.7593 (3.2917)
Constant	–19.2142 (53.3211)	108.0061* (56.9401)	10.6337 (64.6409)	148.2406** (59.8066)	37.3777 (54.4439)	125.4494 (82.2772)
R^2	0.2072	0.1572	0.3246	0.2172	0.3525	0.1492
Observations	48	48	48	48	48	48

Notes: Standard errors are in parentheses. The notation ** denotes 5% significance level and * denotes 10% significance level.

TABLE 5—THE RESULTS OF FIXED-EFFECT OLS: SPECIFICATION 2

5.1 WITH OBSERVATIONS FROM ALL ROUNDS

5.1.1 compliance rate = $x_{it}\beta + \mu_i + \varepsilon_{it}$

	tax-o1	tax-o2	aud-o1	aud-o2	fin-o1	fin-o2
No-voting stage	48.7943** (7.0349)	25.2179** (6.2625)	74.4603** (5.7591)	87.0896** (7.0307)	78.6984** (7.4092)	66.1462** (7.3389)
Round	–1.9091** (0.2792)	–1.5862** (0.2594)	–1.5677** (0.2328)	–2.4774** (0.2778)	–2.7511** (0.2987)	–1.4411** (0.3019)
Round^2	0.0730** (0.0270)	0.0257 (0.0250)	0.0743** (0.0225)	0.0780** (0.0269)	0.0227 (0.0288)	0.0840** (0.0291)
True income	–0.0714* (0.0365)	–0.0210 (0.0338)	–0.0581* (0.0306)	–0.0304 (0.0360)	–0.0108 (0.0389)	–0.0544 (0.0395)
Vote for high rate and high rate applied	56.2613** (7.1228)	25.2134** (6.4092)	98.5984** (6.0166)	104.4199** (8.0389)	104.3767** (8.0205)	78.2969** (7.5658)
Vote for low rate and high rate applied	64.1375** (7.6806)	27.1639** (6.7233)	107.5370** (6.3472)	108.6910** (7.4548)	112.4620** (7.6626)	81.1167** (8.0693)
Vote for high rate and low rate applied	48.5925** (7.4526)	24.4185** (6.4255)	72.9005** (5.9725)	80.1643** (7.8192)	84.4265** (8.1158)	63.5668** (7.5077)
Vote for low rate and low rate applied	48.2400** (7.5063)	24.7237** (6.5631)	70.6964** (6.0432)	72.4612** (6.9912)	85.3423** (7.6126)	63.5590** (7.5820)
Observations	960	960	960	960	960	960
S2p1	0.0383	0.5696	0.0187	0.4058	0.0563	0.5427
S2p2	0.9218	0.9279	0.4675	0.0400	0.8106	0.9984
S2p3	0.8817	0.8852	0.1899	0.0000	0.0773	0.4915
S2p4	0.9602	0.8343	0.6525	0.1165	0.1988	0.5644

Notes: The dependent variable is compliance rate. Standard errors are in parentheses. The notation ** denotes 5% significance level and * denotes 10% significance level.

S2p1 is the *p*-value of the test for the equality between the compliance rates of individuals voting for a low rate and the compliance rates of those voting for a high rate given the actual rate being applied is high during the voting stage.

S2p2 is the *p*-value of the test for the equality between the compliance rates of individuals voting for a high rate and the compliance rates of those voting for a low rate given the actual rate being applied is low during the voting stage.

S2p3 is the *p*-value of the test for the equality between the compliance rates of individuals voting for a low rate and the actual rate being applied is low and the compliance rates of individuals in

the no-voting stage.
S2p4 is the *p*-value of the test for the equality between the compliance rates of individuals voting for a high rate and the actual rate being applied is low and the compliance rates of individuals in the no-voting stage.

