Florida Tax Review

Volume 20 Article 1

Can Audits Encourage Tax Evasion?: An Experimental Assessment

Emily Satterthwaite
University of Toronto

FLORIDA TAX REVIEW

Volume 20 *2016* *Number 1*

CAN AUDITS ENCOURAGE TAX EVASION?: AN EXPERIMENTAL ASSESSMENT

by

Emily Satterthwaite[*]

Abstract

Governments and tax administrators around the world rely on the premise that audits will deter tax evasion. This Article presents experimental evidence that this premise may be, at least in part, misguided. Counterintuitively, I find that audits presented as random may induce taxpayers to cheat more. Where audits were described as being conducted at random, participants increased their levels of evasion in the tax periods immediately following the audit. This effect, however, did not plague nonrandom audits. When a separate group of participants faced audits that were presented as being nonrandom—participants were told that detected evasion would "flag" a participant for one or more future audits—participants cheated less in the periods immediately following the audit. Overall, average compliance in the nonrandom audit condition systematically and significantly dominated average compliance in the random audit condition. By revealing, under experimental conditions, strong behavioral responses to the way tax audits are presented, this Article highlights the potential enforcement benefits of being more transparent with taxpayers about the nature of audit selection.

---

* Assistant Professor, University of Toronto Faculty of Law. For helpful comments, thank you to Benjamin Alarie, James Alm, Lisa Austin, Joseph Bankman, Wei Cui, Yasmin Dawood, Angela Fernandez, Brian Galle, Andrew Hayashi, Erich Kirchler, Jack Manhire, Susan Morse, Shuyi Oei, Jason Oh, Leigh Osofsky, Mariana Mota Prado, Diane Ring, Ted Seto, Daniel Shaviro, Stephen Shay, Michael Trebilcock, Albert Yoon, and workshop participants at the 2015 Law & Society Association Annual Meeting, the 2015 Junior Tax Scholars' Workshop, the 2015 Canadian Law & Economics Association Meetings, the National Tax Association 108th Annual Conference, and the Tulane Tax Roundtable. Special thanks to Leandra Lederman for feedback and questions. The author is grateful for generous support from the University of Toronto's 2015–16 Connaught New Researcher Award. Dennis (Zhenyu) Luo provided outstanding research assistance. All errors and omissions are mine.

INTRODUCTION

The legal obligation to self-assess individual income and calculate the resulting tax liability arises annually in most developed taxing jurisdictions.[1] Individuals typically join the ranks of taxpayers upon entering the workforce and remain there until death intercedes. Over this stretch of time, an individual taxpayer's experience of self-assessment is likely to be affected by a range of possible interactions with the taxing agency, regardless of whether or not self-assessment is mediated by a tax preparer, an accountant, or an attorney.

The repeat nature of self-assessment and the corresponding potential for audit is an obvious structural feature of the U.S. system of income tax administration. However, until quite recently, there has been a relative dearth of empirical research[2] examining the question of how an individual's tax compliance trajectory changes over time.[3] One possible explanation is that the standard theoretical model of an individual's tax compliance decision focuses on a discrete number of variables—the random audit probability, penalty rate, tax rate, and earned income subject to tax—as determinants of the expected utility from declaring none, some, or all of one's income to the tax authorities.[4] The past experience of being audited is generally not

1. *See, e.g.*, I.R.C. §§ 6012(a)(1)(A), 7701(a)(23).

2. There is, however, a theoretical literature that extends and modifies the standard model to incorporate the possibility of taxpayer learning and changing behavior in response to audits. *See infra* Part I.A.

3. *See, e.g.*, Barbara Kastlunger et al., *Sequences of Audits, Tax Compliance, and Taxpaying Strategies*, 30 J. ECON. PSYCHOL. 405, 406 (2009) ("To date, paying taxes as a repeated duty with random audits or audits following a specific pattern has not been studied empirically."). A key exception is a 1993 experimental paper by James Alm et al., which compares the overall performance (aggregate compliance over 20 rounds) of eight different random and nonrandom (endogenous) audit selection rules. It does not focus in particular on post-audit compliance patterns, but it is notable for its use of endogenous audit rules (EARs), and this Article's treatment experimental condition exploits a slightly different version of Alm et al.'s "conditional future audit" rule. *See* James Alm et al., *Tax Compliance with Endogenous Audit Selection Rules*, 46 KYKLOS 27, 37 (1993) [hereinafter Alm et al., KYKLOS] (defining the conditional future audit rule as initial five percent random audits followed, in the event of underreporting, by certain audits for two consecutive periods); *see also infra* Part I.B (containing detailed discussion of the more recent experimental research on the dynamics of tax compliance).

4. Michael G. Allingham & Agnar Sandmo, *Income Tax Evasion: A Theoretical Analysis*, 3 J. PUB. ECON. 323, 324 (1972) (the baseline model's exogenous parameters are actual income (which "is known by the taxpayer but not by the government's tax collector"), rate of tax on declared income, probability that

included in the standard model.

Why? The reason is straightforward. The lived experience of being audited is merely the probabilistic manifestation of a known random variable: the audit rate. As the audit rate already has been included in the taxpayer's expected utility function, the rational taxpayer is assumed to have taken it into account in her optimal compliance decision, *ex ante*.[5] In the standard model of tax compliance, the experience of a past audit should be wholly irrelevant to a taxpayer's post-audit compliance behavior.[6]

Notwithstanding the predictions of the standard model, there is a growing body of empirical evidence suggesting that the experience of being audited *is* relevant to an individual's subsequent tax compliance choices.[7] As between the two possible directional outcomes, however, there is evidence pointing in both directions: in some studies, audits appear to bolster post-audit compliance levels. In others, audits are associated with subsequent drops in compliance. The ambiguous findings are harder to gloss over because there are a set of plausible rationales that might explain each of the two outcomes.

On the first outcome, an increase in post-audit compliance for a given taxpayer is intuitively consistent with the "direct deterrence" rationale for

---

the taxpayer will be audited, and a penalty rate (which is constrained to be higher than the tax rate) on undeclared income detected in an audit. The taxpayer's decision variable is declared, or reported, income.). *See also* T.N. Srinivasan, *Tax Evasion: A Model*, 2 J. PUB. ECON. 339, 340 (1973).

5. Note that Allingham and Sandmo relax the assumption of a one-time audit at the very end of their paper and explore the dynamic case for both the "myopic" and the "consistent" individual. *See* Allingham & Sandmo, *supra* note 4, at 332–38.

6. This statement would need to be qualified to take into account any departures from or refinements to the standard model, such as taxpayer uncertainty about the probability of audit, the use by the government of a nonrandom audit strategy, or a time dimension in which the audit rate is not constant across time. With respect to the first example, taxpayer uncertainty about the audit rate, it is easy to imagine that experience of an audit (or knowledge of another individual being audited) might induce the taxpayer to update her expectations about the true audit rate. See discussion of this possibility in Henrik Jacobsen Kleven et al., *Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark*, 79 ECONOMETRICA 651, 681–82 (2011) (using a variation on the standard model to analyze taxpayers' dynamic response to audits over time; noting that "[b]ecause audits are rare events for a taxpayer, they are likely to provide new information and therefore lead to a change in the perceived detection probability"). *See also* Michael W. Spicer & Rodney E. Hero, *Tax Evasion and Heuristics: A Research Note*, 26 J. PUB. ECON. 263, 264 (1985); *infra* Part I.A.

7. *See infra* Part I.B–C.

audits.[8] Each enforcement action seeks to right the wrong of past tax evasion while also deterring evasion on a going-forward basis for that taxpayer. More specifically, the direct deterrence account implies that the experience of an audit may make the possibility of being audited again in the future more salient to the taxpayer. The taxpayer may rely on the availability heuristic and decide to increase her level of post-audit compliance.[9] An alternative pathway for direct deterrence suggests that there may be something about the experience of being audited and living through any consequences that changes a taxpayer's appetite for risk.[10] She may become more averse to the risk of an audit, and therefore may be expected to choose higher post-audit levels of compliance to maximize the expected value of her (new) utility function.

---

8. James Alm et al., *Getting the Word Out: Enforcement Information Dissemination and Compliance Behavior*, 93 J. PUB. ECON. 392 (2009) ("Taxpayer audits are a central feature of the voluntary compliance mechanism in the personal income tax system of many countries, largely because more frequent audits are thought to reduce tax evasion. Audits are thought to have a *direct* deterrent effect on those individuals actually audited; perhaps of more importance, audits are also thought to have an *indirect* deterrent effect on individuals not actually audited. However, the magnitude of these deterrent effects depends critically on the taxpayer being informed of enforcement efforts. As Plumley (1996) notes, '[i]t is generally believed . . . that many taxpayers would perceive increased auditing by the IRS as an increase in their chances of being audited, and that they would improve their voluntary compliance as a result.'") (footnote omitted).

9. Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 COGNITIVE PSYCHOL. 207, 208 (1973) ("A person is said to employ the availability heuristic whenever he estimates frequency or probability by the ease with which instances or associations could be brought to mind. To assess availability it is not necessary to perform the actual operations of retrieval or construction. It suffices to assess the ease with which these operations could be performed, much as the difficulty of a puzzle or mathematical problem can be assessed without considering specific solutions.").

10. One might expect this to be true also by association—when a close family member, colleague, or friend is audited, the effects of the experience likely will be felt beyond just that specific individual. This Article does not explore that proposition experimentally, but Osofsky has argued that particularly in sectors where compliance norms are weak, "concentrated enforcement" is an optimal strategy for maximizing revenue collection given scarce auditor resources. *See* Leigh Osofsky, *Concentrated Enforcement*, 16 FLA. TAX REV. 325, 327–28 (2014) ("The intuition behind concentrated enforcement is that, under certain circumstances, concentrated enforcement can increase compliance as a result of (1) increasing marginal returns to enforcement and (2) psychological factors that can support concentrated enforcement.").

But findings in the opposite direction—that audits result in decreased compliance—can be explained by an equally plausible set of accounts. First, the experience of being selected for random audit may be associated with a logical error on the part of taxpayers: they may misperceive their chances of experiencing a second random audit in the following period as being lower than in the prior period. As explained by Kastlunger et al., "[m]isperception of chance describes the mistaken estimation of the appearance of an event with a known probability distribution as more or less likely depending upon recent occurrences."[11] Second, a decrease in compliance after an audit may stem from the motivation to compensate for a loss of resources associated with fines, penalties, or back taxes owing after an audit.[12] This would only apply to taxpayers who were found to be noncompliant upon audit, and the effect would be likely to increase in relation to the magnitude of the amounts owing after an audit. Nonetheless, the "loss repair" mechanism could easily coexist with the misperception of chance mechanism.

A third account is more difficult to test but has been discussed in other theoretical literature: the experience of an audit may induce taxpayer "learning." A taxpayer may be able to draw valuable conclusions from her past audit experience concerning what kinds of underreporting are likely to be detected by the tax authority. She also may learn that an audit is not as terrible of an experience as she feared. In the former case, this learning may increase her confidence that she can tailor her future evasion to escape detection, leaving less compliance within less-easily-detectable areas.[13] In

11. Kastlunger et al., *supra* note 3, at 406–07 (citing Daniel Kahneman & Amos Tversky, *Judgement Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124 (1974)).

12. Boris Maciejovsky et al., *Misperception of Chance and Loss Repair: On the Dynamics of Tax Compliance*, 28 J. ECON. PSYCHOL. 678, 688 (2007) (explaining the loss-repair hypothesis as relating to "[the] negative consequences of detected tax evasion, and the resulting fines, [that] might induce taxpayers to engage in risky behavior, for instance, by exhibiting non-compliance in subsequent filings"). The authors tested their loss-repair hypothesis by comparing "the drop of compliance rates for honest taxpayers, i.e., those who fully declared their income, with dishonest ones, i.e., those who underreported." *Id.* They ultimately rejected it because they did not find "any systematic differences in the compliance rates of honest and dishonest participants after audits. Taken together, these findings suggest that the main mechanism, underlying the bomb crater effect, is misperception of chance. Loss repair was not observed." *Id.*

13. *See* James Andreoni et al., *Tax Compliance*, 36 J. ECON. LITERATURE 818, 842–43 (1998); *see also* R.C. Bayer, *A Contest with the Taxman—The Impact of Tax Rates on Tax Evasion and Wastefully Invested Resources*, 50 EUR. ECON. REV. 1071, 1099 (2006).

the latter case, audit learning may reduce her compliance by dampening her aversion towards future audits. Fourth, and finally, among audited taxpayers found to be compliant, audits may breed resentment and undermine confidence in the tax enforcement system, causing a post-audit response of lower compliance. Enforcement might play the perverse role of "crowding out" a taxpayer's intrinsic motivation to voluntarily comply.[14]

To better understand the factors that affect post-audit changes in compliance, this Article uses a novel experimental setting to test the following hypothesis: the use of random audits invites taxpayers to misperceive the chances of repeat audits, so a post-audit drop in compliance levels can be expected in response to random—but not nonrandom—audits. To the extent that a post-audit compliance "bomb crater" effect is observed in response to random audit, such an effect can be ameliorated by replacing random audits with an endogenous audit selection rule. Put differently, the hypothesis implies that decreased post-audit compliance levels will occur in the presence of random audits but not audits that are conditioned on prior audits in which cheating has been detected. Where audits are random, taxpayers are vulnerable to the impulse to "play the odds" after an audit, even though nothing has changed that should affect their rational calculus of the risk of audit. Where audits are "targeted" towards a particular group of taxpayers by being conditional on detected cheating (e.g., endogenized), misperception of chance is unlikely to occur: taxpayers are on notice that they are more likely to be selected for future audit. Their compliance levels should rise, consistent with a deterrence effect.

Why is this hypothesis worth testing? First, it offers a new way to probe the validity of the "misperception of chance" account for post-audit drops in compliance. It also has the attractive potential to address at least some of the inconsistencies in the dynamic audit literature and provide potentially useful insights for tax administrators tasked with setting enforcement policy. If random audits alone are plagued by post-audit increases in cheating while endogenous audits as associated with the desired deterrent effect, policymakers can better weigh the costs and benefits of using each type.

To test the above hypothesis, I used a stylized experimental setting that allowed me to carefully compare the effect on tax compliance of two audit rules.[15] The first rule functions as the "control" experimental condition

14. For a detailed summary of the research on voluntary compliance crowd out, see generally Leandra Lederman, To What Extent Does Enforcement Crowd Out Voluntary Tax Compliance? (Jan. 13, 2017) (unpublished manuscript) (on file with *Florida Tax Review*).

15. Both audit rules were chosen to reflect the reality of a resource-

by using random audits: it assigns audits based on independent draws from a uniform probability distribution of ten percent. This implies that the period in which a participant experiences her first audit is determined at random and does not occur during a specific period. The second audit rule, the nonrandom "endogenous audit rule" (EAR) functions as the "treatment" experimental condition. It begins exactly like the control experimental condition, in that each participant is subjected to audits drawn from a uniform ten percent probability distribution. Once a participant is randomly selected for audit,[16] her next-period experience in the experiment will depend on the audit's outcome. If she reported all of her income, she is "returned" to the random ten percent audit pool (e.g., she continues to be subject to audits based on independent draws from a uniform probability distribution of ten percent). If she failed to report all her income, she is re-audited in the immediately succeeding period. If the second audit reveals that the participant again has not reported all her income, she is audited in the following period as well. This continues until one of two events occurs: the participant reports all of her income, or she is audited a total of six times. In the event of the former, the participant is returned to the random ten percent audit pool. In the event of the latter, there are no further audits.[17]

To perform the experiment, I exploited a flexible platform that is increasingly common in social science research but has not been used in the tax compliance context. I collected survey responses from a sample of 199 participant workers on Amazon's Mechanical Turk (MTurk) task-completion platform. To probe the robustness of prior laboratory studies conducted in

---

constrained taxing authority: the average audit rate and average number of audits is held constant across the two audit rules.

16. Here, I assume the participant is randomly selected for audit at some point over the 60 periods such that the endogenous audit treatment condition "kicks in" as described. It is of course possible that a given participant will—by pure chance—never be selected randomly for audit in any of the periods, in which case there would be no possibility for endogenous audits. To be absolutely clear about the structure of the EAR: there is no specific period in which a participant in the EAR experimental condition experiences an audit. Like the control condition, each participant in the EAR begins by being randomly selected for audit (at a uniform probably of ten percent). Only once a given participant has been randomly selected for audit does the EAR take effect.

17. This six-audits-per-participant cap is *not* explicitly disclosed to participants who are randomly assigned to the treatment condition. *See* Part II.C.4.b for further discussion. The six-audit cap is designed to mechanically ensure an average audit rate of ten percent that is equal to the audit rate in the control experimental condition. See *infra* note 95 and accompanying text for details about this "probability-forcing" feature of the design.

Europe, I framed the experiment as taking place in the U.S. taxpaying context, and I employed several methods to ensure that only U.S. resident MTurk workers were included in my participant sample.[18]

My inquiry yields four main results. First, I confirm the existence of a bomb crater effect in response to random audits. U.S. resident MTurk participants subjected to random audits exhibit behavior consistent with the bomb crater effect found in three European bricks-and-mortar laboratory studies. Following a random audit, participants' compliance rate, on average, drops eight percentage points from the non-post-audit average.

Second, I show that a second consecutive random audit can overcome the initial bomb crater effect by increasing the "post-two-audit" average compliance rate in excess of the non-post-audit average.

Third, I confirm the hypothesis that the bomb crater effect does not plague an audit rule in which past noncompliance "flags" taxpayers for future audits. In particular, my results stand for the importance of audit presentation: the way in which a taxing agency's audit rule is presented (in the sense of how it is described) to participants as being random versus endogenous had an outsized impact on respondents' average post-audit compliance levels. Simply announcing to participants—via one line of text embedded in a series of instructions—that evasion will trigger one or more future audits had a large and significant positive effect on average post-audit compliance. The deterrent effect of the EAR announcement was measured at more than four percentage points.

Fourth, the performance of the particular EAR tested here is inconclusive: an investment by a tax administrator in costly repeat audits in the context of a preannounced EAR may not be justified. I found that average compliance did not increase with each additional consecutive audit, although it also did not fall significantly. Further research—both experimental and using controlled field trials—to understand the dynamic effects of various EARs is warranted.

