Tulane Economics Working Paper Series

Declaring income versus declaring taxes in tax compliance experiments: Does the design of laboratory experiments affect the results?

Stephan Muehlbacher
Karl Landsteiner University of Health Sciences

Andre Hartmann
University of Vienna

Erich Kirchler
University of Vienna
erich.kirchler@univie.ac.at

James Alm
Tulane University
jalm@tulane.edu

Working Paper 2210
November 2022

Abstract

Laboratory experiments are frequently criticised, in part because of the sensitivity of the results to specific features of the design. This paper addresses an important question regarding the key aspect of the experimental environment: How should the dependent variable – participants' choices – be operationalised? For the specific context of laboratory research on income tax compliance, we compare the effects of the two most common operationalisation types: the declaration of gross income versus the declaration of tax payment. It is found that compliance is higher when participants indicate their tax payment than when they declare their income. It is also discovered that the effects of the three policy parameters of the economic model (tax rate, audit probability, and fine rate) are stronger when participants declare their taxes than when they declare their income. These results are relevant for interpreting prior and future experimental evidence on tax compliance and can explain some contradictory previous findings. More broadly, this study suggests that the results of laboratory experiments may depend on specific features of the experimental design, which proposes a strong need for more systematic methodological research.

Keywords: Laboratory experiments; experimental design; tax compliance; tax rate; audit probability; fine rate
JEL codes: B41; C90; C91; H26

Declaring income versus declaring taxes in tax compliance experiments: Does the design of laboratory experiments affect the results?

Stephan Muehlbacher[a]*, Andre Hartmann[b], Erich Kirchler[b, c] and James Alm[d]

[a]Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences, Krems, Austria; [b]Faculty of Psychology, University of Vienna, Vienna, Austria; [c]Institute for Advanced Studies, Vienna, Austria; [d]Department of Economics, Tulane University, New Orleans, USA

ABSTRACT
Laboratory experiments are frequently criticised, in part because of the sensitivity of the results to specific features of the design. This paper addresses an important question regarding the key aspect of the experimental environment: How should the dependent variable – participants' choices – be operationalised? For the specific context of laboratory research on income tax compliance, we compare the effects of the two most common operationalisation types: the declaration of gross income versus the declaration of tax payment. It is found that compliance is higher when participants indicate their tax payment than when they declare their income. It is also discovered that the effects of the three policy parameters of the economic model (tax rate, audit probability, and fine rate) are stronger when participants declare their taxes than when they declare their income. These results are relevant for interpreting prior and future experimental evidence on tax compliance and can explain some contradictory previous findings. More broadly, this study suggests that the results of laboratory experiments may depend on specific features of the experimental design, which proposes a strong need for more systematic methodological research.

KEYWORDS
Laboratory experiments; experimental design; tax compliance; tax rate; audit probability; fine rate

JEL CODES
B41; C90; C91; H26

CONTACT: Stephan Muehlbacher; stephan.muehlbacher@kl.ac.at; Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences, Krems, Austria.

Introduction

In the past decades, laboratory experiments (lab experiments hereinafter) have become a widespread method in economics and economic psychology. Lab experiments are often criticised for various reasons, and a particular concern is whether the results are robust to changes in the experimental design (for a detailed discussion, see Falk and Heckman 2009; Frechette and Schotter 2015; Levitt and List 2007). The similar issue is raised specifically for laboratory experiments on tax compliance (Alm, Bloomquist, and McKee 2015; Torgler 2002). A recent meta–analysis by Alm and Malézieux (2021) seeks to clarify some of the potential effects of experimental designs. Still, many questions remain unsolved. Indeed, there is little consensus about the 'state-of-art' of the experimental design for studying income tax compliance (Malézieux 2018; Muehlbacher and Kirchler 2016).

Perhaps the most significant element of the experimental environment is the operationalisation of the dependent variable(s). If the invariance principle – 'tacitly assumed' (Tversky and Kahneman 1986, 253) in the traditional normative decision theory – holds, it should not matter how preferences are elicited. However, there is overwhelming evidence for its violation. Such research finds, for instance, a disparity between willingness-to-pay versus willingness-to-accept (Horowitz and McConnell 2002), framing effects in risky choice (Steiger and Kühberger 2018) and in public good experiments (Cookson 2000), differences in eliciting responses directly ('hot play') or by choice from a presented vector of strategies ('cold play', see Brandts and Charness 2011), different levels of cooperation in 'give–some', 'take–some', 'keep–some', or 'leave–some' in social dilemma games (Haesevoets et al. 2019; Van Dijk and Wilke 2000), and only weak correlations among different methods to elicit risk preferences (Frey et al. 2017). For a specific context of tax compliance, possible effects of the

operationalisation have hardly been studied so far, although there are reasons to assume that those possible effects affect experimental results (Muehlbacher and Kirchler 2016). Hence, the question arises as how tax compliance should be operationalised in the laboratory.

In principal, two different approaches are discussed in the field. In some studies, compliance has been measured by asking participants to declare the amount of income they report to a tax authority (e.g. Alm, Jackson, and McKee 1992; Kirchler et al. 2009; Mittone 2006). In other researches, the task is designed to indicate the amount of taxes participants are paying (e.g. Bazart and Pickhardt 2011; Casal et al. 2016; Torgler and Schaltegger 2005). Table 1 shows the standard instructions for these two operationalisation types.

[Table 1 near here]

A meta–analysis of 70 lab experiments on tax compliance (Alm and Malézieux 2021) reports that the instruction of Type I (the income declaration) is applied in 67% of the studies, and 33% of the studies apply the instruction of Type II (the tax declaration). The results on the effects of operationalisation types are somewhat inconclusive. In the main analysis, operationalisation has no significant impact on compliance. However, it has the largest effect of all predictor variables studied in an additional analysis conducted as a robustness check (see Table 13 in Appendix of Alm and Malézieux 2021) based on a smaller sub–set of studies for which information about operationalisation types and policy variables (tax rate, audit probability, and fine rate) are available. When controlling for the effects of the policy variables, instructing participants to declare the amount of taxes to pay (Type II) is related to higher

compliance than when the task is to declare the amount of gross income (Type I). These contradictory results are difficult to interpret. One of the possible explanations is that the effect of operationalisation probably interacts with other features of the experimental design. Such moderation effects of the experimental environment are not considered in the meta–analysis, given its intended focus. Due to the ambiguous results, it is necessary to address the effect of operationalisation types more systematically.