5.1.2 EXPLAINING THE FIXED EFFECTS— $\mu_i = \omega_i \gamma + e_i$

	tax-o1	tax-o2	aud-o1	aud-o2	fin-o1	fin-o2
Female	8.3639 (6.3056)	9.4822 (6.9573)	9.0035 (7.0568)	1.0146 (9.0424)	–2.5492 (6.5287)	6.8734 (7.7846)
Risk-taking	–3.6490** (1.3097)	–2.6982 (1.7391)	–5.0947** (1.3646)	–3.1759* (1.6326)	–1.6045 (1.3287)	–3.6317** (1.3824)
Age	1.4779 (2.2420)	–3.8845 (2.3125)	1.5865 (3.0427)	–6.4532** (2.8647)	–3.0534 (2.6938)	–6.0367 (3.9025)
Econ	–12.0665 (7.9450)	–7.0663 (8.3026)	–18.4542** (8.7210)	–6.9603 (7.2656)	–10.7037* (5.9716)	–0.3103 (8.9036)
Donation	4.1571 (3.3175)	–1.0278 (4.5263)	5.0191 (3.5286)	5.2785 (3.8099)	12.5102** (2.5313)	–0.4064 (3.1086)
Constant	–14.7684 (44.3925)	92.7122* (49.2523)	–7.4335 (59.3146)	142.4662** (58.4443)	53.8909 (54.4632)	134.3394* (77.6668)
R^2	0.2105	0.1701	0.3719	0.2181	0.3057	0.1761
Observations	48	48	48	48	48	48

Notes: Standard errors are in parentheses. The notation ** denotes 5% significance level and * denotes 10% significance level.

5.2 WITH OBSERVATIONS FROM ROUNDS 6–15

5.2.1 compliance rate = $x_{it}\beta + \mu_i + \varepsilon_{it}$

	tax-o1	tax-o2	aud-o1	aud-o2	fin-o1	fin-o2
No-voting stage	74.7112** (8.9083)	22.6162** (8.4120)	3.9181 (8.4090)	71.9083** (9.9987)	62.7907** (10.1861)	51.0277** (10.2742)
Round	–1.4657* (0.7795)	–1.2826* (0.7231)	–2.5573** (0.6831)	–1.9557** (0.8073)	–1.8683** (0.8753)	–0.7899 (0.8742)
Round^2	0.0206 (0.1491)	–0.0438 (0.1378)	0.2536* (0.1310)	0.0075 (0.1536)	0.1791 (0.1673)	0.1754 (0.1670)
True income	–0.0655 (0.0499)	–0.0310 (0.0466)	–0.0519 (0.0453)	–0.0345 (0.0514)	–0.0584 (0.0561)	–0.0646 (0.0574)
Vote for high rate and high rate applied	76.2453** (9.7121)	21.9978** (8.5085)	27.3720** (9.5312)	94.4057** (11.1669)	87.0091** (10.9874)	67.7520** (10.7343)
Vote for low rate and high rate applied	87.1136** (9.7146)	27.5008** (9.0173)	36.5426** (9.1129)	96.3604** (10.9253)	89.0562** (11.2565)	73.1142** (11.4498)
Vote for high rate and low rate applied	72.9388** (9.6830)	25.1127** (8.5727)	5.0919 (9.1898)	64.6172** (9.9439)	64.7639** (10.7200)	54.7727** (10.2857)
Vote for low rate and low rate applied	73.7174** (9.2473)	25.5563** (8.8532)	5.5447 (8.6202)	59.7863** (10.2271)	64.9876** (10.7012)	50.8343** (10.5289)
Observations	480	480	480	480	480	480
S2p1	0.0411	0.2491	0.1016	0.7794	0.7479	0.4307
S2p2	0.8746	0.9237	0.9160	0.3974	0.9674	0.4909
S2p3	0.8400	0.5432	0.6983	0.0113	0.6798	0.9712
S2p4	0.7352	0.6286	0.8170	0.2409	0.7619	0.5640

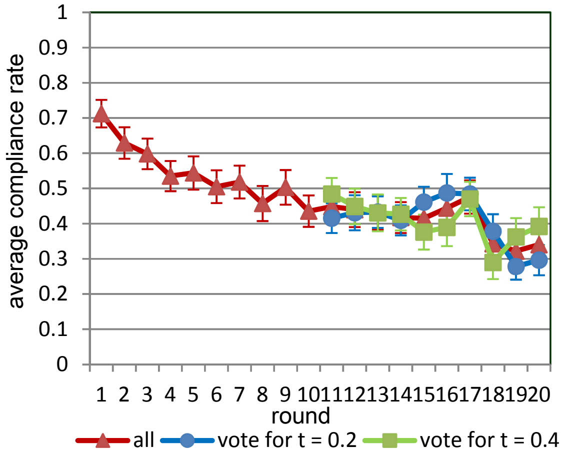
Notes: The dependent variable is compliance rate. Standard errors are in parentheses. The notation ** denotes 5% significance level and * denotes 10% significance level.

