

## BYU Law Review

---

Volume 2018 | Issue 3

Article 6

---

Winter 12-31-2018

# Does Enforcement Reduce Voluntary Tax Compliance?

Leandra Lederman

Follow this and additional works at: <https://digitalcommons.law.byu.edu/lawreview>

Part of the [Tax Law Commons](#)

---

### Recommended Citation

Leandra Lederman, *Does Enforcement Reduce Voluntary Tax Compliance?*, 2018 BYU L. Rev. 623 (2019).

Available at: <https://digitalcommons.law.byu.edu/lawreview/vol2018/iss3/6>

This Article is brought to you for free and open access by the Brigham Young University Law Review at BYU Law Digital Commons. It has been accepted for inclusion in BYU Law Review by an authorized editor of BYU Law Digital Commons. For more information, please contact [hunterlawlibrary@byu.edu](mailto:hunterlawlibrary@byu.edu).

# Does Enforcement Reduce Voluntary Tax Compliance?

*Leandra Lederman\**

*Governments generally use enforcement methods, such as audits and the imposition of penalties, to deter noncompliance with tax laws. Although this approach is consistent with most economic modeling of tax compliance, some scholars caution that enforcement may backfire, “crowding out” taxpayers’ intrinsic motivations to pay taxes to such an extent that they reduce their tax payments. This Article analyzes the existing evidence to determine if this occurs. In fact, field studies suggest that enforcement tools, such as audits, are very effective deterrents. A few recent studies have found that audits have a negative effect on the subsequent tax payments of those found compliant on audit. This outcome, while perhaps initially surprising, is consistent with the deterrence model; a favorable outcome after audit may lower that taxpayer’s perceived likelihood of subsequent audit and the perceived magnitude of any sanction.*

---

\* William W. Oliver Professor of Tax Law, Indiana University Maurer School of Law, Bloomington. The author thanks Arun Advani, Matthias Kasper, Sarah Lawsky, Rebecca Morrow, Shu-Yi Oei, Jim Repetti, Diane Ring, and Emily Satterthwaite for comments on earlier drafts. She is very grateful to Emily Satterthwaite for sharing analyses of her original data. She is also grateful for helpful discussions with Michael Alexeev, Bill Black, Devin Blankenship, Brian Broughman, Nickolas Cole, Wendy Epstein, Adam Feibelman, Judith Freedman, Bruno Frey, Dawn Johnsen, Marjorie Kornhauser, Austen Parrish, Eric Rasmusen, Steve Sheffrin, Holger Spamann, Jeff Stake, Indiana University’s Law & Economics Lunch Bunch, and attendees at presentations at the Boston College Tax Policy Workshop, the Critical Tax Conference, the University of Florida’s Tax Policy Colloquium, the Maurer Law School’s Summer Faculty Workshop, the National Tax Association’s Annual Conference on Taxation, Northwestern Law School’s Colloquium on Advanced Topics in Taxation, the Oxford University Centre for Business Taxation, the Tulane Law School Workshop on Regulation and Coordination, and the Vienna University of Economics and Business. Nicole Alonzo, Sean Hamner, Rustie Mills, James Ruley, and Mike Tenenboym provided valuable research assistance.

## CONTENTS

INTRODUCTION .....	624
I. CROWDING OUT IN THEORY AND PRACTICE .....	626
A. Rewards .....	627
B. Punishment .....	633
II. DETERRENCE AS A TOOL FOR FOSTERING TAX COMPLIANCE .....	642
A. The False Puzzle and “Tax Morale” .....	643
1. The importance of opportunity to evade .....	647
2. Other factors affecting tax compliance .....	650
B. The Effectiveness of Particular Deterrence Tools .....	655
1. Audits and audit threats .....	655
2. Penalties .....	662
III. POST-AUDIT REDUCTION OF COMPLIANT BEHAVIOR .....	667
A. The “Bomb-Crater Effect” in Laboratory Experiments .....	667
B. Field Experiments on Post-Audit Compliance .....	677
CONCLUSION .....	692

## INTRODUCTION

The United States (U.S.) federal income tax requires taxpayers to self-report income on a tax return, which provides an opportunity for cheating.<sup>1</sup> The Internal Revenue Service (IRS) generally operates from the intuitive perspective that enforcing the tax laws—by auditing a percentage of taxpayers and imposing penalties where appropriate—fosters compliance.<sup>2</sup> However, a number of scholars have argued that tax enforcement may “crowd

---

1. Tax noncompliance includes unintentional mistakes, legal tax avoidance, and intentional evasion. The Internal Revenue Service distinguishes between filing, reporting, and payment noncompliance. *Understanding the Tax Gap*, I.R.S. Fact Sheet FS-2005-14 (Mar. 2005), <https://www.irs.gov/newsroom/understanding-the-tax-gap>. This Article focuses on tax evasion and generally on intentional underreporting, and thus underpayment, of taxes due.

2. *IRS Audits*, IRS, <https://www.irs.gov/businesses/small-businesses-self-employed/irs-audits> (stating that the purpose of audits is to “to verify the reported amount of tax is correct”) (last updated Aug. 23, 2018). Various things may trigger an audit, including indicators of noncompliance, audits of related taxpayers, and (less often) random selection. *See id.*

out” intrinsic motivations to comply with the tax laws.<sup>3</sup> For example, Benno Torgler has stated, “When monitoring and penalties for noncompliance are intensified, individuals notice that extrinsic motivation has increased, which . . . crowds out their intrinsic motivation to comply with taxes.”<sup>4</sup> Some of these scholars argue that enforcement ultimately will have the perverse effect of *reducing* voluntary tax compliance.<sup>5</sup>

Because enforcement resources are scarce, with tax collectors such as the IRS quite underfunded,<sup>6</sup> it would be valuable to know whether decreasing enforcement could increase tax compliance, even in particular situations. Accordingly, this Article analyzes the empirical literature on deterrence and crowding out, in order to determine if there are contexts in which enforcement may backfire.

As there is some evidence in non-tax contexts that external motivators may displace internal ones, Part I of the Article examines that literature so as to gain greater understanding of the

3. See, e.g., Michael Doran, *Tax Penalties and Tax Compliance*, 46 HARV. J. ON LEGIS. 111, 135 (2009) (“[H]arsh tax penalties . . . should not be aimed at taxpayers who respond to tax-compliance norms because the harsh penalties will undercut the norms and crowd out compliance . . .”); Lars P. Feld & Bruno S. Frey, *Tax Compliance as the Result of a Psychological Tax Contract: The Role of Incentives and Responsive Regulation*, 29 LAW & POL’Y 102, 105 (2007); Dan M. Kahan, *Trust, Collective Action, and Law*, 81 B.U. L. REV. 333, 338 (2001) (“Material incentives can also suppress reciprocal cooperation by *crowding out* moral motivations.”); Yair Listokin & David M. Schizer, *I Like to Pay Taxes: Taxpayer Support for Government Spending and the Efficiency of the Tax System*, 66 TAX L. REV. 179, 203 (2013) (“[A] range of sanctions can be explored for noncompliance, including not just cash payments, which are the most likely to crowd out pro-social motivations and may even signal that noncompliance is acceptable so long as the non-complier pays for the privilege . . .”).

4. Benno Torgler, *Moral Suasion: An Alternative Tax Policy Strategy? Evidence from a Controlled Field Experiment in Switzerland*, 5 ECON. GOV. 235, 236 (2004).

5. See Dan M. Kahan, *The Logic of Reciprocity: Trust, Collective Action, and Law*, 102 MICH. L. REV. 71, 83 (2003) (“[E]vidence suggests that when taxpayers are exposed to information highlighting the penalties for evasion, they respond . . . by contributing *less*.”); cf. Marjorie E. Kornhauser, *Normative and Cognitive Aspects of Tax Compliance: Literature Review and Recommendations for the IRS Regarding Individual Taxpayers*, in 2 NATIONAL TAXPAYER ADVOCATE 2007 ANNUAL REPORT TO CONGRESS 138, 151 (“The commoditization of a behavior crowds out positive normative influences on that behavior. Thus, setting a price or giving an economic incentive for behavior motivated by social, non-pecuniary motives such as reciprocity can actually decrease the desired behavior.”).

6. See, e.g., Leandra Lederman, *The IRS, Politics, and Income Inequality*, 150 TAX NOTES 1329, 1330–32 (2016) (showing IRS budget cuts and discussing their impact); *General Report of the Internal Revenue Service Advisory Council*, IRS (2014), <https://www.irs.gov/tax-professionals/2014-irsac-general-report> (“In IRSAC’s view, the IRS is in the midst of an existential funding crisis.”).

crowding-out phenomenon. Part II turns to the tax context, analyzing the empirical literature on the effects of sanctions on tax compliance, which generally finds that enforcement techniques, such as audit threats, increase compliance. Part III focuses on the extent to which crowding out occurs in the tax-compliance context, discussing what experimental taxpaying games and field experiments involving actual taxpayers have shown.