Clarifying any potential effects of operationalisation is important for interpretating previous and future experimental results, and such research may help to improve the experimental designs. Further, although at presence declaring the gross income is a more realistic task as it mimics the taxpaying situation of self–employed taxpayers, declaring the tax payment may soon increase in its importance due to the digitalisation of the tax reporting process. In countries like Germany and the US, for instance, the electronic filing systems are becoming more popular. Such applications automatically compute and display the tax obligation that results from the information provided by users, so that the focus is turned to on the tax payment through the electronic tax reporting . Playing around with the input until one is satisfied with the resulting due is facilitated by such tax filing software. Hence, when this becomes a standardised procedure, the declaration of tax payment may increase its importance in the operationalisation of lab experiments.

A simple effect of operationalisation would not be overly problematic for the interpretation of experimental results, as it would merely shift the observed mean tax compliance up- or down. The interpretation of the treatment effects studied in an experiment would remain the same. More worrisome would be if the operationalisation types interacted with other variables, because this could mean that the existing experimental evidence is biased by false positive or negative results. The observation of

such interaction effects would seriously challenge the validity of a large part of the published results, such as the findings on the effects of the three policy parameters of the traditional economic model on income tax evasion (Allingham and Sandmo 1972). Although the meta–analysis (Alm and Malézieux 2021) reveals that across all studies the tax rate has a negative effect on compliance and that the audit probability and the fine rate have a positive effect on compliance, the observations made in each individual experiment are quite volatile. Indeed, the extant literature (e.g. Kirchler et al. 2010; Muehlbacher and Kirchler 2016) reveals that the tax rate has a negative effect in 71% of the experiments, a positive effect in 14%, and no effect in 14%; for the audit probability, a positive in 91% and no effect in 9% are observed; and for the fine rate, a positive in 50% and no effect in 50% are found. A more recent review of the literature (Malézieux 2018) reports similar divergent effects of the three policy parameters. There are several explanations for these (partly) contradicting results. However, if the size and/or direction of these effects depends on how tax compliance are operationalised, this can well explain the puzzling findings. Thus, in addition to examining the main effect, it seems crucial to explore possible interaction effects that the operationalisation types might have with other parameters of the experimental design, such as the tax rate, the audit probability, and the fine rate.

In this paper, we report the results of an incentivised laboratory experiment in which we compare the effects of the two most common operationalisation types on tax compliance. Further, we analyse whether the operationalisation of compliance moderates the effects of three policy variables (tax rate, audit probability, and fine rate) and income. We also conduct a web–based experiment without monetary incentives as a pilot study, which produces the similar results (the pilot study is summarised in Section A in Supplemental material). Before summarising and discussing the analysis, we

present the theoretical consideration about how operationalisation can affect tax compliance in the lab as well as our hypotheses.

Theoretical framework

Framing effects and similar violations of the invariance principle has led to the development of alternative and more realistic theories about how decisions are made. Among the more prominent theories is the prospect theory (Kahneman and Tversky 1979), which is also frequently applied to explain taxpayers' reporting decisions (Fochmann and Wolf 2019; Schepanski and Shearer 1995; Yaniv 1999). The core assumption in the prospect theory is that decision outcomes are evaluated as deviations from a neutral reference point, which may be constituted, for instance, by the status quo, expectations or external stimuli provided in the decision environment. The reference point determines whether outcomes are perceived as gains or losses, and, thus, affects the propensity for a risky choice such as evading taxes (for an overview of this paradigm in the tax compliance research, see Muehlbacher 2021).

Assumptions about the reference point are therefore crucial for applying the prospect theory in the analysis of tax compliance. For self–employed taxpayers, three potential reference points are discussed in the extant literature (Bernasconi, Corazzini, and Seri 2014; Copeland and Cuccia 2002; Dhami and Al–Nowaihi 2007; Kirchler et al. 2009; Kirchler and Maciejovsky 2001; Martinez–Vazquez, Harwood, and Larkins 1992; Muehlbacher and Kirchler 2013; Schepanski and Shearer 1995): gross income, expected net income, and current asset position (the last of which corresponds to the taxpayers' withholding–position that depends on whether and how much advance tax payments are made), each leading to different predictions for behaviour. Which of these potential reference points taxpayers actually employ is unclear, but the scarce empirical evidence suggests expected net income (Copeland and Cuccia 2002; Kirchler et al. 2009) and

current asset position (Kirchler and Maciejovsky 2001; Schepanski and Shearer 1995) as the most likely choices.

Similar issues arise when analysing tax compliance in lab experiments. Again, participants can either use the gross income or the expected net income as the reference in their reporting decisions. The third option – the current asset position, however, is slightly different because in experiments, typically, no prepayments are made that lead to over (or under) withholding of taxes. Hence, the current asset position in an experimental setting is either (again) the gross income or participants status quo when coming to the lab and before receiving the experimental payment (i.e. 0 or any other pre–experimental financial status). Which reference point is applied in the lab has not been systematically investigated. However, it seems likely that the operationalisation of tax compliance affects the decision frame and subsequently the willingness to take the risk of evading taxes (as hypothesised in Muehlbacher and Kirchler 2016). In the following, we propose two different mechanisms that lead to competing predictions. For the first mechanism, it is assumed that the operationalisation type determines participants' reference point by putting the focus either on the gained income or the loss caused by the tax payment. The second mechanism draws on the assumption that the different numerical information provided by the two types of instructions (the gross income versus the tax payment) is used as the reference.

We consider operationalisation Type I first: instructing participants to declare their income puts the focus on the gains achieved. The reference point used for this positive decision frame is (most likely) the participants' pre–experimental status quo. Then the choice is between a sure gain – the net income after paying a full tax due – and a risky but potentially larger gain obtained by evading taxes. For strictly positive outcomes, the prospect theory predicts risk aversion, which makes tax compliance more

likely to occur.[1] Turn to the operationalisation Type II instructions: instructing participants to declare their taxes. The focus here is on the loss that occurs due to the tax payment, which implies that the reference point is the gross income before paying taxes. The choice is between a strictly negative outcome – a sure loss by being compliant – and a potentially smaller loss by evading taxes. In the loss domain, the willingness to take the risk of evading taxes is high. To summarise the first mechanism, we suggest that operationalisation affects the reference point by putting the focus either on gains or on losses. If this is true, it means a consequence that Type I (the income declaration) produces a higher level of compliance than Type II (the tax declaration).

