Article:

Kasper, M. and Rablen, M. orcid.org/0000-0002-3521-096X (2023) Tax compliance after an audit: higher or lower? Journal of Economic Behavior & Organization, 207. pp. 157-171. ISSN 0167-2681

https://doi.org/10.1016/j.jebo.2023.01.013

Journal of Economic Behavior and Organization 207 (2023) 157–171

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Tax compliance after an audit: Higher or lower?☆

Matthias Kasper[a,*], Matthew D. Rablen[b]

[a] University of Vienna, Department of Applied Psychology, Austria
[b] University of Sheffield, Department of Economics, United Kingdom

ARTICLE INFO

Article history:
Received 15 March 2022
Revised 19 October 2022
Accepted 15 January 2023

JEL classification:
H26
H83
C9

Keywords:
Tax audits
Probability misperception
Learning
Endogenous audit rules
Censoring
Uncertain audit probability
Tax evasion game

ABSTRACT

What is the compliance effect of experiencing a tax audit? Empirical studies typically report a positive effect, while laboratory experiments frequently report a negative effect. We show experimentally that whether a tax audit increases or decreases subsequent compliance hinges on the balance of learning opportunities, misperception of audit risk, and the confounding effect of censoring. After an audit, taxpayers lower their perceived risk of audit – consistent with a bomb-crater effect – when audit selection is exogenous. However, for an endogenous audit rule under which taxpayers can learn to reduce their audit risk by reporting higher income, learning effects outweigh probability misperception, resulting in an increase in post-audit tax compliance. Finally, we show that accounting for censoring effects can eliminate on its own the negative post-audit compliance effect frequently observed in laboratory experiments.

1. Introduction

Tax audits are an essential instrument in achieving compliance. While the threat of audit may itself alter compliance behavior, what effect, if any, does carrying out an audit have upon the future compliance of its target (the post-audit effect)?

Under conventional assumptions on risk aversion, the seminal Allingham-Sandmo (AS) analysis of tax compliance, Allingham and Sandmo (1972), predicts a zero post-audit effect for audited taxpayers who do not receive a fine, and a positive post-audit effect for fined taxpayers. Thus, empirically, the average post-audit effect is predicted to be positive, but potentially difficult to distinguish statistically from zero in samples where a high proportion of audits do not lead to fines. This prediction of a positive average post-audit effect is broadly in line with a growing empirical literature on post-audit effects, but discordant with experimental findings, which are instead consistent with a (weakly) negative average post-audit effect. Such seemingly contradictory findings in the existing evidence base on the post-audit effect are consequential for

☆ Acknowledgements: This study has been approved by the Institutional Review Board of the Department of Work, Economic, and Social Psychology at the University of Vienna (2022/W/001). We thank two anonymous reviewers, Robert Böhm, Linda Dezső, Aidas Masiliunas, and participants of IAREP (Kristiansand) for helpful comments. We also thank the Vienna Center for Experimental Economics (VCEE), University of Vienna, for use of their laboratory. Kasper acknowledges the University of Vienna, Faculty of Psychology, for financial support.

* Corresponding author.
E-mail addresses: matthias.kasper@univie.ac.at (M. Kasper), m.rablen@sheffield.ac.uk (M.D. Rablen).

https://doi.org/10.1016/j.jebo.2023.01.013
0167-2681/

tax enforcement policy: audits are costly, so determining how many to do and how best to allocate them are key policy questions. If, as the experimental evidence suggests, auditing taxpayers reduces their subsequent compliance, this weakens the case for auditing, and strengthens the case for alternative compliance measures, e.g., enhanced taxpayer support. Also, tax administrations should seek to maximize the perceived risk of audit while at the same time minimizing the number of audits they actually perform. If, however, auditing increases compliance, as suggested in the empirical literature, tax administrations should seek to maximize the number of audits they actually perform.

In this paper we examine the apparent downwards bias of laboratory findings on the average post-audit effect relative to the empirical literature. We seek to trace the origins of this difference to three principal factors:

(a) *Learning*: Learning here refers to the Bayesian notion of the acquisition of knowledge from probabilistic signals (i.e., audits) regarding an objective state. In the canonical tax evasion game (TEG) all aspects of enforcement are fixed through time and taxpayers have perfect information over these aspects, including (i) the stochastic process determining audit selection; (ii) the audit technology (effectiveness); (iii) tax rates; and (iv) the penalties for detected non-compliance. In this environment being audited is an uninformative signal, ruling out learning effects, and the mechanisms by which a non-zero post-audit effect can arise are accordingly limited. By contrast, in empirical settings, experiencing an audit invariably does offer opportunities for learning – for both the taxpayer and the tax authority. If, for instance, the taxpayer experiences uncertainty as to their audit probability, being audited will – under any reasonable updating process – cause the taxpayer to increase their posterior belief about the audit probability. Similarly, the tax authority gains a detailed understanding of the taxpayer's financial affairs, and thereby of their likely true income in the near future. While both of these effects act to raise the post-audit effect, other considerations could act to decrease it, such as if penalties and/or audit effectiveness transpired to be lower than expected.[1]
A close examination of empirical studies that report a positive average post-audit effect reveals that, in each case, the principal contributor to the effect would not exist in the canonical laboratory TEG. For instance, Advani et al. (2021) – who report a positive effect in audit data from the UK – attribute their finding to learning by the tax authority as to the likely true income of the taxpayer in the near future. This effect is not, however, present in the standard TEG, the design of which precludes the tax authority (experimenter) from exploiting learning possibilities arising from behavior observed during the experiment. Kotsadam et al. (2021) attribute a positive post-audit effect to audited taxpayers acquiring improved knowledge of tax law. Again, such an effect would not exist in the standard TEG, in which tax rates and penalties are known.

(b) *Probability misperception*: Beginning with Mittone (2006), scholars observing a negative average post-audit effect in laboratory experiments have often conjectured that this may be due to taxpayers perceiving (incorrectly) that they become less likely to be audited in the period immediately following an audit (the "bomb-crater effect"). Why might such downwards misperception be more salient in the laboratory environment than in the field? One possible answer to this question – which we investigate in this paper – is that, in the field, taxpayers perceive that they can exercise judgement and skill over their audit risk through their choice of how much income to declare: the higher the declaration, the lower the risk. By contrast, audit selection is exogenous to individual characteristics in the standard TEG, rendering the audit probability a more abstract quality. This abstract quality may result in taxpayers falling back on focal, but statistically inappropriate, heuristics.

(c) *Censoring effects*: A compliant taxpayer cannot further raise their compliance in response to being audited (ceiling effect), and a fully non-compliant taxpayer cannot further lower their compliance in response to not being audited (floor effect). Both of these two effects act to systematically lower estimates of the post-audit effect, for they narrow the observed differences in compliance behavior between rounds that follow an audit and rounds that do not. While censoring effects also apply in the field, their impact may be of greater magnitude in the laboratory. This is for two principal reasons. First, as censoring effects arise only when compliance is at (or sufficiently close to) the extremes of the spectrum, they are expected to be stronger when a high proportion of income reports are towards the extremes of the compliance spectrum. This is typically the case in laboratory tax experiments: reports implying either zero- or full-compliance account for some 65 percent of all reporting decisions in TEGs, according to the meta-analysis of Alm and Malézieux (2021). Second, in the field, the targeting of most (in some cases all) operational audits is risk-based, such that they are systematically less likely to fall upon compliant taxpayers. This limits the scope for the ceiling effect to bias downward estimates of the post-audit effect. In contrast, in the canonical TEG, audit selection is undirected.

We investigate these effects in a laboratory experiment. This allows us, in a way that is impossible in the field, to vary the audit rule, rigorously elicit incentivized beliefs at the individual level, and observe accurately taxpayers' true compliance levels. The control treatment is a standard TEG in which there is perfect information, and therefore nothing can be learned from being audited. The true audit probability is fixed throughout the rounds and we remind participants of its value in each and every round. We employ an incentivized belief elicitation technique to garner direct evidence as to whether audit belief falls in the round after an audit, as the bomb-crater effect supposes. In contrast, prior experimental literature, while

[1] Consistent with this point, recent research finds that taxpayers who receive minimal forms of audit that are plausibly ineffective at detecting noncompliance – so-called "desk" or "correspondence" audits – exhibit a strictly negative post-audit effect (Erard et al., 2020; Kasper and Alm, 2022a).

conjecturing a bomb-crater effect, has never demonstrated such an effect in elicited beliefs. We find strong evidence of such a bomb-crater effect, and link this tightly to an observed negative average post-audit effect.

We then implement two alternative treatments in which being audited offers opportunities for taxpayers to learn about a payoff salient parameter. The first such treatment allows learning over the probability of audit, for the experimental instructions give participants only partial knowledge. Specifically, in this treatment participants learn that they face a constant but unknown audit probability. Audit selection remains exogenous, however, such that participants cannot influence their audit risk. The second such treatment implements an endogenous audit rule whereby audits are targeted systematically on the taxpayers with the lowest levels of compliance. Being audited allows participants to learn that they are less compliant than (most) other participants in the experiment. Accordingly, participants can learn, and apply judgement as to, the level of compliance required to avoid being selected for audit.