The Article concludes that enforcement generally has a strong, positive effect on tax compliance and that audits are a very productive tool for a tax collector such as the IRS. While there is some evidence that audits may result in reduced tax payments by self-employed taxpayers who, on audit, are found not to owe anything, that result is consistent with the deterrence model.

### I. CROWDING OUT IN THEORY AND PRACTICE

A basic economic approach suggests that the more you pay for something the more of it you receive—a price effect.<sup>7</sup> Conversely, penalties should reduce the penalized behavior. Crowding-out theory predicts the opposite result—paying (or paying more) for something will reduce the amount received, and penalizing it will increase it.<sup>8</sup> Although the crowding-out prediction is counterintuitive, scholars have made it in several contexts, as discussed below.<sup>9</sup>

Most of the studies testing crowding-out effects involve rewards rather than sanctions.<sup>10</sup> In an economic sense, rewards and sanctions are variations on the same theme, with a fine or other sanction simply a negative reward.<sup>11</sup> However, they are not

---

7. See Bruno S. Frey & Felix Oberholzer-Gee, *The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out*, 87 AM. ECON. REV. 746, 746 (1997).

8. Marianne Premberger & Theresa M. Marteau, *When Do Financial Incentives Reduce Intrinsic Motivation? Comparing Behaviors Studied in Psychological and Economic Literatures*, 32 HEALTH PSYCHOL. 950, 953 (2013) (“The definition of ‘motivation crowding out’ in economic research includes any effect that is opposite to the relative price effect of standard economic theory: rewards decreasing behavior levels, and penalties increasing behavior levels.”).

9. See *infra* Part I.A–B.

10. Tor Helge Holmås et al., *Does Monetary Punishment Crowd Out Pro-Social Motivation? A Natural Experiment on Hospital Length of Stay*, 75 J. ECON. BEHAV. & ORG. 261, 262 (2010).

11. See Brian Galle, *The Tragedy of the Carrots: Economics and Politics in the Choice of Price Instruments*, 64 STAN. L. REV. 797, 808 (2012) (“Society can use either carrots or sticks

actually identical. That is, a reward transfers wealth in one direction and a fine in the other.<sup>12</sup> In addition, people's psychological responses to rewards and sanctions are often different.<sup>13</sup> Although the tax-enforcement context typically focuses on fines rather than rewards, it is helpful to start with rewards because that literature is more developed. First, this Part briefly discusses rewards. It then turns to sanctions.

### A. Rewards

There are three strands of literature outside of tax that are highly relevant to the hypothesis that rewards can reduce intrinsic motivation. The first strand of literature involves blood donation, the second is the psychology literature on intrinsic motivation for performing tasks, and the third is the economics literature on "motivation crowding out." With respect to blood donation, Richard Titmuss's 1970 book, *The Gift Relationship*, argued that (1) paying for blood donations decreases the donations' quality because payment creates incentives for those with infectious diseases or drug addictions to conceal those conditions and (2) payment is economically inefficient.<sup>14</sup> Titmuss's inefficiency argument was that there is more wasted blood and higher administrative costs when donors are compensated.<sup>15</sup>

---

interchangeably to get externality producers to 'internalize' the marginal effects of their decisions on others. . . . These are equivalent . . . from the perspective of my single decision . . . because of opportunity costs.").

12. See *id.* at 801 ("Relative to present policy, a carrot transfers wealth from taxpayers to its recipients, while sticks have the opposite effect.").

13. *Id.* at 808-09 (citing JONATHAN BARON, *THINKING AND DECIDING* 269-70 (4th ed. 2007)).

14. See RICHARD M. TITMUSS, *THE GIFT RELATIONSHIP* 246 (1970) (referring to paid donations as resulting in wasted blood). Titmuss based his blood quality hypothesis on reports from American doctors to that effect, *id.* at 112-15, 117-18, contrasting that with the United Kingdom (U.K.), where virtually all blood donation was voluntary, *id.* at 89.

15. *Id.* at 246 (stating that the American system "is five to fifteen times more costly than the voluntary system in Britain"); see also Claudia Niza et al., *Incentivizing Blood Donation: Systematic Review and Meta-Analysis to Test Titmuss' Hypotheses*, 32 *HEALTH PSYCHOL.* 941, 941 (2013).

Titmuss's assertions were controversial<sup>16</sup> and heavily studied empirically.<sup>17</sup> Studies generally found that compensation had no effect on the quantity of blood donated<sup>18</sup> or that payment increased donation,<sup>19</sup> which is a price effect. The principal exception is that two studies found evidence of a crowding-out effect among female participants who were offered a small cash reward.<sup>20</sup>

Independent of Titmuss's work, a psychology literature developed hypothesizing that rewarding a behavior can undermine an individual's intrinsic motivation to continue that behavior once the reward is removed.<sup>21</sup> A classic example involves an experiment in which nursery school children who were promised and rewarded with an attractive certificate for drawing with markers in a first session spent less time drawing and drew less well in a second, unrewarded session than children who had not received a

---

16. Niza et al., *supra* note 15, at 941 (“[Titmuss’s] arguments raised a heated discussion, particularly among economists, with criticism of Titmuss for his narrow view of market forces and lack of empirical support.”).

17. Several studies used meta-analysis, which allows data from studies on the same topic to be combined, effectively increasing the sample size. See Judy Cameron et al., *Pervasive Negative Effects of Rewards on Intrinsic Motivation: The Myth Continues*, 24 BEHAV. ANALYST, Spring 2001, at 1, 5.

18. *Id.*

19. See Nicola Lacetera et al., *Economic Rewards to Motivate Blood Donations*, 340 SCI. MAG. 927, 927 (2013) (“Observational studies that control for confounding factors have examined 14 incentive items ranging from small coupons to a paid day off work. All were found to increase blood donations.”); Christian Weidmann et al., *Monetary Compensation and Blood Donor Return: Results of a Donor Survey in Southwest Germany*, 41 TRANSFUSION MED. & HEMOTHERAPY 257, 257–58, 260 (2014) (finding in Southwest Germany that compensation increased the likelihood a donor would return during the first and second year after the first donation but had no statistically significant effect in the third year).

20. See Nicola Lacetera & Mario Macis, *Do All Material Incentives for Pro-Social Activities Backfire? The Response to Cash and Non-Cash Incentives for Blood Donations*, 31 J. ECON. PSYCHOL. 738, 740 (2010) (“[F]emale donors are almost twice as likely to declare that they would stop donating or donate less often [when offered cash] compared to male donors; the gender difference is much smaller for the in-kind payment, however.”); Carl Mellstrom & Magnus Johannesson, *Crowding Out in Blood Donation: Was Titmuss Right?*, 6 J. EUR. ECON. ASS’N. 845, 847 (2008) (Swedish study finding that a \$7 payment reduced the percentage of female subjects donating blood from 52% to 30%).

21. See Promberger & Marteau, *supra* note 8, at 950 (citing, as the start of that literature, Edward L. Deci, *Effects of Externally Mediated Rewards on Intrinsic Motivation*, 18 J. PERSONALITY & SOC. PSYCHOL. 105 (1971)); see also Bruno S. Frey & Reto Jegen, *Motivation Crowding Theory*, 15 J. ECON. SURVEYS 589, 589 (2001) (“The basic idea that rewards . . . may crowd out intrinsic motivation emanates from two quite different branches of literature in the social sciences.”).

certificate in the first session.<sup>22</sup> One way to look at this is that compensation reframes something that used to be fun as work, leading the recipient to expect compensation for it in the future.<sup>23</sup>

The idea that rewards might reduce desirable behavior was controversial in this context, too.<sup>24</sup> A large literature developed that tested when rewards might have an undesirable effect.<sup>25</sup> Competing meta-analyses reached different results with respect to whether a tangible reward for doing an interesting activity undermined or increased motivation.<sup>26</sup> However, they generally found that a tangible, expected reward given simply for performing a particular task – of which the certificate given to the children is an example – had a negative effect on intrinsic motivation.<sup>27</sup>

In 1997, Bruno Frey, citing the psychology literature, published an article that launched the economics literature on “motivation crowding out.”<sup>28</sup> In that article, Frey argued:

When a work activity is supported by both high work morale and external intervention, a ‘psychologically’ unstable situation

22. Mark R. Lepper & David Greene, *Undermining Children’s Intrinsic Interest with Extrinsic Reward: A Test of the “Overjustification” Hypothesis*, 28 J. PERSONALITY & SOC. PSYCHOL. 129, 133–34 (1973).