5.2.2 EXPLAINING THE FIXED EFFECTS— $\mu_i = \omega_i \gamma + e_i$

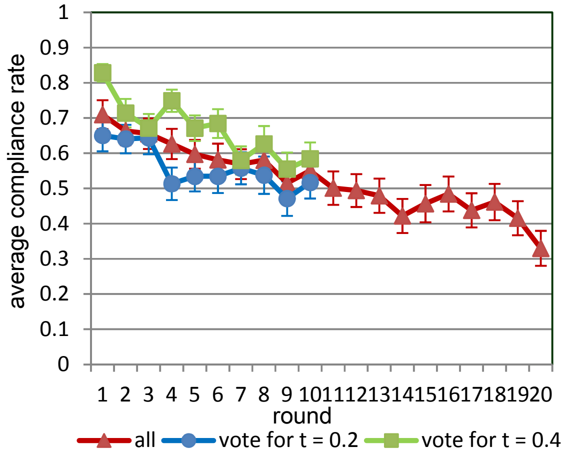
	tax-o1	tax-o2	aud-o1	aud-o2	fin-o1	fin-o2
Female	7.7901 (7.2812)	8.0537 (7.7089)	8.2590 (6.5986)	1.4941 (9.8810)	–4.4413 (7.2149)	7.3322 (8.2876)
Risk-taking	–4.0927** (1.4258)	–2.9184 (1.8893)	–4.6532** (1.3870)	–3.6504** (1.6583)	–1.6259 (1.4608)	–3.4780** (1.4949)
Age	1.7750 (2.6774)	–4.5520* (2.6575)	0.5030 (3.3149)	–6.5995** (2.9240)	–2.0628 (2.7446)	–5.6270 (4.1519)
Econ	–13.9489 (8.3869)	–8.1020 (9.4512)	–19.1291* (9.9774)	–6.0455 (7.4512)	–17.3898** (6.1020)	0.7433 (9.8124)
Donation	5.3429 (3.8596)	–0.4906 (4.9596)	5.4025 (4.0889)	5.4934 (4.0529)	13.9405** (2.6216)	–0.8120 (3.2936)
Constant	–18.7289 (53.2364)	107.8417* (56.7977)	11.9952 (64.5199)	146.6068** (59.8118)	37.4910 (54.4349)	125.2916 (82.1052)
R^2	0.2052	0.1565	0.3228	0.2145	0.3524	0.1479
Observations	48	48	48	48	48	48

Notes: Standard errors are in parentheses. The notation ** denotes 5% significance level and * denotes 10% significance level.

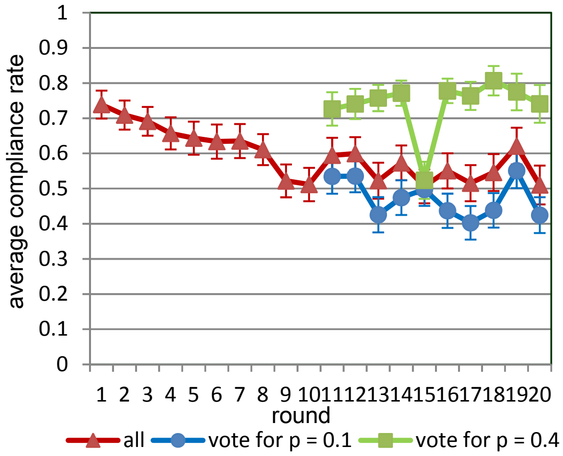
Figure 1. Average Compliance Rates Conditional on Subjects' Voting Decisions