In both the aforementioned learning treatments we observe evidence of positive learning effects: when the audit probability is uncertain, elicited audit beliefs fall less after an audit than in the control treatment, signalling the existence of a learning effect acting against the bomb crater effect. Our point estimate for the average post-audit effect remains negative, but less so than in the control treatment, and no longer significantly different from zero statistically. Under an endogenous audit rule we observe strong evidence that, after experiencing an audit, participants revise upwards their beliefs about the compliance levels of other participants, consistent with learning. The resulting post-audit effect is positive, consistent with evidence from the field.

As a final analysis, we re-evaluate the results from all three treatments with an alternative methodology intended to mitigate possible censoring effects. This methodology controls for the effects of censoring by analyzing subsamples of the data that exclude potentially censored observations, as described in Section 4.4. Comparing the results under the alternative and original methodologies allows us to measure the extent to which censoring effects may be responsible for downwardly biasing our estimates of the post-audit effect. In the control and uncertain audit probability treatments – both of which employ undirected audit selection and in which reported income was frequently everything or nothing – we find evidence of large censoring effects. In the control treatment, for instance, we find that – on its own – correcting for censoring effects switches the estimates sign of the post-audit effect from negative to positive. By contrast, we find that the endogenous audit rule treatment – in which audits are directed towards noncompliant taxpayers – is immune to censoring effects.

As well as mattering for tax enforcement policy, the methodological faultline in the existing evidence base regarding the size of the post-audit effect also matters in another sense. Specifically, it raises old questions as to the external validity of experimental evidence in the study of tax compliance. For instance, Slemrod (2019) chooses to omit any discussion of evidence from laboratory experiments in his (otherwise comprehensive) review of the tax compliance and enforcement nexus, while (Advani et al., 2021, p. 7) note pointedly that "...lab experimental evidence" of a negative post-audit effect "...is not reflected in real-world settings." Thus, another important reason for studying the seemingly contradictory findings in the existing evidence base is to understand their implications, if any, for the veracity of experimental approaches to the study of tax compliance. Our findings do not point to a lack of veracity of the experimental approach. Rather, as we converge the experimental setting closer to that found in the field, so our results converge towards those found in the field. Specifically, we replicate a negative post-audit effect in the standard TEG, but observe that it can be made to disappear as participants are given the opportunity to learn about aspects of the experimental environment by being audited. In this sense, our findings chime with the prior literature investigating the external validity of tax compliance experiments, which finds that experimental compliance outcomes are predictive of real-world tax compliance behavior (Bloomquist, 2009; Alm et al., 2015). This finding is also echoed by Dai et al. (2018) in the related context of (non-tax) fraud. Our results do, though, carry implications for the design of TEGs, and the interpretation of their results, a theme we take up in the conclusion.

Our analysis contributes to a considerable literature investigating tax compliance experimentally (see, e.g., Alm and Malézieux (2021) for a review) and connects more specifically to much smaller experimental literatures on the effects on tax compliance of an uncertain audit probability and of endogenous audit rules. The former literature considers whether compliance outcomes are higher or lower with uncertain audit probability (e.g., Friedland, 1982; Alm et al., 1992; Choo et al., 2016), but does not consider the post-audit effect specifically, and does not elicit beliefs. The literature on the use of endogenous audit rules (e.g., Alm et al., 1993; Tan and Yim, 2014) has a similar focus. Although it too does not systematically elicit beliefs to measure learning, the dynamics of reporting in these experiments are nonetheless suggestive of learning effects.

The plan of the paper is as follows. Section 2 provides a detailed review of evidence relating to the post-audit effect in the context of individual tax compliance; Section 3 describes the experiment; Section 4 develops a simple model of play in the experiment, which we use to form null hypotheses; Section 5 presents the results; and Section 6 concludes. Proofs are in the appendices.

2. Post-audit compliance effects

Audits can have two types of effect on compliance. The *pre-audit* effect refers to compliance induced prior to an audit occurring by the threat of subsequent detection and punishment of non-compliance. The *post-audit* effect, by contrast, refers to the subsequent compliance response arising from actually carrying out the threat. Here we shall focus on the post-audit effect and, in particular, the "own" post-audit effect, i.e., the post-audit compliance response of the audited taxpayer. We note, however, evidence that audits exert a (positive) "cross" post-audit effect on the subsequent compliance of unaudited taxpayers who observe the audit of another taxpayer (e.g., Rincke and Traxler, 2011; Boning et al., 2020; Drago et al., 2020).[2]

The seminal AS analysis of tax compliance predicts that the pre-audit effect is (weakly) positive. When strictly positive its magnitude is strictly increasing in both the probability of audit and the level of fines. As, however, the post-audit effect is a dynamic property of compliance, it is not, strictly speaking, captured by the (static) AS model. To make predictions, one may instead consider a minimal two-period repeated version of the AS model in which a myopic taxpayer chooses tax compliance according to the static AS model in each period. There are two cases to distinguish. First, if the taxpayer is treated with an audit in the first period, but not fined on account of being compliant, the taxpayer enters the second period with the same initial wealth as they would if, counterfactually, the audit had not occurred. In this case, second period compliance is the same in both the treatment and counterfactual, i.e., a zero post-audit effect. Second, if the first-period audit detects non-compliance, such that the taxpayer receives a fine, they enter the second period with a lower initial wealth than would have been the case in the counterfactual scenario in which no audit occurred. In this case the model predicts higher compliance in the second period relative to under the counterfactual, i.e., a positive post-audit effect. The negative shock to wealth induced by the fine generates a pure wealth effect. This acts to increase risk aversion (and so compliance) under the assumption of the AS model of decreasing absolute risk aversion.

While, in respect of the pre-audit effect, there is a considerable body of evidence generally supportive of the aforementioned predictions of the AS model (e.g., Slemrod et al., 2001; Fellner et al., 2013; Alm, 2019; Slemrod, 2019), the evidence for the AS prediction of a (weakly) positive post-audit effect is more nuanced. Consistent with this prediction is a growing body of empirical evidence that uncovers either a strictly positive average post-audit effect (Kleven et al., 2011; DeBacker et al., 2018a; 2018b; Løyland et al., 2019; Beer et al., 2020; Boning et al., 2020; Kotsadam et al., 2021; Mazzolini et al., 2022; Sánchez, 2022; Advani et al., 2021), or a zero average post-audit effect (e.g., Gemmell and Ratto, 2012; Best et al., 2021). Two recent experiments also report a zero average post-audit effect (Choo et al., 2013; Kasper and Alm, 2022a). There remains, however, an important strand of the experimental literature that – contrary to the predictions of the AS model and the empirical evidence – finds that the average post-audit effect is negative (Mittone, 2006; Maciejovsky et al., 2007; Kastlunger et al., 2009; Mittone et al., 2017).

Studies analyzing field data from the UK (Gemmell and Ratto, 2012) and US (Beer et al., 2020) offer a possible explanation of a negative average post-audit effect: when restricting attention to the subset of audited taxpayers who are found to be compliant, they find a negative post-audit effect. Other recent studies, however, have questioned this finding, instead reporting a zero post-audit effect among this subset of taxpayers (e.g., Mazzolini et al., 2022; Advani et al., 2021). Advani et al. (2021) is notable in this regard as it employs the same (albeit extended) dataset of Gemmell and Ratto (2012), but an alternative identification strategy. The laboratory experiment of Kasper and Alm (2022b) finds that although taxpayers do reduce their compliance substantially after declaring their entire income, this effect occurs whether or not an audit is experienced. Other scholars (e.g., Maciejovsky et al., 2007; Kastlunger et al., 2009) have sought to explain the negative effect as arising from (i) taxpayers seeking to recoup past fines by lowering their future compliance ("loss repair"); or (ii) taxpayers perceiving (incorrectly) that they become less likely to be audited in the period immediately following an audit (the "bomb-crater effect"). Mittone et al. (2017), however, reject both such explanations. Thus, the validity – both internal and external – of the negative post-audit effect observed in laboratory experiments remains questioned. We replicate the negative findings referred to here, but show that they can be made to disappear as the experimental environment is made more realistic in certain aspects.

3. Experimental design, procedure, and sample

To investigate the post-audit effect, our experiment implements the fundamental elements of voluntary income tax reporting, following the standard procedure of tax compliance experiments (Alm and Jacobson, 2007). The experiment was conducted at the Vienna Center for Experimental Economics (VCEE) in February and March 2022. Participants were recruited via ORSEE (Greiner, 2015). The sample size ($N = 273$, across 12 experimental sessions) is determined by our budget constraint.[3]

In each experimental round, an amount between 2000 and 3500 Experimental Currency Units (ECU) is drawn randomly. All participants then receive an income equal to this amount. Participants do not know the number of rounds they will play. Once participants have received their income, they decide how much income to report to the tax administration (between 0 ECU and their true income). The tax system parameters are calibrated according to the considerations detailed in Alm (2019). Specifically, in all treatments, reported income is taxed at a rate of 25 percent and the fine for noncompliance is twice the evaded tax. The audit probability, audit selection mechanism, and the amount of information that is available to participants to assess the risk of audit vary across treatments, as shall be described below.

[2] We also restrict attention to individual taxpayers, though there exists a growing literature on the sign of the post-audit effect (own and cross) for corporations (DeBacker et al., 2015; Li et al., 2018; Kotsogiannis et al., 2021).