23. DOUGLAS G. MOOK, CLASSIC EXPERIMENTS IN PSYCHOLOGY 172–74 (2004) (in chapter 28, “Lepper et al. on the Costs of a Reward”).

24. Edward L. Deci et al., *A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation*, 125 PSYCHOL. BULL. 627, 627 (1999) (“Although the ever-expanding field of research that began with exploration of the effects of extrinsic rewards on intrinsic motivation has moved in these numerous directions, the original finding of the undermining of intrinsic motivation by tangible extrinsic rewards has continued to be the focus of considerable controversy.”).

25. *Id.* at 627–28 (examining “motivational theories, attributional theories, and behavioral or cognitive-behavioral theories”).

26. *Id.* at 631–32 (describing four meta-analyses conducted by 1995); see also Cameron et al., *supra* note 17 (2001 meta-analysis). See *supra* note 17 for a brief description of meta-analysis.

27. See Cameron et al., *supra* note 17, at 15–16, 21 (finding small but statistically significant negative effect both for high-interest tasks and for tangible rewards overall); Judy Cameron & W. David Pierce, *Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis*, 64 REV. EDUC. RES. 363, 394 (1994) (finding the negative effect only for a tangible, expected reward given for the act of performing the task); Deci et al., *supra* note 24, at 631–32 (citing Rummel and Feinberg (1988); Wiersma (1992); and Tang & Hall (1995), all finding that rewards generally undermined intrinsic motivation); *id.* at 644 (finding that performance-contingent rewards have a significantly negative effect on intrinsic motivation but not on self-reported interest in the task).

28. Bruno S. Frey, *On the Relationship Between Intrinsic and Extrinsic Work Motivation*, 15 INT’L J. INDUS. ORG. 427, 429 (1997).



arises. The agent is 'over motivated' as she would do the work even if one (or both) motivations were reduced. A rational actor responds by reducing that motivation which is under her control, i.e. she lowers her intrinsic work motivation. Intrinsic motivation is partially or totally substituted by externally controlled extrinsic work motivation.<sup>29</sup>

Frey argued that two conditions are required for motivation crowding out to occur: (1) the worker must be sufficiently intrinsically motivated at the outset and (2) conditions perceived as *controlling the worker's behavior* must be present.<sup>30</sup> Frey further argued that although rewards can crowd out intrinsic motivation, particularly where the reward is contingent on performance, workers "perceive rewards as less restrictive to their self-determination than commands, which are felt to intrude directly into the agents' realm of self-determination."<sup>31</sup>

Harvey James developed a model in which when a reward was sufficiently large, it overpowered the internal motivation of an intrinsically motivated individual by becoming so salient as to be perceived as a control mechanism.<sup>32</sup> Frey summarized that "[e]xternal intervention *crowds out* intrinsic motivation if the individuals affected perceive the intervening individuals to be controlling. . . . External intervention *crowds in* intrinsic motivation if the individuals concerned perceive it as supportive (or informative in a positive way)."<sup>33</sup> Other scholars have also found crowding out of extrinsic motivation in the work context.<sup>34</sup> Frey

---

29. *Id.* at 430 (single quotation marks in the original).

30. *Id.* at 431-32.

31. *Id.* at 432.

32. Harvey S. James, Jr., *Why Did You Do That? An Economic Examination of the Effect of Extrinsic Compensation on Intrinsic Motivation and Performance*, 26 J. ECON. PSYCHOL. 549, 563 (2005) ("If total compensation is too large, then the salience of the extrinsic reward might be so overwhelming that the agent is rationally compelled to perceive the compensation as a mechanism of control, thus resulting in MCO [motivation crowding out].").

33. Bruno Frey, *Crowding Out and Crowding In of Intrinsic Preferences*, in REFLEXIVE GOVERNANCE FOR GLOBAL PUBLIC GOODS 78 (Eric Brousseau et al. eds., 2012) (emphasis altered).

34. See Ernst Fehr & Simon Gächter, *Fairness and Retaliation: The Economics of Reciprocity*, 14 J. ECON. PERSP. 159, 168 (2000); Uri Gneezy & Aldo Rustichini, *Pay Enough or Don't Pay at All*, 115 Q. J. ECON. 791, 793-94 (2000) (finding that college students got more questions correct when offered higher pay per correct answer, but students that paid a flat

and his coauthors have found a negative effect of rewards, consistent with crowding-out theory, in other contexts, too.<sup>35</sup> For example, offering Swiss residents compensation to have a nuclear waste repository near them decreased the percentage who were willing to have the facility there.<sup>36</sup>

These literatures may be applicable in other contexts, as well. For example, Marianne Promberger and Theresa M. Marteau examined the experiments in the psychology and economic literatures to determine when they might “predict[] financial incentives to crowd out motivation for health-related behaviors.”<sup>37</sup> In so doing, they reported the main findings of the psychology literature as follows: “Tangible, expected rewards reduce behavior for interesting tasks (where behavior is initially high). No reduced persistence in task for previously dull tasks (where little time is initially spent on task).”<sup>38</sup> They also noted that the economics literature finds that increasing the *size* of an incentive usually has a price effect, meaning that a greater reward produces more of the rewarded activity.<sup>39</sup>

With respect to tax, at least one scholar has proposed rewarding tax compliance.<sup>40</sup> However, tax authorities generally do not use rewards for compliance, focusing instead on sanctions for noncompliance. Some countries use lotteries to foster tax

fee for participating got more questions correct than the students offered the lowest fee per correct answer).

35. Iris Bohnet et al., *More Order with Less Law: On Contract Enforcement, Trust, and Crowding*, 95 AM. POL. SCI. REV. 131, 132 (2001) (paper coauthored with Frey, finding in a model of contract formation that trust is crowded in with weak enforcement and crowded out with moderate enforcement).

36. Frey & Oberholzer-Gee, *supra* note 7, at 749 (finding that 50.8% of those surveyed agreed when not offered compensation but that only 24.6% agreed once offered compensation). The significance level is not reported. *See id.*

37. Promberger & Marteau, *supra* note 8, at 950.

38. *Id.* at 951 tbl.1. They also noted that the psychology literature looks at tasks the subject is intrinsically motivated to do, gives one group a reward for doing the task, and compares subjects' persistence at the task after removal of the reward. *Id.*

39. *Id.*

40. *See* Joshua D. Rosenberg, *The Psychology of Taxes: Why They Drive Us Crazy, and How We Can Make Them Sane*, 16 VA. TAX REV. 155, 220–21 (1996) (discussing the benefits of reducing tax liability as a reward for disclosing transactions that would be difficult to deter otherwise); *cf.* Sarah B. Lawsky, *Fairly Random: On Compensating Audited Taxpayers*, 41 CONN. L. REV. 161 (2008) (discussing the possibility of compensating individuals subject to random audits).

compliance, but the lottery participants are consumers who send in their receipts, while the target of the enforcement is merchants.<sup>41</sup> There are few studies of what effects rewards would have in the context of tax compliance, which presumably is not a “high-interest task.” However, a recent field study in Germany included a reward intervention.<sup>42</sup>

The German study, by Dwenger et al., involves a small church tax. The authors explain that “the church tax is compulsory and noncompliance represents a violation of tax law, but the church highlights the good cause and encourages overpayments that are defined as donations.”<sup>43</sup> Moreover, the base for the tax is the taxpayer’s taxable income *as reported* to the government, making compliance easily verifiable, although the law was not enforced prior to the study.<sup>44</sup>

Dwenger et al. found that, before their intervention, 20.9% of the 40,000 Protestant-church-member taxpayers in their sample had paid at least the full amount of tax due—with 55.5% of those who paid paying precisely the amount due and the remaining 79.1% evading, with 91.9% of the evaders paying nothing.<sup>45</sup> Dwenger et al. then assigned these taxpayers randomly to one of twelve treatments or the control group.<sup>46</sup> Unlike the studies discussed above, where the reward was offered to 100% of those eligible, the reward in the church tax study was being entered into

---

41. See Nirmal Sandhu, *A Lottery Lesson from Thailand*, TRIB. (India) (May 19, 2003), <http://www.tribuneindia.com/2003/20030519/biz.htm#2> (“The government in Thailand now offers a handsome prize of \$7,000 each week to 11 winners, who scribble their names on the back of the receipts and send them to the Finance Ministry.”); see also WORLD BANK, WORLD DEVELOPMENT REPORT 2005: A BETTER INVESTMENT CLIMATE FOR EVERYONE 110 (2005), [http://siteresources.worldbank.org/INTWDR2005/Resourses/complete\\_report.pdf](http://siteresources.worldbank.org/INTWDR2005/Resourses/complete_report.pdf) (describing a similar program in China).

42. See Nadja Dwenger et al., *Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany*, 8 AM. ECON. J.: ECON. POL’Y 203, 205 (2016).