Research along these lines has clear relevance for tax enforcement policy debates. Despite the exceedingly high rates at which income subject to withholding and substantial information reporting is declared voluntarily (99%),[19] there are persistent pockets of low voluntary compliance among

18. For further discussion of the use of MTurk as a new experimental setting and MTurkers as a new subject pool, see Appendix A: Discussion of MTurk Sample.

19. INTERNAL REV. SERV., TAX GAP FOR TAX YEAR 2006: OVERVIEW, 1–3 (2012) ("For example, the net misreporting percentage, or NMP, (defined as the net misreported amount as a ratio of the true amount) for amounts subject to substantial information reporting and withholding is 1%; for amounts subject to substantial

hard-to-tax groups of individuals.[20] Less than half of income not subject to information reporting is voluntarily declared.[21] Moreover, governments have had little success in using nontraditional means, such as cohort-specific shaming techniques or public disclosure of non-filers, to influence taxpayers' underlying preferences, motivations, and attitudes towards taxation.[22] Such

---

information reporting but no withholding, it is 8%; and for amounts subject to little or no information reporting, such as business income, it is 56%."). However, scholars have argued that once the near certainty of detection due to information reporting is taken into account, a significant part of the "puzzle" of why people report their income at such high rates given the low chance of enforcement can be solved. *See* Leandra Lederman, *Tax Compliance and the Reformed IRS*, 51 U. KAN. L. REV. 971 (2003).

20. INTERNAL REV. SERV., *supra* note 19; *see also* Susan Cleary Morse et al., *Cash Businesses and Tax Evasion*, 20 STAN. L. & POL'Y REV. 37, 67 (2009) ("Cash business owners rely on parallel cash economies to underreport receipts and thereby evade income, employment and sales taxes . . . . Evasion seems best explained by opportunity, including the low-perceived likelihood of detection and penalty, and by peer norms. The perceived equity of the tax system has less importance, and the complexity of the tax law does not appear to play a significant role.").

21. INTERNAL REV. SERV., *supra* note 19, at 3 (only 44% for income subject to no information reporting is declared voluntarily by taxpayers). *See* James Alm et al., *Why Do People Pay Taxes?* 48 J. PUB. ECON. 21 (1992); Leandra Lederman, *Statutory Speed Bumps: The Roles Third Parties Play in Tax Compliance*, 60 STAN. L. REV. 695, 697 (2007); Leandra Lederman, *The Interplay Between Norms and Enforcement in Tax Compliance*, 64 OHIO ST. L. J. 1453, 1460 (2003); Joel Slemrod, *Cheating Ourselves: The Economics of Tax Evasion*, 21 J. ECON. PERSP. 25 (2007); Lederman, *supra* note 14, at 9.

22. *See* Michael Hallsworth, *The Use of Field Experiments to Increase Tax Compliance*, 30 OXFORD REV. ECON. POL'Y 658 (2014); Günter Schmölders, *Fiscal Psychology: A New Branch of Public Finance*, 12 NAT'L TAX J. 340 (1960) (presaged subsequent work published in German using survey research by suggesting that at the societal and national levels, the aggregate of individuals' motivational postures yields tax morale); Erzo F.P. Luttmer & Monica Singhal, *Tax Morale* (Nat'l Bureau of Econ. Research, Working Paper No. 20458, 2014). Building indirectly on Schmölders, a recent paper by Luttmer and Singhal, *supra*, at 1, define tax morale as the "nonpecuniary motivations for tax compliance as well as factors that fall outside the classic expected utility framework, including departures from standard rationality assumptions." Luttmer & Singhal, *supra*, at 2–3, identify from the empirical literature five potential mechanisms for the formation of individual-level tax morale: intrinsic motivation (individuals intrinsically want to comply, or "feel guilt or shame for failure to comply"), reciprocity ("the willingness to pay taxes in exchange for benefits that the state provides to them or to others even though their pecuniary payoff would be higher if they didn't pay taxes"), peer effects

failures have highlighted the urgency of gaining a better understanding of the forces that determine compliance behavior. Interestingly, they have also spurred a back-to-basics focus on tax administration, including the thoughtful and detail-oriented design of audit policies.[23] A better understanding of the factors that can shape taxpayers' motivations and attitudes towards tax compliance would allow governments to proactively influence overall tax morale through better-tailored audit policies or other pro-compliance strategies.[24]

("[i]ndividuals may be influenced by peer behavior and the possibility of social recognition or sanctions"), cultural or social norms (which "can affect the strength of these intrinsic motivations, reciprocal motivations, or the sensitivity to peers") and, finally, "deviations from standard expected utility maximization, such as loss aversion." Such departures from rationality include loss aversion, systematic overweighting of small-probability events such as audit or detection, and other cognitive errors or distortions. In a different strand of the tax morale literature, individual-level tax morale is linked to and incorporates an individual's "motivational posture," which in turn includes subjective constructs about tax compliance as well as the individual's internalization of societally shared beliefs. *See* Valerie Braithwaite, *Dancing with Tax Authorities: Motivational Postures and Non-compliant Action*, *in* TAXING DEMOCRACY: UNDERSTANDING TAX AVOIDANCE AND EVASION 18 (Valerie Braithwaite ed., 2003) (describing a continuum of five postures, two positive and three negative towards voluntary compliance: commitment and capitulation (positive); resistance, disengagement, and game playing (negative)).

23. A recent literature review of field (e.g., natural) experiments regarding tax compliance behavior observed that seemingly incidental factors like the presentation of audits and timing of interactions with the taxing authority appear to exert an outsized impact on taxpayers' compliance behavior in a variety of settings. *See* Hallsworth, *supra* note 22, at 673 ("This use of behavioural science suggests a greater focus on *how* tax authorities intervene and deal with taxpayers more generally. As Hallsworth et al. (2014) show, significantly different results can be obtained by the way in which a policy (e.g. sending reminder letters) is implemented, even though the application of the policy itself is not randomized (everyone receives a letter). Compliance may therefore be affected by a range of apparently incidental factors such as timing, framing, complexity, tone, visual presentation, and so on.") (footnotes omitted).

24. *See* Braithwaite, *supra* note 22, at 282 ("Many of the papers in this volume are supportive of the proposition that outcomes of this kind are achieved through direct intervention in the human system, that is, going further than changing law and administrative procedures to improving the quality of the relationship between the tax office and the taxpayer. Actions might include the tax authority being reasonable and clear in its day-to-day communication with taxpayers, listening to taxpayers, treating them with respect, responding to concerns, and following through purposefully to elicit compliance. Those who engage with the human

One note is in order about generalizing findings from experimental tax compliance research to real-world taxpayer behavior. "External validity" is a measure of the transferability of experimental results obtained in a simulation setting such as a university laboratory or online survey platform to the real world.[25] There is a longstanding debate among prominent tax

---

dimension of compliance in this relational way are likely to reap a double reward. As tax officers listen to taxpayers to better understand the reasons that underlie resistance to compliance, it is unlikely that the feedback they receive is solely related to their localized compliance problem. The functional lines of a tax office are meaningful within the organisation to those familiar with its operations, but are relatively meaningless to most outsiders who are likely to see the tax office and the tax system as one entity. Thus, while listening to the taxpaying community, taxpayers are faced with a reality that is not bounded by a localized compliance problem, and in the process find a bridge to engage with the broader issue of institutional integrity. They learn, through the eyes of those outside, how a localized compliance problem sits alongside compliance problems elsewhere in the organisation to create a picture of institutional integrity overall.").

An important strand of this literature on the dynamics of taxpayers' subjective experiences with the tax system (albeit one that is somewhat tangential to my focus here on sequential audits) explores the relationship between taxpayers' subjective feelings about the tax system over time and various outcome variables. Taxpayers across a variety of jurisdictions (the United States, Australia, Europe) report feelings of incompetence and anxiety with regard to taxes, with such attitudes most pronounced among the young. *See, e.g.*, Adrian Furnham, *Understanding the Meaning of Tax: Young People's Knowledge of the Principles of Taxation*, 34 J. SOCIO-ECON. 703, 709–10 (2005) ("The results of the study showed first, as may be expected, knowledge of tax grows with age. However, it is equally interesting to note that for each question even the majority of the 15-year olds did not have a full understanding . . . . It is not interesting to demonstrate cognitive development but it is important to notice when full understanding occurs. This begs two further central questions: by what age are children/adolescents able to fully grasp the principle of tax; and second what experiential factors (i.e. schooling, shopping) [are] likely to facilitate that understanding."); Margaret McKerchar, *Understanding Small Business Taxpayers: Their Sources of Information and Level of Knowledge of Taxation*, 12 AUSTL. TAX F. 25, 40 (1995) (concluding that "the small business taxpayers included in the study did not have a satisfactory level of knowledge of taxation as it affected their businesses and may be unintentionally non-compliant as a result. The analysis of the sample population as a whole demonstrates that variations in levels of knowledge existed between turnover ranges, the length of time in business, and across the range of taxes . . ."). Complexity is cited as the central culprit for tax-related anxiety in surveys, which in turn has been linked to political disengagement on issues of fiscal policy. *See* McKerchar, *supra,* at 27.

25. WEBLEY ET AL., TAX EVASION: AN EXPERIMENTAL APPROACH 44 (1991) ("The artificial nature of experiments is probably the commonest criticism in

compliance researchers about the external validity of tax compliance research performed under laboratory conditions. This Article does not intervene in this larger debate. Instead, I address the "experimental robustness" dimension of external validity with reference to the bomb crater effect and its possible cures.[26] There is broad consensus among scholars that variation in the context of experiments designed to test a particular hypothesis unambiguously increases the experimental robustness of results found to be consistent across contexts. It is true that even robust experimental findings may fail to accurately describe taxpayers' behavior when observed in real taxpaying situations. Nonetheless, an important phenomenon's subsequent replication or falsification in diverse experimental settings, through the use of new subject samples, or in the presence of variation in the details of the experiment, can provide a valuable contribution.[27] In the case of a replication, it can substantially increase

---

both psychology and economics and this usually leads to comments about their lack of external validity. Brookshire, Coursey, and Shulze (1987) reckon that when experimental economists present their work at conferences the typical response is to question whether the results apply to 'real-world' settings."); Francesco Guala & Luigi Mittone, *Experiments in Economics: External Validity and the Robustness of Phenomena*, 12 J. ECON. METHODOLOGY 495, 495–96 (2005) ("External validity, in its original formulation, is a 'question of *generalizability*: to what populations, settings, treatment variables, and measurement variables can [an experimental effect] be generalized?'") (citation omitted).

26. Guala & Mittone, *supra* note 25, at 506–07 ("What does this tell us about the generality of the bomb crater effect? The most plausible answer is that we are dealing with a robust bias, which tends to arise whenever subjects have to do with probabilistic reasoning of this kind. To establish robustness is to establish a sort of generality, to a set of situations that are somehow similar to the ones in which the phenomenon has been observed. Robustness invites 'generic' confidence, in the sense that it is no evidence that the phenomenon will occur in all circumstances, and provides no precise indication of the situations in which it will occur and those in which it will not. Stylized facts from the real world invite caution: there are reasons to believe, for instance, that erratic behaviour such as the one observed in the tax experiments (a variance further exacerbated by the crater effect) may not arise in real-world circumstances. Some governments take erratic tax payments as indicators of possible evasion, and therefore check erratic taxpayers more often than the others. This strategy (if known to taxpayers, which is an empirical hypothesis of course) may be enough to attenuate or even eliminate the bomb crater effect.") (footnote omitted).

27. Experiments also allow researchers to finely tune a set of experimental conditions that may be hard to control in the real world. Researchers can test specific hypotheses that may not be amenable to testing using naturally arising data. *See* WEBLEY ET AL., *supra* note 25, at 44 ("[Using the experimental approach] we can

researchers' confidence in the plausibility of the phenomenon and can strengthen the policy case for conducting controlled field experiments in collaboration with tax agency officials.

The Article proceeds as follows. Part I reviews the existing empirical research on audit dynamics and taxpayer compliance patterns. Part II outlines the MTurk experiment. Part III presents the results of the experiment. Part IV concludes with a discussion of next steps for researching audit dynamics as well as some of the potential policy implications of the study.

I. LITERATURE ON THE DYNAMIC COMPLIANCE EFFECTS OF AUDITS

A. The Standard Model of Tax Compliance

The standard "rational expected utility" model of tax compliance decision-making was developed over forty years ago and presents the choice to self-report one's income as an optimal portfolio allocation problem.[28] Individuals are posited to have a choice between a risk-free (reported, after-tax income) asset and a risky (unreported, nontaxed income) asset. A taxpayer's utility, and her resulting compliance rate, is modeled as a function of a number of parameters: her income, the rate of tax on that income, the probability of audit, the penalty applicable in the case of audit (usually as a percentage or multiple of back taxes owed), and the taxpayer's degree of risk

---

isolate the variables we are interested in and study the effects of our manipulations knowing that these are not the result of extraneous factors. . . . [A]rtificial environments can be more easily manipulated by the experimenter than real-life situations; complex tasks can be condensed into manageable abstractions; people can be put into situations where ordinarily they cannot be observed . . . .").

28. Allingham & Sandmo, *supra* note 4, at 324 ("The tax declaration decision is a decision under uncertainty. The reason for this is that failure to report one's full income to the tax authorities does not automatically provoke a reaction in the form of a penalty. The taxpayer has the choice between two main strategies: (1) He may declare his actual income. (2) He may declare less than his actual income. If he chooses the latter strategy his payoff will depend on whether or not he is investigated by the tax authorities. If he is not, he is clearly better off than under strategy (1). If he is, he is worse off. The choice of a strategy is therefore a non-trivial one."); Srinivasan, *supra* note 4, at 339 ("[T]he optimum . . . proportion of income to be understated will be derived as a function of true income, probability of detection of understatement and the properties of the tax function."); Shlomo Yitzhaki, *A Note on Income Tax Evasion: A Theoretical Analysis*, 3 J. PUB. ECON. 201, 201 (1974) ("In the Allingham- Sandmo model the taxpayer should pay a penalty rate π on the undeclared income. This assumption leads to the conclusion that when the tax rate increases there will be two opposing effects, an income and a substitution effect.").

aversion.[29] Extensions of the standard framework have incorporated additional factors relating to the decision to report income, including psychological dimensions such as guilt and shame,[30] as well as the use of nonrandom EARs rather than random audits.[31] Expected utility models

29. The comparative statics of voluntary compliance in the Allingham-Sandmo model are unambiguously positive for the following parameters: audit rates, detection rates upon audit, and fines upon detection. An increase in any of these parameters, holding all other parameter values in the model constant, implies higher rates of voluntary compliance by the representative taxpayer. Practically speaking, things are more ambiguous. The predicted results have been observed in experimental settings, but the effects are often small and not infrequently reverse direction. For a selection of studies examining the consequences of increasing or decreasing fines for evasion, see WEBLEY ET AL., *supra* note 25; C. Blackwell, *A Meta-Analysis of Incentive Effects in Tax Compliance*, *in* DEVELOPING ALTERNATIVE FRAMEWORKS EXPLAINING TAX COMPLIANCE 97–112 (James Alm et al. eds., 2010); Odd-Helge Fjeldstad & Joseph Semboja, W*hy People Pay Taxes: The Case of the Development Levy in Tanzania*, 29 WORLD DEV. 2059 (2001). A prominent study of the effects of changing the audit rate and the detection rate is Carol M. Fischer et al., *Detection Probability and Taxpayer Compliance: A Review of the Literature*, 11 J. ACCT. LITERATURE 1 (1992). Relatedly, research on how income levels affect tax compliance has yielded inconsistent results across countries and experimental designs. *See, e.g.*, Vital Anderhub et al., *On the Interaction of Risk and Time Preferences: An Experimental Study*, 2 GERMAN ECON. REV. 239 (2001); Chang-Gyun Park & Jin Kwon Hyun, *Examining the Determinants of Tax Compliance by Experimental Data: A Case of Korea*, 25 J. POL'Y MODELING 673 (2003).

30. Brian Erard & Jonathan Feinstein, *The Role of Moral Sentiments and Audit Perceptions in Tax Compliance*, 49 PUB. FIN. 70 (Supp. 1994).

31. Michael Graetz et al., *The Tax Compliance Game: Toward an Interactive Theory of Law Enforcement*, 2 J. L., ECON., & ORG. 1, 4–5 (1986) (modeling the use of EARs that "introduces the IRS as a strategic actor in a game-theoretic approach that allows the IRS to condition its audit rules on the reports it receives from taxpayers. . . . [T]he behavioral framework of tax non-compliance makes it an especially appropriate case for a theoretical construct that explicitly takes into account the interactions and responses of the cognizant law enforcement agency.") (footnote omitted); Mark D. Phillips, *Deterrence vs. Gamesmanship: Taxpayer Response to Targeted Audits and Endogenous Penalties*, 100 J. ECON. BEHAV. & ORG. 81 (2014); Jennifer Reinganum & Louis L. Wilde, *Equilibrium Verification and Reporting Policies in a Model of Tax Compliance*, 27 INT'L ECON. REV. 739 (1986) (exploring a different model in which "the taxpayer observes his true income . . . . Based upon true income, the taxpayer conveys a statement of reported income to the IRS. Since the IRS does not observe the taxpayer's true income, it must make some conjecture about the type of taxpayer who would report a given level of income. Based on the level of income reported and these conjectures,

represent only one strand of theoretical work applicable to this topic: others include the role of peer pressure, civic responsibility, citizen's beliefs and attitudes, and a variety of other factors.[32] Others have sought to leverage and test these theories using agent-based computational models.[33]

---

the IRS chooses a level of effort to be devoted to investigating the taxpayer. This effort will be assumed to generate a particular probability that the taxpayer's true income will be verified, with the property that greater effort leads to a greater probability of verification; investigation does not imply certain apprehension."); Jennifer F. Reinganum & Louis L. Wilde, *Income Tax Compliance in a Principal-Agent Framework*, 26 J. PUB. ECON. 1, 2–3, 9 (1985) (using a model that "incorporate[s] the information content of a taxpayer's report into the IRS' choice of an audit policy" and comparing "an *alternative* audit policy which uses the information conveyed by the taxpayer's report (albeit in a rather crude manner) to the standard random audit policy").