[3] The participant pool has a slightly larger percentage of female participants (61 percent) than male participants, and the pool includes students and non-students. The mean age is 26 years with a range from 19 to 61 years. Most participants hold at least a high-school degree (63 percent) and study business/economics (12 percent). Only 16 percent state that they had previously participated in a study on tax compliance. Moreover, 35 percent indicate that they self-prepared a tax return in the past.

3.1. Treatments

Participants are randomly assigned to one of three treatments:

1. *Control (T0)* (85 participants, 30 reporting decisions);
2. *Uncertain audit probability (T1)* (89 participants, 30 reporting decisions);
3. *Endogenous audit rule (T2)* (99 participants, 25 reporting decisions).[4]

The *control* treatment is the canonical TEG, where the random audit probability is constant at 0.1 and announced to participants.

The uncertain audit probability treatment introduces scope for learning. Participants can learn regarding the audit probability based on the audit selection outcomes arising during the experiment. The treatment is otherwise identical to the control treatment, but rather than participants knowing the audit probability, they know only that it has been drawn randomly from the interval $p \in [0.05, 0.15]$, with each value in this range being equally likely. The actual audit probability in this treatment is 0.11.

The endogenous audit rule treatment also allows scope for participants to learn, but this time regarding the behavior of the other participants. Specifically, as in the control treatment, the audit probability in this treatment is constant at 0.1 and known to the participants. Participants are not randomly selected for an audit in this treatment, however. Instead, the 10% of the participants who reported the lowest amounts of income are selected for audit.[5] As all participants receive the same amount of income in a given round, this corresponds to the bottom 10% of the compliance distribution for every reporting decision.[6]

We measure three dependent variables in every round: reported income (in ECU), perceived audit probability ("audit belief") (in percentage points), and, in T2 only, perceived mean compliance of other participants (in percentage points). The precise procedure employed to elicit beliefs is detailed in section 3.2.[7] To avoid framing effects (priming), we vary the order in which participants (i) declare an amount of income; (ii) state their audit belief; and (iii) state their belief regarding the mean compliance of other participants.

3.2. Procedure

The experimental procedure is as follows. At the beginning of the experiment participants are informed that their information is private and that it is impossible to identify individual participants. Subsequently, they learn about the tax compliance game, including the tax system parameters as well as the audit selection mechanism. Participants then learn about the compensation mechanism. Each participant receives (i) a show-up fee of € 5 (in treatments T0 and T1) and of € 15 in T2; and (ii) an additional amount that reflects decisions made in the experiment.[8] Specifically, this second component of compensation is based on:

1. the after-tax income in a randomly selected round (up to € 12.25);
2. the accuracy of the elicited belief for the perceived audit probability in a randomly selected round, as described below (up to € 2);
3. the accuracy of the elicited belief for the mean compliance of other participants in a randomly selected round, as described below (up to € 2).[9]

The first component of this additional compensation incentivizes compliance choices. The second and third components incentivize the elicitation of beliefs. We elicit beliefs in the form of an interval, following the *most likely interval* method of Schlag and van der Weele (2015). Under this method, the payoff for belief elicitation reflects both (i) the width of the interval – narrower stated intervals are associated with higher payoffs; and (ii) whether or not the "true" unknown value over which beliefs are being elicited lies in the interval.[10]

[4] This treatment comprises only 25 reporting decisions to reduce the duration of this treatment. Please see footnote 8 for details.

[5] This method of audit selection is related closely to that considered in (i) Bayer and Cowell (2009) and Bayer (2022), which allows for a more general form of interdependence between audit probabilities; (ii) Alm and McKee (2004), in which audit selection is based on deviations from the average report; and (iii) Andreoni and Gee (2012, 2015), in which punishments in a public goods game are targeted at the lowest contributors.

[6] If (owing to ties) the least compliant taxpayers account for more than 10% of the sample in a round, the 10% of taxpayers to be audited is determined by a random draw over the set of least compliant taxpayers.

[7] For a further recent experimental contribution on tax compliance to utilize incentivized belief elicitation, see Dezső et al. (2022).

[8] The show-up fee is higher in the endogenous audit rule treatment because this treatment takes longer. Specifically, in this treatment all participants have to wait in every round until the slowest player has made their compliance decision. Moreover, participants additionally have to state their beliefs about the mean compliance of the other participants in every round.

[9] We elicit beliefs about the compliance of other participants only in the endogenous audit rule treatment.

[10] Specifically, this component of compensation is determined according to.

Selecting single rounds at random for each aspect of remuneration – the so-called *pay one* approach to experimental compensation (Charness et al., 2016) – addresses a possible hedging strategy that could otherwise arise in the endogenous audit rule treatment (Blanco et al., 2010).[11]

After learning about the compensation mechanism, participants read a detailed introduction to the experimental task and an example of the tax compliance decision. Subsequently, participants must answer correctly five questions on the audit probability, the audit selection mechanism, the effect of not being audited, the effect of being audited and found to be compliant, and the effect being audited and found to be noncompliant before proceeding to a practice round. In the practice round participants face the same tax system parameters as in the experiment, but they are not compensated for this round. To facilitate compliance decisions throughout the experiment, we program a calculator that shows participants how declared income translates into after-tax income, conditional on being audited or not.

Participants then proceed to the experiment. After completing the final round, they answer a short survey and some demographic questions. Treatments T0 and T1 last approximately 50 minutes, whereas the endogenous audit rule treatment T2 lasts approximately 90 minutes. The mean compensation in the control treatment T0 is € 14.20; it is € 14.30 in treatment T1, and € 21.20 in treatment T2.[12]

4. Hypotheses

In this section we formulate a model that encompasses our three experimental treatments. We then use this model to derive a set of null hypotheses regarding play in the experiment.

4.1. Theoretical framework

We develop the null hypotheses in a standard theoretical environment. Paralleling the discussion of the repeated AS model in Section 2, we suppose participants choose compliance in each round as the solution to a static maximization problem.

For a given participant in round n, let $I_n \in [2000, 3500]$ denote the amount of experimental income received, and $R_n \in [0, 1]$ be the fraction of income declared for taxation. Further, let the random variable $\tilde{p}(n|R_n)$ denote a participant's belief, conditional on reporting R_n, as to the probability that they will be audited in round $n \geq 1$. This belief evolves according to Bayesian updating, and the initial (prior) belief $\tilde{p}(1|R_1)$ is consistent with the experimental instructions. In the control treatment $\tilde{p}(n|R_n)$ is therefore a degenerate distribution, i.e., a Dirac measure at the "true" announced audit probability. In the learning treatments, however, $\tilde{p}(n|R_n)$ is non-degenerate. In the uncertain audit probability treatment the audit belief is unaffected by the participant's choice of R_n, such that $\tilde{p}(n|R_n)$ reduces to $\tilde{p}(n)$. In the endogenous audit rule treatment, however, the participant has a lower perceived risk of audit when reporting more income, hence the dependence on R.

To capture the possible variation in compliance between rounds as a consequence of the wealth shocks generated by fines, let ζ_n be the participant's initial wealth at the start of round n. ζ_n reflects wealth accrued outside the laboratory prior to the experiment, along with expected "paper" gains and losses to wealth arising from outcomes in prior experimental rounds (realized gains and losses occur only after the experiment).[13] Given the description of payoffs and parameter values in Section 3, it follows that wealth when audited (w_n^a) and when not audited ($w_n^{\neg a}$) is given by

$$w_n^a(R_n) = \zeta_n + \{1 - 0.25R_n - 0.50[1 - R_n]\}I_n; \qquad w_n^{\neg a}(R_n) = \zeta_n + [1 - 0.25R_n]I_n.$$

Let $U(.)$ denote a differentiable utility function over income (wealth) satisfying $U' > 0$, $U'' < 0$, and decreasing absolute risk aversion.[14] If a participant behaves as if they maximize expected utility, as supposed in the frameworks of, e.g., Allingham and Sandmo (1972) and Yitzhaki (1974), their choice of income declaration in round n is then the solution to

$$\max_{R_n} \mathbf{E}\{\tilde{p}(n|R_n^*)U(w_n^a(R_n)) + [1 - \tilde{p}(n|R_n^*)]U\left(w^{\neg a}(R_n)\right)\}, \tag{1}$$

where the maximum is attained for R_n^*, and $\mathbf{E}\{.\}$ is an expectations operator over the random variable $\tilde{p}(n|R_n^*)$. Note that (1) assumes that, in forming audit beliefs, participants have rational expectations over their private reporting decision: thus audit belief is $\tilde{p}(n|R_n^*)$ and does not change if the participant deviates and chooses $R_n \neq R_n^*$ (this consideration is relevant in the endogenous audit rule treatment only). Owing to (1) being linear in probabilities, we may reduce it to

$$\max_{R_n} \overline{p}(n|R_n^*)U(w_n^a(R_n)) + [1 - \overline{p}(n|R_n^*)]U\left(w^{\neg a}(R_n)\right), \tag{2}$$

where $\bar{p}(n|R_n^*) \equiv \mathbf{E}(\tilde{p}(n|R_n^*))$ is the mean audit belief. For the maximization in (2) we then have the following result.

[11] The potential for hedging arises in the endogenous audit rule treatment as, if the participant biases upwards their stated belief of the risk of audit then, in the event that their true belief is an underestimate and leads to them being selected endogenously for audit, the participant is more likely to get a high payoff in the belief task.