43. *Id.* at 204. Most adults do not regularly attend church but remain members. *Id.* at 207 (“[B]etween 0.8% and 8.8% of eligible church members regularly attend church services in our sample parishes.”). Adults can avoid tax liability by leaving the church. See Justin Huggler, *Compulsory Income Tax on Christians Drives Germans Away from Protestant and Catholic Churches*, TELEGRAPH (Jan. 30, 2015), <http://www.telegraph.co.uk/news/world-news/europe/germany/11380968/Compulsory-income-tax-on-Christians-drives-Germans-away-from-Protestant-and-Catholic-churches.html>.

44. Dwenger et al., *supra* note 42, at 204.

45. *Id.* at 216.

46. *Id.* at 205.

a lottery for a monetary prize, a small probability of being recognized in a local newspaper for tax compliance, or both.<sup>47</sup> Dwenger et al. found that the prospect of a reward or recognition had no significant effect on donors' likelihood of donating but did significantly positively affect the likelihood that they increase their donation.<sup>48</sup> However, the reward/recognition intervention had a negative effect on baseline evaders, who statistically significantly both (1) increased their likelihood of evading (by 1.27%) and (2) reduced the likelihood they increase the amount paid (by nearly 16%).<sup>49</sup>

### B. Punishment

Turning to sanctions, which are much more widely used than rewards in the tax context, an argument that Bruno Frey has advanced in several articles is that deterrence can crowd out "tax morale"—the intrinsic motivation to comply<sup>50</sup>—as discussed further below.<sup>51</sup> For example Frey argues:

To some extent, paying one's taxes is a "quasi-voluntary" act attributable to an intrinsic motivation to contribute to the burden of taxation. . . . An unfair, inconsiderate way of treating taxpayers—punishing honest taxpayers by error—tends to undermine this tax morale. The net effect of using punishment in an effort to establish legal behavior is counterproductive if the

---

47. *Id.* at 213 ("For all these compliance rewards, the probability of winning the reward is close to zero . . . ; for the monetary reward treatment the notification makes explicit that the probability of winning is 1/1,000.").

48. *Id.* at 227. There was a 48.34% probability of an increased donation, which was statistically significant at  $p < 0.10$ . See *id.* at 221 tbl.2 (reporting coefficients and robust standard errors). *P*-value is the likelihood a finding is by chance, with  $p < 0.10$  denoting a likelihood under 10%. See Jonathan J. Darrow, *Pharmaceutical Efficacy: The Illusory Legal Standard*, 70 WASH. & LEE L. REV. 2073, 2113 (2013). "By convention, a *p*-value calculated to be 5% or less ( $p=0.05$ ) is considered to be 'statistically significant,' but other *p*-values could also be used." *Id.* (footnote omitted).

49. See Dwenger et al., *supra* note 42, 227. These were statistically significant at  $p < 0.10$  and  $p < 0.05$ , respectively. See *id.* at 221 tbl.2.

50. Bruno S. Frey, *Deterrence and Tax Morale in the European Union*, 11 EUR. REV. 385, 391 (2003).

51. See *infra* Part II.A.

relative price effect of the punishment is smaller than the crowding-out effect.<sup>52</sup>

Thus, it is important to examine the extent to which punishment may have the perverse effect of increasing noncompliant behavior.

There are several studies examining this possibility in non-tax contexts. For example, a laboratory study by Ernst Fehr and Simon Gächter found a crowding-out effect of fines in a work context, where an “employer” could offer a “worker” a “wage” and “effort level.”<sup>53</sup> In the study, the employer could impose a fine if he or she found the worker shirking; the worker faced a one-third chance that effort-level would be checked.<sup>54</sup> The study found that except at low “rent” levels (defined as wages minus the cost of the worker’s effort), the fine treatment *reduced* effort levels compared to the baseline (no-fine) control—a crowding-out effect.<sup>55</sup>

Perhaps the most well-known field experiment finding that sanctions increase undesired behavior<sup>56</sup> is “A Fine Is a Price,” by Uri Gneezy and Aldo Rustichini, involving a study of Israeli day-care centers.<sup>57</sup> The authors had the day-care centers introduce a fine of ten New Israeli Shekels (NIS) per child for each instance where a parent was ten or more minutes late picking up a child.<sup>58</sup>

---

52. Bruno S. Frey, *Punishment – and Beyond*, 5 CONTEMP. ECON. 90, 92 (2011) (citation omitted).

53. Fehr & Gächter, *supra* note 34, at 168, 170-71.

54. *Id.* at 170.

55. *Id.* at 170-71. The statistical significance is not reported there. *See id.* In addition, the article does not explore whether the result is driven by workers who were not themselves sanctioned, workers who were, or both. This remains unclear in a related, unpublished article. *See* Ernst Fehr & Simon Gächter, *Do Incentive Contracts Crowd Out Voluntary Cooperation?* 12-17 & tbl.3 (Univ. of S. Cal. Ctr. for Law, Econ. & Org., Research Paper No. C01-3, Feb. 2001), <https://ssrn.com/abstract=229047>. Studies in the tax context suggest that experiencing the sanction may affect behavior. *See infra* Part III. Subsequent experiments by Fehr & Gächter “indicate that a significant part of the undermining effect is due to the negative framing of the explicit incentive.” Ernst Fehr & Simon Gächter, *Do Incentive Contracts Undermine Voluntary Cooperation?* 1 (Inst. for Empirical Research in Econ., Univ. of Zurich, Working Paper No. 34, Apr. 2002).

56. *See* Ariel Rubinstein, *Discussion of “Behavioral Economics,”* in 3 ADVANCES IN ECONOMICS AND ECONOMETRICS: THEORY AND APPLICATIONS, NINTH WORLD CONGRESS 250 (Richard Blunell et al. eds., 2006) (“The behavioral economics literature has wholeheartedly adopted the paper.”).

57. Uri Gneezy & Aldo Rustichini, *A Fine Is a Price*, 29 J. LEGAL STUD. 1 (2000).

58. Based on the authors’ conversion of NIS 1400 to \$380, *see infra* text accompanying note 71, NIS 10 would be approximately \$2.71.

The fine was due at the end of the month, with the parents' regular bill.<sup>59</sup> The study found that the lateness incidence not only *increased*,<sup>60</sup> the increased frequency of lateness persisted after the fine was removed.<sup>61</sup>

Uri Gneezy, Stephan Meier, and Pedro Rey-Biel later explained the result:

One interpretation of this result is *information*: the parents did not initially know how important it was to arrive on time. The contract specified that they should pick their children up on time but failed to specify the penalty if they did not. The distribution of the parents' beliefs regarding how bad it was to be late may have included bad scenarios (for example, "the teacher will make my child suffer"). Once a small fine was imposed, the contract was complete in that being late was priced. The relatively small fine signaled to parents that arriving late was not that important. This new piece of information—that it was not so bad to be late—did not disappear once the fine was removed.<sup>62</sup>

Ariel Rubinstein critiqued the day-care center study<sup>63</sup> because his experience in Israel—that rules are rarely enforced and lateness is easily excused—made it difficult for him to believe that the

59. Gneezy & Rustichini, *supra* note 57, at 5 ("Payment was made to the principal of the day-care center at the end of the month. Since monthly payments are made to the owner during the year, the fines were added to those amounts.").

60. *Id.* at 3, 7 fig.1. The authors contrasted blocks of four weeks each. They found no time trend in the control group, but the difference between Block 1 (before the fine was introduced in the treatment group) and each of the other fine-condition blocks was significant at  $p < 0.001$ . See *id.* at 17 app. B.

61. *Id.* at 3, 8. The authors apparently did not examine whether parents who were late in a particular week had been fined in the previous week (or at all). See Uri Gneezy & Aldo Rustichini, *The Second Day-Care Center Study* (2005), <http://arielrubinstein.tau.ac.il/papers/WC05/GR1.pdf> (explaining that the study reflected "only the average data per daycare per week," not "individual parent data").

62. Gneezy et al., *When and Why Incentives (Don't) Work to Modify Behavior*, 25 J. ECON. PERSP. 191, 194 (2011) (emphasis added). Of course, schools have limited legitimate means to make children suffer (and for good reason). However, they are not obligated to reward children who are picked up late. A friend who taught at a K-12 school said that to deter parental lateness, the teachers would have the children ready on time, sitting and waiting in their coats. That way, a late parent would experience a bored child.