For the second mechanism that we propose, it is assumed that participants draw on a number stated in the instruction when choosing a reference for their reporting decision. In the instruction of Type I, participants read a number indicating their gross income (e.g. ECU1,000) and are instructed to declare this amount. If this number serves as the reference point, then the reporting decision occurs in the loss domain, and thus, in opposition to our earlier prediction, tax evasion would be more likely to take place. By contrast, the number emphasised in the instruction of Type II is the legally prescribed tax payment (e.g. ECU400, given a tax rate of 40%), and participants are expected to pay this amount. Using the number indicating the legally prescribed tax payment as the reference point results in a mixed gamble with one outcome in the gain domain ('successful' evasion) and another in the loss domain (fined evasion). For mixed gambles, the prospect theory predicts a particularly strong risk aversion. Accordingly, tax compliance should be exceptionally high. Hence, if participants apply the number

[1] Focussing on income might also mean that the expected net income serves as the reference point. However, this seems less likely to occur because necessary calculations would require substantial cognitive efforts.

provided in the instructions as the reference point, lower compliance with the instruction of Type I than that with the instruction of Type II is to be expected.

To summarise, the two proposed mechanisms lead to opposing predictions. If different instructions change the focus made in the decision frame, it will be expected that Type I yields higher compliance than Type II. If, however, a number indicated in the instruction serves as the reference in reporting decisions, Type I will result in lower compliance than Type II. Although, from this theoretical viewpoint, the direction of effects is unclear (suggesting a two–sided hypothesis), a recent meta–analysis (Alm and Malézieux 2021) suggests that the instruction of Type I lead to lower compliance than the Type II instruction. Therefore, we propose a directional hypothesis H1:

Operationalising tax compliance in laboratory experiments by using the Type I instruction (the income declaration) leads to a lower level of compliance than the Type II instruction (the tax declaration).

Four further hypotheses, H2–H5 (see Supplemental material), concern the potential interaction effects of operationalisation with other parameters in the experimental design. The core question these hypotheses are addressing is whether instructions moderate the effects of these parameters. We test such potential interaction effects for four determinants of compliance: tax rate, audit probability, fine rate, and income.[2] All the hypotheses, the procedure implemented in the experiment, and the

[2] Regarding the bomb crater effect and the corresponding results, an additional hypothesis H6 is detailed in Subsection B3 in Supplemental material.

strategy for the statistical analyses were pre-registered on the website of the Open Science Framework (https://osf.io/gnjha).[3]

Method

The hypotheses of this research are tested in the incentivised laboratory experiment, in which we manipulate the operationalisation of tax compliance, the three policy parameters of the economic model and experimental income. Additionally, a web–based experiment is conducted as a pilot study ($N = 467$) which produces the similar results. This pilot study is reported in Section A in Supplemental material. Sections D and E in Supplemental material include the materials from both the pilot and the main studies. All data are posted in the online repository (https://osf.io/a8q2f/).

Participants

A prior power analysis (see Section C in Supplemental material) suggested a sample of 494 participants, but due to budget constraints our final sample is $N = 365$. The mean age is $M = 23.42$ ($SD = 5.57$) years, 60% are female, 39% are male, 0.3% indicate 'other' as gender, and 0.8% leaves this question blank (for the analysis, the latter two values are replaced with the mode, which is female, and 3 missing values regarding age are replaced by the median of 22).

---

[3] Note that at the time of the pre-registration, a preliminary version of the meta-analysis by Alm and Malézieux (2021) did not include the analysis of operationalisation types, and so our original hypothesis H1 was non-directional.

The experimental design and procedure

Participants were recruited at the University of Vienna in 2019. Participation was remunerated by the amount between €2–12. To determine the payoff, one of the16 experimental rounds was drawn randomly, and income from this round was converted to Euro (ECU200 = €1). In addition, participants received a show–up fee of €2. The average total payoff was M = €9.45 (SD = 1.98, range: 2–12) for spending about M = 15.31 (SD = 3.15, range: 9.82 – 35.57) minutes in the lab. Course credits were not provided for participation.

The basic experimental design was similar to those standard tax compliance experiments. The two operationalisation types (the income declaration versus the tax declaration) were implemented as the between–subjects treatments. The policy parameters were manipulated within–subjects. These parameters included the tax rate (20% versus 40%), the fine rate (0.5 versus 1.5 times the evaded amount, in addition to paying the missing tax due), and the audit probability (1% versus 15%). The level of income varied endogenously (between ECU1,000 and 2,000) with participants' performance in a real–effort task (or the 'slider task', see Gill and Prowse 2012) which had to be completed at the beginning of each round.

Two test (pilot) rounds were followed by 16 experimental rounds which were subject to the analyses. The policy variables varied in each round with each possible combination occurring twice (for details see Table B1 of Section B in Supplemental material). After the real–effort task was completed, the screen showed an information sheet about the total gross income, the tax due, the tax rate, the audit probability, and the fine rate of the actual round. Finally, a tax report for the respective round by declaring either their gross income (the income declaration) or by stating their tax payment (the tax declaration) was provided to participants. After completion,

participants were informed whether they were audited and whether a fine had to be paid. An audit occurred only in round eight.

The dependent variable was relative compliance in percent; this was calculated by dividing the responses by the total income or the total tax due (multiplied by 100). For the technical reason, it was possible to indicate a number larger than the actual gross income or the tax due, which would yield a relative compliance of > 100%. The values from 101% to 110% were set to 100%, and the values above 110% were treated as missing observations (as stated in our pre-registration). There were only five such observations in the 16 experimental rounds from four participants. A relative compliance level between 101% and 110% was observed two times, and a relative compliance level above 110% was observed three times.

The original material and an English translation can be found in Section E in Supplemental material. The data are available in the online repository (https://osf.io/a8q2f/).

The results

Regardless of operationalisation types, averaged relative compliance is M = 39.41 (SD = 42.55) across all n = 5828 observations from the 16 experimental rounds. For the first test of H1, compliance is averaged across the 16 experimental rounds separately for each participant, regardless of the varying levels in the policy variables in each round. This overall compliance indicator is lower in the experimental condition in which participants are instructed to declare their gross income (M = 36.19, SD = 31.69) than when they are asked to declare their tax payment (M = 42.68, SD = 27.43), $t(363)$ = -2.09, p = .037, d = 0.22.