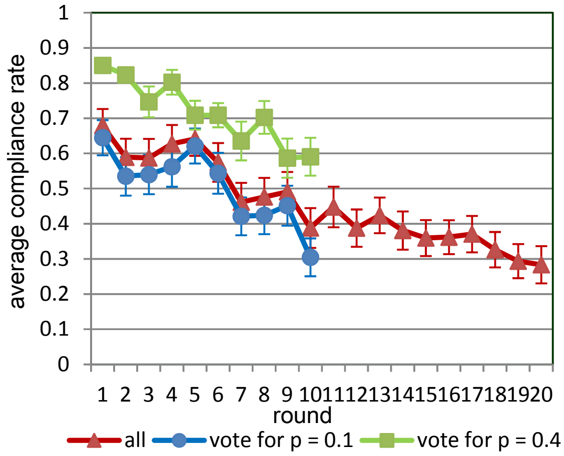
(a) The tax-o1 treatment



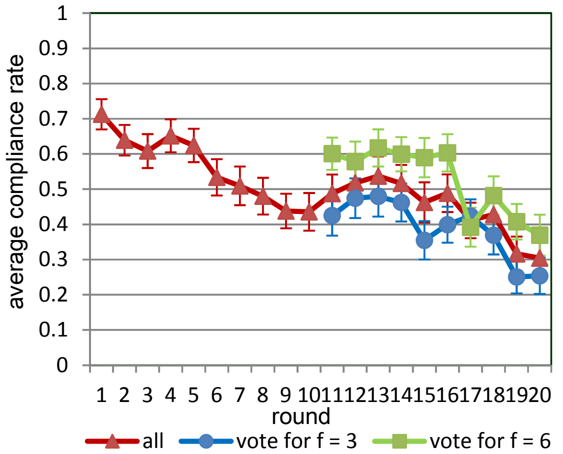
(b) The tax-o2 treatment



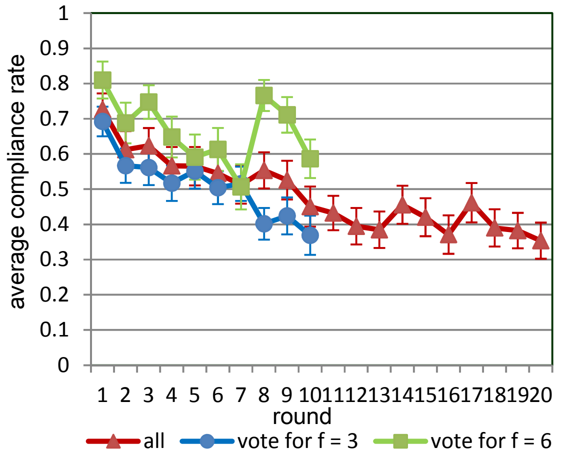
(c) The aud-o1 treatment



(d) The aud-o2 treatment

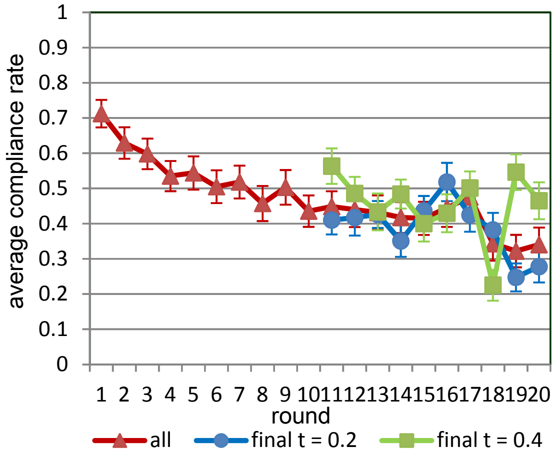


(e) The fin-o1 treatment

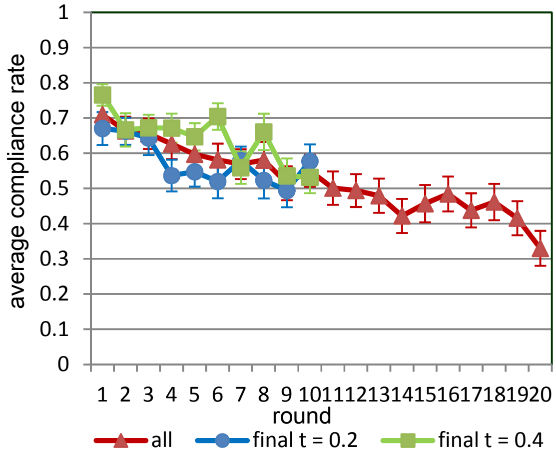


(f) The fin-o2 treatment

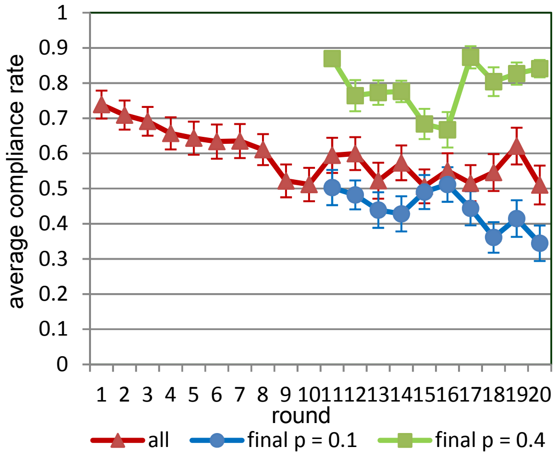
Figure 2. Average Compliance Rates Conditional on the Outcome Being Applied