32. *See* Joel Slemrod, *Introduction* to WHY PEOPLE PAY TAXES: TAX COMPLIANCE AND ENFORCEMENT 2 (Joel Slemrod ed., Univ. of Mich. Press 1992) ("Economists have traditionally viewed the problem as one of rational decision making under uncertainty—cheating on your taxes is a gamble that either pays off in lower taxes or, with some probability, subjects you to sanctions. From this perspective, one would expect that noncompliance would respond to changes in both the likelihood that an act of evasion will be detected and punished, as well as the severity of the punishment. Social scientists from other disciplines have stressed the narrowness of this view and focused on the importance of peer pressure, civic responsibility, and other factors."); Marco R. Steenbergen et al., *Taxpayer Adaptation to the 1986 Tax Reform Act: Do New Tax Laws Affect the Way Taxpayers Think About Taxes?*, *in* WHY PEOPLE PAY TAXES, *supra*, at 32 ("Taxpayers systematically changed their attitudes and beliefs during the tax season, but the biggest impact in our study came from the social influence of discussions about taxes during the season, not from evaluations related to the TRA. The perceived attitudes of individuals with whom our respondents discussed taxes had consistently strong impacts on inhibitors and commitment related compliance. Furthermore, discussants' attitudes had as much influence on postseason beliefs as taxpayers' preseason beliefs!"); *see also* BRUNO S. FREY, NOT JUST FOR THE MONEY: AN ECONOMIC THEORY OF PERSONAL MOTIVATION 45 (1997) ("An important case is provided by the fundamental attitude enshrined in tax laws. In some countries it is assumed that all citizens want to cheat on taxes, and they therefore have to prove that they have paid all what they legally owe. In other countries it is assumed that the citizens are prepared to pay their "fair share', and the tax administration has to prove that this is not the case. Even if it suspects that not all taxable income has been declared, it tends to attribute it to an error on the taxpayer's side rather than an attempt purposely to cheat on taxes.").

33. J. T. Manhire, *There Is No Spoon: Reconsidering the Tax Puzzle,* 17 FLA. TAX REV. 623, 633 (2015) ("This Article attempts insights into the effect of audits on compliance, both locally and globally, by way of an agent-based model.

The standard expected utility model generally does not incorporate the repeat-play aspect of tax compliance. Unless there is uncertainty about the audit strategy used by the tax administrator or an exogenous change to the parameters included in the utility function, the taxpayer's optimal choice of compliance will be static. As noted above, the experience of an audit is merely the concrete occurrence of a probabilistic event. In the no-frills standard model, the audit itself should not matter.

Some theoretical research relaxes the no-uncertainty assumption. In the context of agents' contributions to a public good, Miller and Andreoni employ an adaptive learning algorithm (roughly, a replicator strategy that adjusts each agent's contribution to a public good over time relative to the group mean) to reconcile anomalous laboratory findings with existing theories of public good contribution.[34] Free riding on public goods can be seen as an analogy to the case of tax evasion whereby taxpayers experience direct benefits from the goods financed by taxation. Relatedly, Bloomquist models the tax compliance of small business owners using an evolutionary dynamics approach calibrated with results from laboratory experiments and real-world random audits to find that full-compliance-oriented taxpayers (categorized as "honest" taxpayers) decline over time.[35] However, adaptive

---

Like all models, the one developed here is a generalization. It does not seek to fully replicate the U.S. tax administration system, but instead to generate certain characteristic phenomena and core dynamics to help understand the effects of audits on both individual compliance decisions and system-level compliance patterns."); Kim M. Bloomquist, Agent-Based Simulation of Tax Reporting Compliance 4 (July 20, 2012) (unpublished Ph.D. dissertation, George Mason University), http://ebot.gmu.edu/handle/1920/7927.pdf?sequence=1&isAllowed=y ("It is the goal of this study to develop an ABM for use in assessing alternative strategies to improve tax reporting compliance. The model incorporates detailed tax return information and empirically based taxpayer behaviors. Key institutional features of the tax system (e.g., paid preparers, employers, and third-party information reporting) are included in the model as well.").

34. John H. Miller & James Andreoni, *Can Evolutionary Dynamics Explain Free Riding in Experiments?*, 36 ECON. LETTERS 9, 10–11 (1991) ("The model used here is based on *replicator dynamics*. . . . [R]eplicator dynamics are based on the notion that strategies that perform relatively well will replace those that perform relatively poorly. . . . [T]he model mirrors prominent psychological theories of decision making based on 'reference points' . . . . [A]ccording to these theories, people adjust their decisions more rapidly the more the outcome falls short of some reference point, such as the average.") (citations omitted).

35. Kim Bloomquist, *Tax Compliance as an Evolutionary Game: An Agent Based Approach*, 39 PUB. FIN. REV. 25, 27–28 (2011) ("An advantage of agent-based models over laboratory experiments is the ability to collect data from large populations of heterogeneous agents.").

learning frameworks are relevant to cases only where the taxpayer does not know the true rate at which random audits are performed. In most laboratory settings, this information is explicitly given to participants to constrain participants' possibly wildly-divergent assumptions about the audit environment that might confound the researcher's results.

B. The Ambiguous Empirical Evidence on Post-Audit Compliance Patterns

Almost twenty years ago, the lack of consensus on models for understanding tax evasion, or for explaining the significance of audit experience on subsequent tax compliance, fueled calls for research on the dynamics of compliance behavior.[36] At that time, only a few studies were available, the most relevant being Alm et al.'s comparison of the performance of random and EARs in the context of a 20-period lab experiment.[37]

Researchers have responded to this call. In the intervening years, scholars have begun to investigate the dynamic effects of different patterns of audits in a taxpayer's lifecycle. These studies have yielded mounting evidence gathered in a variety of settings that the experience of an audit can affect a taxpayer's post-audit compliance posture. But in which direction? Does an audit increase or decrease post-audit compliance? In some settings for some populations, audits appear to increase post-audit compliance, and in others, they appear to decrease it. To highlight this lack of consensus, I briefly survey the studies that have contributed to our knowledge in this area.

1. Studies Showing That Audit Experience Is Associated with Higher Post-Audit Compliance

In support of the proposition that audits can work as successful

36. In a 1998 review of the economics of tax compliance, Andreoni et al. stated that "[c]learly, more research is needed both to confirm whether there is any specific deterrent effect of an audit and to uncover the reasons for the presence or absence of such an effect. This is an important area, because the econometric results [of the two studies] to date suggest that the use of the 'stick' to enforce compliance with tax laws may not have any long-run impact." *See* Andreoni et al., *supra* note 13, at 844.

37. *See* Alm et al., KYKLOS, *supra* note 3 (introducing an experimental condition featuring a conditional future audit rule pursuant to which a taxpayer is audited again in the event that underreporting was found on audit; this study's EAR is a variant of this rule). This paper builds on Alm et al., KYKLOS by zeroing in on one particular dimension: post-audit compliance behavior.

deterrents, two recent studies using U.S. data have found a positive association between an audit and an increase in the audited taxpayer's self-reported income after the audit. First, Beer et al. use Internal Revenue Service (Service) income tax return microdata to study the effect of being selected for audit.[38] Overall, average reported income (from which voluntary compliance is estimated) increased following the audit. However, they present evidence that audits seemed to affect post-audit reported income of different taxpayers differently. Audited taxpayers who were found to have a deficiency (e.g., whose evasion was detected) increased their reported taxable income in post-audit years, consistent with a direct deterrence response. However, they observed an opposite reaction from those who had reported truthfully in the period of the audit: they responded by reporting less income in the period after the audit.[39]

Second, DeBacker et al. exploit the Service's National Research Program of randomized audits to follow the trajectory of the reported income of taxpayers who experienced an audit for up to six years after the audit.[40]

---

38. SEBASTIAN BEER ET AL., U.S. IRS TAXPAYER ADVOCATE SERV., AUDIT IMPACT STUDY 68, 70 (2016), http://www.taxpayeradvocate.irs.gov/Media/Default/Documents/2015ARC/ARC15_Volume2_3-AuditImpact.pdf (finding results consistent with long-term revenue implications of audits; specifically, "[t]hree years after an audit, the average small business taxpayer reports around 20 percent more income. The indirect long-term effect thus clearly adds to the static gain of additional tax assessments. However, by differentiating the response of compliant and noncompliant taxpayers, we find that there is scope for improving the revenue efficiency of audits. Our more nuanced analysis of the behavioral response to an audit shows that taxpayers who receive a positive additional recommended tax assessment increase their subsequent reporting of taxable income dramatically (+120 percent), while those who receive no additional tax assessment actually report less (-35 percent). There are several plausible explanations for this finding. The positive impact of audits on the former . . . group is likely due to some kind of specific deterrent effect.") (footnote omitted).

39. *Id.* at 91 (discussing differential effect). This finding suggests that, depending on taxpayers' pre-audit compliance postures, there may be a crowd-out effect. *See generally* Lederman, *supra* note 14. In this Article, when using the term "voluntary compliance," I will follow convention in using it to describe taxpayers' pre-audit reporting behavior. *See* J.T. Manhire, *What Does Voluntary Compliance Mean?: A Government Perspective*, 164 U. PA. L. REV. ONLINE (2015).

40. Jason DeBacker et al., Once Bitten, Twice Shy? The Lasting Impact of IRS Audits on Individual Tax Reporting 4 (Oct. 28, 2015) (unpublished manuscript), https://www.aeaweb.org/conference/2016/retrieve.php?pdfid=28 (the authors describe their use of an enormous amount of data—all tax filers from 2000 to 2012, plus those audited through the Service's National Research Program (NRP)—as follows: "[t]o test for the long-term effects of audits on tax reporting, we

For taxpayers with income that is generally not third-party reported, the authors find an increase of 14% in voluntarily reported income after a random audit as compared to pre-audit voluntarily reported income. However, this large near-term positive effect on reported income (actual "compliance" is not observed, because audits are not repeated) is not sustained over time: the effect becomes negative by the fifth and sixth post-audit years.

In addition to these U.S. studies, there has been important research on dynamic compliance in other national contexts. In a multi-stage randomized field experiment involving a large sample of both self-employed and third-party employed Danish taxpayers, Kleven et al. found that being thoroughly audited had a strong effect on declarations of non-third-party-reported income in the tax year following the audit, but, as predicted by the authors' model, almost no effect on third-party reported income.[41]

Another recent set of studies use U.K. individual tax data that includes the results of taxpayers selected for audit. Advani et al. examine the effects of a randomized audit program applicable to individuals who are subject to self-assessment (e.g., individuals for whom withholding may not correctly determine tax, such as pensioners, the self-employed, partners in a business partnership, high earners, company directors, or individuals with capital gains or income from U.K. land or property).[42] Comparing audited taxpayers

---

use data from the Service's National Research Program (NRP). The NRP began conducting random audits of individual tax filers starting in tax year 2001, and began conducting annual random audits starting in 2006. To these data, we merge returns from the universe of filers from 2000 to 2012, allowing us to examine the impact of audits on individual taxpaying behavior for a period of up to six years after an audit. These data are instrumental in addressing the impact of legal enforcement on subsequent behavior for several reasons. First, the Service conducts intermittent audits and keeps systematic records of them. Second, the Service provides accurate data on subsequent tax payments in each year following the audit, even when there is no audit. Third, these data comprise a panel of the entire population of individual taxpayers over time, allowing for rigorous empirical analysis.") (footnote omitted).

41. *See* Kleven et al., *supra* note 6, at 689, 691 (finding also that threat-of-audit letters had a substantial increase on voluntarily reported income that was not subject to third-party reporting, and noting that "[f]or self-reported income, our empirical results fit remarkably well with the basic AS model: tax evasion is substantial and responds negatively to an increase in the perceived probability of detection coming from either a prior audit or a threat-of-audit letter," but concluding on the basis of extremely high declarations of third-party reported income, "[g]iven that audits are very costly and eliminate only a part of tax evasion, enforcement resources may be better spent on expanding third-party reporting than on audits of self-reported income.") (footnote omitted).

42. Arun Advani et al., *How Long-Lasting Are the Effects of Audits?* (Tax

to a set of three control groups, they find that the effect of an audit on reported tax liability becomes significant in the first year after the audit and increases in size up to year four.[43] Gemmell and Ratto use an earlier version of the U.K. dataset relied upon by Advani et al. to investigate whether taxpayers randomly selected for audit and found to be compliant displayed systematically different post-audit reporting behavior than randomly audited taxpayers found to be noncompliant. To make this determination, they observed reported income for the audited (treatment) group and compared it to reported income of a non-audited control group that was matched across relevant selection criteria to the treatment group. They found that taxpayers increased their reported income after an audit that detected evasion, in line with other results. But (as noted in the next subpart), where an audit did not detect evasion, the authors' measure of compliance (reported income post-audit as compared to control group) fell.[44] The authors hypothesize that "the predicted responses by taxpayers faced with random auditing are likely to depend both on their expected, endogenous probability of audit, and on the expected or perceived 'quality' of those audits."[45] Advani et al. point out that Gemmell and Ratto's results cannot be viewed as causal (i.e., an audit verdict of compliance is not necessarily causally related to increased compliance),[46] but the finding that "successful" audits (i.e., those that find unreported income) are associated with increased compliance is consistent with their hypothesis that higher-quality audits will have a "preventative" effect on

---

Admin. Research Ctr., Discussion Paper 011-15, 2015), at 2, 9, 12 (this group of self-assessment individual taxpayers were selected for audit randomly with a probability of 0.03%; the authors provide the following explanation as to why they focused on random audits rather than HM Revenue & Custom's targeted audit programs: "[w]e focus on random rather than targeted audits because it enables us to say something about the population of self-assessment taxpayers rather than just the selected group that faces a targeted audit.").

43. *Id.* at 17, 19 (finding results that "suggest that there is a large and persistent impact of audits on reported tax liability that reaches around 26 per cent by the fourth year following the tax year to which the audit relates.").

44. *See* Norman Gemmell & Marisa Ratto, *Behavioral Responses to Taxpayer Audits: Evidence from Random Taxpayer Inquiries*, 65 NAT'L Tax J. 33 (2012).

45. *Id.* at 34.

46. *See* Advani et al., *supra* note 42, at 4 (noting that Gemmel and Ratto "distinguish between taxpayers found to be non-compliant and those found to be compliant, arguing that the former are likely to increase their subsequent compliance while the latter could reduce their compliance" but critiquing "this distinction between compliant and non-compliant taxpayers [as] endogenous, making it hard to interpret the comparison with an unconditionally randomly selected control group as causal.").

future evasion for that particular taxpayer.

In addition, laboratory studies have found that participants with a history of being audited at random in prior rounds showed higher post-audit compliance levels.[47] There is also survey evidence in different national contexts suggesting that being audited is generally something that spurs taxpayers to increase compliance. A survey of Dutch taxpayers who had received corrections for errors on their taxes indicated that the corrections reinforced the taxpayers' beliefs that substantial (but not insubstantial) underreporting would be detected, with the implication that the taxpayers' compliance would increase.[48] Finally, a large survey of Canadian taxpayers conducted by mail in Toronto found a positive association between self-reported evasion and the experience of being audited in the past.[49]

---

47. Spicer & Hero, *supra* note 6, at 266 (discussing a ten-round experiment in which they found evidence that "taxpayers do not engage in optimizing strategies in making tax evasion decisions. Instead, they apparently use heuristics or 'rules of thumb'. In particular, the experience of being audited appears to lower levels of tax evasion even where the likelihood of an audit is completely random. Our analysis suggests, then, that random audits may lead to significantly higher levels of compliance among those audited, a result not predicted by conventional economic models."). This study experimentally investigates that claim and finds evidence to the contrary. *See also* WEBLEY ET AL., *supra* note 25, at 40–41.

48. Dick J. Hessing et al., *Does Deterrence Deter? Measuring the Effect of Deterrence in Field Studies and Experimental Studies, in* WHY PEOPLE PAY TAXES, *supra* note 32, at 300–01 ("The experience of detected evasion did change—though inconsistently—the subsequent perceptions of deterrence (see tables 2 and 3). While the experience of detection had no effects on the perceived certainty and perceived severity for evading F 500, the effects were significant for evading F 5,000: those who had been corrected (without suspicion of intentional evasion) showed higher perceived certainty of punishment.").

49. Neil Brooks & Anthony Doob, *Tax Evasion: Searching for a Theory of Compliant Behaviour*, *in* SECURING COMPLIANCE: SEVEN CASE STUDIES 120, 154 (Martin L. Friedland, ed., 1990) ("One of the most surprising findings of a number of surveys including ours, is that those who admit evading tax are more likely to report that they have been questioned by Revenue Canada [the Canadian tax agency] about their return or that they had been subject to a formal audit than those who reported compliance."). While the direction of causality in this relationship is not clear—respondents were not asked whether contact with Revenue Canada preceded their evasion activity or followed it—it is suggestive of a number of explanations for decreased post-audit compliance. "First, it might reflect a negative reaction to an audit or other form of tax department intervention. Stringent assessment may lower willingness to co-operate." *Id.* The explanation the authors find most plausible, however, is that "an audit is not as terrifying or as thorough as taxpayers anticipate. Consequently, once taxpayers have been subject to some form of intervention they are no longer deterred from evading by the threat of an audit." *Id.* Brooks and Doob

2. Studies Showing That Audit Experience Is Associated with Lower Post-Audit Compliance

On the other side, outcomes consistent with a decrease in post-audit compliance have been observed both experimentally and in studies that use tax agency microdata. Mittone labeled this the "bomb crater" effect because of the similarity to the behavior of infantrymen in the First World War, who were known to shelter in bomb craters in the belief that spots that had already been hit by bombs would be unlikely future targets.[50]

Laboratory studies conducted in Europe find that participants systematically change their compliance behavior in response to an audit, and the direction is negative.[51] I discuss these studies in detail in the next section

---

also mention reverse causality: those who exploit an opportunity to evade may be more likely to be audited because of the nature of their returns. *Id.* For discussion of the cross-disciplinary concerns about using survey evidence to study tax evasion, see WEBLEY ET AL., *supra* note 25, at 33 ("Given the difficulties involved in obtaining access to confidential tax returns or actually observing people completing tax forms, it is not surprising that most research into tax evasion relies on self-report. Economists are generally somewhat suspicious of this kind of data; they think it 'soft', which means it is 'not to be trusted', if not to say 'fishy'. . . . [W]e find a different perspective in criminology, however. . . . [O]n the other hand again, social psychologists are doubtful but 'most investigators using such measures have simply acknowledged an awareness of the scepticism surrounding the liability and validity of self-reported data' (Sobell 1976, 2).").