[12] This study was preregistered under https://osf.io/u2ts5/?viewonly=8300d55fc6f84ec18319749bcb04cd14, where the data and the code, as well as the instructions and screenshots of the experimental task, are available.

[13] Consistent with this point, Imas (2016) shows how experimental subjects react systematically to such paper gains and losses, not only to realized gains and losses.

[14] The requirement for decreasing absolute risk aversion is formally that $-U''(W)/U'(W)$ is a decreasing function of W.

Proposition 1. *For the maximization problem in (2) it holds that, at a maximum $R_n = R_n^* \in [0, 1)$, the average post-audit effect is strictly positive.*

Proof. See Appendix A. □

Proposition 1 clarifies that the model predicts a positive average post-audit effect for compliance at interior optima. In the absence of opportunities for learning (as in the control treatment), wealth effects are the sole driver of the post-audit effect. As audits do not generate a wealth effect for compliant taxpayers, a zero post-audit effect applies in this case, but the average post-audit effect is still positive on account of noncompliant taxpayers. When learning is possible, post-audit compliance can be influenced not only by wealth effects, but also by updating. In particular, experiencing an audit raises the posterior audit belief under Bayesian updating. In the uncertain audit probability treatment, the posterior belief updates as a function of the frequency of audits a participant experiences, whereas in the endogenous audit rule treatment updating of the posterior audit belief is via updating of beliefs regarding the reporting behavior of the other participants. Ceteris paribus, the higher the average reporting of the other participants, the higher is the audit belief $\bar{p}(n|R_n^*)$ at given R_n^*. The proposition clarifies that such updating of beliefs is predicted to induce a positive post-audit effect.

4.2. Post-audit effect

To map Proposition 1 to testable hypotheses regarding the experimental treatments, we first develop some notation. An experimental observation is denoted $\{i, n, T\}$, where i indexes the participant, n indexes the experimental round, and T indexes the treatment. The full set of such observations we denote by Φ. Φ may be partitioned by treatment, $T \in \{0, 1, 2\}$, as $\Phi = \bigcup_T \Phi_T$, and by participant, $i \in \{1, \ldots, 273\}$, as $\Phi = \bigcup_i \Phi_i$. To isolate the set of observations relevant to identifying the post-audit effect we define a set $A \subset \Phi$:

$$A = \{\{i, n, T\} \in \Phi : a_i(n-1) = 1, a_i(n-2) = 0\};$$

where $a_i(n) \in \{0, 1\}$ is an indicator variable for the audit of participant i in round n. Thus, A is the set of all observations that follow an audit round, with a further restriction, however, that rules out observations that follow two-or-more consecutive audits. This restriction is to mitigate possible confounding if, when a participant is audited in consecutive rounds, the post-audit effect of the subsequent audit(s) differs systematically from that of the first (we can neither rule-in nor rule-out such an effect as too few instances of consecutive audits are observed).

Let $R_i(n)$ be the fraction of income declared for taxation by participant i in round n. The change between consecutive rounds in participant i's reporting behavior we denote by

$$\Delta R_i(n) = R_i(n) - R_i(n-1). \tag{3}$$

Using the differences in (3) directly does not distinguish between reporting responses owing to audit and responses that occur irrespective of whether an audit occurs. If, for instance, a participant reduces their compliance by a fixed amount after each round, irrespective of audit, use of (3) would lead to the erroneous attribution of a (negative) post-audit effect. To prevent such mis-attribution, we measure the deviation of $\Delta R_i(n)$ from a participant's mean reporting response:

$$\Gamma_i^R(n) = \Delta R_i(n) - \overline{\Delta R}_i, \tag{4}$$

where, letting $\mathbf{E}_z$ denote an expectation (average) over z,

$$\overline{\Delta R}_i = \mathbf{E}_{\Phi_i}(\Delta R_i(n)). \tag{5}$$

We then measure the average post-audit effect under treatment T as $\overline{PA}_T$:

$$\overline{PA}_T = \mathbf{E}_{\Phi_T \cap A}\left(\Gamma_i^R(n)\right); \tag{6}$$

For a given treatment T, $\overline{PA}_T$ computes across participants the average of the Γ_i^R in rounds that follow an audit.

Applying the model in Section 4.1 to $\overline{PA}_T$ we have the following (null) hypotheses:

Hypothesis 1. (post-audit effect)

(a) In the control treatment, the average post-audit effect will be positive:

$$\overline{PA}_0(\Phi) > 0.$$

(b) In both the uncertain audit probability and endogenous audit rule treatments, the average post-audit effect will exceed that measured in the control treatment:

$$\overline{PA}_1(\Phi) > \overline{PA}_0(\Phi); \qquad \overline{PA}_2(\Phi) > \overline{PA}_0(\Phi).$$

Hypothesis 1(a) follows directly from Proposition 1. It is tested with a one-sample one-tailed t-test. A rejection of this hypothesis in favor of $\overline{PA}_0(\Phi) < 0$ would be tempting to interpret as overt evidence of probability misperception, but it may also be symptomatic of loss repair. Unlike the previous literature, however, we are able to use elicited audit beliefs to help distinguish between these alternative explanations. In particular, if a negative post-audit effect is accompanied by downwards updating of measured audit beliefs, then the evidence favors a probability misperception interpretation. If, however, a negative post-audit effect is not accompanied by downwards updating of measured audit beliefs, then the evidence favors alternative interpretations.

Hypothesis 1(b) also follows directly from Proposition 1. Positive learning effects, not present in treatment T0, are predicted to drive up the estimated post-audit effect relative to that treatment. We test the hypothesis with two-sample t-tests.

4.3. Updating and learning

To measure the extent of belief updating, we examine the updating of beliefs over the experimental rounds. Let $\{p_{il}, p_{iu}\}$ be, respectively, the elicited upper and lower bounds for the probability of audit, and $\{\mu_{il}, \mu_{iu}\}$ be the analogous elicitations for the average reporting of other participants. In both cases, we estimate a participant's mean belief as the mid-point between the elicited upper and lower bounds, i.e.,

$$\bar{p}_i(n) = \frac{p_{il} + p_{iu}}{2}; \qquad \overline{\mu}_i(n) = \frac{\mu_{il} + \mu_{iu}}{2}.$$

The change between consecutive rounds in participant i's beliefs $\{\bar{p}_i(n), \overline{\mu}_i(n)\}$ are therefore

$$\Delta\bar{p}_i(n) = \bar{p}_i(n) - \bar{p}_i(n-1); \qquad \Delta\overline{\mu}_i(n) = \overline{\mu}_i(n) - \overline{\mu}_i(n-1);$$

which we correct, analogous to the procedure in (4), as

$$\Gamma_i^p(n) = \Delta\bar{p}_i(n) - \overline{\Delta\bar{p}_i}; \qquad \Gamma_i^\mu(n) = \Delta\overline{\mu}_i(n) - \overline{\Delta\overline{\mu}_i}.$$

The average updating effects $\left\{\overline{\Delta p_T}, \overline{\Delta\mu_T}\right\}$ are then defined as

$$\overline{\Delta p_T} = \mathbf{E}_{\Phi_T\cap A}\left(\Gamma_i^p(n)\right); \qquad \overline{\Delta\mu_T} = \mathbf{E}_{\Phi_T\cap A}\left(\Gamma_i^\mu(n)\right).$$

With respect to the model in Section 4.1, our hypotheses regarding $\left\{\overline{\Delta p_T}, \overline{\Delta\mu_T}\right\}$ are as follows:

Hypothesis 2. (updating and learning)

(a) In the control treatment, the average updating of audit belief will be zero, consistent with the absence of learning:

$$\overline{\Delta p_0} = 0.$$

(b) In the uncertain audit probability treatment, the average updating of audit belief will exceed that measured in the control treatment, consistent with learning:

$$\overline{\Delta p_1} > \overline{\Delta p_0}.$$

(c) In the endogenous audit probability treatment, the average updating of reporting beliefs will be positive, consistent with learning:

$$\overline{\Delta\mu_2} > 0.$$

(d) In all treatments, the individual post-audit effect, $\Gamma_i^R(n)$, will correlate positively with individual updating of beliefs (over audit probability in T0 and T1, and over the average compliance of others in T2).

Hypothesis 2(a) follows as participants already have perfect information regarding the audit probability, hence audit is an uninformative signal. Under Bayesian updating, an uninformative signal leaves beliefs unchanged. We test the hypothesis using a two-tailed one-sample t-test. A rejection of Hypothesis 2(a) would not indicate learning – which is ruled out a priori in the control treatment – but would rather indicate that a psychological process distinct from Bayesian learning is acting on beliefs. In particular, if Hypothesis 2(a) were rejected in favor of the alternative hypothesis that audit belief falls after an audit ($\overline{\Delta p_0} < 0$), consistent with a bomb-crater effect, this would indicate the presence of probability misperception.

In the uncertain audit probability, learning is possible. Thus, under Bayesian updating, audit beliefs should rise after an audit. This would heighten the response of beliefs to audit relative to the response observed in the control treatment, hence Hypothesis 2(b). We test this hypothesis with a one-tailed two-sample t-test. An analogous argument applies in the endogenous audit rule treatment, hence Hypothesis 2(c), which is tested with a one-tailed one-sample t-test.[15] We test

[15] We write Hypothesis 2(c) against the benchmark of zero, rather than against the control treatment as – following feedback from the pilot experiments that the sessions were taking too long – we opted not to elicit beliefs over the average compliance of other participants in treatments T0 and T1.