63. See Uri Gneezy & Aldo Rustichini, *Reply to Ariel Rubinstein's Critique of "A Fine Is a Price"* (2005), <http://arielrubinstein.tau.ac.il/papers/WC05/GR.pdf> (responding to discussant remarks at the World Congress of the Econometric Society); Rubinstein, *supra* note 56 (published proceedings of that World Congress).

procedures the study reported were actually followed.<sup>64</sup> Rubinstein observed that the authors did not verify the reports of the research assistant who asked the assistant teacher each week who had been late the previous week. Rubinstein also stated that he was not permitted to talk to the teachers. He adds that one of the coauthors of the day-care center article “agreed that he should delay publication until a new experiment with better monitoring of data collection is conducted. Eventually, the paper was published as is.”<sup>65</sup>

Gneezy and Rustichini responded to Rubinstein that (1) they did not create the fine of NIS 10, the day-care centers’ central organization provided that option; (2) it is common practice for managers in the day-care centers to record which parents were late; and (3) neither the managers nor the research assistants knew the authors’ hypothesis. They also stated that they offered to allow Rubinstein to participate in a follow-up study, but he declined.<sup>66</sup>

Gneezy and Rustichini did release a follow-up paper, and it reports the same general results.<sup>67</sup> “The Second Day-Care Center Study” is dated September 2005, but it appears preliminary; it contains three pages and concludes by stating that “[a] more detailed analysis of the data and the result is in preparation.”<sup>68</sup> This paper states that the authors received approval from the managers of the day-care centers for Rubinstein to call them. However, Rubinstein states, “Although the authors report that they are still working on the data and that they have asked the teachers for permission to be interviewed by me, the author who directed the

---

64. Rubinstein, *supra* note 56, at 250. He also found it “impossible . . . to imagine that Israeli teachers would have kept even roughly accurate records of late arrivals” in the bustle of pick-up. *Id.*

65. *Id.* Gneezy and Rustichini explained that Rubinstein “advised us to withdraw the paper from the *Journal of Legal Studies* before the publication. Since we were fully convinced that the design, the method, the data collection and the result were sound . . . , we eventually declined to follow his advice.” Gneezy & Rustichini, *supra* note 63.

66. Gneezy & Rustichini, *supra* note 63.

67. *See* Gneezy & Rustichini, *supra* note 61, at 3. The results are statistically significant at  $p < 0.00001$ . *Id.*

68. *See id.* Although the study states that “in the new study we also collected the individual parent data,” disaggregated results are not reported.

experiment has just informed me that he has lost the names of the kindergarten teachers who participated in the experiment."<sup>69</sup>

If we assume that Gneezy and Rustichini's data are accurate despite Rubinstein's critiques, what do their results show? First, it is important to observe that the NIS 10 fine in the day-care center was very low.<sup>70</sup> Gneezy and Rustichini reported that babysitters earned NIS 15 to 20 per hour and the monthly per-child day-care fee was NIS 1400 (approximately \$380).<sup>71</sup> Thus, the fine was only half to two-thirds of an hour of babysitter care, and without having to arrange for a babysitter. Second, other factors that may be relevant to the ability to generalize from this study include the fact that the study involves individuals and not firms; the fine applied in every instance of wrongdoing;<sup>72</sup> the specific context of the study, involving parental lateness in picking up children from day care; and even local culture.<sup>73</sup>

We also do not know whether a particular subgroup of parents is driving this result. That is, did all parents react similarly to the

69. Rubinstein, *supra* note 56, at 251.

70. See Leandra Lederman, *The Interplay Between Norms and Enforcement in Tax Compliance*, 64 OHIO ST. L.J. 1453, 1461 n.38 (2003) (noting how low the fine was); Stephan Meier & Matthew Stephenson, *When Do Monetary Incentives Backfire?*, HARV. ECON. REV. (Apr. 13, 2016), <http://harvardecon.org/?p=3283> ("The evidence from the daycare (and elsewhere) showed that incentives did not always work as intended. Of course, if the incentives are large enough they likely will . . .").

Lin and Yang developed a theoretical model to explain Gneezy and Rustichini's results. See Chung-cheng Lin & C.C. Yang, *Fine Enough or Don't Fine at All*, 59 J. ECON. BEHAV. & ORG. 195 (2006). In their model, there is a social norm of being on time rather than late. *Id.* at 197. Once a fine is introduced, the psychological cost of violating the social norm decreases, reducing the effectiveness of the social norm. In addition, because the psychological cost of lateness is affected by how many other parents are late, there is a "bandwagon effect" that further increases the number of parents arriving late. *Id.* at 205. This results in a new norm that makes lateness more acceptable, such that when the fine is removed, parental lateness does not return to the pre-fine level. *Id.* at 206. Lin and Yang's model therefore counsels setting any fines used high enough to be effective.

71. Gneezy & Rustichini, *supra* note 57, at 4-5.

72. See Gneezy & Rustichini, *supra* note 67, at 1 ("One of the coauthors interviewed some parents at random and was told that the fine was indeed collected at the end of the month.").

73. Barak Ariel found that deterrence letters had no effect on Israeli corporate taxpayers, and moral suasion letters actually increased their noncompliance. See Barak Ariel, *Deterrence and Moral Persuasion Effects on Corporate Tax Compliance: Findings from a Randomized Controlled Trial*, 50 AM. J. CRIM. 27, 41-45, 58 (2012). Other studies have found audit threat letters to increase tax payments, see *infra* Part II.B.1, and that moral suasion letters have no effect or a slight positive one, see *infra* note 163.



imposition of the fine, or did those who were actually fined (or not fined) increase the frequency of their lateness? In other words, how did the effects on those *experiencing* an enforcement action differ from *regime-level* effects? In the original study, Gneezy and Rustichini “had only the average data per daycare per week,”<sup>74</sup> which would preclude that analysis. In “The Second Day-Care Center Study,” they “collected the individual parent data,”<sup>75</sup> but they do not report separate results for parents who were fined from those who were not fined (or not fined for a while).<sup>76</sup>

Given both Rubinstein’s critique and the lack of detail in “The Second Day-Care Center Study,” a logical next question is whether Gneezy and Rustichini’s result exists in other contexts. There are few comparable studies. However, a 2010 study in Norway took advantage of a natural experiment in another lateness context, so-called “bed blocking.” That study examined whether fines imposed by hospitals on the owners of long-term care facilities (municipalities) for not having a bed available for a hospital patient transferring to long-term care affect the likelihood patients over age sixty-seven will stay in the hospital after their planned discharge date.<sup>77</sup> Hospitals can impose these fines on the long-term care facility’s owners if the patient is in the hospital ten days after the hospital’s stated discharge date (seven days in Oslo).<sup>78</sup>

---

74. Gneezy & Rustichini, *supra* note 67, at 1.

75. *Id.*

76. In “The Second Day-Care Center Study,” the authors distinguish between “three periods, a first of four weeks, a second of 11 weeks (from the 5th to the 15th) and a final third period, weeks 16th to 18th.” *Id.* They introduced the fine at the beginning of the fourth week—part way through the first period—and removed it at the beginning of the sixteenth week. *Id.* They report that “[i]n the centers where fines were introduced, the increase in the number of delays is higher in the families with higher delays in the first period.” *Id.* This could be a function of a lateness-prone effect, *see infra* note 95 and accompanying text, or it could suggest that those fined responded with greater delays. However, without more details, it is impossible to know, as the first period includes three weeks without fines and one week after fining began.

It is also unclear if any of the ten day-care centers studied (six treated and four control, Gneezy & Rustichini, *supra* note 67, at 1) had been treated with the fine condition in the previous study and, if so, whether some of the same parents were still using that center.

77. Holmås et al., *supra* note 10, at 262, 263 (“[W]e . . . restrict our sample to individuals older than 67 who are discharged from hospital to a long-term care institution. For these patients, it is quite clear that the hospital [length of stay] is directly influenced by the municipalities since they cannot be discharged from the hospital before an institution accepts . . . them.”).

78. *Id.* at 262.

The study examined two large hospitals, one that used fines and one that did not. A “natural experiment took place in 2004, where parts of Hospital F’s [the Fine hospital’s] catchment area were transferred [by the government] to Hospital NF [the No-Fine hospital], while parts of Hospital NF’s catchment area were transferred to Hospital F.”<sup>79</sup> Thus, some municipalities stayed in the fine (or no-fine) area before and after 2004, and some moved from fine to no-fine status (or the reverse) in 2004.<sup>80</sup>

The study’s results were consistent with crowding-out theory: “Patients living in municipalities that change[d] hospital[s] from F to NF stay[ed] approximately 2.3 days shorter at hospital after their potential discharge date, while patients changing from NF to F stay[ed] close to 3 days longer.”<sup>81</sup> They found that the effects were symmetric: delays increased when the fine was introduced and decreased when the fine was removed (unlike in the day-care center study).<sup>82</sup> They also explained that “the daily penalty rate [of 1600 Norwegian Krone (NOK)/day; 2000 in Oslo<sup>83</sup>] is . . . well below the estimated average daily operation cost of 7[000]–8[000] NOK in Norwegian hospitals and can therefore be characterised as a relatively small fine.”<sup>84</sup>

Thus, these field experiments suggest that a low fine may be viewed as a (favorable) price, such that, at least in the context of lateness, the fined behavior actually increases. A low fine may also signal that the fined behavior is not so bad.<sup>85</sup> Yet a low price may

79. *Id.*

80. *Id.* at 262–63.