Figure 1 depicts average relative compliance separately for each of the 18 rounds (the 2 test rounds and the 16 experimental rounds) by the operationalisation types. Note that in each round, a different combination of the policy variables (tax rate, audit probability, and fine rate) is implemented. To test the effects of the between–subjects factor operationalisation type (H1), the policy variables manipulated within–subjects, the endogenously varying level of experimental income and their potential interactions (H2–H5), and a mixed–effects regression with relative compliance as criterion are estimated. To account for the repeated measurements, the regression model includes a random intercept for individuals. All experimental treatments are effect–coded for the analyses.

The results are summarised in Table 2.[4] The type of operationalisation affects compliance as hypothesised (H1) states, with higher compliance in the tax declaration condition than that in the income declaration condition. All policy variables and experimental income have the expected effects (Alm and Malézieux 2021). Also, participants' gender and age affect compliance similarly as reported from prior studies (e.g. Hofmann et al. 2017). Further, compliance slightly declines with the progressing of the experimental rounds. Whether an audit occurs in a previous round does not negatively affect compliance. The significant interaction effects indicate that the effects of the policy variables vary in association with the operationalisation types (thus, supporting H3–H5). However, for experimental incomes, no such interaction effects are observed (i.e. rejecting H2).[5] The similar results as with the OLS regression detailed in

[4] A simpler regression model including only the main effects produces the similar results.

[5] We also explored for higher order interactions and found some significant three-way interactions. These do not change the interpretation of the results presented in Table 2. A summary of this analysis can be found in Subsection B2 in Supplemental material.

Table 2 are obtained by means of a logistic mixed–effects regression analysis with dichotomised compliance (1 = full compliance; 0 = evasion) as the dependent variable. An exception is that the interaction between the operationalisation type and the fine rate is not significant in this additional analysis (for a summary see Table B2 in Supplemental material).

[Figure 1 near here]

[Table 2 near here]

To explore the nature of the observed interaction effects, all experimental rounds, which share the same factor level of the analysed policy variables (e.g. there are eight rounds with a tax rate of 20%), are averaged and compared to the average observed for the other factor levels (e.g. a tax rate of 40%) by an univariate repeated measures analysis of variance (ANOVA) with the operationalisation types as the between–subjects factor and with the respective policy variable as the within–factor. This analysis is repeated for each of the three policy variables. Table 3 shows the means and standard deviations of averaged compliance for each level of the policy variables separately for the two operationalisation types.

Regarding the effect of tax rate (H3), the ANOVA, as in the regression analysis, yields significant main effects of operationalisation types, $F(1, 363) = 4.37$, $p = .037$, $\omega_p^2 = .09$, and tax rate, $F(1, 363) = 17.28$, $p < .001$, $\omega_p^2 = .02$, as well as the interaction effect of these two experimental manipulations, $F(1, 363) = 63.20$, $p < .001$, $\omega_p^2 = .08$. How this interaction affects compliance is depicted in Figure 2. Two dependent sample t-tests conducted as post hoc analysis show that a higher tax rate resulted in less compliance only when the task for participants is to indicate the tax, $t(183) = -7.48$, $p <$

.001, d = -0.55. When they are instructed to indicate their gross income, a higher tax rate has a small, positive effect on compliance, $t(180) = 3.27$, $p = .001$, $d = 0.24$.

[Table 3 near here]

[Figure 2 near here]

A similar analysis is conducted to explore the interaction effect of operationalisation types and audit probability (H4). This analysis shows, once again, a significant main effect of operationalisation types, $F(1, 363) = 4.41$, $p = .037$, $\omega_p^2 = .03$, a significant effect of audit probability, $F(1, 363) = 335.69$, $p < .001$, $\omega_p^2 = .31$, and a significant interaction effect between these two factors, $F(1, 363) = 25.82$, $p < .001$, $\omega_p^2 = .03$. Figure 3 shows the observed interaction effect. Post hoc tests reveals that the effect of audit probability is stronger in the tax declaration condition, $t(183) = 14.84$, $p < .001$, $d = 1.09$, than in the income declaration condition, $t(180) = 10.75$, $p < .001$, $d = 0.80$.

[Figure 3 near here]

The ANOVA for the interaction between the fine rate and operationalisation (H5) yields the significant main effects of operationalisation types, $F(1, 363) = 4.39$, $p = .037$, $\omega_p^2 = .06$, the fine rate, $F(1, 363) = 131.40$ $p < .001$, $\omega_p^2 = .15$, and a significant interaction effect between these two factors, $F(1, 363) = 9.68$, $p = .002$, $\omega_p^2 = .01$. This interaction is depicted in Figure 4. The post hoc analysis shows that the effect of the fine rate is slightly stronger in the tax declaration condition, $t(183) = 8.75$, $p < .001$, $d = 0.65$, than in the income declaration condition, $t(183) = 7.58$, $p < .001$, $d = 0.56$.

Figure 4 about here

Before conducting the above analyses, we first check the frequency distribution of relative compliance. We notice that in 6.6% of all observations, participants showed a level of relative compliance that is exactly equal to the tax rate in this round. This occurs in 13% of all observations in the income declaration condition but only once in the tax declaration condition (for details see Subsection B4 in Supplemental material). A possible explanation for this is that some participants in the income declaration condition had misunderstood the instruction: instead of declaring the gross income, they stated the tax payment and thought that they would comply with the given tax rate by doing so. Because this occurs mainly in the income declaration condition, it may have amplified the observed effect of operationalisation types reported in the following. Hence, we conduct robustness tests after correcting the data for this problem and gain largely the same results. These are reported in Subsection B4 in Supplemental material.