50. Luigi Mittone, *Dynamic Behaviour in Tax Evasion: An Experimental Approach*, 35 J. SOCIO-ECON. 813, 823–24 (2006); *see also* Guala & Mittone, *supra* note 25, at 505 ("They say that troops under heavy enemy fire hide in the craters of recent explosions, for they believe it is highly unlikely that two bombs will fall exactly in the same spot within a short time period. Something similar seems to happen in the tax experiments: immediately after each audit, tax payments fall sharply (i.e. evasion increases).").

51. *See* Guala & Mittone, *supra* note 25, at 505; Kastlunger et al., *supra* note 3, at 417 ("It is shown that the effectiveness of audits and fines—suggested by the standard economic model as the most relevant determinants deterring from tax evasion—cannot completely be confirmed. Especially, the finding of the bomb crater effect, shows that, rather than increasing or strengthening compliance, audits can lead taxpayers to develop strategies to ''escape' and thus have the opposite than expected effect."); Maciejovsky et al., *supra* note 12, at 684–85 (finding post-audit compliance bomb craters in both experiments); Mittone, *supra* note 50, at 823–24 ("Even if the trends are highly unstable and apparently follow some sort of random walk, there is a sort of constancy in the rounds immediately after a fiscal audit, which is almost always followed by a systematic increase in tax evasion. This

because their designs and findings are important to my own experimental design.

However, in addition to experimental findings, there is evidence that audits do not necessarily increase compliance,[52] and in some cases can result in a decline in tax compliance. As noted above, Gemmell and Ratto found that a particular audit result—a "clean" audit in which the randomly audited taxpayer was found to be compliant—was associated with a post-audit decrease in reported income.[53] Another study (unrelated to that mentioned above) by DeBacker et al. uses U.S. corporate tax return data to show that corporations tend to take more aggressive reporting positions and to pay less tax following an audit.[54] And Beer et al.'s finding that compliant taxpayers reported less income following an audit in which their honesty was confirmed suggests that crowding out might be at least partially responsible for a post-audit bomb crater effect.[55]

Outside the context of income tax compliance, Bergman and Nevarez find that compliance with the value-added tax by businesses drops following

---

increase generally has its lowest peak in correspondence to the round immediately after the fiscal audit, and sometimes lasts for more than one round. This shall be called the 'bomb crater effect' . . . the subjects decide to evade immediately after a fiscal audit because they believe that it cannot happen twice in the same place (time).") (footnote omitted); *infra* Part I.C.

52. ERICH KIRCHLER, THE ECONOMIC PSYCHOLOGY OF TAX BEHAVIOR 119 (Cambridge Univ. Press 2007) (discussing two older studies using U.S. taxpayer data: (1) Brian Erard, *The Influence of Tax Audits on Reporting Behavior*, *in* WHY PEOPLE PAY TAXES, *supra* note 32, at 114; and (2) Susan B. Long & Richard D. Schwartz, *The Impact of IRS Audits on Taxpayer Compliance: A Field Experiment in Specific Deterrence*, 1987 ANN. MEETING L. & SOC'Y ASS'N)); Andreoni et al., *supra* note 13, at 834. The studies identified taxpayers who had randomly experienced an audit twice in a row. Neither found significant deterrent effects on evasion in subsequent periods, but the results were sensitive to the specifications used by the researchers.

53. *See* Gemmel and Ratto, *supra* note 44.

54. Jason DeBacker et al., *Legal Enforcement and Corporate Behavior: An Analysis of Tax Aggressiveness After an Audit*, 58 J. LAW & ECON. 291, 316 (2015) ("Using data on the taxpaying behavior of US corporations, our study shows that firms tend to pay less in taxes after audits. In particular, firms become increasingly more tax aggressive for a few years after audits and then increase their tax payments before they are re-audited. Although this corporate behavior appears similar to that of individual taxpayers in the lab experiments mentioned above (in which individuals misperceived audit risk), we show that our corporate results may be driven by informed response to audit risk and a strategic Bayesian updating process.").

55. BEER ET AL., *supra* note 38, at 70.

an audit.[56] In a randomized field experiment involving Ecuadorean businesses, Carillo et al. study the effect of providing to businesses tax agency notifications of discrepancies with third-party revenue reports.[57] While this study did not involve an audit *per se*, the impact on taxes collected of firms' post-notification adjustments to reported revenues was found to be minimal: the firms simultaneously offset about 96% of their increases in reported revenue with increases in reported costs. Taken together, these findings suggest that a real-world bomb crater effect relating to tax agency compliance efforts may be a genuine concern for enforcement policy. As discussed in the following subpart, experimental work has sought to gain traction on finding the underlying mechanism behind individuals' post-audit compliance choices.

C. Prior Dynamic Audit Experiments

This section discusses in detail the design and results of the three most relevant experimental studies on the dynamics of tax compliance, both to situate my own study in the existing literature and to illustrate the contributions of this Article's unique experimental design.

By way of overview, the key parameters described for each experiment are the number and type of participants in each experimental condition, the number of rounds of repeat tax reporting, the amount of and manner in which income is assigned to or "earned" by the participants, the audit rate, the "rule" that determines the pattern of audits, the tax rate, the penalty rate, and the incentive structure facing the participants. Because disclosure of this design information (e.g., whether participants have full information or whether they use audit experience to help them update their

---

56. Marcelo Bergman & Armando Nevarez, *Do Audits Enhance Compliance? An Empirical Assessment of VAT Enforcement*, 59 NAT'L TAX J. 817, 821 (2006) ("The most puzzling result is that, on average, audits do not affect the level of future individual compliance. On the contrary, the gap between pre- and post-audit compliance rates for certain groups widens. On average, the non-audited have better compliance than audited taxpayers after enforcement.").

57. Paul Carrillo et al., *Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement* 30 (Harvard Bus. Sch., Working Paper No. 15-026, 2014) ("When firms are notified by the tax authority about detected revenue discrepancies, they amend their returns and increase reported revenues, closely matching the amount indicated when it is provided. However, they offset much of this higher declared revenue by an increase in declared costs, resulting in only small changes in their reported profits, implied tax evasion, and corresponding corporate tax liabilities. This is true even when adjustments are in the tens or hundreds of thousands of dollars.").

beliefs about the audit rate being random or endogenous) is so important to interpreting results, I also try to note the extent to which information is disclosed to participants.

1. Mittone (2006)

a. Experimental Design. Mittone's study is broadly pitched as an investigation of whether the presence of various psychological factors in a repeat-play experimental environment might impact the predictions of the standard expected utility model of tax compliance. Mittone used subjects recruited through bulletins at the Department of Economics at the University of Trento, Italy, and carried out the experiment using a computerized tax compliance program in the department's experimental laboratory. Thirty participants (15 males, 15 females) were recruited for each of eight experimental conditions, for a total of 240 participants. All of the experimental conditions involved repeated tax reporting over 60 periods. In all but one experiment, the subjects were divided into two groups that experienced audits at different intervals (rounds 13, 31, 34, 48, and 54 for the first group, and rounds 3, 24, 27, 40, 46, and 50 for the second group). The audits were distributed such that the effective audit rate was 6% for rounds 1 to 21, 10% for rounds 22 to 40, and 15% for rounds 41 to 61. The audit probabilities were disclosed in real time to the participants via a reminder that popped up on the screen before the audit rate changed in each interval of the experiment.

When a participant was selected for audit, her tax reporting in the current period and the past three periods was checked. The amount of income assigned to participants—out of which they had to decide how much to report—was 0.51 Euro cents from round 1 to round 48, and then decreased to 0.36 Euro cents from rounds 49 to 60. The tax rate applicable to this income was 20% in the first third of the periods, 30% in the second third of the periods, and 40% in the last third of the periods. The penalty rate was 4.5 times evaded taxes (over the maximum of four periods that were checked in the audit). Changing the values of key expected utility parameters over the course of the experiment was necessary to explore one of Mittone's questions: whether participants' compliance choices in a repeat-play environment with differences across periods in the expected returns from evasion were consistent with the standard expected utility model.

Further to his interest in testing the role of psychological factors in a repeat-compliance setting, Mittone inserted elements in some of his experimental conditions to test participants' sensitivity to "moral" factors. For instance, did the knowledge that taxes collected would be used for redistribution, both with and without reference to a specific public good (in

the sixth experimental condition, a scholarship fund), have an effect on compliance behavior? In other experimental conditions, he tested the role of "context": did the simulation of a fiscal environment—rather than a generic gambling situation—affect participants' compliance behavior? Did the use of a visual depiction of the probability of audit (a "probability wheel") have any effect on compliance outcomes? For the purposes of understanding the effects of audits over time, these varying experimental conditions are helpful to probe the robustness of the effects observed in a particular experimental setting.

b. Results. Mittone found that the trends of participants' compliance decisions were highly unstable with reference to the predictions of the standard model (participants "apparently follow some sort of random walk").[58] He conjectures that one reason for this may be because the participants found the expected value from evasion as compared to the sure choice of full reporting of income too difficult to compute.[59] However, he did find one striking regularity: in the rounds immediately after an audit, compliance levels dropped. The median percentage of tax paid in the round before the tax audit was much higher than the median percentage of tax paid in the round after the tax audit. Furthermore, "sometimes this effect may have some sort of echo, so that some subjects persist in evading for two or three rounds after the audit."[60] Thus, Mittone found that audits may have the unintended effect of increasing evasion.

In Mittone's study, the bomb crater effect was robust across the majority of the experimental conditions: "[the bomb crater effect] is influenced neither by the tax yield redistribution, nor by the context, nor by the fiscal audit system adopted, nor by the method used to inform the

58. Mittone, *supra* note 50, at 823 (finding that participants' behavior was consistent with the model in that there was an inverse relationship between the sure choice value and the amount of tax evaded, but that it was inconsistent because the expected value from evasion is negatively correlated with the amount of tax evaded).

59. *Id.* ("By combining the existence of an apparently rational behaviour, based on the inverse relationship between the sure choice value and the amount of tax evaded, with the continuously changing structure of the subjects' choices, one may suppose that some form of adaptive dynamic behaviour is taking place. It seems that the subjects ignore the trend of the expected value from evasion (maybe because it is too difficult to compute) and that they 'explore' the space of their alternatives by changing their choices in each round.").

60. *Id.* at 824 ("The echo effect seems to be influenced in some way (i.e., it is more or less compressed in time) by the different systems of fiscal audits introduced into the experiment, that is to say, by the audit retroactivity extent (from 0 to 10 rounds before the active round when the audit effectively took place).").

experimental subjects of the probability of their being investigated (i.e., using numbers—percentages—or the visual device)."[61]

In addition to the presence of the bomb crater effect and its persistence across one or more periods, Mittone looked at whether the two subgroups in each experimental condition—those that were audited for the first time in period 13 versus those first audited in period three—exhibited different levels of post-audit evasion. Measuring evasion by total instances of underreporting in and total tax yield from each experimental condition subgroup, he found that the members of the subgroups that were audited later (period 13) evaded more than those first audited early in the experiment (period three).[62] This was true in all experimental conditions except that in which the reporting exercise was framed as a game, not as a taxpaying experiment. Mittone interprets these results as showing that early audits may cause participants to learn to comply as they become more risk averse in the context of audits.[63] He referred to this as an "echo" effect: audits experienced early in a participant's experimental "tax life" reinforce compliance over a longer time period than audits experienced late in a participant's tax life.[64]

Mittone found further evidence consistent with this learning phenomenon in the last experimental condition, in which audits departed from the pattern fixed for the other seven experimental conditions. In this experimental condition, audits were concentrated either in the first half of the experimental life of the taxpayer or in the second half of the experimental life of the taxpayer (e.g., there were no audits in periods 31–60 for the former group, and no audits in periods 1–30 in the second group). Mittone summarizes the results as being "quite clear. Those subjects who learned in the first half of their experimental lives that fiscal audits are very uncommon became risk takers . . . ." By contrast, "the subjects [in the group that was audited early] learned that fiscal audits were very frequent and consequently also learned to be risk adverse [sic], maintaining this virtuous behavior for the entire experiment."[65]

c. Possible Explanations. What might explain these findings? All participants in all experiments were told the probability of audit, so there was no uncertainty: the probability distribution from which the audit event would

61. *Id.* at 826.

62. *Id.* at 826–27.

63. *Id.* at 827 ("One notes that being subjected to an audit at the beginning of the experiment produces a sort of risk aversion effect.").

64. *Id.* at 824–28; *see also* Kastlunger et al., *supra* note 3, at 407 (discussing these results).

65. Mittone, *supra* note 50, at 828.

be drawn was known, assuming that the participants did not systematically doubt the truth of the stated audit rate. As mentioned in the introduction, Mittone identified two explanations relevant to his experimental context.

The first is misperception of chance: taxpayers make systematic errors in understanding the probability of random audits.[66] The taxpayer in such a scenario mistakenly believes that her chances of being audited are conditional rather than independent and identically distributed (IID) draws from an audit lottery. After experiencing an audit, she concludes that her chances of being audited again in the next period are lower than if she had escaped audit in the prior period. Based on this belief, the hypothesis implies that the taxpayer decreases her level of subsequent compliance. The bomb crater effect implies that one-time random audits are experienced by taxpayers as next-period enforcement shelters rather than as deterrents to future evasion.[67]

The second possible explanation for the bomb crater effect is that taxpayers are more risk-seeking after an audit as part of an attempt to recoup their economic losses from the audit.[68] Such losses typically result from the fines or penalties imposed on taxpayers for nonpayment of tax owed on unreported income. In addition, for a taxpayer who was counting on her evasion going undetected, the amount of back taxes owed (on which fines or penalties are typically based) could also be experienced as a loss. If taxpayers have not budgeted or mentally accounted for the risk of incurring liabilities for the payment of back taxes and penalties upon audit, they may be more likely to "gamble" to restore themselves to their pre-audit economic position.

---

66. *Id.* at 829–30 (noting that the results are consistent with other observed instances of misapprehension of probabilities and "logical errors" by participants in experiments (e.g., the Allais and Ellsberg paradoxes)).

67. The bomb crater effect is similar in many instances to the effect of "gamblers' fallacy." Gamblers' fallacy refers to the belief that a random event that has not occurred for a period of time becomes more likely the longer it does not occur (even though draws are IID random). *See* Daniel Chen et al., *Decision-Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires*, 131 Q. J. ECON. 1181 (2016). However, this would imply that, in the absence of an audit, participants' compliance rates would rise over time. Mittone and others do not find evidence consistent with such a trend.

68. Others have noted that the bomb crater effect may also be observed in taxpayers that are fully compliant—a random audit might cause them to retreat from a pro-compliance posture. The implied level of risk aversion of the taxpayer after an audit might decrease due to psychological factors including anger and umbrage at being targeted, whether or not they are found to be fully compliant upon audit. *See generally* Lederman, *supra* note 14.

The next study discussed seeks to distinguish between the two effects in understanding the origin of the bomb crater effect.

2. Maciejovsky et al. (2007)

Maciejovsky et al. seek to disentangle the two possible causes of the bomb crater effect suggested by Mittone's research: misperception of chance and loss repair.[69] They hypothesize that audits influence future behavior in a time-dependent way: "the effect of audits crucially depends on the time lag between past audits, and the naïve generation of taxpayers' expectations as when the next audit is likely to occur."[70] In particular, they expected that compliance following an audit would look like a quadratic function with the time lag because the audit served as the exponent.[71]

The authors undertook two laboratory experiments to explore the prediction of the "three-way interaction effect of audits, fines, and the time lag between audits on tax compliance."[72] The participants for both studies

69. *See* Maciejovsky et al., *supra* note 12, at 679 ("In this paper, we investigate the dynamics of tax compliance in a stylized experimental setting with the aim of shedding new light on the competing mechanisms for the bomb crater effect: misperception of chance and loss repair. For this purpose, we conducted two experiments that varied audit probabilities, sanctions, and the time lag between past audits. Although a stylized experimental setting might not be the perfect vehicle for studying 'real world' tax behavior, it provides a satisfactory tool for evaluating the *relative* merits of competing explanations in a controlled environment (Alm, 1999).").

70. *Id.* at 681 ("We conjecture that audits constitute not only a reaction to past tax compliance, but audits and their consequences represent also a cause for future behaviors. Particularly, we hypothesize that the effect of audits crucially depends on the time lag between past audits, and the naïve generation of taxpayers' expectations as when the next audit is likely to occur. Naïve reasoning might predict that the probability of consecutive, yet independent, audits is low; a violation of independence according to probability theory . . . .").

71. *Id.* ("[I]f participants are prone to misperception of chance, we predict that (a) non-compliance is significantly lower immediately after an audit and (b) compliance returns gradually to its baseline level, as observed during an audit. This gradual adjustment process is predicted to be a quadratic function.").

72. *Id.* at 682 ("Taken together, we predict that reactions to audits and sanctions are strongest immediately after experiencing a fine and fade out in subsequent filings. . . . [W]e test these predictions by varying the audit probability and the severity of fines. We predict that in the case of frequent audits, compliance will be low immediately after an audit and will increase rapidly over the course of subsequent tax filings. In the case of less frequent audits, compliance is expected to be low after an audit and will increase comparatively slowly. This pattern is assumed

were students from the University of Vienna.

a. Experiment One

i. Experiment One Design. The first experiment required participants to earn income each period in a trading market that used a computerized, continuous double auction. Participants were given both cash to buy assets as well as five risky assets that they could choose to trade on this auction market. Trading would generate taxable sales revenue; alternatively, participants could hold the assets to generate taxable dividend income. The motivation for this setup was to probe the robustness of experimental findings in which participants were provided with "windfall earnings." The trading market allowed the researchers to test whether the experience of having more actively "earned" taxable income within the experimental setting—rather than simply being told how much income one has earned—would affect the results. In the income declaration phase, participants were required to report separately their taxable sales revenues and their taxable dividends.