Hypothesis 1(d), which is implied by the causal structure of the theoretical model, with a t-test applied to the separate correlation coefficients, $\{r_{\Gamma_i^R,\Delta\bar{p}_i}, r_{\Gamma_i^R,\Delta\overline{\mu}_i}\}$. This is equivalent to testing the null hypothesis $\beta_1 = 0$ in the linear regressions $\Gamma_i^R = \beta_0 + \beta_1 \Delta\bar{p}_i + \varepsilon_i$ (T0, T1) and $\Gamma_i^R = \beta_0 + \beta_1 \Delta\overline{\mu}_i + \varepsilon_i$ (T2).

4.4. Censoring

Under the theoretical model in Section 4.1, two distinct forms of censoring effect can arise:

1. *Ceiling effect*: In post-audit rounds, A, the model predicts that compliance may increase as a result of wealth effects and/or an increase in audit belief. This gives rise to censoring at the *top* of compliance spectrum, for taxpayers who were fully-compliant in the audit round cannot increase their subsequent compliance. If, using (4), we rewrite (6) as
$$\overline{PA}_T = \mathbf{E}_{\Phi_T\cap A}(\Delta R_i(n)) - \mathbf{E}_{\Phi_T\cap A}\left(\overline{\Delta R_i}\right), \tag{7}$$
then this source of censoring biases both expectations in (7) downwards. As, however, these two expectations enter $\overline{PA}_T$ with opposing sign, it might appear at first blush that the overall direction of the bias induced on $\overline{PA}_T$ is indeterminate in sign. In actuality, however, the bias must be negative in sign. To see this, note that $\mathbf{E}_{\Phi_T\cap A}(\Delta R_i(n))$ is computed solely on (potentially downwards biased) observations belonging to A. By contrast, $\mathbf{E}_{\Phi_T\cap A}\left(\overline{\Delta R_i}\right)$ is only partially constructed from A, for $\overline{\Delta R_i}$ also reflects observations from outside A, to which the ceiling effect is not predicted to apply. Thus, $\mathbf{E}_{\Phi_T\cap A}(\overline{\Delta R_i})$ is predicted to be less sensitive to bias only acting on elements of A than is $\mathbf{E}_{\Phi_T\cap A}(\Delta R_i(n))$. As a result, the ceiling effect is predicted to impart an overall downward bias on $\overline{PA}_T$. We establish this intuitive argument formally in Appendix B.
2. *Floor effect*: In all other rounds, $\Phi \setminus A$, the model predicts that the absence of audit in the prior round may cause compliance to fall as a result of wealth effects and/or a decrease in audit belief. This gives rise to censoring at the *bottom* of compliance spectrum, for taxpayers who were fully non-compliant in the prior round cannot decrease their subsequent compliance. As information in $\Phi \setminus A$ only enters $\overline{PA}_T$ through $\overline{\Delta R_i}$, $\mathbf{E}_{\Phi_T\cap A}(\Delta R_i(n))$ is predicted to be unaffected by the floor effect. Rather, the potential for bias (upwards) arises only in $\mathbf{E}_{\Phi_T\cap A}\left(\overline{\Delta R_i}\right)$. Thus, as $\mathbf{E}_{\Phi_T\cap A}\left(\overline{\Delta R_i}\right)$ enters $\overline{PA}_T$ negatively, such upward bias in $\mathbf{E}_{\Phi_T\cap A}\left(\overline{\Delta R_i}\right)$ translates into a downward bias on the post-audit effect $\overline{PA}_T$.

As, under the model, both the floor and ceiling effects act to bias downwards estimates of the post-audit effect, the theoretical prediction for the effect of censoring is unambiguous. To investigate whether indeed censoring effects bias downwards our estimate of the post-audit effect in this way – and, if so, by how much – we construct an alternative measure designed to mitigate the effects of censoring. To diminish the ceiling effect, we exclude observations from those post-audit rounds where full compliance was observed in the audit round. Formally, the set of observations excluded is written as $A \cap \Phi^{R=1}$, where
$$\Phi^{R=1} = \{\{i, n, T\} \in \Phi : R_i(n-1) = 1\}$$
is the set of observations that follow an observation with full compliance. Similarly, to dampen the floor effect, we exclude observations belonging to $\Phi \setminus A$ where additionally no income was reported in the prior round. This is, formally the set $(\Phi \setminus A) \cap \Phi^{R=0}$, where
$$\Phi^{R=0} = \{\{i, n, T\} \in \Phi : R_i(n-1) = 0\}$$
is the set of observations that follow an observation with full non-compliance. Under the procedure above, the reduced observation set is
$$\Phi^C = \Phi \setminus \left(A \cap \Phi^{R=1}\right) \setminus \left((\Phi \setminus A) \cap \Phi^{R=0}\right).$$

We then compute estimates of the post-audit effect in an identical manner to that outlined in Section 4.2, except that Φ is everywhere replaced with Φ^C. The resulting "censoring corrected" estimate we denote by $\overline{CCPA}_T$:
$$\overline{CCPA}_T = \mathbf{E}_{\Phi_T^C\cap A}\left(\Gamma_i^R(n)\right),$$
where $\Gamma_i^R(n)$ is constructed using $\overline{\Delta R_i} = \mathbf{E}_{\Phi_i^C}(\Delta R_i(n))$.

We then have the following hypotheses:

Hypothesis 3. (censoring)

(a) In the control and uncertain audit probability treatments, the effects of censoring on changes in reporting behavior will bias downwards estimates of the average post-audit effect:
$$\overline{CCPA}_0 > \overline{PA}_0; \qquad \overline{CCPA}_1 > \overline{PA}_1.$$
(b) In the endogenous audit probability treatment, there will be a zero censoring effect:
$$\overline{CCPA}_2 = \overline{PA}_2.$$

Hypotheses 3(a)-(b) are both tested with a two-sample t-test. While Hypothesis 3(a) follows directly from the prior discussion, Hypothesis 3(b) requires further comment. Owing to the endogenous nature of audit selection in treatment T2, audits are targeted systematically at non-compliant taxpayers, such that the set $A \cap \Phi^{R=1}$ is expected to be empty. Accordingly, a zero ceiling effect is predicted. Also, if participants believe that reporting $R_i = 0$ will lead to certain audit (a correct belief given how actual play in the treatment realized) then $R_i = 0$ is never a prediction of the model. Thus, the set $(\Phi \setminus A) \cap \Phi^{R=0}$ is also predicted to be empty, with the implication of a zero floor effect.

5. Results

In this section we describe the results of the experiment in respect of the post-audit effect, and the roles of probability misperception, learning, and censoring.

5.1. Post-audit effect

The results for our measure of the post-audit effect, $\overline{PA}_T$, across the three experimental treatments are illustrated in Fig. 1.

According to Hypothesis 1(a) a positive post-audit effect will be observed in the control treatment. The results, however, reject this hypothesis decisively (p <.001), for we find a strong negative post-audit effect $\overline{PA}_0$=−13.2. That is, average compliance falls 13.2 percentage points in the round after an audit. Our point estimates for the post-audit effect in the two learning treatments (T1 and T2), however, are higher than in the control treatment, in line with Hypothesis 1(b). Specifically, we obtain $\overline{PA}_1$ =−2.8 and $\overline{PA}_2$=19.2, such that we cannot reject the hypothesis $\overline{PA}_1 > \overline{PA}_0$ (p <.01) or the hypothesis $\overline{PA}_2 > \overline{PA}_0$ (p <.001) at conventional levels. Moreover, although not one of our initial hypotheses, it is apparent that the uplift in the post-audit effect (relative to the control) is larger in the endogenous audit rule treatment than in the uncertain audit treatment (p <.001). As a consequence, although $\overline{PA}_2$ is positive (p <.001), the sign of $\overline{PA}_1$ is unclear: as seen in Fig. 1, the confidence intervals cross zero. To begin to interpret these findings, we now turn to the results on elicited beliefs.