Discussion

Laboratory research in general, and the research on tax compliance in particular, is frequently criticised for various reasons. Apart from the controversial discussion about the external validity of experimental results, the criticism also concerns the lack of consensus in the experimental design. This research is an attempt to systematically examine this issue by focussing on the most important aspect of experimental environments: How is tax compliance operationalised in the lab? It is found that instructing participants to declare their gross income, as most reported in tax compliance experiments, results in lower compliance than when instructing participants

to declare their tax payment (H1). Further, and probably even more relevant for the experimental tax research, it is discovered that the type of operationalisation substantially alters the effects which are well–documented by some previous experiments (Alm and Malézieux 2021; Kirchler et al. 2010) and the cornerstone of economic theories about tax compliance – the impact of tax rates, audit probabilities, and fine rates (Allingham and Sandmo 1972). All interaction effects detected in this research are of similar nature, that is, the expected effects are stronger when participants declare their tax payment than when declaring their income. This well supports H3, H4, and H5). For the tax rate, we even observe a reversed effect, higher compliance due to a higher tax rate, in the income declaration condition. For the income declaration, there is no interaction effect in the main analysis; thus H2 is not supported. Such a moderation effect is only drawn from some of the additional analyses conducted as robustness checks.

Our main finding regarding lower compliance when participants declare their income can be explained by assuming that participants draw on the numbers stated in the experimental instruction when choosing their reference point for the reporting decision. Using the gross income as the reference creates a negative decision frame that makes susceptible to tax evasion. By contrast, using the tax payment as the reference frames the decision as a mixed gamble which enhances compliance. However, how can the observed interaction effects be explained? Regarding the tax rate, a simple explanation for the interaction with operationalisation is that the numbers used in the instruction vary with tax rates in the tax declaration but not in the income declaration condition. The resulting tax due participants have to declare is higher or lower depending on the given tax rate, but the income they are instructed to declare remains the same regardless of the tax rate. Similarly, the stronger effect of audit probabilities

produced by operationalising compliance as the tax payment can be explained by the fact that it is easier to compute the consequence of getting caught in this experimental condition. The fine rate is communicated as a multiple of the evaded tax amount, so calculating the costs of getting caught is simpler than in the income declaration task. Hence, the observed interaction may result from simplifying the decision problem – the tax payment declaration task by the instruction of the second type operationalisation.

These interaction effects may explain some contradictory findings from the previous experiments. A review of empirical studies (Kirchler et al. 2010) finds that particularly for the effects of income, tax rate, and fine rate, the experimental results are puzzling. The most ambiguous prior findings are on the tax rate, for which, as in this study, both positive and negative effects have been detected. Different approaches in operationalising compliance can explain why some experiments gain the expected effects and others do not, or even reverse effects. Thus, the present research is highly relevant for evaluating and interpreting prior and future experimental evidences about tax compliance.

The findings of this research also have broader implications beyond the scope of researching tax compliance, and they may be relevant for designing experiments in other fields (for more recommendations for good experimental designs, see Camerer 2011). This study adds to a growing body of literature by addressing the critique on the generalisability of experimental findings (Frechette and Schotter 2015; Levitt and List 2007). Its focus is on the operationalisation of the dependent variable – the key element of experimental designs. This research complements the previous effort to systematically improve and refine operationalisation (e.g. Brandts and Charness 2011; Frey et al. 2017; Horowitz and McConnell 2002; Van Dijk and Wilke 2000). In line with the prior research, it demonstrates that the operationalisation of experimental

designs matters. Although we consider that it can be beneficial to use a certain amount of 'pluralism' to assess the replicability and generalisability of experimental results, the importance of systematic research on such an issue cannot be stressed enough.

There are some limitations of this research. The experiment is based on a within–subjects design, which makes the manipulation of the policy variables more transparent for participants. A replication of our findings by applying a between–subjects design will be useful. Further, our sample consists mainly of students from various disciplines. Although employing student participants as the subjects remains common and although the study by Alm, Bloomquist, and McKee (2015) demonstrates that student responses are similar to those of real world taxpayers, further studies should try to replicate our findings with more experienced participants who are familiar with the task of declaring the gross income in view of their practice in the real world. Perhaps, the difference between the two operationalisation types is even stronger among this population.

It is bit surprising that a substantial number (13%) of participants in the income declaration condition declared an amount of income that mirrored exactly the tax rates of the experiment. The most plausible explanation is that participants misunderstood the instruction and thought to comply with the given tax rate by declaring the corresponding amount (20% and 40%, respectively) of income. After re-checking the applied materials, however, we find no obvious confusing information that could have caused the problem. Another explanation is that the most recent tax related experiment conducted in the same laboratory (by another research group) a few weeks before this research had implemented the tax declaration method of operationalising compliance. Since participants are recruited from the same database, it cannot be ruled out that some of the subjects might have already participated in that earlier experiment and that they

were used to declare the tax payment to indicate the level of compliance. Nevertheless, the robustness checks we conduct to correct for this problem in the data produce largely the same results as the original analyses (see Subsection B4 in the Supplemental material).

Which type of operationalisation should be used in future experiments on income tax compliance? The type of the income declaration corresponds to the situation of self–employed taxpayers in the real world, and hence it may have more external validity than the tax declaration. However, the income declaration also seems to be more complicated for participants, as it requires more calculations when evaluating the outcome of reporting decisions. Indeed, a non–negligible number of participants in this study seemed to have misunderstood the instruction for this operationalisation. Further, the expected effects of the policy parameters – tax rate, audit probability and fine rate – are weaker in the income declaration condition than in the tax declaration condition. To some degree, this indicates a better internal validity for the use of the tax declaration because these behavioural responses are more in line with the theory and the prior empirical results. In addition, although the tax declaration is at present different from the real world setting for most taxpayers, this difference can well change soon given that digitalisation will change the procedure of submitting tax files to authorities. As noted in the introduction, the tax declaration task may increase in its importance when electronic filing systems become more broadly used.

In sum, the results of this paper indicate that any conclusion for real world tax advice, which is drawn from experiments in the tax research, must be taken cautiously. Additional research is needed on how other types of operationalisation – in tax compliance laboratory experiments and beyond – may affect experimental results and their interpretation.

References

Allingham, Michael G., & Sandmo, Agnar 1972. "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics* 1(3–4): 323–338.

Alm, James, Bloomquist, Kim M., & McKee, Michael 2015. "On the External Validity of Laboratory Tax Compliance Experiments." *Economic Inquiry* 53(2): 1170–1186.

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

Alm, James, & Malézieux, Antoine 2021. "40 Years of Tax Evasion Games: A Meta–Analysis." *Experimental Economics* 24(3): 699–750.

Bazart, Cecile, & Pickhardt, Michael 2011. "Fighting Income Tax Evasion with Positive Rewards." *Public Finance Review* 39(1): 124–149.