81. *Id.* at 265. This was statistically significant at  $p < 0.01$ . *Id.* at 265 tbl.5.

82. *Id.* at 266. They note that “in the daycare study it was the same agents (parents) that had the fine introduced and then subsequently removed, while in our study some agents had the fine introduced while others had the fine removed.” *Id.* It is unclear whether length of delay changed immediately after the fine was introduced/removed, or not until the hospitals actually experienced the change. *See generally id.*

83. At the October 2018 exchange rate of 1 NOK = \$0.12 (rounded to the nearest cent), 1600 NOK is \$192.00. *See XE Currency Converter*, XE, <http://www.xe.com/currencyconverter/convert/?Amount=1&From=NOK&To=USD> (last visited Oct. 18, 2018).

84. Holmås et al., *supra* note 10, at 266 (citation omitted).

85. Some U.S. day-care centers report using late fees that increase as lateness rises (such as \$1 per minute per child) and/or increase with repeated lateness. *See Please Give Me Your Late Pick Up Policy*, DAYCARE.COM, <http://www.daycare.com/forum/archive/index.php/t-21692.html> (Daycare.com Forum > Main Category > Daycare Center and Family Home Forum > Please Give Me Your Late Pick Up Policy) (last visited Oct. 18, 2018).

not necessarily increase undesired behavior in all contexts. Lior Strahilevitz's article on the commodification of California carpool lanes presents an interesting example. San Diego was having a problem with traffic congestion.<sup>86</sup> HOV (high-occupancy vehicle) lanes were not incentivizing carpooling enough.<sup>87</sup> In addition, approximately 15% of drivers using the HOV lanes were solo drivers violating the law and risking a \$271 fine.<sup>88</sup>

San Diego then instituted an "ExpressPass" option that allowed solo drivers to pay a small amount<sup>89</sup> to use the HOV lanes (renamed HOT lanes—presumably meaning "High Occupancy Toll"<sup>90</sup>). It also increased patrolling for violators from once a month to three times per week.<sup>91</sup> The changes had two effects, both positive: unauthorized use of the special lanes decreased<sup>92</sup> and carpooling increased.<sup>93</sup>

This outcome is not necessarily inconsistent with the negative results small fines had in other studies because, unlike in the other studies, the baseline included a large fine for violations (\$271 in 1996 dollars, equivalent to \$433 in 2018<sup>94</sup>). The main intervention was the addition of a comparatively small per-use price, although patrolling also increased—akin to increasing the audit rate.<sup>95</sup> The

---

86. Lior Jacob Strahilevitz, *How Changes in Property Regimes Influence Social Norms: Commodifying California's Carpool Lanes*, 75 IND. L.J. 1231, 1250 (2000).

87. *Id.* ("Although a vehicle needed only two occupants in order to constitute a carpool, the HOV lanes tended to be underutilized." (footnotes omitted)).

88. *Id.* at 1256; *see also id.* at 1257 n.141.

89. A toll amount is displayed at the entrance to the HOT lanes, and it rises with the volume of traffic in the HOT lanes. Tolls averaged \$1.95 to \$2.26 in 1998. *Id.* at 1251 & n.111. In 2000, the tolls ranged from \$0.50 to \$8.00 per trip but only exceeded \$4.00 in the event of extreme congestion. *Id.* at 1252 n.111.

90. *Id.* at 1251 n.108.

91. *Id.* at 1257 n.143.

92. *Id.* at 1256 (reporting that HOV violations decreased from about 15% of drivers in the HOV lanes to approximately 3% of drivers in the HOT lanes).

93. *Id.* at 1234. Data suggested that "the new carpools consisted mostly of drivers who had neither used an ExpressPass nor participated in a carpool during the previous year." *Id.* at 1256. It is not clear from the article what percentage of these drivers had previously been fined.

94. *Inflation Calculator*, SMART ASSET, <https://smartasset.com/investing/inflation-calculator#uEsf3XMVYx> (last visited Oct. 18, 2018).

95. Strahilevitz argues that the increased patrolling does not explain the entire increase in compliance. He states that compliance did not fully correspond to enforcement because there was at least one month of relatively low enforcement/high compliance and one of high enforcement/lower compliance. Strahilevitz, *supra* note 86, at 1258–59. However, drivers cannot perfectly observe enforcement, and they may also respond with a lag, such as by increasing compliance the month after they perceive a "crackdown."

intervention increased traffic in the special lanes, perhaps analogous to the increased instances of lateness once the Israeli day-care centers imposed a small fine, although, in the HOT lanes context, the increase was desired. The HOT lanes also increased carpooling, which would be analogous in the day-care center study to a small minority of parents increasing their commitment to being on time, so as to avoid paying the fine.

This raises the question of whether the increased instances of lateness once the day-care centers imposed fines were by many different parents (generally supporting a change in the lateness norms) or instead were by a “lateness-prone” subgroup, which would suggest that some parents might—like San Diego carpoolers—feel that their responsible behavior was now quantifiable in money savings. And, in fact, Gneezy and Rustichini report in their second study:

[F]amilies who are more likely to be late in the first period are also more likely to be late in the second and third: this is true in the non fined centers as well. In the centers where fines were introduced, the increase in the number of delays is higher in the families with higher delays in the first period.<sup>96</sup>

Thus, results reported in “The Second Day-Care Center Study” support the notion that a subgroup of people—here, those prone to be late—drive much of the result.

Nonetheless, it is hard to draw firm conclusions from so few studies. That is not only because of the questions Rubinstein has raised about the Israeli day-care center study,<sup>97</sup> but also because “one must recognize that many novel and surprising experimental results might not be robust—not because of falsification or something egregious, but merely due to the mechanics of the problem.”<sup>98</sup> As Zacharias Maniadis et al. explain, “[t]he common reliance on statistical significance as the sole criterion leads to an

96. Gneezy & Rustichini, *supra* note 61, at 3. This result is analogous to a Dutch study that found that taxpayers evading in one year were more likely to do so in a later year. See *infra* note 408 and accompanying text.

97. See *supra* text accompanying notes 63–69.

98. See Zacharias Maniadis et al., *One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects*, 104 AM. ECON. REV. 277, 278 (2014) (“[O]ur theoretical model suggests that many surprising new empirical results are likely not recovering true associations.”).

excessive number of false positives”<sup>99</sup> because “the statistical power of the test and the fraction of tested hypotheses that are true associations” also matter.<sup>100</sup> A bias toward publication of surprising or unexpected results exacerbates the problem.<sup>101</sup>

Moreover, in addition to the small number of studies, the day-care center and hospital studies are non-tax studies, so they do not bear directly on the question of whether sanctions crowd out taxpayers’ intrinsic motivation to pay taxes. The next Part discusses this theory and argues that, in the tax context, the evidence from empirical studies suggests that deterrence generally is effective.

## II. DETERRENCE AS A TOOL FOR FOSTERING TAX COMPLIANCE

Noncompliance with a tax saves that taxpayer money, at least up front, but it carries risks. The government imposing the tax typically audits taxpayers and imposes penalties for detected noncompliance.<sup>102</sup> Accordingly, economic models of tax compliance generally consider tax evasion a decision under uncertainty.<sup>103</sup> The basic idea is that the existence of a penalty for noncompliance deters tax evasion but only in light of the probability that a penalty will actually be imposed. For example, a taxpayer who omits from income an amount resulting in understated tax of \$1000 and who faces a 1% chance of audit that will detect the evasion and a 20% penalty in addition to the tax if detected (\$1200 in total)<sup>104</sup> faces an “expected” (probabilistic) cost of \$12.

---

99. *Id.* at 288.

100. *Id.*

101. *Id.*

102. See I.R.C. § 6662(a), (b) (2012) (imposing a 20% penalty for such things as negligence or substantial understatement of tax); § 6663(a) (“If any part of any underpayment of tax required to be shown on a return is due to fraud, there shall be added to the tax an amount equal to 75 percent of the portion of the underpayment which is attributable to fraud.”). When penalties are imposed, taxpayers likely experience the resulting tax bill, which will include interest, as high. See Lederman, *supra* note 70, at 1466 & n.67.

103. See, e.g., Michael G. Allingham & Agnar Sandmo, *Income Tax Evasion: A Theoretical Analysis*, 3 J. PUB. ECON. 323, 323–26 (1972); T.N. Srinivasan, *Tax Evasion: A Model*, 2 J. PUB. ECON. 339, 340 (1973).

104. These are realistic figures. See *infra* note 108 (official IRS audit rate for individuals is just under 1%); *supra* note 102 (20% penalty).