5.2. Updating and learning

We now describe our results regarding the updating of beliefs, as depicted in Fig. 2. Hypothesis 2(a) is that there should be no updating of audits beliefs in the control treatment. As seen in Fig. 2, however, our point estimate for updating is $\overline{\Delta p}_0$=−4.3. That is, average audit belief falls by 4.3 percentage points in the round after an audit. Accordingly, we reject the null hypothesis of no updating at conventional levels of significance (p <.001), but cannot reject the alternative hypothesis that audit belief falls after an audit. Moreover, at the participant level, changes in audit belief, $\Delta \bar{p}_i$, correlate positively with changes in reporting, per Hypothesis 2(d) (p <.001). Accordingly, the results support the inference that the negative post-audit effect found in the control treatment traces to probability misperception. This is a remarkable finding considering

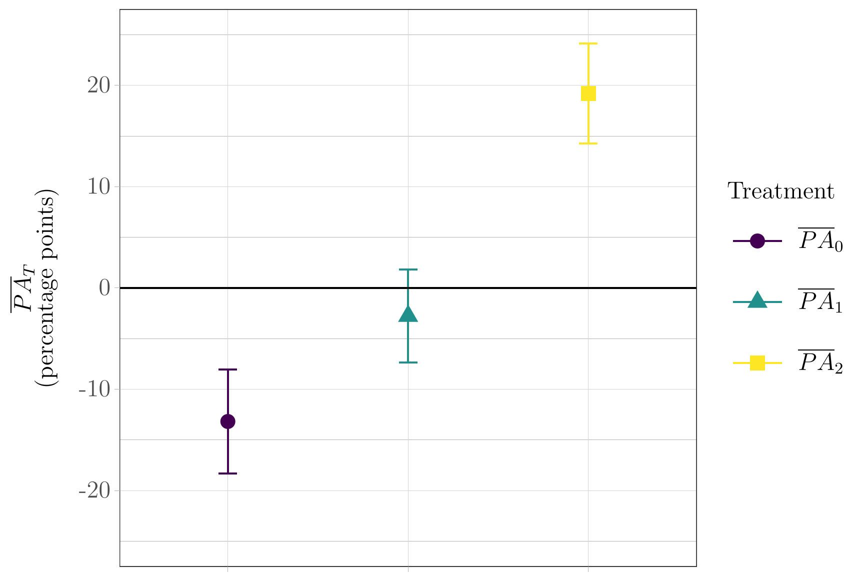

Fig. 1. Average post-audit compliance effect by treatment. $T = 0$ is the control treatment, $T = 1$ is the uncertain audit probability treatment, and $T = 2$ is the endogenous audit rule treatment.

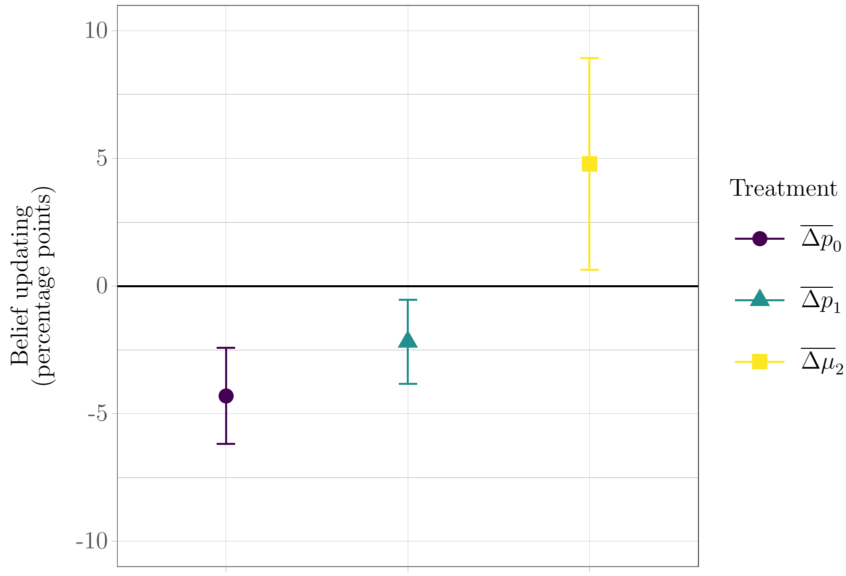

Fig. 2. Average updating of beliefs by treatment. In $T = 0$ (control treatment) and $T = 1$ (uncertain audit probability treatment) updating is of audit belief. In $T = 2$ (endogenous audit rule treatment) updating is of belief regarding mean compliance of other participants.

that the true audit probability did not vary throughout the experimental rounds, that participants were informed of the true probability at the beginning of each and every round, and that belief elicitation was incentivized. As such, it provides compelling evidence of a bomb-crater effect in compliance behavior.

In the treatments which allow for learning we observe the predicted positive learning effects. According to Hypothesis 2(b), learning in the uncertain audit treatment will heighten (relative to the control) the updating of audit belief following an audit. Consistent with this hypothesis, we estimate a higher point estimate $\overline{\Delta p_1}$=−2.2, and this uplift is statistically significant (p <.05), albeit only at the 5 percent level. The positive learning effect is, however, insufficiently strong to countervail fully the negative effect of probability misperception, such that the overall updating of beliefs, $\overline{\Delta p_1}$, remains negative (p <.01). At the participant level, changes in audit belief, $\Delta \bar{p}_i$, correlate positively with changes in reporting (p <.01). Consistent with Hypothesis 2(c), we observe a positive learning effect (regarding the average compliance of other participants) in the endogenous audit rule treatment: in the round after an audit, beliefs $\overline{\mu}_i$ increase on average by 4.8 percentage points ($\overline{\Delta \mu_2}$=4.8, p <.001). Changes in compliance belief, $\Delta\overline{\mu}_i$, correlate positively with changes in reporting ($r_{\Gamma_i^R, \Delta\overline{\mu}_i}$=.1), albeit there is too much noise in these data to establish statistical significance for this correlation at conventional levels (p <.23).

As a final point we note that, as in treatments T0 and T1, audit belief under endogenous audit selection, $\overline{\Delta p_2}$, also falls after an audit ($\overline{\Delta p_2}$=−5.2). One interpretation of this finding is that the probability misperception witnessed under exogenous audit selection is robust to the switch to endogenous selection. Such an interpretation, however, misses that, under endogenous selection, the fall in audit belief is rational, given that participants choose on average to increase their compliance. That is, as audited participants subsequently increased their compliance, they were objectively less likely to be audited in the subsequent round. In contrast, under exogenous selection, the fall in audit belief witnessed after audit has no rational basis. Accordingly, although we cannot rule out the presence of probability misperception under endogenous audit selection, the evidence points to a reduced role for this effect.

5.3. Censoring

The final set of results we consider are those relating to the possible effects of censoring. Our findings, which indicate a prominent role for censoring effects, are illustrated in Fig. 3, which depicts, side by side, our corrected and uncorrected estimates of the post-audit compliance effect.

According to Hypothesisa0~3(a), there should be a negative censoring effect in treatments T0 and T1, such that the corrected measure $\overline{CCPA}$ will exceed the uncorrected measure $\overline{PA}$. As seen in the figure, this hypothesis is supported firmly by these data. In the control treatment, the corrected measure $\overline{CCPA}_0$ =7.3 is discernibly higher than (and of a different sign to) the uncorrected estimate $\overline{PA}_0 = -13.2$ (p <.001). A similar picture applies also in the uncertain audit treatment, where $\overline{CCPA}_1 = 13.8 > \overline{PA}_1 = -2.8$ (p <.001). The scale of these effects is considerable and confirms our concern in the introduction that, owing to a combination of (i) the high frequency of extreme reporting behavior $R \in \{0, 1\}$ in laboratory tax experiments; and (ii) the use of undirected audit selection mechanisms, such experiments are highly susceptible to

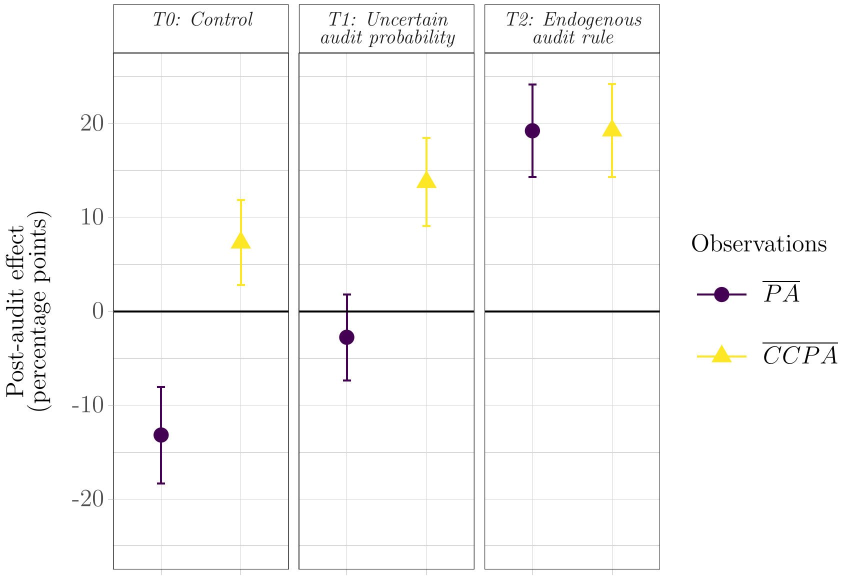

Fig. 3. The effects of censoring on the estimated post-audit effect. The $\overline{CCPA}$ are our estimates of the post-audit effect, corrected for censoring. For comparison, the $\overline{PA}$ are our estimates of the post-audit effect, uncorrected for censoring, in Fig. 1.

negative censoring effects.[16] In particular, as field-audits are risk-based – systematically targeting taxpayers likely to be non-compliant – it is likely that reporting behavior in laboratory tax experiments is impacted by censoring effects to a greater extent than is reporting behavior in the field.

As, even in the control treatment, the censoring-corrected post-audit effect is positive, does this call into question our earlier finding that probability misperception exerts a downward bias on estimates of the post-audit effect? The evidence suggests not: in the control treatment, average audit belief continues to fall after an audit in the restricted sample Φ^C. The fall is less pronounced – a 2.6 percentage point fall when using Φ^C compared to a 4.3 percentage point fall when using the full sample Φ – but the effect remains significant at the one percent level. Thus, in the absence of probability misperception, we should expect to have estimated a still more positive corrected post-audit effect.