Bernasconi, Michele, Corazzini, Luca, & Seri, Raffaello 2014. "Reference Dependent Preferences, Hedonic Adaptation and Tax Evasion: Does the Tax Burden Matter?" *Journal of Economic Psychology* 40: 103–118.

Brandts, Jordi, & Charness, Gary 2011. "The Strategy Versus the Direct–Response Method: A First Survey of Experimental Comparisons." *Experimental Economics* 14(3): 375–398.

Camerer, Colin F. 2011. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.

Casal, Sandro, Kogler, Christoph, Mittone, Luigi, & Kirchler, Erich 2016. "Tax Compliance Depends on Voice of Taxpayers." *Journal of Economic Psychology* 56: 141–150.

Cookson, Richard 2000. "Framing Effects in Public Goods Experiments." *Experimental Economics* 3(1): 55–79.

Copeland, Phyllis V., & Cuccia, Andrew D. 2002. "Multiple Determinants of Framing Referents in Tax Reporting and Compliance." *Organizational Behavior and Human Decision Processes* 88(1): 499–526.

Dhami, Sanjit, & Al–Nowaihi, Ali 2007. “Why Do People Pay Taxes? Prospect Theory Versus Expected Utility Theory.” *Journal of Economic Behavior & Organization* 64(1): 171–192.

Falk, Armin, & Heckman, James J. 2009. “Lab Experiments are a Major Source of Knowledge in the Social Sciences.” *Science* 326(5952): 535–538.

Fochmann, Martin, & Wolf, Nadja 2019. “Framing and Salience Effects in Tax Evasion Decisions – An Experiment on Underreporting and Overdeducting.” *Journal of Economic Psychology* 72: 260–277.

Frechette, Guillaume R. & Schotter, Andrew 2015. *Handbook of Experimental Economic Methodology*. New York, NY: Oxford University Press.

Frey, Renato, Pedroni, Andreas, Mata, Rui, Rieskamp, Jörg, & Hertwig, Ralph 2017. “Risk Preference Shares the Psychometric Structure of Major Psychological Traits.” *Science Advances* 3(10): e1701381.

Gill, David, & Prowse, Victoria 2012. “A Structural Analysis of Disappointment Aversion in a Real Effort Competition.” *The American Economic Review* 102(1): 469–503.

Haesevoets, Tessa, Van Hiel, Alain, Van Assche, Jasper, Bostyn, Dries, & Reinders Folmer, Chris 2019. “An Exploration of the Motivational Basis of Take–Some and Give–Some Games.” *Judgment and Decision Making* 14(5): 535–546.

Hofmann, Eva, Voracek, Martin, Bock, Christine, & Kirchler, Erich 2017. “Tax Compliance Across Sociodemographic Categories: Meta–Analyses of Survey Studies in 111 Countries.” *Journal of Economic Psychology* 62: 63–71.

Horowitz, John K., & McConnell, Kenneth E. 2002. “A Review of WTA/WTP Studies.” *Journal of Environmental Economics and Management* 44: 426–447.

Kahneman, Daniel, & Tversky, Amos 1979. “Prospect Theory: An Analysis of Decision Under Risk.” *Econometrica* 47(2): 263–292.

Kirchler, Erich, & Maciejovsky, Boris 2001. “Tax Compliance Within the Context of Gain and Loss Situations, Expected and Current Asset Position, and Profession.” *Journal of Economic Psychology* 22(2): 179–194.

Kirchler, Erich, Muehlbacher, Stephan, Hoelzl, Erik, & Webley, Paul 2009. “Effort and Aspirations in Tax Evasion: Experimental Evidence.” *Applied Psychology: An International Review* 58(3): 488–507.

Kirchler, Erich, Muehlbacher, Stephan, Kastlunger, Barbara, & Wahl, Ingrid 2010. “Why Pay Taxes? A Review of Tax Compliance Decisions.” In *Developing*

Alternative Frameworks for Explaining Tax Compliance, edited by James Alm, Jorge Martinez–Vazquez, & Benno Torgler, 15–31. Oxon: Routledge.

Levitt, Steven D., & List, John A. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *The Journal of Economic Perspectives* 21(2): 153–174.

Malézieux, Antoine 2018. "A Practical Guide to Setting Up Your Tax Evasion Game." *Journal of Tax Administration* 4(1).

Martinez–Vazquez, Jorge, Harwood, Gordon B., & Larkins, Ernest R. 1992. "Withholding Position and Income Tax Compliance: Some Experimental Evidence." *Public Finance Review* 20(2): 152–174.

Mittone, Luigi 2006. "Dynamic Behaviour in Tax Evasion: An Experimental Approach." *Journal of Socio-Economics* 35(5): 813–835.

Muehlbacher, Stephan 2021. "Prospect Theory and Tax Compliance." In *The Cambridge Handbook of Compliance* (Cambridge Law Handbooks), edited by Benjamin Van Rooij & Daniel Sokol, 541–550. Cambridge: Cambridge University Press.

Muehlbacher, Stephan, & Kirchler, Erich 2013. "Mental Accounting of Self-Employed Taxpayers: On the Mental Segregation of the Net Income and the Tax Due." *FinanzArchiv: Public Finance Analysis* 69(4): 412–438.

Muehlbacher, Stephan, & Kirchler, Erich 2016. „Taxperiments: About the External Validity of Laboratory Experiments in Tax Compliance Research." *Die Betriebswirtschaft (DBW) / Business Administration Review* 76(1): 7–19.

Schepanski, Albert, & Shearer, Teri 1995. "A Prospect Theory Account of the Income Tax Withholding Phenomenon." *Organizational Behavior and Human Decision Processes* 63(2): 174–186.

Steiger, Alexander, & Kühberger, Anton 2018. "A Meta–Analytic Re-Appraisal of the Framing Effect." *Zeitschrift für Psychologie* 226(1): 45–55.

Torgler, Benno 2002. "Speaking to Theorists and Searching for Facts: Tax Morale and Tax Compliance in Experiments." *Journal of Economic Surveys* 16(5): 657–683.

Torgler, Benno, & Schaltegger, Christoph A. 2005. "Tax Amnesties and Political Participation." *Public Finance Review* 33(3): 403–431.