There were 18 rounds of the experiment: a trading period followed by an income declaration for tax purposes and then a possible audit (to avoid end-game gambling, participants were not told this in advance—participants were told there would be a 20% probability that the game would be terminated in each of the last five periods). Participants were also told that the tax rate on income earned from either source was 50%. They were further told that the audit probability would either be 15% in one experimental treatment or 30% in the other treatment. The periods for the audits were randomly selected: audits took place after periods 3 and 12 for the low audit condition, and after periods 3, 5, 8, 12, and 16 for the high audit condition. There were two penalty rates to which participants were assigned: 50% or 100% of unreported income (rather than a percentage of back taxes owing on the unreported income). Finally, there were five possibilities for distributions on the return to the risky assets (each with a likelihood of 20%): 0, 30, 60, 90, or 120 experimental currency units (ECUs). Although the discussion of the experiment does not say so explicitly, the implication is that participants were permitted to keep a reward based on their after-tax and after-penalty positions at the end of the experiment, because ten ECUs was said to equal seven Euro cents.

ii. Experiment One Results. The study's results suggest that audits have a stronger impact on compliance than fines. The bomb crater effect seemed to be driven by misperceptions of chance far more than by loss-

---

to be particularly pronounced if sanctions are high.").

repair efforts on the part of participants. The hypothesized quadratic relation between audit probability and time lag between audits was borne out only in the high audit condition (30%)—this resulted in a bomb crater effect of approximately eight percent, and compliance increased back to pre-audit levels over a few periods after the initial bomb crater. For the low-audit condition, there was also a drop in compliance after an audit (approximately five percent from pre-audit levels), but it was smaller and compliance continued to fall in the subsequent three periods rather than increase.

Importantly, in the high- and low-sanctions analysis, there was no evidence that participants attempted to repair their loss from an audit and resulting fine by increasing evasion in the next period. However, they did sell significantly more assets in the following period, suggesting that this was the channel through which they sought to reclaim their cash position.[73]

b. Experiment Two

i. Experiment Two Design. The researchers had concerns about participants' belief in the true randomness of audits per the pre-experiment random assignment process described above. Therefore, they ran a second experiment in which participants themselves rolled a die to determine audit. A post-experiment questionnaire confirmed that participants were confident in the randomness of the audits.

For this experiment, there was just one level of audit probability (33.3%) and one level of sanctions (taxes owing *plus* 50% of unreported income). Taxes, rather than being 50%, were 40%. And there was no work effort required in gaining the endowment that participants had to declare for tax purposes in each period: endowments were randomly selected from a uniform distribution ranging between 80 ECU and 120 ECU, where one ECU was equivalent to ten Euro. This higher ECU-to-euro ratio was due to the way the reward was calculated for this experiment: "[a]t the end of the experiment, one of the 20 filing periods was randomly selected. The participants were paid their earnings of that particular period in cash . . . ."[74] Thus, cash payouts were made on the basis of one period, rather than the aggregate experience of the taxpayer over all 20 periods.

ii. Experiment Two Results. The results of the second experiment reinforce the pattern observed in the high-audit condition in the first experiment: there was a significant drop in compliance (six percent,

73. *Id.* at 685 ("Rather than repairing a loss by cutting one's tax share, participants seemed to have attempted to replenish their cash holdings by selling their assets.").

74. *Id.* at 687.

significant at five percent level) immediately after an audit as compared to the pre-audit level of compliance. The post-audit bomb crater was followed by a gradual increase of compliance to pre-audit rates via the hypothesized quadratic adjustment within two filing periods. This suggests that the method by which participants are assigned their taxable income is not determinative: earning income within the game through actual work effort versus simply being endowed with an arbitrary amount of income yielded the same experimental results. Nor was the reward condition a significant influence on participants' compliance.

In addition, the results reinforce the conclusion that the main mechanism for the bomb crater effect is misperception of chance. This was demonstrated by looking at the compliance behavior post-audit of participants who evaded as compared to participants who did not evade: there were fines only for the first group, but not the second. However, both groups exhibited similar reductions in compliance after an audit. Such reductions are consistent with the conclusion that loss repair (of back taxes owing or fines due upon audit) does not explain the bomb crater effect. These findings are also notable because the authors do not find evidence of crowding out of post-audit compliance among non-evaders.

3. Kastlunger et al. (2009)

Kastlunger et al. frame their study as exploring the experiential "life span" of tax compliance behavior in the context of the obligation to file annually repeatedly:

> The experience of being audited, and of possible fines, may influence further decisions on cooperation or evasion. Taxpayers' compliance can either be enforced through the auditing outcome or lead to increased evasion in the following years: taxpayers might assume that a subsequent audit is unlikely and therefore, take the risk to cut taxes in the subsequent filing period. Moreover, previous fines could fuel the motivation to compensate the experienced loss by saving taxes in the future.[75]

Thus, the authors sought to test a number of hypotheses relating to

75. Kastlunger et al., *supra* note 3, at 406 ("It should be interesting, what patterns of audits are likely to deter tax evasion, no matter whether compliance decreases owing to misperceived probability or to the motivation to compensate for an experienced loss.").

how patterns of audits affected taxpayers' post-audit compliance, as well as to study further the mechanism for the bomb crater effect (misperception of chance versus loss repair). They hypothesized that operant conditioning theory—by which behavior is either reinforced by rewards or deterred by punishment—might explain a number of features of the echo effect observed by Mittone. (Note that the echo effect broadly refers to the effect where a series of audits that took place in the early periods of a participant's experimental life span increased the participant's overall compliance more effectively than audits that took place later in the life span.)

The participant pool for all experimental conditions was undergraduate business students at the Faculty of Economics, University of Trento. The students were recruited via announcements on a bulletin board at the Faculty. Students in both studies were paid on the basis of their performance, as described below, and were instructed that the experiment simulated a real taxpaying context, but with protections for anonymity (in particular they were given a code to use in place of their name when entering the laboratory that was recorded in the file on the computer).

a. Experimental Design. The number of students participating was 120, with 30 students in each of four experimental conditions (a control condition and three non-control conditions). There were 60 rounds of tax reporting in each experimental condition, and nine audits took place (although in different sequences) across the 60 rounds. Participants were endowed with a constant income in each round of 1,000 experimental currency units. The following information was disclosed to participants: the tax rate is 20%, the audit rate is 15% (nine audits in 60 rounds), and the penalty for evasion, which would be detected upon audit, is the "evaded sum multiplied by three."

In this study, the experimental conditions tested various patterns of audits, as follows:

- The "control" experimental condition distributed the audits randomly over 60 rounds.
- The first experimental condition concentrated all nine audits in the first 20 rounds, and there were two sequences of continuous audits: audits occurred after the first, second, and third rounds; they also occurred after rounds nine and ten. There were no audits after round 20.
- In the second experimental condition, all audits were concentrated in the middle third of the rounds, e.g., between rounds 20 and 39. Here, there were also two sequences of continuous audits:

a sequence of three audits after rounds 20, 21 and 22, as well as a sequence of two audits after rounds 28 and 29.

- In the third experimental condition, like the first, all audits were placed in the first 20 rounds, but took place regularly after every second round (e.g., after rounds two, four, six, etc.).

b. Results

i. Effect on Overall Compliance. The authors used a random effects probit model with compliance (versus noncompliance) as the dependent variable. Previous-round tax payments, occurrence of an audit in the previous round, payment of fines in the previous round (presumably also a dummy variable), as well as the experimental conditions were the independent variables. The interaction effects of experimental conditions one through three with an audit in the previous round were also included to reveal the effect of the different audit patterns on post-audit compliance (e.g., the bomb crater effect).

Even though there were no significant differences in mean tax payments across the four experimental conditions, a comparison of overall tax payments including fines in the four conditions showed that participants in the third experimental condition paid significantly less taxes and fines than participants in the other three experimental conditions. The authors suggest that this is because the audit pattern in the third experimental condition was transparent.

ii. Bomb Crater Effect. In the control condition, compliance decreased after an audit, consistent with the bomb crater prediction. Consistent with the authors' hypothesis that repeat audits could reverse the bomb crater effect, experimental conditions one and two were found to be associated with higher post-audit compliance. Assuming there were no systematic differences between the participants across experimental conditions, this finding implies a causal relationship between the treatments in experimental conditions one and two—e.g., two sets of repeated audits—and higher levels of post-audit compliance. And, because experimental conditions one and two varied the placement of the repeated audits as between the first third and the second third of the participants' experimental lives, the temporal positioning of the repeat audits does not appear to be important to reversing the bomb crater effect. Interestingly, however, the coefficient on the interaction between experimental condition two and compliance immediately post-audit is higher than for the interaction between experimental condition one and post-audit. This suggests that the immediate effect of the later repeated audits, in rounds 20–39, is stronger than earlier audits. The authors do not discuss this in depth.

iii. Echo Effect. Does the temporal positioning of the series of audits matter not just for post-audit compliance but for compliance across the subsequent periods of the participants' experimental tax lives? Here, the authors break up their results from experimental conditions one and two into ten-round sections, and look at average compliance in each section. The expected effect—that sequences of earlier audits condition participants for higher levels of post-audit compliance—was weaker than expected. The "conditioning" effect of the audit sequences was short-lived regardless of whether the sequences came early or in the middle of the participants' experimental tax lives. In both experimental conditions, compliance increased with the sequences of audits. This increase was smaller in experimental condition two than in experimental condition one, consistent with some of the ideas of operant conditioning theory and the notion that participants already had a base of experience from which it was harder to get them to deviate. However, in experimental condition one, after not being audited for more than ten rounds, compliance began to free fall, getting lower and lower with each ten rounds of not being audited. This was also seen in experimental condition two, but the free fall in compliance seems likely to have been cut short by the end of the experiment.

As a result of these findings of a weaker-than-expected echo effect, the authors ran a second study with strong similarities to experimental condition two. Here, they sought to test an audit sequence they hypothesized could be more efficient than in the prior study: they added a single follow-up audit after a stretch of having no audits following the two concentrated sequences. Consistent with their hypothesis, they confirmed that they could extend the echo effect by having at least one audit at a later stage of the participant's tax life.

D. Limitations of Prior Dynamic Audit Studies

The findings of the three studies discussed above can be summarized as follows: first, the bomb crater effect following a single (random) audit appears to be an experimental regularity for undergraduate participants in a laboratory setting. Second, the bomb crater effect appears to be driven not by an effort to repair losses from fines of penalties from the prior-period audit, but by a misperception of chance on the part of the participants. Third, the bomb crater effect can be reversed: participants' post-audit compliance can be increased—for a time—by a sequence or two sequences of audits. Last, this post-audit bump in compliance caused by repeat audits can be reinforced and possibly extended, if single audits are performed in later rounds.

There are a number of aspects of these studies, however, that give rise

to further research questions. First, the samples are quite small, at approximately 30 students in each experimental condition. Even doubling the sample size would allow for more fine-grained experimental conclusions to be drawn, as controls could be added for some of the demographic characteristics of the participants.

Second, as is true for most tax compliance studies, the sample of participants is drawn from university undergraduates. Further, they were undergraduates in Italy and Austria, raising the question of whether different populations of experimental subjects might respond differently, particularly because the setting of the experiments is introduced as a simulation of the "real" taxpaying environment. This framing—or use of a "loaded" versus a "neutral" set of instructions to introduce the experiment—may increase its realism but may give rise to variation across different countries where the institutions and norms for taxpaying are different.[76] The fact that the results were broadly similar for the Italian undergraduates versus the Austrian undergraduates speaks to experimental robustness and suggests that this effect may be small, but it does raise the question of whether different behavior would be observed if the experiment were run with a different national framing. Relatedly, it raises the question of whether the effects are robust to nonstudent samples. Recent research regarding the context of experimental tax compliance studies suggests that undergraduates behave similarly to nonstudents and even to actual taxpayers.[77] However, the availability of a different sample of participants presents an opportunity to compare outcomes and further evaluate experimental robustness.

Third, for the studies that examine the effect of repeat audits on post-audit compliance levels,[78] the analysis of the results does not include the marginal effect of each additional audit. Although Kastlunger et al.

---

76. James Alm et al., *Institutional Uncertainty and Taxpayer Compliance*, 82 AM. ECON. REV. 1018, 1022 (1992) (differentiating between "neutral" and "loaded" terminology in experimental tax compliance research).

77. James Alm et al., *On the External Validity of Laboratory Tax Compliance Experiments*, 53 ECON. INQUIRY 1170, 1182–83 (2015) ("Our analysis suggests two main conclusions regarding the external validity of tax compliance experiments. Both conclusions are consistent with the result that students and nonstudents behave largely the same," but noting "a caveat: care must be taken when the policy treatment may incorporate a substantial level of external experience. We find that students respond differently to the presence of tax liability uncertainty, and our conjecture is that this may be the result of nonstudent subjects having more experience with this specific phenomenon in the field. Regardless, however, we still find that the *changes* in compliance behavior in response to institutional changes (treatments) of these pools (if not always their *levels*) largely parallel each other.").

78. Kastlunger et al., *supra* note 3; Mittone, *supra* note 50.

confirmed Mittone's finding that repeat audits reversed the bomb crater effect, the question of the marginal effect on compliance of each additional audit remains unanswered. Given the expense of audits, this has clear importance for tax administrators and those who study enforcement policy. Differentiating the effect on compliance of two audits in a row versus three audits in a row—versus even four audits in a row—is a natural next step in this line of research.

Fourth, the three existing studies compare random audits to predetermined sequences of audits. For each subgroup within each experimental condition (or each subgroup within it), the period in which the audits occur does not vary across participants. The actions of the experimental taxing authority are invariant to the behavior of the taxpayer: that is, audits are exogenously determined, not endogenously determined. While this makes a lot of sense as an initial approach to assessing the existence of the bomb crater effect, to determining its possible causes, and to discerning whether an echo effect exists or can be induced by certain pre-specified audit patterns, it is unrealistic as a reflection of plausible audit policies a tax administrator might choose for two reasons. First, it is unlikely that a tax administrator would adopt an audit policy in which taxpayers were always audited, say, the third, fourth and fifth time they filed a return. The pattern of such policies would be too transparent and vulnerable to taxpayer gaming (certainly, accountants or tax preparers would be in a position to notice the pattern). Second, random audits are likely more the exception than the rule. While tax administrators typically engage in some random audits (for research purposes, at the very least), audits are often endogenous, meaning they are conditional on the tax agency's observation of that specific taxpayer's behavior.[79] As Alm et al. note, endogenous audits may arise from information disclosed on the taxpayer's return (e.g., the Service's "Discriminant Index Function") or from information gleaned from a previous audit.[80] Using a variant on the conditional future audit rule of Alm et al. to

79. *See also* Mark D. Phillips, *Deterrence vs. Gamesmanship: Taxpayer Response to Targeted Audits and Endogenous Penalties*, 100 J. ECON. BEHAV. & ORG. 81 (2014). For discussion of the use of random audits in the Service's National Research Program, see DeBacker et al., *supra* note 40, at 24–25 ("As noted above, an NRP audit is random, and the letters NRP subjects receive explicitly mention that they were selected for a random audit used for IRS research purposes," but noting that one explanation for the significant effects of audits on some filers "may be that filers misunderstand the nature of the NRP audit.").

80. An example of the former type is the cutoff audit rule explored theoretically by Reinganum and Wilde and experimentally by Alm et al. *See* Alm et al., KYKLOS, *supra* note 3, at 31–34; Reinganum & Wilde, *supra* note 31. The cutoff rule is roughly similar to the DIF tax return selection system used by the Service,

extend the experimental inquiry of the bomb crater and echo effects into the realm of EARs is a natural avenue for further research.

To summarize, the provocative experimental results from the three studies detailed above motivate my research project. Building on these studies, I seek to test the hypothesis that post-audit "bomb craters" in compliance will be present in response to random audits but not in response to an endogenous audit in which audit selection is determined by past detected evasion. My approach offers three experimental innovations: (1) the use of a larger sample in (2) a new institutional context with (3) an online experimental participant pool not yet used in the context of tax compliance research, dynamic or otherwise. My contributions are intended to speak more to experimental robustness than to external validity; however, design details that intentionally hew to the U.S. taxpaying context are intended to suggest the potential for future controlled field experiments on the dynamics of compliance.

II. EXPERIMENTAL DESIGN

This section describes in detail the survey experiment that I use to explore the relationship between audit patterns and subsequent tax compliance.[81]

A. Goals

This experiment brings together two related, but as yet distinct, literatures within experimental tax compliance: the new literature on the bomb crater effect with the now classic literature on EARs.[82] Specifically, the survey experiment tests the hypothesis that the bomb crater effect occurs in the presence of random audits but not endogenous audits that are conditional on past detected evasion. Confirmation of this hypothesis would support the proposition that misperception of chance is driving the bomb crater effect, and would provide tax administrators with new information

---

and one of its points of attraction is cost: it does not depend (necessarily) on repeat audits.

81. I refer interchangeably to the MTurk HIT survey, an experiment, or both. This is because it is structured on MTurk as a survey task, with the added dimension of randomized experimental conditions.

82. For discussion of the newer experimental literature, *see supra*, Part I.C. With respect to the classic literature on EARs, I am referring primarily to Alm et al., KYKLOS, *supra* note 3, as well as theoretical work on endogenous audits by Reinganum & Wilde, *supra* note 31, and Graetz et al., *supra* note 31.

about the possible unintended effects of presenting audits as "random" versus nonrandom audit policies.

The experimental design is informed by best practices for experimental economic research generally and tax compliance experiments in particular.[83] In that regard, it seeks to minimize unnecessary complexity by limiting the number of varying experimental parameters.[84] With reference to the categories of experimental research laid out by Roth, this experiment (like its predecessors that were reviewed in the preceding part) lies in the "searching for facts" category: it seeks to uncover empirical regularities in participants' behavior that may not be amenable to explanation by existing theoretical models.[85]

B. Setting

MTurk is staffed by workers—called alternatively providers, "turkers," or "MTurkers"—who can sign up easily on the website to complete "human intelligence tasks" (HITs) in exchange for payment.[86] HITs are posted on the workers' portal of MTurk by MTurk "requesters." This portal displays all HITs available for possible acceptance by a worker. Each HIT has a title and a short description; the compensation for completing the task is also listed.

Compensation on MTurk is at the discretion of the requester—there is no minimum, and it can be calculated in a number of ways. For instance, MTurkers might be compensated for completing a HIT on a per minute basis, but compensation can also include a quality- or merit-based component. The merit pay encourages MTurkers to complete a task quickly, or accurately, or is designed to otherwise incentivize thoughtful strategic behavior on the part of the MTurk participant. In general, tasks with merit pay are advertised as such on the worker's portal.