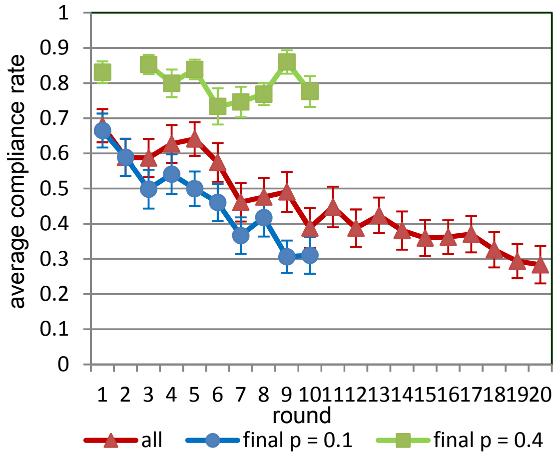
(a) The tax-o1 treatment



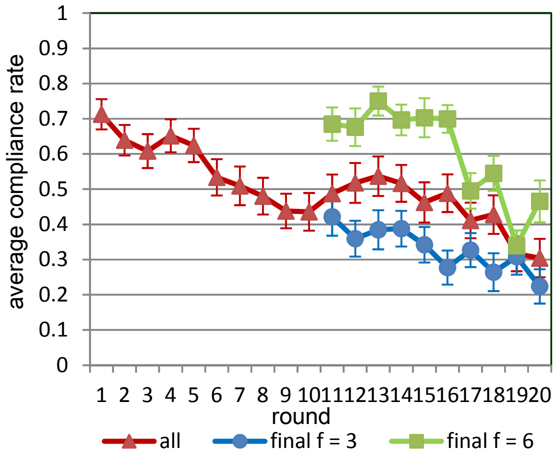
(b) The tax-o2 treatment



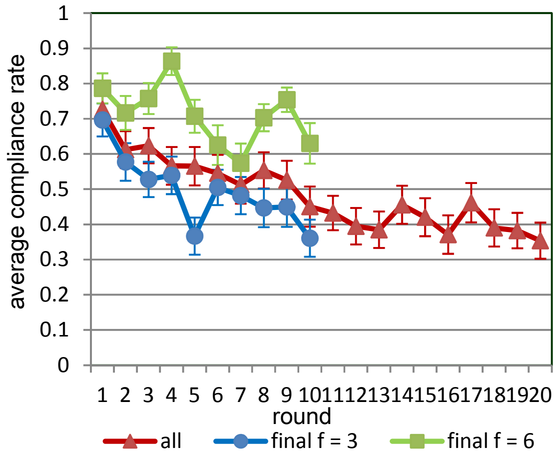
(c) The aud-o1 treatment



(d) The aud-o2 treatment



(e) The fine-o1 treatment



(f) The fine-o2 treatment

REFERENCES

Alm, James, Mark B. Cronshaw, and Michael McKee. 1993. "Tax Compliance with Endogenous Audit Selection Rules," *Kyklos*, 46(1), 27–45.

Allingham, Michael G. and Agnar Sandmo. 1972. "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics,* 1, 323–338.

Alm, James, Betty R. Jackson, and Michael McKee. 1992a. "Institutional Uncertainty and Taxpayer Compliance," *American Economic Review,* 82(4), 1018–1026.

Alm, James, Betty R. Jackson, and Michael McKee. 1992b. "Estimating the Determinants of Taxpayer Compliance with Experimental Data," *National Tax Journal,* 45(1), 107–114.

Alm, James, Betty R. Jackson, and Michael McKee. 1993. "Tax Exchange, Collective Decision Institutions, and Tax Compliance," *Journal of Economic Behavior and Organization,* 22(3), 285–303.

Alm, James, Gary H. McClelland, and William D. Schulze. 1992. "Why Do People Pay Taxes?" *Journal of Public Economics,* 48(1), 21–38.

Alm, James, Gary H. McClelland, and William D. Schulze. 1999. "Changing the Social Norm of Tax Compliance by Voting," *Kyklos*, 52, 2, 141-171.