This very basic model has a number of simplifying assumptions. It assumes that taxpayers know the odds of detection and sanction levels and that they are risk-neutral. However, individuals may overestimate the likelihood that noncompliance will be detected and fined, which would tend to increase compliance by raising the expected cost of cheating. Corporate taxpayers are more likely to have sophisticated tax advisors who know the odds. Individuals are also less likely to be risk-neutral than corporate taxpayers.<sup>105</sup> Risk aversion would tend to increase compliance.<sup>106</sup>

#### A. The False Puzzle and “Tax Morale”

Even if we were to assume that all audits detect all non-compliance and result in the imposition of a penalty – which is not true – federal income tax penalties may be too low in light of existing audit rates (and vice versa) to make compliance the wealth-maximizing choice. That is, penalties are a fraction (in the United States, typically 20%) of understated tax.<sup>107</sup> The IRS reports audit rates for U.S. individuals of under 1%,<sup>108</sup> making the expected value

105. See Joel Slemrod, *The Economics of Corporate Tax Selfishness*, 57 NAT’L TAX J. 877, 882 (2004) (“[T]he assumption of risk aversion seems unsatisfactory for a large publicly-held firm, because presumably the shareholders hold diversified portfolios, implying that the firm should behave as if it is risk-neutral, even if its shareholders are not.”).

106. See, e.g., Allingham & Sandmo, *supra* note 103, at 329 (“We can then conclude that when actual income varies, the fraction declared increases, stays constant or decreases according [sic] as relative risk aversion is an increasing, constant or decreasing function of income.”); Paul J. Beck & Woon-Oh Jung, *Taxpayer Compliance Under Uncertainty*, 8 J. ACCT. & PUB. POL’Y 1, 18 (1989) (“Risk-averse taxpayers were shown to have incentives to increase their reported income (compliance) as the tax rate increases when penalties were proportional to the amount of underpaid taxes.”).

107. See *supra* note 102.

108. See DEP’T OF THE TREASURY, INTERNAL REVENUE SVC., INTERNAL REVENUE SERVICE DATA BOOK, 2015, at 23 (Mar. 2016), <https://www.irs.gov/pub/irs-soi/15databk.pdf> (reporting that 0.8% of individual tax returns and 1.3% of corporate returns were examined for fiscal year (FY) 2015); DEP’T OF THE TREASURY, INTERNAL REVENUE SVC., INTERNAL REVENUE SERVICE DATA BOOK, 2014, at 26 (Mar. 2015), <https://www.irs.gov/pub/irs-soi/14databk.pdf> (reporting that 0.9% of individual tax returns and 1.3% of corporate returns were examined for FY 2014). These are official IRS audit rates. The National Taxpayer Advocate has explained that the IRS makes additional “compliance contacts with taxpayers that it does not consider to be ‘real’ audits. These types of contacts . . . include math error corrections, Automated Underreporter (AUR) (a document matching program), identity and wage verification, and Automated Substitute for Return (ASFR) (a non-filer program).” NTA Blog: “Real” vs. “Unreal” Audits and Why This Distinction Matters, TAXPAYER ADVOC. SERV.

of any penalty less than 1% of its nominal value, even assuming that audits detect all cheating and always result in the imposition of a penalty.<sup>109</sup>

So, in the example above, the expected cost of cheating was \$12.<sup>110</sup> By contrast, the cost of compliance was the full \$1000 tax. For the expected cost of cheating to equal the cost of compliance, “an audit rate of 1% would require a \$99,000 penalty in this example.”<sup>111</sup> Given these low levels of audit rates and penalties, it seems that a rational wealth-maximizing taxpayer should cheat whenever possible. Thus, at first cut, the basic economic model would seem to predict that no one pays any taxes.<sup>112</sup> Yet the IRS estimates an overall voluntary compliance rate with U.S. federal taxes of almost 82% of dollars due.<sup>113</sup>

Many scholars have relied on this type of contrast to argue that deterrence does not explain tax compliance.<sup>114</sup> They argue that the

---

(July 6, 2018), <https://taxpayeradvocate.irs.gov/news/nta-blog-real-vs-unreal-audits-and-why-this-distinction-matters>. “When adding these ‘unreal’ audit numbers to the ‘real’ ones, the IRS’s combined coverage rate jumps to over six percent.” *Id.* An audit rate of 6% would make the expected cost of cheating a still low \$72 (6% of \$1200).

Taxpayers are selected for audit in various ways. *See supra* note 2. The IRS’s whistleblower program plays a small part in that process. IRS, IRS WHISTLEBLOWER PROGRAM FISCAL YEAR 2016 ANNUAL REPORT TO THE CONGRESS 10 tbl.1, [https://www.irs.gov/pub/whistleblower/fy16\\_wo\\_annual\\_report\\_final.pdf](https://www.irs.gov/pub/whistleblower/fy16_wo_annual_report_final.pdf) (reporting that 240 claims were made in 2014, 204 in 2015, and 761 in 2016).

109. Even if a taxpayer erroneously believed that the audit rate is (a grossly inflated) 30% and the penalty rate is 50%, for example, while that would increase the expected cost of cheating to \$450 in the example (30% of \$1500), it remains lower than the \$1000 cost of compliance.

110. That is, 1% of \$1200. *See supra* text accompanying note 104. If the taxpayer’s behavior meets the burden of proof for fraud, the taxpayer would owe \$1750 (the \$1000 plus a 75% penalty). At a 1% audit rate, the expected value of the penalty is only \$17.50.

111. Lederman, *supra* note 70, at 1465 (footnotes omitted).

112. *See* Dick J. Hessing et al., *Does Deterrence Deter? Measuring the Effect of Deterrence on Tax Compliance in Field Studies and Experimental Studies*, in *WHY PEOPLE PAY TAXES* 291, 293 (Joel Slemrod ed., 1992) (“Given the current mild sanctions and low probability of detection, this kind of [economics] approach would predict that virtually everyone should be evading tax . . .”).

113. *See* INTERNAL REVENUE SVC., *TAX GAP ESTIMATES FOR TAX YEARS 2008–2010*, at 3 (Apr. 2016), <https://www.irs.gov/pub/newsroom/tax%20gap%20estimates%20for%202008%20through%202010.pdf> (overall 81.7% voluntary compliance rate).

114. Dan M. Kahan, *Signaling or Reciprocating? A Response to Eric Posner’s Law and Social Norms*, 36 U. RICH. L. REV. 367, 377 (2002) (“Tax compliance rates – which vary dramatically across nations – seem to bear no connection to enforcement levels.”); *see, e.g.*, James Alm et al., *Economic and Noneconomic Factors in Tax Compliance*, 48 KYKLOS 3, 3 (1995) (“[T]he real puzzle [of tax compliance] is ‘Why is there so little cheating?’”) (emphasis removed); *cf.* Eric

high level of voluntary compliance is a puzzle that the economic model cannot explain.<sup>115</sup> For example, Bruno Frey has claimed:

Empirical research has convincingly established that it is impossible to account for the extent of tax paying by only considering the expected punishment. The crucial question is not why people do not pay their taxes, but rather “Why is there so little cheating?” To some extent, paying one’s taxes is a “quasi-voluntary” act attributable to an intrinsic motivation to contribute to the burden of taxation. Risk aversion is not able to account for the extent of taxes paid in the United States and Switzerland.<sup>116</sup>

Although Frey does not fully reject the deterrence model, he argues that it is inapplicable in some countries where, instead, “tax morale” is what prompts compliance.<sup>117</sup> Although tax morale generally means simply an intrinsic compliance motivation,<sup>118</sup> Frey sources it in “*participatory governments*” that have “(semi-) direct democracies, in which the citizens have initiative and referendum rights.”<sup>119</sup> He points to California, Oregon, and Switzerland, and states that “*civic virtue* emerges which, with respect to taxation, shows up as tax morale,” contrasting that with “*exploitative governments* ruling their people in an authoritarian or even

Posner, *Law and Social Norms: The Case of Tax Compliance*, 86 VA. L. REV. 1781, 1782 (2000) (“A widespread view among tax scholars holds that law enforcement does not explain why people pay taxes.”).

115. See Alm, *supra* note 114, at 3; *infra* note 116 and accompanying text.

116. Frey, *supra* note 52, at 92 (citations omitted) (quoting Alm, *supra* note 114); see also Ronald G. Cummings et al., *Tax Morale Affects Tax Compliance: Evidence from Surveys and an Artefactual Field Experiment*, 70 J. ECON. BEHAV. & ORG. 447, 449 (2009) (“[E]xtreme . . . risk aversion would be required to explain observed . . . compliance. Other factors must be at work.”); Frey, *supra* note 50, at 387 (making a similar argument). Feld and Frey estimated the 1995 voluntary compliance rate for Switzerland at 77.7%. Lars P. Feld & Bruno S. Frey, *Tax Evasion in Switzerland: The Roles of Deterrence and Tax Morale*, in TAX EVASION, TRUST AND STATE CAPACITIES 134 (Nicolas Hayoz & Simon Hug eds., 2007).