---

83. THE HANDBOOK OF EXPERIMENTAL ECONOMICS (John H. Kagel & Alvin E. Roth eds., 1995); WEBLEY ET AL., *supra* note 25; Guala & Mittone, *supra* note 25; Alvin E. Roth, *Laboratory Experimentation in Economics: A Methodological Overview*, 98 ECON. J. 974 (1988).

84. Maximizing complexity is clearly not recommended, but oversimplification of tax evasion experimental conditions has also been criticized. For instance, economist Frank Cowell argues that experimental complexity is not necessarily an enemy in the context of tax evasion: complexity is a feature of tax systems, so designing an experiment that is unrealistically simple may exacerbate concerns about external validity. WEBLEY ET AL., *supra* note 25, at 123–27.

85. Guala & Mittone, *supra* note 25, at 495; Roth, *supra* note 83.

86. Alek Felstiner, *Working the Crowd: Employment and Labor Law in the Crowdsourcing Industry*, 32 BERKELEY J. EMP. & LAB. L. 143, 148, 161 (2011).

The HIT representing this survey experiment was titled "Academic Survey About Taxes." Its description advertised it as having a merit-pay component (see discussion below for details on calculation of merit pay "bonus" amount) and indicated that the task "asks you to make a series of tax choices and then reflect on your experience as a taxpayer." This reflects a suggested best practice on MTurk that requesters should be as specific and upfront about the nature of the task as they can. The rationale is that this transparency facilitates better matching of workers with tasks, thereby increasing the quality of the work product.

C. Structure

Here, I summarize the six key elements of the experiment: (1) eligibility of participants, (2) assignment of eligible participants to experimental conditions, (3) instructions, (4) the details of the control versus treatment experimental conditions, (5) the periodic tax reporting process, including the "reward calculator," and (6) the post-experiment questionnaire.

1. Eligibility

The validity of my experimental results rests on ensuring only the intended MTurk workers respond to the survey—that is, high-quality MTurk workers who are U.S. residents. This is important to avoid contaminating my data with responses from an unintended participant pool.

Following the best practices recommended by Kuziemko et al. for screening survey participants in the MTurk environment,[87] I place a number of checks in place to achieve the objective of surveying only U.S. resident individuals who will complete the survey in a careful and serious manner and will be prevented from retaking the survey.

First, I limit eligible participants to MTurkers who have minimum MTurk performance ratings. In particular, I require anyone who accepts the HIT to have a past MTurk completion rate of more than 90%. This requirement increases the likelihood that the worker will be an individual who will take the task seriously and will continue with it to completion, if not to earn the wage then to maintain his or her high MTurk task completion rating (a high rating is a common requester requirement).

Second, participants are required to take and pass a short quiz after

87. Ilyana Kuziemko et al., *How Elastic are Preferences for Redistribution? Evidence from Randomized Survey Experiments*, 105 AM. ECON. REV. 1478 app. at 1–21 (2015), https://eml.berkeley.edu/~saez/kuziemko-norton-saez-stantchevaAER15.pdf.

completing the instruction phase of the survey. This feature is intended both to ensure that participants comprehend the setup of the survey and to alleviate concerns about robots, or "bots." A bot is a hacker or hacker-created algorithm that breaches the MTurk system and exploits the online tasks by clicking through to completion and illegitimately claiming a wage (this is exceedingly uncommon, but it is a concern that may be raised when using MTurk for social science research).

Third, as a supplemental layer of security against bots, I replicate the feature of Kuziemko et al. that requires workers to manually type in a code before they can claim their reward.[88] In our HIT, the worker is required to type the word "banana" before their results can be approved for payment.

Fourth, to make sure only U.S. residents are granted access, I use Amazon's functionality to directly select for U.S.-based workers. This means that only those who have submitted directly to Amazon (as is required when signing up for a MTurk account) a U.S. address and social security number, and who have had those sources of identification verified by Amazon, are eligible. In addition, there is a special screen in the task itself that requires the worker to certify that he or she is a resident of the U. S. Finally, I limit IP addresses to U.S.-only.

Fifth, to prevent participants from skipping steps by advancing without completing all fields, I use popup windows as a "progress blocker" when a worker tries to advance prematurely.

Last, I seek to address the risk that a given MTurker will complete the survey more than once, thereby biasing the results. Here, I use Amazon's "once only" functionality to ensure that workers who have accepted the HIT in the past are blocked from accepting it again.

2. Assignment to Experimental Conditions

Before the experiment begins with the instructions and the period one reporting exercise, each participant is assigned to one of two experimental conditions. This is accomplished through a random-number-generator process embedded in the HIT code.[89]

---

88. *Id.* at 9–11.

89. Javascript has a built-in math library which includes a random number generator that generates a pseudo-random number from zero to one. To assign participants to experimental conditions, a worker who draws a number between [0,0.5] is assigned to the control condition; if the number drawn is between [0.5,1] th e worker is assigned to the treatment. For more information, see *Math.random()*, MO ZILLA DEVELOPER NETWORK, https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects/Math/random (last visited Dec.

3. Instructions

The instructions to the survey experiment embody a "full disclosure" approach to experimental design. I offer here a brief explanation of my use of this approach. Full information about the simulated taxpaying environment is provided to participants to ensure that as many parameters as possible can be held constant across the two experimental conditions. To the extent that all relevant parameters are uniform except the participant's assignment to the control versus experimental conditions, the credibility of a causal interpretation of any differences in the outcome variable—the voluntary compliance rate—is bolstered. Conversely, where participants are left to rely on their own expectations of relevant parameters, and further where these parameters are not observable to the researcher and thus cannot be "controlled for" in the subsequent data analysis, interpreting differences in compliance rates across the two experimental conditions becomes more difficult.

The instructions disclose all relevant information for the tax-reporting exercise over the 60 periods: how income is "earned," the tax rate, the average audit rate, the penalty rate, and how the performance-based component of the participants' participation is calculated. The instructions are common to both experimental conditions except for one line in the "audit" screen of instructions that informs participants that they will be flagged for future audit if an audit reveals unreported income. Thus, the treatment should be viewed as both the EAR and the delivery of the extra line of instructions.

After an MTurk worker selects and qualifies for the HIT, she will be asked a series of demographic questions, which will be immediately followed by a series of six "instruction screens" that provide the necessary information to perform the task. The content of the instruction screens can be summarized as follows:

- Period: the experiment lasts a minimum of 62 periods; these periods are intended to correspond to successive tax years (tax periods).[90]

13, 2016); *Pseudorandom Number Generator: Mathematical Definition*, WIKIPEDIA https://en.wikipedia.org/wiki/Pseudorandom_number_generator#Mathematical_definition (last visited Dec. 13, 2016).

90. The experiment actually lasts for 60 rounds; as a result, from the standpoint of the participant, it ends early. Use of 60 rounds is the standard in all prior dynamic audit experiments. The goal of setting expectations for the participant that the experiment will last longer than 60 rounds is to avoid "gambling" during the

- Income source: in each tax period, participants are endowed with an amount of income that ranges each period from $800 to $1200. This income is described as "extra" income, in the sense that it is income from self-employment, or earned in addition to participants' other sources of "regular" wage or investment income. Participants are told that while they do have a legal obligation to report and pay taxes on this income, the payer is not required to inform the government of the payment.
- Tax rate: 30%.
- Probability of audit: Ten percent. Specifically, participants are told, "[a]udits occur with an average probability of 10 percent. This means that the audit rate, averaged over all taxpayers, is 10 percent." They are also told that they may be selected for audit even if they reported all their income, and that audits detect unreported income with certainty.
    - Only in the treatment experimental condition are participants given a further piece of information: the announcement of the EAR (see below).
- Delay upon audit: in the event of audit, the experiment will be paused for between 5 and 30 seconds to simulate the hassle and time cost of complying with an audit request.
- Penalty for underreporting: 100% of unpaid taxes in the prior period only.[91] If an audit reveals unreported income, participants will have to pay tax on this unreported income, plus a penalty equal in amount to (i.e., 100% of) the tax.
- Reward structure: participants' monetary reward has a fixed component of three dollars that was visible as the wage when they clicked on the HIT, but also a "bonus" component. The bonus component is based on the participants' after-tax and after-penalty income (i.e., her income totaled across all periods, less taxes paid and fines assessed). This amount, which is defined in the instructions as "accumulated net income" is calculated and displayed prominently for the participant after she reports her income in each period of the task.[92]

---

ultimate round or set of rounds.

91. This is somewhat unrealistic because there is generally a three-year statute of limitations on audit. I.R.C. § 6501; Reg. § 301.6501(a)–1.

92. This follows the design of Mittone, *supra* note 50, which features a calculator box at the bottom of the computer "game" used by participants. Note also that the "bonus" component of the reward reflects taxpayers' real-world incentives as embodied in the standard expected utility model.

With the exception of the piece of information about the audit rule in the fourth bullet point above, these instructions are common to all participants in the HIT, regardless of their assignment to an experimental condition.

As noted above, participants are required to take a quiz before progressing to the start of the survey: this is considered a best practice where the setup of a survey experiment is complicated. The quiz is multiple choice but requires correct answers before the participant can proceed. It asks about all of the key parameters of the experiment: tax rate, audit rate, penalty rate, number of periods, and whether information is confidential or not.

How realistic are the background conditions of the experiment? This is a version of the "external validity" question discussed in the introduction to this Article. Even without explicitly claiming that laboratory results are externally valid to real-world decision-making situations, researchers and policymakers' abilities to extrapolate experimental results to predict outcomes outside the laboratory grows weaker as the differences between an experimental environment and the "real world" setting it seeks to mimic grow stronger. Here, I tried wherever possible to use realistic parameter values. A tax rate of 30% and a taxpaying life span of 60 years are plausible values for many U.S. taxpayers. Although the Service primarily uses nonrandom audits pursuant to the DIF (discriminant index function) system, random audits do exist (such as for research purposes).[93] Thus, the experimental conditions described below are broadly in line with reality.

The audit and the penalty rate, however, are not as realistic. With respect to the audit rate, measures of the overall average probability of being audited is about one percent for individual taxpayers, and estimates of the average probability of audit conditional on a taxpayer underreporting her income are likely higher but not much higher.[94] A ten percent audit rate was chosen despite these facts: the opportunity to choose compliance "post-audit" was the key question in the study, so having more "post-audit" periods rather than fewer—while still keeping the audit rate constant across the two experimental conditions—outweighed the drawbacks of having a higher-than-realistic audit rate. Future studies that check the robustness of these

93. *See* J.T. Manhire, *Toward a Perspective-Dependent Theory of Audit Probability for Tax Compliance Models*, 33 VA. TAX REV. 629, 642 (2014) (citing the Internal Revenue Manual and noting that "given G's [the government's] selection algorithms, such as the [Service's] Discriminant Index Function score . . . the chances of finding underreported tax with the current algorithmic selection criteria will always be better than random.").

94. *Id.* at 644.

results to a lower audit rate would be helpful. Similarly, a penalty rate of 100% is higher than most taxpayers would face under current U.S. law. These divergences, I believe, were justified because a sufficiently large number of observations needed to be generated.

4. Details of the Two Experimental Conditions

a. Control Condition: Random Audits. Participants assigned to the first experimental condition (the "control") experience a random audit process: it is deployed through independent draws from a uniform constant probability of ten percent. If an audit occurs in one reporting period, the participant faces an identical audit draw in the following period, regardless of whether undeclared income was detected. After any audit, the participant is simply returned to the pool and subjected again to the random IID audit lottery.

b. Treatment Condition: The Endogenous Audit Rule. Participants assigned to the second experimental condition experience audits according to an EAR. As noted above, but emphasized here, the instructions for this experimental condition are identical to the control condition, with one exception: the use of endogenous audits is announced to participants upfront. Participants are given an additional piece of information in the instructions, as follows:

> Audits occur with a probability of 10 percent. If you do not report all of your income, and you are chosen for audit, your unreported income will be detected with certainty. Moreover, you will be 'flagged' for audit in subsequent periods.

The mechanism for the flagging—the specifics of the EAR including the probability that the "flag' will be heeded by the taxing authority—is not disclosed. In terms of its design, the EAR is initially identical to that of the control condition. Participants first are chosen for audit through the random IID process—with the same uniform ten percent probability—that is used in the control experimental condition.

However, once a participant is selected for audit through this random process, her experience in subsequent reporting periods is conditional on the outcome of the prior period's audit. And, to ensure an average audit rate of ten percent, the EAR incorporates a cap of no more than six audits per participant over the 60 rounds of the experiment. The algorithm for the EAR is as follows:

1. Initially, each participant is subjected to a random IID audit rule, pursuant to which audits are drawn from a uniform probability distribution of ten percent (equivalent to the "control" condition's audit rule).
2. Once the participant is randomly chosen for audit, if the audit reveals that she reported all her income truthfully, she will be returned to the random IID audit lottery.
3. If the initial audit reveals underreported income, she will be audited again in the immediately succeeding period.
4. If the repeat audit reveals evasion, there is another repeat audit. This conditional audit process continues until the participant has reached the cap of six audits. After the sixth audit, the participant faces no more audits (note that there is no disclosure on this point).
5. Note that, following any audit other than the sixth, where an audit reveals truthful reporting, the participant is returned to the pool of participants selected for audit through the ten percent IID random process.

Why is it important to cap the number of audits at six for the EAR? This approach adopts a "fixed auditor resource" approach to audits across time for a given participant in the study, rather than drawing audits from a specific distribution according to a random process and is similar to that of Collins and Plumlee.[95] Collins and Plumlee compare joint tax reporting and work-effort decisions under conditions of differing tax rates and audit schemes. To keep the audit probability fixed across the three audit schemes evaluated (random, cutoff, and conditional audit rules), they hold the absolute number of audits fixed at two audits out of every ten taxpayers.

The fixed-auditor-resource approach has two key advantages over an EAR without a cap. First, it is realistic in that it evokes the real-world resource constraint faced by taxing agencies. Limited resources are available for audit, and the use of a cap is one way of simulating a fixed agency

95. Julie H. Collins & R. David Plumlee, *The Taxpayer's Labor and Reporting Decision: The Effect of Audit Schemes*, 66 ACCT. REV. 559 (1991). Collins and Plumlee "fix at a constant level the total amount of auditing that the taxing authority can employ in each [audit] scheme [random, cutoff or conditional]. With the assumption of particular audit cost parameters, this study could be considered somewhat analogous to the situation in models referenced above where audit resources are fixed by a binding budget constraint at a level below the equilibrium." *Id.* at 561. In Collins and Plumlee's experimental design, the "probability-forcing" mechanism is to select two out of ten participants for audit in each session of the experiment regardless of the audit scheme used. *See id.* at 565–68. Only the method of selection is varied depending on the audit rule.

budget. Second, interpreting compliance results across experimental conditions—that is, comparing compliance behavior under a random audit rule versus an EAR—becomes more difficult when one of the parameters that typically influences compliance is also endogenous. Simply performing repeat audits on taxpayers found to be noncompliant implies that the overall audit rate would be much higher than if a random audit rule at a fixed percentage were followed. In short, holding the audit rate fixed across experimental conditions makes it easier to isolate the effects of the treatment condition itself, rather than trying to disentangle the effect of the treatment from the effect of an increase in the average rate of audit.[96]

5. Periodic Tax Reporting Process

The participant in the MTurk survey proceeds through screens representing each of 60 audit periods. Each tax year screen has a space into which the participant must type an amount.

After entering the reported amount of income (which is constrained to be between zero dollars and the amount assigned for that reporting period), the taxpayer must click a button, after which she will see one of two screens.

If there is no audit, a "summary" screen appears that lists separately the reported income for that period, the tax assessed on that income, any unreported income, any taxes assessed on that unreported income (this will be zero dollars because no audit has occurred), any penalties assessed on unreported income (again, it will be zero dollars because no audit has occurred), and the resulting net income for that period. In the upper right hand corner of this "summary" screen is a yellow box that tallies the accumulated net income (over all periods) of the participant; this design follows Mittone.[97] This yellow box is called the "reward calculator" because the bonus component of the participant's reward will be based on the accumulated net income listed in the yellow box.

If there is an audit, a different screen pops up before the "summary" screen. The screen says "Audit!" and there is a time-lapse bar that tolls for a short period, per the instructions. After the tolling concludes, the summary screen described above appears. In the event that the participant had unreported income, there will be a value for the two blanks labeled "taxes assessed on unreported income" and "penalties assessed on unreported income."

---

96. The importance of holding key parameters fixed across experimental conditions is much discussed in the experimental tax compliance literature. S*ee, e.g.*, Alm et al., *supra* note 8; Kastlunger et al., *supra* note 3.

97. *See supra* Part I.C.1.

6. Post-Experiment Questionnaire

Webley et al. make a strenuous case for surveying participants in tax compliance experiments after the experiment has concluded.[98] Surveys should be designed to gain insight into the participants' experience during the experiment, as well as some aspects of their taxpaying history and subjective postures towards audits.[99] A number of prominent studies, including the Kastlunger et al. study detailed above, use post-experiment questionnaires.

At the conclusion of the 60-period reporting and audit experiment, and after the final tax reporting "summary" screen is shown, each participant is required to answer a survey that includes the following prompts:

- Please provide any reactions to the task you just completed (text box).
- Do you feel that your responses in this task were similar to how you would report your income on a tax return, in "real life"? Why or why not? Remember, all answers are anonymous and confidential (text box).
- How many times have you filed an annual personal income tax return in the past (dropdown menu)?

98. WEBLEY ET AL., *supra* note 25. Frank Cowell's chapter of the Webley book advocates for the use of a follow-up questionnaire to "the experimental session with some sort of a 'what did you think you were doing at the time' questioning" to gauge subjects' awareness and gain at least anecdotal insight into possible mechanisms behind counterintuitive or counter-rational behavior. Frank A. Cowell, *Tax Evasion Experiments: An Economist's View*, *in* WEBLEY ET AL., *supra* note 25, 123, 127.

99. Cowell notes in particular that post-experimental questionnaires seem "to add enormously to the value of the experiment. To some extent it offsets the unease felt by many about a fundamental issue in the experimental approach to investigating economic behavior: the problem that the subjects' awareness that they are in an experiment substantially alters the nature of their responses." Cowell, *supra* note 98, at 127.