Tversky, Amos, & Kahneman, Daniel 1986. "Rational Choice and the Framing of Decisions." *Journal of Business* 59(4): 251–278.

Van Dijk, Eric, & Wilke, Henk 2000. “Decision-Induced Focusing in Social Dilemmas: Give–Some, Keep-Some, Take–Some, and Leave-Some Dilemmas.” *Journal of Personality and Social Psychology* 78(1): 92–104.

Yaniv, Gideon 1999. “Tax Compliance and Advance Tax Payments: A Prospect Theory Analysis.” *National Tax Journal* 52(4): 753–764.

Table 1. Two types of operationalising tax compliance in lab experiments.

Type 1. Declaration of income	Type 2. Declaration of taxes
Your gross income: 1,000 ECU Your tax due (40%): 400 ECU *Please indicate how much of your income you declare (any amount between 0–1,000 ECU)*: _____ ECU	Your gross income: 1,000 ECU Your tax due (40%): 400 ECU *Please indicate how much of your tax due you pay (any amount between 0–400 ECU)*: _____ ECU

Table 2. Summary of a linear mixed–effect regression with relative compliance as dependent variable.

Relative compliance		
Intercept	23.91	***
	(6.64)	
Round	-0.26	***
	(0.08)	
Income[a]	-1.03	**
	(0.39)	
Last period audited[b]	2.92	†
	(1.69)	
Age	0.92	***
	(0.28)	
Gender[c]	-9.04	**
	(3.14)	
Operationalization type[d]	3.43	*
	(1.52)	
Tax rate[e]	-1.65	***
	(0.38)	
Audit probability[f]	11.56	***
	(0.38)	
Fine rate[g]	5.31	***
	(0.38)	
Operationalization type × Tax rate	-2.95	***
	(0.38)	
Operationalization type × Audit probability	3.26	***
	(0.38)	
Operationalization type × Fine rate	1.52	***
	(0.38)	
Operationalization type x Income	0.33	
	(0.38)	
Operationalization type x Last period audited	-0.53	
	(1.69)	
n of observations	5828	
N of subjects	365	
Marginal R^2	.24	

Notes: Dependent variable is relative compliance measured in overall 16 rounds; BIC = 56545.60.
Random effects (individual): Intercept *SD* = 27.99; Residual *SD* = 27.96.
Numbers in Column 2 are the regression coefficients, standard errors in parentheses.
[a]Experimental income is standardised by the person means and standard deviations. [b]Dummy coded with 0 = no audit in previous round, 1 = audit in previous round. [c]0 = female, 1 = male; values of two participants who indicated 'other' as gender or did not answer the question were replaced with the sample mode (female). [d]Effect coded with -1 = the income declaration, +1 = the tax declaration. [e]Effect coded with -1 = 20%, +1 = 40%. [f]Effect coded with -1 = 1%, +1 = 15%. [g]Effect coded with -1 = 0.5 times evaded sum, +1 = 1.5 times evaded sum.
† $p \leqslant .10$. * $p \leqslant .05$. ** $p \leqslant .01$. *** $p \leqslant .001$.

Table 3. Descriptive statistics for relative compliance by the operationalisation type and the manipulated policy variables.

	Income declaration (n = 181)	Tax declaration (n = 184)
Policy variable	M (SD)	M (SD)
Tax rate		
20%	34.75 (33.38)	47.20 (28.51)
40%	37.64 (31.07)	38.16 (28.75)
Audit probability		
1%	27.71 (32.61)	27.69 (31.44)
15%	44.68 (34.22)	57.72 (29.87)
Fine rate		
0.5	32.30 (31.67)	35.88 (29.72)
1.5	40.07 (33.16)	49.47 (29.05)

Notes: N = 365. Numbers indicate means for relative compliance averaged across the eight experimental rounds with similar levels of the respective policy variable (e.g., all rounds with a tax rate of 20%).

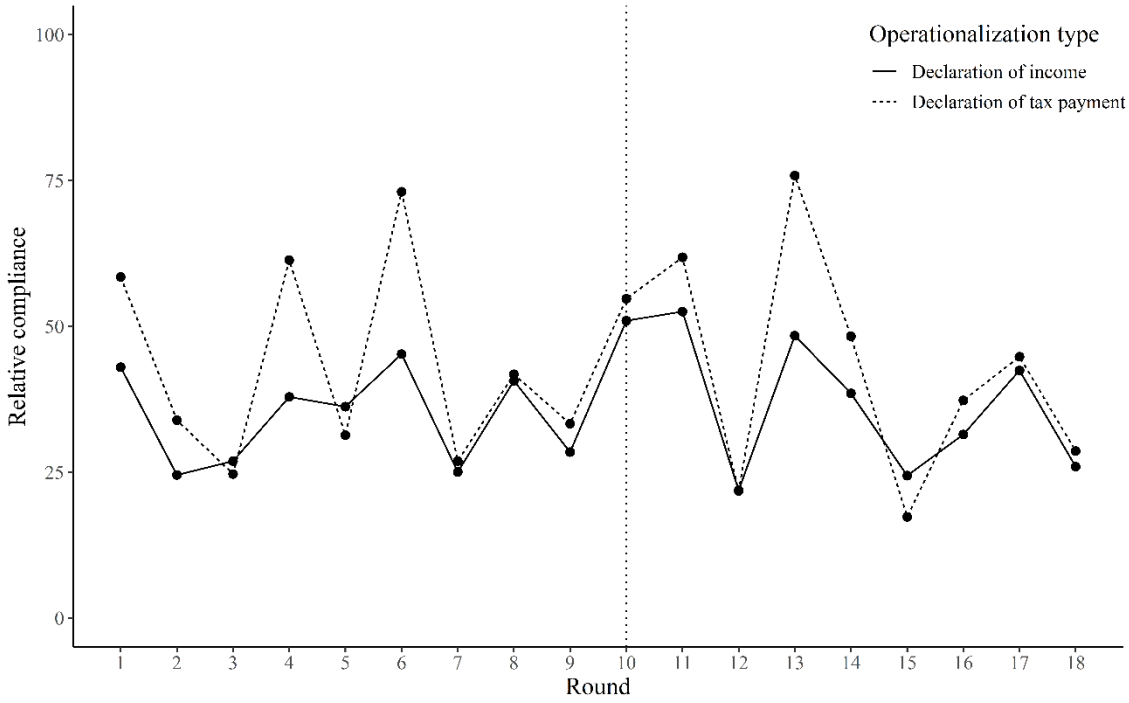

Figure 1. Relative compliance in each experimental condition and round.