The endogenous audit rule treatment shares with practice in the field that audits are directed systematically towards non-compliant taxpayers. So long as most taxpayers are at least partially compliant, it should also deter full non-compliance through certain audit. For these reasons, Hypothesis 3(b) is that this treatment will be immune to the effects of censoring. Consistent with this reasoning, we cannot reject the hypothesis $\overline{CCPA}_2 = \overline{PA}_2$ at conventional levels (p <.50). As anticipated, all audits fell on participants that were non-compliant. The set $A \cap \Phi^{R=1}$ of excluded observations is thus empty, thereby ruling out a ceiling effect of the type discussed in Section 4.4. The audit mechanism did not deter all reporting of zero income, but the frequency of this behavior was limited to the extent that the set of excluded observations $(\Phi \setminus A) \cap \Phi^{R=0}$ contains just one entry. The resulting floor effect is, therefore, too small to be significant statistically.

6. Conclusion

The sign and size of the post-audit effect is an important ingredient in determining optimal enforcement of the tax system. Prior literature has struggled to agree on the sign of the effect, however, with field studies generally pointing to a positive effect, but experimental findings often pointing to a negative effect.

As discussed in the Introduction, much of the existing field literature has the characteristic that findings indicating a positive post-audit effect are interpreted as signifying both (i) the presence of a rational deterrence effect, operating via learning; and (ii) the absence of a bomb-crater effect. Proceeding from this perspective, the apparent presence of the bomb-crater effect in laboratory outcomes must then be interpreted as evidence of a failure of external validity with respect to laboratory experiments. The perspective suggested by our findings is rather different. It is that evidence – both from the field and the laboratory – will simultaneously comprise effects due to learning, probability misperception (driving a bomb-crater effect), and censoring. It is only the sizes of these competing effects that may differ systematically between the field and the laboratory. This view is supported by the observation that, as we harmonize conditions in the laboratory towards those in the field (a process which would be expected to converge the sizes of the effects discussed with those observed in

[16] The reason our censoring effects are so large may be understood by computing the post-audit effect just on the excluded sample $\Phi_T \setminus \Phi^C$. Doing this for T0 yields a post-audit effect of -28.1, more than twice as negative as the post-audit effect we measure for the full sample. For T1 the negative post-audit effect for the sample of observations $\Phi_1 \setminus \Phi^C$ is even more stark at -32.4. This is more than 11 times more negative than the average post-audit effect for the full sample, a result comparable to the findings reported in Kasper and Alm (2022b).

the field), our findings in the laboratory indeed converge towards those in the field. Moreover, from this alternative vantage, one might go as far as to argue that, in light of the systematic differences between laboratory and field settings in the existing literature, it would be more surprising were the outcomes from the experimental literature convergent with those from field studies, than were they divergent.

We also sense that, although it plays a role, probability misperception has so far been given undue weight in attempts to explain the negative post-audit effect found in tax experiments. An alternative effect, as yet barely discussed in the literature, that may nonetheless be driving much of the observed effect is censoring. Specifically, we do find the tendency for laboratory participants, when facing an audit probability that is exogenous to their own actions, to subjectively lower their perceived probability of audit in the round after an audit. As this finding applies despite us regularly reminding participants of the true (and constant) audit probability, and even providing financial incentives to indicate the correct audit probability, we would speculate that probability misperception might be even more prevalent in experimental designs without these features. That said, we also find evidence of censoring effects that seem of even greater quantitative significance. Such censoring effects act to amplify the downward bias already present in estimates of the post-audit effect due to probability misperception.

We finish with the implications of our analysis. For tax compliance researchers, our findings underscore the need for care when comparing *quantitative* predictions across studies from the field and the laboratory. We have demonstrated that, quantitatively, the results of the tax experimental game are sensitive to design choices. Accordingly, to the extent that design choices result in a wedge between enforcement in the laboratory versus in the field, a divergence in quantitative findings is to be anticipated. By contrast, our *qualitative* findings – that the post-audit effect is mediated by opportunities for learning, by probability misperception, and by the effects of censoring – are largely robust across treatments and consistent with results in the field. For practitioners in tax administrations, seemingly the most eye-catching feature of our results is the extent to which the post-audit effect under an endogenous audit rule exceeds that in the remaining treatments (in which audit selection is exogenous). Unpicking the mechanisms that may be driving this effect – including a possible role for social norms – is important before precise policy prescriptions can be made, but one possibility is that it is desirable to create a strong endogenous link between reporting and audit probability. Such an endogenous link may give the audit probability a more tangible quality, such that taxpayers seek to exercise skill and judgement over it through their reporting decision. By contrast, an exogenous audit probability has a more abstract quality, and is, accordingly, prone to (unhelpful) downward misperception. We hope future researchers will be inspired to take up these loose threads.

Declaration of competing interest

None.

Appendix A. Proof of Proposition 1

The first order condition for an interior maximum, $R_n \in (0,1)$, is given by

$$0.25\Big\{\bar{p}(n,R_n^*)U'(w_n^a(R_n)) - [1-\bar{p}(n,R_n^*)]U'(w_n^{-a}(R_n))\Big\}I_n = 0. \tag{A.1}$$

The second derivative of expected utility with respect to R is

$$D_n \equiv 0.25^2\Big\{\bar{p}(n,R_n^*)U''(w_n^a(R_n)) + [1-\bar{p}(n,R_n^*)]U''(w_n^{-a}(R_n))\Big\}I_n < 0.$$

Thus, the second order condition for a maximum, $D<0$, holds. Using the implicit function theorem in (A.1) we then have

$$\frac{\partial R_n}{\partial \bar{p}(n,R_n^*)} = \frac{[U'(w_n^{-a}(R_n)) + U'(w_n^a(R_n))]I_n}{4D_n} > 0. \tag{A.2}$$

An audit in round n will cause $\bar{p}\big(n+1,R_{n+1}^*\big)$ to update above $\bar{p}(n,R_n^*)$. From (A.2), it follows that income reporting will be higher than if no audit in round n had occurred. If the taxpayer is found to be non-compliant in round n, and therefore pays a fine, we have $\zeta_{n+1} < \zeta_n$. This too raises reporting in $n+1$ relative to if no audit in round n had occurred, for the wealth effect under decreasing absolute risk aversion makes the participant poorer, and therefore more risk averse.

Appendix B. Sign of Ceiling Effect

From (5), $\overline{\Delta R_i} = \mathbf{E}_{\Phi_i}(\Delta R_i(n))$. Partitioning Φ_i as $(\Phi_i \cap A)\cup(\Phi_i \setminus A)$ we may rewrite $\mathbf{E}_{\Phi_i}(\Delta R_i(n))$, and hence $\overline{\Delta R_i}$, as

$$\overline{\Delta R_i} = \frac{1}{|\Phi_i|}\Big[|\Phi_i \cap A|\mathbf{E}_{\Phi_i\cap A}(\Delta R_i(n)) + |\Phi_i \setminus A|\mathbf{E}_{\Phi_i\setminus A}(\Delta R_i(n))\Big]. \tag{B.1}$$

Hence, using (B.1), we have

$$\begin{aligned}\mathbf{E}_{\Phi_T\cap A}\big(\overline{\Delta R_i}\big) &= \mathbf{E}_{\Phi_T\cap A}\left(\frac{1}{|\Phi_i|}\Big[|\Phi_i \cap A|\mathbf{E}_{\Phi_i\cap A}(\Delta R_i(n)) + |\Phi_i \setminus A|\mathbf{E}_{\Phi_i\setminus A}(\Delta R_i(n))\Big]\right)\\ &= \frac{1}{|\Phi_i|}\left[\begin{array}{l}|\Phi_i \cap A|\mathbf{E}_{\Phi_T\cap A}\big(\mathbf{E}_{\Phi_i\cap A}(\Delta R_i(n))\big)\\ +|\Phi_i \setminus A|\mathbf{E}_{\Phi_T\cap A}\big(\mathbf{E}_{\Phi_i\setminus A}(\Delta R_i(n))\big)\end{array}\right].\end{aligned} \tag{B.2}$$

Noting that $\mathbf{E}_{\Phi_T\cap A}\left(\mathbf{E}_{\Phi_i\cap A}(\Delta R_i(n))\right) = \mathbf{E}_{\Phi_T\cap A}(\Delta R_i(n))$, (B.2) reduces to

$$\mathbf{E}_{\Phi_T\cap A}\left(\overline{\Delta R}_i\right) = \frac{1}{|\Phi_i|}\left[\begin{array}{c}|\Phi_i\cap A|\mathbf{E}_{\Phi_T\cap A}(\Delta R_i(n))\\ +|\Phi_i\setminus A|\mathbf{E}_{\Phi_T\cap A}\left(\mathbf{E}_{\Phi_i\setminus A}(\Delta R_i(n))\right)\end{array}\right]. \tag{B.3}$$

Substituting (B.3) into (7) we finally obtain

$$\overline{PA}_T = \frac{1}{|\Phi_i|}\left\{[|\Phi_i| - |\Phi_i\cap A|]\mathbf{E}_{\Phi_T\cap A}(\Delta R_i(n)) - |\Phi_i\setminus A|\mathbf{E}_{\Phi_T\cap A}\left(\mathbf{E}_{\Phi_i\setminus A}(\Delta R_i(n))\right)\right\}. \tag{B.4}$$

It follows from (B.4) that the marginal effect on $\overline{PA}_T$ of an increase in $\mathbf{E}_{\Phi_T\cap A}(\Delta R_i(n))$ is given by $|\Phi_i| - |\Phi_i\cap A| > 0$. Thus, a ceiling effect that decreases $\mathbf{E}_{\Phi_T\cap A}(\Delta R_i(n))$ also decreases $\overline{PA}_T$.