- Have you interacted with the IRS after filing your taxes (yes/no)?
- If yes, what kind of interaction was it (correspondence, phone, face-to-face—multiple choice)?
- Did this interaction affect your tax reporting behavior? If yes, how?

The post-experiment questionnaire was designed with several open-ended questions to allow participants to comment on their experience.

III. RESULTS

A. Description of Sample

In September 2015, 199 MTurk workers participated in the survey experiment described in the preceding section. Using a quasi-random number generator to assign participants randomly to experimental conditions, 106 participants were assigned to the control group and 93 participants were assigned to the treatment group. On average, the task took 42 minutes, and the average aggregate reward earned by the participants was $3.26 (computed as a flat wage of three dollars plus "bonus" based on accumulated net income).

The following table summarizes the main demographic characteristics of my sample.[100] Even with random assignment, it is expected that there will be some differences across treatment and control groups, and Appendix B (Tables B1 and B2) confirms that the main results are robust to the inclusion of demographic controls.

100. See Appendix A for a comparison of the MTurk sample to a more nationally representative sample.

Table 1: Demographic Summary Statistics of MTurk Survey Participants

Basic Demographics			**Employment status**		
	Control	***Treatment***		***Control***	***Treatment***
Mean age	35.3	33.7	Full-time employed	59%	60%
U.S. born	96%	98%	Not in labor force	5%	3%
Has kids	28%	18%	Part-time employed	10%	10%
Male	58%	50%	Self-employed or small business owner	21%	17%
Married	34%	25%	Student	1%	3%
			Unemployed and looking for work	5%	7%
Education			**Income**		
	Control	***Treatment***		***Control***	***Treatment***
Some high school	0%	1%	$0-$9,999	7%	4%
High school degree/GED	19%	19%	$10K-$14,999	4%	5%
Some college	17%	27%	$15K-$19,999	9%	5%
2-year college degree	14%	10%	$20K-$29,999	17%	12%
4-year college degree	42%	34%	$30K-$39,999	16%	19%
Master's degree	5%	9%	$40K-$49,999	16%	15%
Professional degree	2%	0%	$50K-$74,999	11%	21%
Doctoral degree	2%	1%	$75K-$99,999	10%	12%
			$100K-$124,999	8%	3%
			$125K-$149,999	1%	1%
			$150K-$199,999	1%	1%
			$200K+	0%	1%

The raw data consists of a spreadsheet with 199 rows containing each participant's responses, and hundreds of columns of variables corresponding

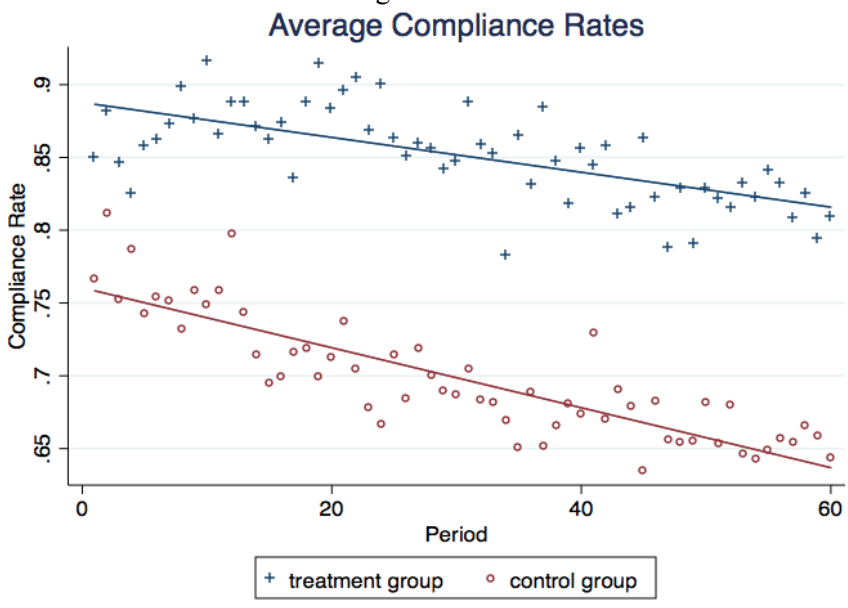
to the assignment of the experimental condition dummy, demographic questions, quiz responses and, for each of the 60 periods, entries for income earned (assigned), income reported, and audit/no audit. To perform further analysis of this data, it was "reshaped" to yield 60 observations for each participant, or a total of 11,940 participant-period observations.

B. Effect of Treatment and Control on Overall Tax Compliance

Which experimental condition performed better in terms of overall tax compliance? To answer this question, I measure tax compliance not as the total dollars declared, but as a fraction: dollars reported for period *t* is the numerator, and income earned in period *t* is the denominator. This is because my design allowed for fluctuations in participants' income across periods: simply reporting that a participant declared $800 of her income would not yield meaningful information about whether she was compliant or evasive (because her income in that period could be anywhere from $800 to $1200 according to a random assignment process).

Figure 1 reports tax compliance results across the 60 periods in the survey. The sample was first split into control and treatment groups, and average compliance within each group for each period was calculated. Thus, each dot on the figure represents period *t*'s average compliance rate for the participants in the relevant experimental condition.

Figure 1:



A few aspects of the pattern in the figure stand out. First, the average compliance rates for the treatment condition are systematically higher, in every period, than the control condition. This is true even in period one, which may seem a bit curious: why would compliance rates differ initially, given that there is no discernable difference in the experiences of participants in the two experimental conditions? The answer appears to lie in the one-line addition to the instructions discussed above (evasion will "flag" you for future audits). Although the first period taxpaying experience of both groups was exactly the same, they had been exposed to different information about the experiment. Therefore, the gap between average compliance for the treatment group and that of the control group can be attributed to a "presentation effect" of audits for the treatment group: endogenous audits that are presented as such appear to have an immediate deterrence effect. On the basis of this figure, the presentation of audits as endogenous—separate and apart from the substantive differences in how a participant experiences an audit rule after the first period—is sufficient to induce a compliance increase of approximately 14% in period one.

Second, the path of compliance is roughly linear for both groups, but it has a decreasing slope, implying that compliance in the experiment erodes over time. The slope of the control group decreases slightly more steeply than that of the treatment group, implying that as time goes on in the random audit scheme, there appears to be more appetite to take risks, whereas that same effect is somewhat attenuated in the treatment group.

Last, there seems to be more variance from the fitted line for the treatment group—this suggests that there may be systematic deviations or some randomness in responses in this group that are not present to the same degree in the control group.

C. Bomb Crater Analysis

1. Visual Analysis

Is there evidence of a bomb crater effect in either of the experimental conditions? My basic hypothesis was that a bomb crater effect of random audits, if present, would be erased in the presence of an EAR. My data was consistent with this hypothesis.

Figure 2 displays compliance rates of participants in the control group, and classifies each participant-period as being one of two statuses: a period immediately following an audit, or a period not immediately following an audit. Each participant-period is therefore either in the "post-audit" pool, or the "non-post-audit" pool. Then, the average compliance rate for each pool is calculated for each period. For the non-post-audit participant-periods, the

average is calculated for all 60 periods. For the post-audit participant-periods, averages were calculated only for periods 2 through 60 (period one has no observations, because it is impossible to be post-audit in period one of the survey).

Figure 2:

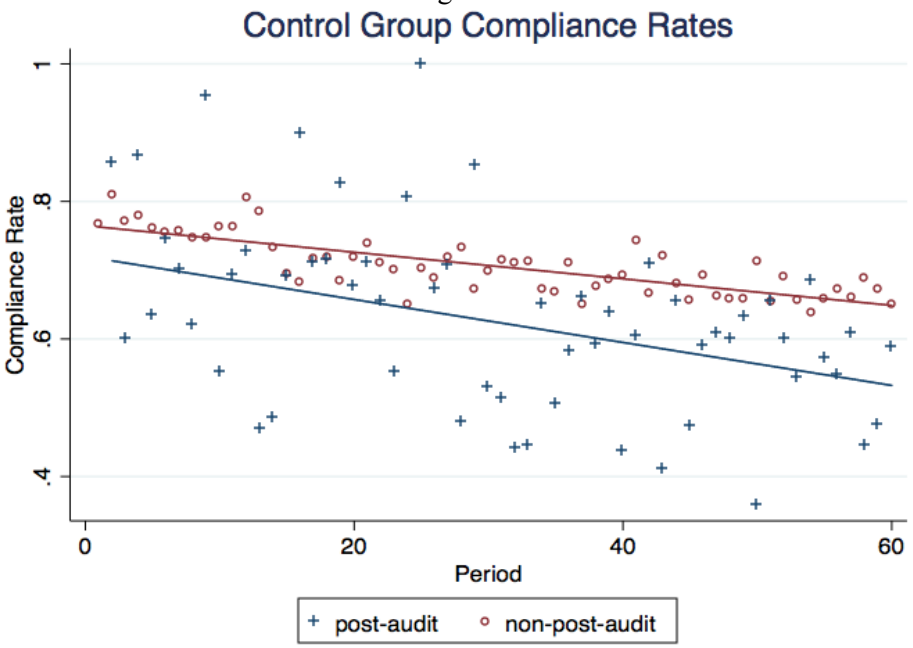


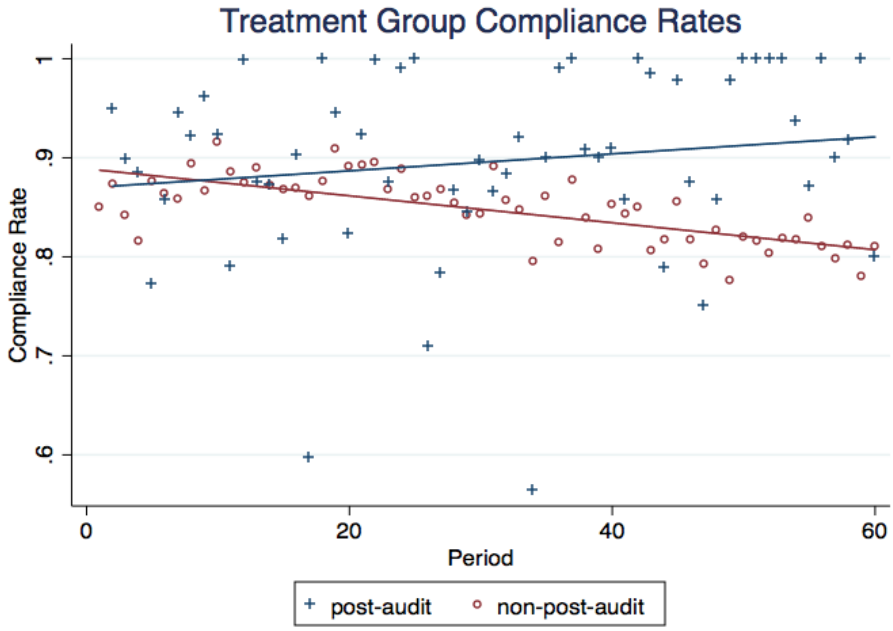
Figure 2 shows that, for all periods in which there were both post-audit and non-post-audit observations, average compliance is lower in a period after an audit than in a period not after an audit. This result is consistent with the presence of a bomb crater effect—on a period-by-period basis, average compliance after an audit is lower than average compliance not after an audit. The downward slope of the trend lines for both groups is similar to that seen in Figure 1, as expected.

Here, it is hard to miss the fact that post-audit average compliance rates display a large variance around the fitted line. In approximately eight periods, post-audit compliance substantially exceeded non-post-audit compliance—indicating that there was no bomb crater effect in these periods, and audits were "working" to foster more honest reporting. Interestingly, all of these instances (in which the bomb crater effect is *not* present) occur during the first half of the experimental tax lives of the participant. This result is broadly consistent with a "learning" story in which taxpayers over

time get a sense of audits as not being conditioned on past underreporting.

Figure 3 below shows the same scatter plot for the treatment group. In this depiction, the position of the lines reverses after about period ten, where post-audit compliance on average by period exceeds non-post-audit compliance. This pattern supports the idea that the EAR may reverse the bomb crater effect.

Figure 3:



Again, however, there is significant variance in the post-audit observations: in approximately the first half of the experiment, there are again eight instances of a reversed trend that is consistent with the bomb crater effect, but in many other periods, average compliance approaches or even reaches one (perfect compliance) following an audit. Such lofty averages are never reached in any period for the non-post-audit group.

2. *Regression Analysis*

Regressions can reveal more about what may be driving the visual trends offered above. A linear regression model was used. The dependent variable is the compliance rate in period t for participant i. In the first specification, the independent variables were created by a period's "post-audit" status (zero, if non-post-audit; one, if post-audit) interact with the

participant's assignment to the experimental condition (either treatment or control). Thus, there are four independent variables: control*non-post-audit, control*post-audit, treatment*non-post-audit, and treatment*post-audit.

In the specifications in the next two columns of Table 2, worker fixed effects and period fixed effects were added.[101]

Table 2: Regression Results with and Without Fixed Effects

	Compliance Rate (baseline)	Compliance Rate (w/ worker fixed effects)	Compliance Rate (w/ period fixed effects)
Treatment (non-post-audit)	0.141***	-	0.142***
	(0.00713)	-	(0.0421)
Control x post-audit	-0.0831***	-0.0812***	-0.0813***
	(0.0182)	(0.0187)	(0.0249)
Treatment x post-audit	0.129***	0.130***	0.120***
	(0.0222)	(0.0285)	(0.0320)
Control (non-post-audit)	0.706***	0.772***	0.706***
	(0.00547)	(0.00138)	(0.0339)
Fixed effects	none	worker	period
Observations	11,940	11,940	11,940

Standard errors in parentheses
* p<0.10 ** p<0.05 *** p<0.01

101. Worker fixed effects isolate variation of compliance rates for a given worker. The table shows the significance only of the effects of being post-audit (as compared to non-post-audit) for treatment and control separately. This is because workers can only be in one category—this regression prevents the pooling of observations across workers. By contrast, period fixed effects control for the period in which the compliance is observed. Adding period fixed effects allows us to check whether the timing within the experimental lifecycle of the taxpayer matters to the trends in compliance that we are observing.

The first column (without fixed effects) shows that each of the four independent variables is highly statistically significant. To calculate the marginal effects, I progressively add the coefficients together (see next table).

The columns with worker and period fixed effects confirm that even when I control for idiosyncratic variation across workers or, respectively, control for the period in which compliance was observed, the effects of the independent variables are still highly statistically significant.

To concretize the magnitude of the bomb crater for the control group and the direct deterrence effect (the reverse of the bomb crater) increase in compliance after an audit for the treatment group, the following table aggregates the marginal effects listed in the regression above.

Table 3: Quantifying the Size of the Bomb Crater

	Marginal Effects (from regression table)	Overall Compliance Rate (%)	
Control, not post-audit	0.701***	70%	Random audit "bomb crater" effect: -8% points
Control, post-audit	-0.081***	62%	
Treatment, not post-audit	0.153***	85%	Endogenous audit direct deterrence effect: +3% points
Treatment, post audit	0.120***	88%***	
***p<0.01			

The summary table highlights not only the magnitude of the bomb crater and deterrent effects but also that these effects are statistically significant—they matter even after accounting for the variance in the compliance rates that was observed in Figure 1 through Figure 3.

What else might be driving the compliance rate patterns we observe in response to random (endogenous) audits in the control (treatment) group? The results of alternative regression specifications appear in Appendix B. Table B2 adds basic demographic controls. Only sex, with being male associated with lower compliance levels, achieves significance at the one percent level. This relationship between sex and compliance is consistent

with prior face-to-face experimental studies using undergraduates (e.g., a participant sample other than MTurk participants), suggesting that higher female compliance may be an experimental regularity.[102] The inclusion of demographic controls does not meaningfully affect the size or the significance of the bomb crater effect (in the control condition) and the deterrent effect (in the treatment condition).

D. Repeat Audit Analysis

As mentioned following the discussion of prior dynamic tax compliance studies above, one of the key issues raised is how repeat audits affect compliance. The prior studies show that repeat audits can reverse the bomb crater effect, and my results above are consistent with those findings. But one of the noted shortcomings of the prior studies is that they do not separately estimate the marginal effects on compliance of an additional audit. This is because the patterns of audits were pre-specified: there were audits, for instance, in periods three, seven, and ten for all participants in a given experimental condition. As a result, it was harder to study how participants' reactions might be affected by audit sequences of varying lengths and at varying times in a taxpayer's lifecycle.

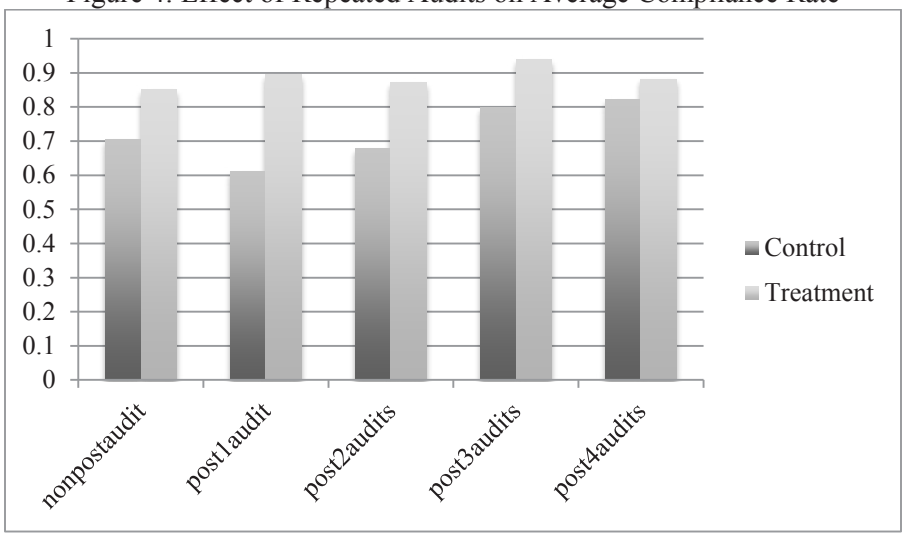
Here, none of the audit patterns are pre-specified. Even the control group does not have a "set" pattern of audits: a random number generator determines when a participant draws an audit. And in the treatment group, the initial audit is random and then it is fully endogenous to the participant's compliance choices. It would be an extremely low-probability event if two audit patterns for two participants were identical. On the other hand, both the control and treatment experimental conditions will contain repeated audits. For the treatment experimental condition, the EAR ensures repeated audits for taxpayers found to be noncompliant upon random audit. For the control experimental condition, applying random audits at ten percent over 60 periods for approximately 100 participants ensures that there will be a number of instances in which audits are performed back-to-back.