Alm, James and Michael McKee. 2004. "Tax Compliance as a Coordination Game," *Journal of Economic Behavior and Organization,* 54(3), 297–312.

Alm, James and Michael McKee. 2006. "Audit Certainty, Audit Productivity, and Taxpayer Compliance," *National Tax Journal*, 59(4), 801–816.

Alm, James, Michael McKee, and William Beck. 1990. "Amazing Grace: Tax Amnesties and Compliance," *National Tax Journal*, 43(1), 23–37.

Alm, James, Isabel Sanchez, and Ana de Juan. 1995. "Economic and Noneconomic Factors in

Tax Compliance,” *Kyklos,* 48(1), 3–18.

Andreoni, James. 1988. “Why Free Ride: Strategies and Learning in Public Goods Experiments,” *Journal of Public Economics,* 37(3), 291–304.

Andreoni, James. 1993. “An Experimental Test of the Public-Goods Crowding-Out Hypothesis,” *American Economic Review,* 83(5), 1317–1327.

Andreoni, James. 1995. “Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation,” *Quarterly Journal of Economics,* 110(1), 1–21.

Andreoni, James, Brian Erard, and Jonathan Feinstein. 1998. “Tax Compliance,” *Journal of Economic Literature* 36(2), 818-860.

Andreoni, James, William Harbaugh, and Lise Vesterlund. 2003. “The Carrot or the Stick: Rewards, Punishments, and Cooperation,” *American Economic Review,* 93(3), 893–902.

Andreoni, James and John H. Miller. 1993. “Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence,” *Economic Journal*, 103, 570–585.

Becker, W., H. Büchner, and S. Sleeking. 1987. “The Impact of Public Transfer Expenditures on Tax Evasion: An Experimental Approach,” *Journal of Public Economics,* 34(2), 243–252.

Bó, Pedro Dal, Andrew Foster, and Louis Putterman (2010), Institutions and Behavior: Experimental Evidence on the Effects of Democracy, *American Economic Review,* 100 (5), December, 2205-2229.

Carpenter, Jeffrey P. 2007. “Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods,” *Games and Economic Behavior,* 60(1), 31–51.

Clark, Jeremy, Lana Friesen, and Andrew Muller. 2004. “The Good, the Bad, and the Regulator: An Experimental Test of Two Conditional Audit Schemes,” *Economic Inquiry,* 42(1), 69–87.

Collins, Julie H. and R. David Plumlee. 1991. “The Taxpayer’s Labor and Reporting Decision: The Effects of Audit Schemes,” *The Accounting Review,* 66(3), 559–576.

Dawes, Robyn M. 1980. “Social Dilemmas,” *Annual Review of Psychology*, 31, 169–193.

Fehr, Ernst and Simon Gächter. 2000. “Cooperation and Punishment in Public Goods Experiments,” *American Economic Review,* 90(4), 980–994.

Fehr, Ernst and Simon Gächter. 2002. “Altruistic Punishment in Humans,” *Nature,* 415, 137–140.

Gächter, Simon, Elke Renner, and Martin Sefton. 2008. “The Long-Run Benefits of Punishment,” *Science,* 322(5907), 1510.

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach. 2006. “The Competitive Advantage of Sanctioning Institutions,” *Science,* 312(5770), 108–111.

Isaac, R. Mark and James M. Walker. 1988a. “Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism,” *Quarterly Journal of Economics,* 103(1), 179–200.

Isaac, R. Mark and James M. Walker. 1988b. “Communication and Free-Riding Behavior: The Voluntary Contributions Mechanism,” *Economic Inquiry,* 26(4), 585–608.

Isaac, R. Mark, James M. Walker, and Susan H. Thomas. 1984. “Divergent Evidence on Free-Riding: An Experimental Examination of Possible Explanations,” *Public Choice,* 43(2), 113–150.

Isaac, R. Mark, James M. Walker, and Arlington W. Williams. 1994. “Group Size and the Voluntary Provision of Public Goods: Experimental Evidence Utilizing Large Groups,” *Journal of Public Economics,* 54(1), 1–36.

Pommerehne, Werner W., Albert Hart and Bruno S. Frey. 1994. “Tax Morale, Tax Evasion, and the Choice of Tax Policy Instruments in Different Political Systems," *Public Finance/Finances Publiques*, 49 (Supplement), 52–69.