References

Advani, A., Elming, W., Shaw, J., 2021. The dynamic effects of tax audits. Rev. Econ. Stat.. doi: 10.1162/rest_a_01101
Allingham, M.G., Sandmo, A., 1972. Income tax evasion: a theoretical analysis. J. Public Econ. 1 (3–4), 323–338.
Alm, J., 2019. What motivates tax compliance? J. Econ. Surv. 33 (2), 353–388.
Alm, J., Bloomquist, K.M., McKee, M., 2015. On the external validity of laboratory tax compliance experiments. Econ. Inq. 53 (2), 1170–1186.
Alm, J., Cronshaw, M.B., McKee, M., 1993. Tax compliance with endogenous audit selection rules. Kyklos 46 (1), 27–45.
Alm, J., Jackson, B., McKee, M., 1992. Institutional uncertainty and taxpayer compliance. Am. Econ. Rev. 82 (4), 1018–1026.
Alm, J., Jacobson, S., 2007. Using laboratory experiments in public economics. Natl. Tax J. 60 (1), 129–152.
Alm, J., Malézieux, A., 2021. 40 years of tax evasion games: a meta-analysis. Exp. Econ. 24 (3), 699–750.
Alm, J., McKee, M., 2004. Tax compliance as a coordination game. J. Econ. Behav. Organ. 54 (3), 297–312.
Andreoni, J., Gee, L., 2012. Gun for hire: delegated enforcement and peer punishment in public goods provision. J. Public Econ. 96 (11), 1036–1046.
Andreoni, J., Gee, L., 2015. Gunning for efficiency with third party enforcement in threshold public goods. Exp. Econ. 18 (1), 154–171.
Bayer, R.-C., 2022. The double dividend of relative auditing – theory and experiments on corporate tax enforcement. J. Public Econ. Theory 24 (6), 1433–1462. doi: 10.1111/jpet.12587
Bayer, R.-C., Cowell, F., 2009. Tax compliance and firms' strategic interdependence. J. Public Econ. 93 (11–12), 1131–1143.
Beer, S., Kasper, M., Kirchler, E., Erard, B., 2020. Do audits deter or provoke future tax noncompliance? Evidence on self-employed taxpayers. CESifo Econ. Stud. 66 (3), 248–264.
Best, M., Shah, J., Waseem, M., 2021. Detection without deterrence: Long-run effects of tax audit on firm behavior. Paper presented at the 8th Annual MannheimTaxation Conference, May 2021, Mannheim.
Blanco, M., Engelmann, D., Koch, A., Normann, H.-T., 2010. Belief elicitation in experiments: is there a hedging problem? Exp. Econ. 13 (4), 412–438.
Bloomquist, K.M., 2009. A comparative analysis of reporting compliance behavior in laboratory experiments and random taxpayer audits. In: 102nd Annual Conference Proceedings. National Tax Association, Denver, CO, pp. 113–122.
Boning, W.C., Guyton, J., Hodge, R., Slemrod, J., 2020. Heard it through the grapevine: direct and network effects of a tax enforcement field experiment. J. Public Econ. 190, 104261.
Charness, G., Gneezy, U., Halladay, B., 2016. Experimental methods: pay one or pay all. J. Econ. Behav. Organ. 131, 141–150.
Choo, L., Fonseca, M., Myles, G., 2013. Lab experiment to investigate tax compliance: audit strategies and messaging. Research Report 308. H.M. Revenue & Customs, London.
Choo, L., Fonseca, M., Myles, G., 2016. Do students behave like real taxpayers in the lab? Evidence from a real effort tax compliance experiment. J. Econ. Behav. Organ. 124, 102–114.
Dai, Z., Galeotti, F., Villeval, M.C., 2018. Cheating in the lab predicts fraud in the field: an experiment in public transportation. Manag. Sci. 64 (3), 1081–1100.
DeBacker, J., Heim, B.T., Tran, A., Yuskavage, A., 2015. Legal enforcement and corporate behavior: an analysis of tax aggressiveness after an audit. J. Law Econ. 58 (2), 291–324.
DeBacker, J., Heim, B.T., Tran, A., Yuskavage, A., 2018. The effects of IRS audits on EITC claimants. Natl. Tax J. 71 (3), 451–484.
DeBacker, J., Heim, B.T., Tran, A., Yuskavage, A., 2018. Once bitten, twice shy? The lasting impact of IRS audits on individual tax reporting. J. Law Econ. 61 (1), 1–35.
Dezső, L., Alm, J., Kirchler, E., 2022. Inequitable wages and tax evasion. J. Behav. Exp. Econ. 96, 101811.
Drago, F., Mengel, F., Traxler, C., 2020. Compliance behavior in networks: evidence from a field experiment. Am. Econ. J.: Appl. Econ. 12 (2), 96–133.
Erard, B., Kirchler, E., Olsen, J., 2020. The specific deterrence implications of increased reliance on correspondence audits. Paper presented at the 2020 IRS/TPC Research Conference, Washington, D.C.
Fellner, G., Sausgruber, R., Traxler, C., 2013. Testing enforcement strategies in the field: threat, moral appeal and social information. J. Eur. Econ. Assoc. 11 (3), 634–660.
Friedland, N., 1982. A note on tax evasion as a function of the quality of information about the magnitude and credibility of threatened fines: some preliminary research. J. Appl. Soc. Psych. 12 (1), 54–59.
Gemmell, N., Ratto, M., 2012. Behavioral responses to taxpayer audits: evidence from random taxpayer inquiries. Natl. Tax J. 65 (1), 33–58.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.
Imas, A., 2016. The realization effect: risk-taking after realized versus paper losses. Am. Econ. Rev. 106 (8), 2086–2109.
Kasper, M., Alm, J., 2022. Audits, audit effectiveness, and post-audit tax compliance. J. Econ. Behav. Organ. 195, 87–102.
Kasper, M., Alm, J., 2022. Does the bomb-crater effect really exist? Evidence from the laboratory. Finanzarchiv 78 (1–2), 87–111.
Kastlunger, B., Kirchler, E., Mittone, L., Pitters, J., 2009. Sequences of audits, tax compliance, and taxpaying strategies. J. Econ. Psych. 30 (3), 405–418.
Kleven, H.J., Knudsen, M.B., Kreiner, C.T., Pedersen, S., Saez, E., 2011. Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. Econometrica 79 (3), 651–692.
Kotsadam, A., Løyland, K., Raaum, O., Torsvik, G., Øvrum, A., 2021. Does perceived risk of future audits explain the behavioral effects of tax enforcement? Paper presented at the 12th Norwegian German Seminar on Public Economics, November 2021, Munich.
Kotsogiannis, C., Salvadori, L., Karangwa, J., Mukamana, T., 2021. Do tax audits have a dynamic impact? Evidence from corporate income tax administrative data. TARC Discussion Paper 035 - 21, University of Exeter.
Li, W., Pittman, J.A., Wang, Z.-T., 2018. The determinants and consequences of tax audits: some evidence from China. J. Am. Tax. Assoc. 41 (1), 91–122.
Løyland, K., Raaum, O., Torsvik, G., Øvrum, A., 2019. Compliance effects of risk-based tax audits. CESifo Working Paper, No. 7616. 10.2139/ssrn.3384307
Maciejovsky, B., Kirchler, E., Schwarzenberger, H., 2007. Misperceptions of chance and loss repair: on the dynamics of tax compliance. J. Econ. Psych. 28 (6), 678–691.
Mazzolini, G., Pagani, L., Santoro, A., 2022. The deterrence effect of real-world operational tax audits on self-employed taxpayers: evidence from Italy. Int. Tax Public Finance 29 (4), 1014–1046.
Mittone, L., 2006. Dynamic behaviour in tax evasion: an experimental approach. J. Socio-Econ. 35 (5), 813–835.
Mittone, L., Panebianco, F., Santoro, A., 2017. The bomb-crater effect of tax audits: beyond the misperception of chance. J. Econ. Psych. 61, 225–243.

Rincke, J., Traxler, C., 2011. Enforcement spillovers. Rev. Econ. Stat. 93 (4), 1224–1234.
Sánchez, G.E., 2022. Non-compliance notifications and taxpayer strategic behavior: evidence from Ecuador. Int. Tax Public Finance 29 (3), 627–666.
Schlag, K.H., van der Weele, J.J., 2015. A method to elicit beliefs as most likely intervals. Judgm. Decis. Mak. 10 (5), 456–468.
Slemrod, J., 2019. Tax compliance and enforcement. J. Econ. Lit. 57 (4), 904–954.
Slemrod, J., Blumenthal, M., Christian, C., 2001. Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. J. Public Econ. 79 (3), 455–483.
Tan, F., Yim, A., 2014. Can strategic uncertainty help deter tax evasion? An experiment on auditing rules. J. Econ. Psych. 40, 161–174.
Yitzhaki, S., 1974. A note on 'Income tax evasion: a theoretical analysis'. J. Public Econ. 3 (2), 201–202.