These "random" repeat audits are essential, as they facilitate the comparison between control and treatment of the performance of repeat audits of varying lengths.

---

102. *See* Spicer & Hero, *supra* note 6, at 265 ("Gender, however, did have a significant effect: men tended to evade more taxes than women. This is consistent with the findings of Spicer and Becker (1980).") (citation omitted).

Figure 4: Effect of Repeated Audits on Average Compliance Rate



Nonpostaudit = no audit in prior period
post1audit = audit in prior period only
post2audits = audit in prior two periods only
post3audits = audit in prior three periods only
post4audits = audit in prior four periods only

1. Description

Figure 4 above shows the effect on compliance of having different durations of audit sequences for each experimental condition (note that the unlabeled Y axis is the average compliance rate, similar to the prior figures above). Here, all audits of fewer than five consecutive periods were categorized: "post1audit" refers to the average compliance rate in periods after a single audit—it excludes any post-audit periods where there have been two audits in a row. (This means that there could have been an audit two or three periods earlier, a period or two periods of no audit, and then the single audit). "Post2audit" refers to average compliance in periods after there have been two consecutive audits but not more than two; "post3audit" refers to average compliance in periods after there have been three consecutive audits but not more than three. These categories achieve a specificity that the "postaudit" variable used above cannot: "postaudit" sweeps in all of the categories in the bar chart (and includes also post5audit and post6audit, which are not listed).

2. Interpretation

The first set of bars (nonpostaudit) shows the overall average compliance rate in each experimental condition when there is no audit in the prior period; this is the "secular" average rate of voluntary compliance across all periods in the absence of an immediately preceding audit. The difference between the nonpostaudit control bar (lighter in grayscale) and treatment bar (darker in grayscale) bars is another way of measuring the "presentation effect" of the EAR (e.g., the additional line in the instructions disclosing that evasion will flag a participant for future audits).

Focusing solely on the control condition (blue/lighter bars), one can see the bomb crater effect after one random audit: the compliance rate drops by almost ten percentage points as compared to the non-post-audit level. Where two random audits occur back-to-back, however, the bomb crater effect reverses, and compliance increases by approximately seven percentage points, almost reaching the non-post-audit level. After three back-to-back random audits, compliance increases dramatically (12 percentage points from level after two consecutive audits). After four repeated random audits, it increases but seems to start plateauing (about two percentage points).[103]

The trajectory of the treatment condition (red/darker bars) is much different. There is no bomb crater effect after one audit (post1audit as compared to nonpostaudit). To the contrary, compliance rises by five percentage points, consistent with the story of participants being responsive to the disclosure about audit endogeneity (i.e., that they have now been flagged if they were found to have evaded in the prior period's audit). Interestingly, after two audits, compliance drops slightly (approximately two percentage points): one interpretation of this is that participants believe that the second audit in a row concludes their "flagging," and they are unlikely to face another audit. However, after the third consecutive audit (at which point they may be disabused of this notion), compliance shows the biggest increase: almost seven percentage points. This is followed by a small decrease after the fourth consecutive audit. It is almost as if the participants in the treatment condition alternate between behaving as if they know they are flagged for audit in the next period (and increase compliance accordingly) and betting that the taxing agency must have tired of repeatedly auditing them.

The takeaway from this repeat-auditing breakdown is that the "payoff" in terms of increased compliance of an EAR similar to that used in the

103. This U-shaped pattern of compliance is consistent with the post-audit behavior of U.S. corporations that was found using Service panel data. *See* DeBacker et al., *supra* note 54.

treatment condition (repeat audits of evaders up to a cap of six audits) may be lower than justifies the resource cost of the additional audits. Average compliance does not increase after two audits in a row; it does after three but decreases after four.

What does this imply for a tax administrator that is looking to these experimental results to inform real-world audit strategies? Setting aside external validity concerns, the answer is not clear. These experimental results may appear to justify endogenous triple audits (e.g., tax authority should audit evaders up to three times in a row before returning them to the random audit pool). However, this is misleading. The overall compliance outcomes of such a strategy need to be compared alongside others: endogenous double-audits, endogenous quadruple audits, and other sorts of endogenous audits, including a "cutoff" rule in which participants reporting lower than a certain absolute amount of income are automatically audited.

These results show that, in this experimental context, the presentation and sequencing of audits have a significant influence on post-audit compliance levels. Therefore, the problem of whether to audit a given taxpayer in a given period cannot be solved simply by weighing the cost of the audit against the additional tax and penalty revenue to be gained from the audit. Future dynamic audit research must help us understand the subsequent trajectory of compliance outcomes that is associated with different possible presentations and sequencings of audits (e.g., alternative EARs).

IV. CONCLUSION

Kim Bloomquist's recent computer science doctoral dissertation includes a review of the extant literature on audit dynamics.[104] He notes that research on the size and duration of any audit effects (increased or decreased compliance) has received little attention, and he concludes that "here is a situation where laboratory experiments may provide the only way to explore this topic in sufficient detail."[105] This Article takes that proposition seriously by combining the experimental study of dynamic audits in a new context (MTurk) with a more realistic treatment condition that features an EAR that, in its non-randomness, is a plausible analog of what a real-world taxing agency might use in response to detected evasion.

This Article provides strong evidence that the bomb crater effect in response to random audits is an experimental regularity. My U.S. results align with those found by researchers in Italy and Austria, even though the experiment was performed in a different setting (MTurk rather than a social

104. Bloomquist, *supra* note 33, at 43–48.

105. *Id.* at 44 (citation omitted).

science laboratory) and used different subjects (adult workers on MTurk rather than undergraduate students). Specifically, I find random audits are associated with a significant bomb crater effect of reduced compliance in the period immediately following the audit. However, a second consecutive audit can—quite dramatically—reverse and overcome this effect. I show that the second audit boosts compliance *above* the level from which it bomb-cratered (i.e., the non-post-audit level). If these experimental results were perfectly generalizable, these findings would suggest that a taxing agency that presented its audits as "random" should try to repeat audit as many taxpayers as possible after detecting evasion in the initial random audit. Of course, such a strategy would run the risk of becoming transparent to taxpayers over time: the "random" audit strategy would be executed (and, once understood, would be experienced) exactly like an EAR in which detected evasion triggered a second audit with certainty.

Thus, the comparison between the EAR specified in this experimental survey and the random audit rule is opportune. The comparison of the two reveals a strong effect of the presentation of the EAR: simply telling participants that audits are not random and that evasion will trigger one or more future audits is sufficient to completely erase the bomb crater effect after the first audit. However, the repeat-audit analysis is inconclusive as to the effects on compliance of adding another audit. Further research on other EARs is necessary, and an experimental setting is a good place to start.

I echo other researchers in voicing the hope that there will be room to move beyond the physical or virtual tax compliance laboratory or other stylized experimental settings. Field experiments have become increasingly common in other national contexts, and state-level field experiments in the U.S. have generated important insights.[106] Agency-level field experimentation holds enormous promise and represents the logical next step in improving our understanding of how taxpayers respond to various types of audits over time.

106. See discussion in Hallsworth, *supra* note 22. For examples, see Joel Slemrod et al., *Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota*, 79 J. PUB. ECON. 455 (2001); Dayanand S. Manoli & Nicholas Turner, *Nudges and Learning Effects from Informational Interventions: Evidence from Notifications for Low-Income Taxpayers* (NBER Working Paper No. 20718, 2015); Ricardo Perez-Truglia & Ugo Troiano, Shaming Tax Delinquents: Evidence from a Field Experiment in the United States (June 1, 2016) (unpublished manuscript), http://ssrn.com/abstract=2558115.

APPENDIX A: DISCUSSION OF MTURK SAMPLE

This Article is the first to use MTurk to study tax compliance behavior. This section seeks to address some of the questions that are likely to arise in connection with relying on MTurk data.[107] To address some of the documented disadvantages of MTurk, it outlines some of the experimental design measures taken to bolster the reliability of the data.

A. Advantages of MTurk: Checks Robustness to Subject Pool Effects

Using MTurk as an alternative to experiments run in a traditional university psychology or behavioral economics laboratory allows us to explore whether there is a "subject pool effect" at issue in dynamic audit experiments. As Alm et al. (2015) note, running an experiment on this different group of participants—workers on MTurk—and finding that neither the bomb crater nor the echo hypotheses could be rejected is one way to bolster the external validity of the findings. Conversely, if the results of prior studies could not be replicated using the new subject pool, external validity concerns might be stronger, or the hypotheses modified. Varying the subject pool in a particular experiment can be seen as a kind of robustness check of the experiment's results.

B. Disadvantages of MTurk: External Validity Issues

The second set of arguments concerns the external validity of MTurk results. Arguments that MTurk is better than traditional lab settings hinge on the higher "representativeness" of MTurkers and the greater "realism" of the MTurk setting. These arguments are more speculative. As I emphasize in the

107. For a discussion of the importance of addressing questions about sample validity when using MTurk for social science research, see Dan Kahan, *What's a "Valid" Sample: Problems with Mechanical Turk Study Samples, Part 1,* CULTURAL COGNITION PROJECT AT YALE LAW SCHOOL BLOG (July 8, 2013, 9:34 AM), http://www.culturalcognition.net/blog/2013/7/8/whats-a-valid-sample-problems-with-mechanical-turk-study-sam.html. Kahan argues for "the invalidity of studies that use samples of Mechanical Turk workers to test hypotheses about cognition and political conflict over societal risks and other policy-relevant facts." Dan Kahan, *Fooled Twice Shame on Who?: Problems with Mechanical Turk Study Samples, Part 2,* CULTURAL COGNITION PROJECT AT YALE LAW SCHOOL BLOG (July 10, 2013, 9:30 AM), http://www.culturalcognition.net/blog/2013/7/10/fooled-twice-shame-on-who-problems-with-mechanical-turk-stud.html. He has three main concerns to which I am preparing more robust responses for a future project.

introduction, I do not seek to intervene in the external validity debate; instead, I seek to add to experimental robustness of past work and make new experimental contributions. However, I briefly engage with these issues here.

1. Representativeness–Demographic Characteristics

Are U.S. resident MTurkers representative of the population of U.S. taxpayers? MTurk workers are certainly more varied in terms of age and tax-filing background than the typical sample of university undergraduates used in undergraduate social science laboratory experiments.[108] Other studies have found that the population of U.S. residents on MTurk has been shown to be comparable to other common survey populations.[109] The table on the following page compares demographic characteristics of this study's MTurk participants to those of other relevant populations.

108. Winter Mason & Siddharth Suri, *Conducting Behavioral Research on Amazon's Mechanical Turk*, 44 BEHAV. RES. METHODS 1 (2012).

109. Kuziemko et al., *supra* note 87, app. at 9 ("[R]espondents to our surveys are not representative of the U.S. population. However, this lack of representativeness does not appear substantially worse than in other surveys.").

Table A1: Comparing This Study's MTurk Sample to Other Populations

		Control Group (n=106)	Treatment Group (n=93)	IRS Stats (Individual Returns by Income, 2013)	American Community Survey 2014 (U.S. Census Bureau)	Current Population Survey 2015 (Bureau of Labor)	U.S. Census 2010
Basic Demographics	Median age	35.3	33.7		37.4		37.2
	U.S. born	96.1	97.8		86.9		0.871
	Has kids	27.8	18.4				
	Male	58.1	50.3		49.2		0.492
	Married	33.9	25.4				
Employment Status	Full-time employee	58.5	59.8				
	Not in labour force	4.72	3.26		36.1	37.3	
	Part-time employee	10.4	9.8				
	Self-employed or small business	20.8	17.4		6.1		
	Student	0.94	3.26				
	Unemployed and looking for work	4.72	6.52		5.8	3.3	
Education	Some high school	0	1.1		7.8	9.6 (less than diploma)	8.9
	High school degree/GED	18.9	18.5		28	24.6	31
	Some college	17	27.2		21.2	14.1	19.3
	2-year college Degree	14.2	9.8		7.9	8.3	8.6
	4-year college degree	41.5	33.7		18.3	17.6	18
	Master's degree	4.7	8.7		11 (grad. or professional)	10.4 (adv. degree)	6.7
	Professional degree	1.9	0				1.4
	Doctoral degree	1.9	1.1				1.2
Income	$0-$9,999	6.6	4.4	16.8			
	$10K-$14,999	3.8	5.4	8.5			
	$15K-$19,999	9.4	5.4	7.9			
	$20K-$29,999	17	12	12.9			
	$30K-$39,999	16.1	18.5	9.8			
	$40K-$49,999	16.1	15.2	7.7			
	$50K-$74,999	11.3	20.7	13			
	$75K-$99,999	10.4	12	8.5			
	$100K-$124,999	7.6	3.3	11 (100K-200K)			
	$125K-$149,999	0.94	1.1	N/A			
	$150K-$199,999	0.94	1.1	N/A			
	$200K+	0	1.1	0.7			

2. Realism

Is the MTurk setting a more "realistic" place to run tax compliance experiments than a laboratory setting? In some sense, yes, but if this aspect of realism itself makes MTurkers non-representative of the larger taxpaying population, "realism" may be a drawback that limits the external validity of the results.

On the positive side, MTurkers are, with high probability, engaged with taxation in a manner that is relevant to my research question. Generally, an MTurker's earnings from completing HITs will be characterized as earnings from self-employment, as they are not employees of Amazon or of HIT requesters.[110] As independent contractors in receipt of such earnings, MTurkers must assess their liability for income taxes (and self-employment taxes, among others).[111] Amazon makes this salient at the very first step in the process of becoming a MTurker: U.S. resident MTurkers are required to provide their taxpayer identification numbers before taking on any HITs.[112] Thus, the assumption of a baseline awareness of and engagement with the activity of tax compliance seems reasonable in the case of MTurkers and may reduce the artificiality of the experimental setting.

On the negative side, however, MTurkers may be a self-selecting subset of micro-entrepreneurs—if there is some correlation between the type of person who signs up to work on MTurk and behavioral characteristics relating to tax such as risk aversion, comfort with uncertainty, or other factors, then the results of the experiment may be limited to this particular subject pool. For instance, one might infer that MTurkers are more comfortable with tax complexity as compared to ordinary taxpayers, or even as compared to ordinary taxpayers who earn income as independent

110. Felstiner, *supra* note 86.

111. Depending, of course, on their country of residence. I, for instance, encounter a slightly different experience as a Canadian-resident MTurker than my U.S. colleagues (e.g., possible withholding and other reporting requirements), although the statement of earnings that I (should) receive from Amazon is identical.

112. Amazon also provides explicit information about tax issues on the MTurk site, and there is robust conversation on the MTurk Reddit forum concerning whether and how to report MTurk income. *See, e.g.*, coffeeturk, *It's 2015, Time to Start Thinking About Taxes*, REDDIT, http://www.reddit.com/r/mturk/comments/2r7pdr/its_2015_time_to_start_thinking_about_taxes/ (last visited Dec. 13, 2016). When registering to work as an MTurker, applicants are required to agree to the following: "I will comply with all applicable laws, including without limitation tax and filing requirements." *See User Registration*, AMAZON MECHANICAL TURK, https://www.mturk.com/mturk/checkregistration (requires sign-in using Amazon.com user information).

contractors. There are no studies examining MTurkers' representativeness as compared to students along the lines of Alm and Bloomquist.[113] And, indeed, there is some evidence that MTurk participants may be unrepresentative in the context of risk-taking situations.[114] For this reason, I position my Article as contributing to the experimental robustness of research on the bomb crater effect and stay away from interpreting my results as a step towards establishing external validity.

APPENDIX B: EMPIRICAL APPENDIX

	(1) Compliance Rate	(2) Compliance Rate	(3) Compliance Rate	(4) Compliance Rate	(5) Compliance Rate	(6) Compliance Rate
Treatment (non-post-audit	0.141***		0.142***	0.127***	0.128***	
	(0.0071)		(0.0421)	(0.0416)	(0.0416)	
Control * post-audit	-0.0831***	-0.0812***	-0.0813***	-0.0745***	-0.0729***	-0.0793***
	(0.0182)	(0.0187)	(0.0249)	(0.0237)	(0.0240)	(0.0188)
Treatment * post-audit	0.129***	0.130***	0.120***	0.113***	0.104***	0.120***
	(0.0222)	(0.0285)	(0.0320)	(0.0319)	(0.0318)	(0.0282)
Control (non-post-audit)	0.706***	0.772***	0.706***	1.181***	1.093***	0.757***
	(0.005)	(0.00138)	(0.034)	(0.206)	(0.129)	(0.002)
Observations	11,940	11,940	11,940	11,520	11,520	11,940
Worker demographic controls	No	No	No	Yes	Yes	No
Period fixed effects	No	No	Yes	No	Yes	Yes
Worker fixed effects	No	Yes	No	No	No	Yes

Standard errors in parentheses

* p<0.10 ** p<0.05 *** p<0.01

The first three columns in the above table match those reported in Table 2. Columns (4), (5) and (6) add additional combinations of worker demographic controls, period fixed effects and worker fixed effects.

113. Alm et al., *supra* note 77.

114. Yanna Krupnikov & Adam Seth Levine, *Cross-Sample Comparisons and External Validity*, 1 J. EXPERIMENTAL POL'Y SCI. 59 (2014).

Table B2: Regression Results with Basic Demographic Controls

	Compliance Rate (w/ demographic controls)
Treatment (not post-audit)	0.145***
	(0.0418)
Post-audit (control group)	-0.0768***
	(0.0246)
Treatment x post-audit	0.107***
	(0.0318)
U.S. born	0.0692
	(0.1100)
Kids	0.0146
	(0.0594)
Male	-0.118***
	(0.0415)
Married	0.0009
	(0.0506)
Constant (control group, not post-audit)	0.699***
	(0.1140)
Fixed effects	none
Observations	11,880

Standard errors in parenthesis
* p<0.10

(Note that three stars indicate statistical significance of the independent variable listed at the one percent confidence level.)

Here, the only demographic characteristic that reached significance was sex: on average, male participants were significantly less compliant than females. Even after including these controls, the effects of the treatment experimental condition and being post-audit in either experimental condition were still highly significant.