Putterman, Louis, Jean-Robert Tyran, and Kenju Kamei. 2011. "Public Goods and Voting on Formal Sanction Schemes," *Journal of Public Economics*, 95 (9–10), 1213-1222.

Sefton, Martin, Robert Shupp, and James M. Walker. 2007. "The Effect of Rewards and Sanctions in Provision of Public Goods," *Economic Inquiry*, 45(4), 671–690.

Spicer, Michael W. and Lee A. Becker. 1980. "Tax Inequity and Tax Compliance: An Experimental Approach," *National Tax Journal*, 33(2), 171–175.

國科會補助計畫衍生研發成果推廣資料表

日期:2013/08/08

國科會補助計畫	計畫名稱: 懲罰效果的兩個實驗研究
	計畫主持人: 徐麗振
	計畫編號: 100-2410-H-004-065-MY2　　學門領域: 數理與數量方法

無研發成果推廣資料

100 年度專題研究計畫研究成果彙整表

<table>
<tr><td colspan="3">計畫主持人：徐麗振</td><td colspan="5">計畫編號：100-2410-H-004-065-MY2</td></tr>
<tr><td colspan="8">計畫名稱：懲罰效果的兩個實驗研究</td></tr>
<tr><td colspan="3" rowspan="2">成果項目</td><td colspan="3">量化</td><td rowspan="2">單位</td><td rowspan="2">備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）</td></tr>
<tr><td>實際已達成數（被接受或已發表）</td><td>預期總達成數(含實際已達成數)</td><td>本計畫實際貢獻百分比</td></tr>
<tr><td rowspan="12">國內</td><td rowspan="4">論文著作</td><td>期刊論文</td><td>0</td><td>0</td><td>100%</td><td rowspan="3">篇</td><td></td></tr>
<tr><td>研究報告/技術報告</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>研討會論文</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專書</td><td>0</td><td>0</td><td>100%</td><td></td><td></td></tr>
<tr><td rowspan="2">專利</td><td>申請中件數</td><td>0</td><td>0</td><td>100%</td><td rowspan="2">件</td><td></td></tr>
<tr><td>已獲得件數</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>0</td><td>100%</td><td>件</td><td></td></tr>
<tr><td>權利金</td><td>0</td><td>0</td><td>100%</td><td>千元</td><td></td></tr>
<tr><td rowspan="4">參與計畫人力（本國籍）</td><td>碩士生</td><td>0</td><td>0</td><td>100%</td><td rowspan="4">人次</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="12">國外</td><td rowspan="4">論文著作</td><td>期刊論文</td><td>0</td><td>0</td><td>100%</td><td rowspan="3">篇</td><td></td></tr>
<tr><td>研究報告/技術報告</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>研討會論文</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專書</td><td>0</td><td>0</td><td>100%</td><td>章/本</td><td></td></tr>
<tr><td rowspan="2">專利</td><td>申請中件數</td><td>0</td><td>0</td><td>100%</td><td rowspan="2">件</td><td></td></tr>
<tr><td>已獲得件數</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>0</td><td>100%</td><td>件</td><td></td></tr>
<tr><td>權利金</td><td>0</td><td>0</td><td>100%</td><td>千元</td><td></td></tr>
<tr><td rowspan="4">參與計畫人力（外國籍）</td><td>碩士生</td><td>0</td><td>0</td><td>100%</td><td rowspan="4">人次</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
</table>

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	無

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

科技部補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估 ■達成目標 □未達成目標（請說明，以 100 字為限） □實驗失敗 □因故實驗中斷 □其他原因 說明：
2. 研究成果在學術期刊發表或申請專利等情形： 論文：□已發表 □未發表之文稿 □撰寫中 ■無 專利：□已獲得 □申請中 ■無 技轉：□已技轉 □洽談中 ■無 其他：（以 100 字為限）
3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限） 本計畫為兩年期計畫，每年計畫包含一個實驗經濟學的研究。第一年計畫探討懲罰，稅率，以及稽核率等財政變數對租稅順從的影響，以及若讓這些變數的高低經由多數決決定，是否能提高租稅順從水準。第二年計畫探討懲罰和獎勵兩種機制在 weakest-link 和 best-shot 兩種賽局中實驗參與者自願捐獻公共財水準的影響。兩年的研究計畫都包含嚴謹的實驗設計與實驗操作，得出的實驗資料亦經過嚴謹的統計分析，兩份研究的結果皆對政府政策提出有意義的建議。