Notes: Rounds 1 and 2 are the test rounds and not subject to the analyses presented in the article. In each round, the policy factors (tax rate, audit probability, and fine rate) and experimental income vary. The vertical line indicates that in round 10 an audit occurs.

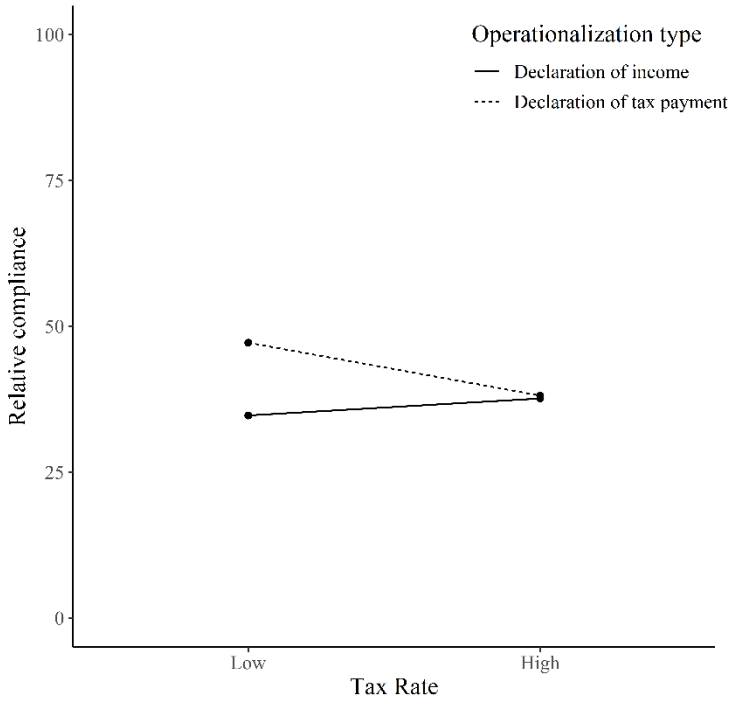

Figure 2. The interaction effect of the tax rate and operationalisation type on relative compliance.

Notes: Compliance in this depiction is relative compliance averaged across the eight experimental rounds with the same tax rate (20% versus 40%).

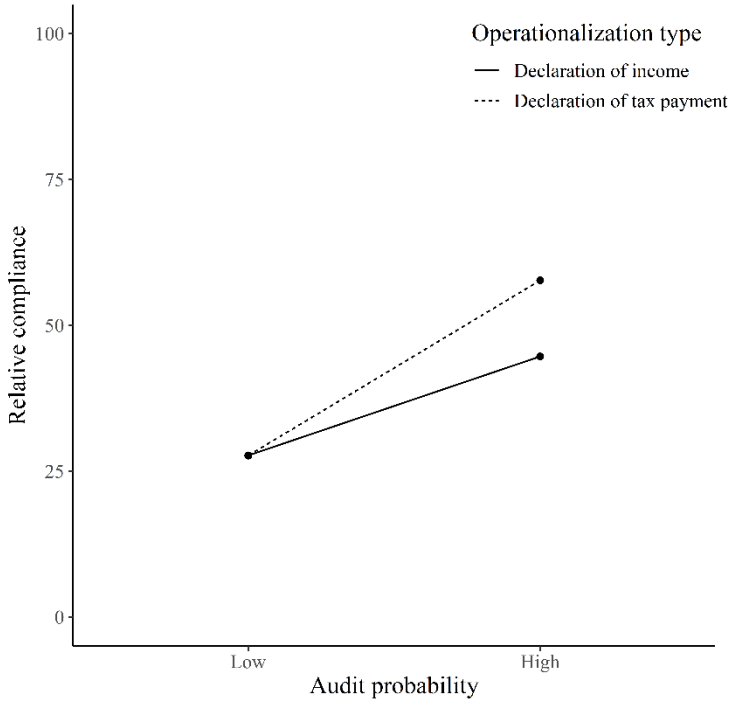

Figure 3. The interaction effect of audit probability and operationalization type on relative compliance.

Notes: Compliance in this depiction is relative compliance averaged across the eight experimental rounds with the same audit probability (1% versus 15%).

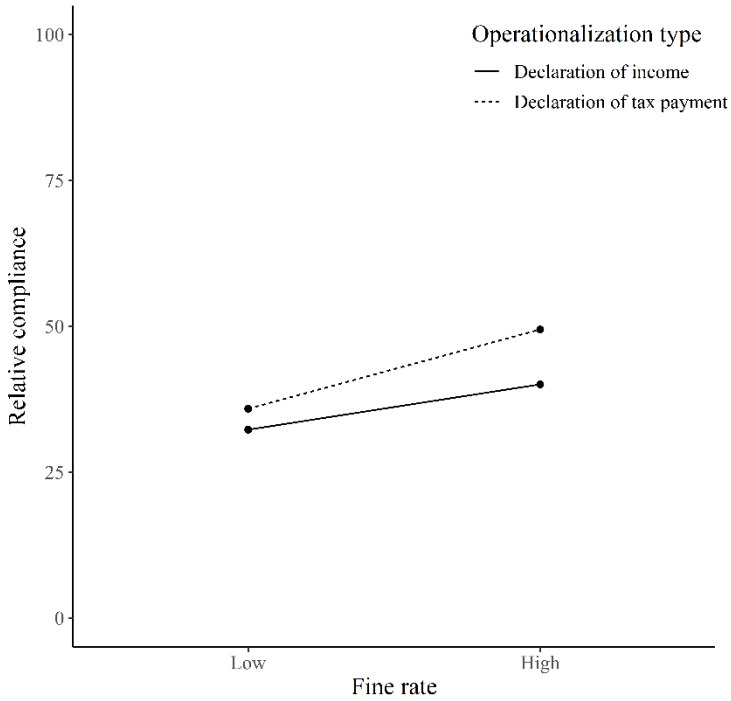

Figure 4. The interaction effect of fine rate and operationalization type on relative compliance.

Notes*:* Compliance in this depiction is relative compliance averaged across the eight experimental rounds with the same fine rate (0.5 versus 1.5 times the amount evaded).