117. See Frey, *supra* note 50, at 388–89 (“[S]eek[ing] to demonstrate that intrinsic motivation in the form of ‘tax morale’ is of substantial importance in explaining tax paying behavior.”).

118. See *id.*; see also, e.g., Lars P. Feld & Bruno S. Frey, *Trust Breeds Trust: How Taxpayers Are Treated*, 3 ECON. GOVERNANCE 87, 88 (2002) (“[T]he existence of an intrinsic motivation to pay taxes . . . has sometimes been called ‘tax morale.’”). This Article does not deny that there are intrinsic motivations to pay taxes. See *infra* text accompanying notes 161–166. The issue this Article tackles is whether deterrence crowds out those motivations, such that it reduces tax compliance.

119. Frey, *supra* note 50, at 390.



dictatorial way” and governments that lie between these poles.<sup>120</sup> Thus, Frey argues that, in places such as Switzerland, the deterrence model does not explain tax compliance; intrinsic motivations do. However, he found in a survey of Swiss cantons that fines increased tax compliance,<sup>121</sup> which is contrary to his prediction and instead supports the deterrence model.

Frey’s reliance on intrinsic motivation as the explanation for high compliance levels starts from the premise that if extremely high-risk aversion – which he rejects as not true in practice<sup>122</sup> – does not explain the observed compliance levels, there must be a missing factor, which he identifies as tax morale.<sup>123</sup> However, tax morale need not be the explanation. In fact, Frey quotes Michael Graetz and Louis Wilde as stating, “the high compliance rate can only be explained either by taxpayers’ . . . commitment to the responsibilities of citizenship and respect for the law or *lack of opportunity for tax evasion*.”<sup>124</sup> These are not the same two possible explanations that Frey presents when he points to tax morale and risk aversion.<sup>125</sup> While there is certainly room for civic commitments and respect for the law as explanations for some compliance, the lack of opportunity for tax evasion—which Frey does not analyze as a possible explanation—explains much tax compliance and is consistent with the deterrence model.<sup>126</sup>

---

120. *Id.*

121. *See id.* at 397 (reporting that increased size of punishment decreases tax evasion); *see also id.* at 387 (stating that only the size of the fine was statistically significant at  $p < 0.05$  level in a model predicting tax evasion in the Swiss cantons).

122. *See id.* at 387. (“The conventionally used measure of risk aversion by Arrow-Pratt would have to be more than 30 in order to account for the present compliance rate in that country [the United States]. The actually measured level of risk aversion reported, however, lies only between 1 and 2.”)

123. *See id.* at 388 (arguing that the deterrence model “is at best incomplete, and may even be wrong” and that tax morale is “the missing factor”).

124. *Id.* at 389 (quoting Michael J. Graetz & Louis L. Wilde, *The Economics of Tax Compliance: Facts and Fantasy*, 3 NAT’L TAX J. 355, 358 (1985)) (emphasis added). On the same page, the Graetz and Wilde article states, “Any economic analysis of the various components of noncompliance must . . . take into account not only . . . institutional constraints on the level of punishment or audits, but also differences in opportunities to understate taxes.” Graetz & Wilde, *supra* at 358.

125. *See supra* text accompanying note 116 (quoting Frey, *supra* note 52, at 92, and referring to intrinsic motivation and risk aversion).

126. *See, e.g.*, James Andreoni et al., *Tax Compliance*, 36 J. ECON. LIT. 818, 821 (1998) (“[W]e may ask, why are so many households honest, and why don’t cheaters cheat by more? Part of the explanation lies in the dramatic increase in information reporting . . . since the

### 1. *The importance of opportunity to evade*

As I have explained in previous work, the stark dichotomy embodied in the purported puzzle sets aside an important feature of the federal income tax system, which is that taxpayers do not have an open opportunity to evade taxes on all of their income.<sup>127</sup> Many sources of income are subject to third-party reporting, and it is much easier for the government to match an information return with a taxpayer's return than to conduct an audit. Thus, practically speaking, information-return matching is a highly effective form of audit not captured in audit statistics.<sup>128</sup> Recent IRS figures show that the voluntary compliance rate with respect to income subject to withholding and substantial information reporting is 99%, and the rate with just substantial information reporting is 93%.<sup>129</sup> Even income amounts subject to partial information reporting have a voluntary compliance rate of 81%. By contrast, income subject to no information reporting has a compliance rate of only 37%.<sup>130</sup> One scholar accurately observed that use of third-party reporting is so

---

1960s."); Henrik Jacobsen Kleven et al., *Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark*, 79 *ECONOMETRICA* 651, 653 (2011) ("[O]ur findings suggest that tax evasion is low, not because taxpayers are *unwilling* to cheat, but because they are *unable* to cheat successfully due to the widespread use of third-party reporting."); Susan Cleary Morse et al., *Cash Businesses and Tax Evasion*, 20 *STAN. L. & POL'Y REV.* 37, 67 (2009) ("Our interview results here generally agree with other research cited above: Tax cheating follows opportunity, not complexity or immorality . . ."); Mark D. Phillips, *Individual Income Tax Compliance and Information Reporting: What Do the U.S. Data Show?*, 67 *NAT'L TAX J.* 531, 563 (2014) ("Using U.S. taxpayer-level data from the 2001 NRP, this paper has shown that the presence and amount of unmatched income are the primary determinants of income tax noncompliance."); Joel Slemrod, *Cheating Ourselves: The Economics of Tax Evasion*, 21 *J. ECON. PERSP.* 25, 37 (2007) ("Line item by line item, there is a clear positive correlation between the rate of compliance and the presence of enforcement mechanisms such as information reports and employer withholding.");

127. See, e.g., Leandra Lederman, *Statutory Speed Bumps: The Roles Third Parties Play in Tax Compliance*, 60 *STAN. L. REV.* 695, 697 (2007) ("An essential missing piece of this seeming puzzle is that the federal income tax law benefits from structural mechanisms that constrain payment with respect to the major sources of income for many people . . ."); Lederman, *supra* note 70, at 1460 ("[T]he eighty-three percent [voluntary compliance] figure is misleading because it is an aggregate comprised of differing levels of compliance that correspond to differences in opportunity to evade tax.");

128. Leandra Lederman, *Tax Compliance and the Reformed IRS*, 51 *KAN. L. REV.* 971, 975 (2003).

129. See *INTERNAL REVENUE SVC.*, *supra* note 113, at 5.

130. *Id.* The fact that this is not zero is discussed *infra* in text accompanying notes 150-152.

widespread that “the notion that deterrence is weak is to some extent an illusion.”<sup>131</sup>

Moreover, a study in Denmark found a near-zero noncompliance rate by individuals for income subject to information reporting, which contrasted with a substantial noncompliance rate by the same individuals for income not subject to third-party reporting.<sup>132</sup> This was true both for individuals with mostly self-reported income and for those with mostly third-party-reported income.<sup>133</sup> In particular, “the evasion rate for self-employment income conditional on third-party reporting is only 0.33%, suggesting that overall tax evasion among the self-employed is large because of the information environment and not because of, for example, different preferences among those choosing self-employment (such as attitudes toward risk and cheating).”<sup>134</sup>

A laboratory experiment also supports the notion that the same individual reports third-party-reported income at a higher rate than other income.<sup>135</sup> Participants in the study had the opportunity to earn income by doing a simple task on the computer.<sup>136</sup> They were then told how much of the income is “matched” (subject to information reporting) and how much is “nonmatched.”<sup>137</sup> Their reports were subject to audit at a rate of 10% or 30%.<sup>138</sup> The study found that the overall compliance rate for matched income was

---

131. Dwenger et al., *supra* note 42, at 204; *see also* Erzo F.P. Luttmer & Monica Singhal, *Tax Morale*, 28 J. ECON. PERSP. 149, 152 (2014) (explaining that, given low evasion on amounts reported to the IRS, “inferring tax morale as residual compliance in the Allingham and Sandmo (1972) model is credible only in settings without third-party reporting”).

132. Kleven et al., *supra* note 126, at 653.

133. *Id.* In Denmark, information received from third-party information reports is included in prepopulated returns that taxpayers are sent to complete. *Id.* at 659.

134. *Id.* at 670. The evasion level in that study for self-reported income was 41.6%. *Id.* at 671. Kleven et al. also found that “the effects of social variables are small and mostly insignificant, whereas variables that capture information (presence and size of self-reported income, self-employment, audit flags, and prior audit adjustments) have very strong effects. This confirms the conclusion that information and traceability are central to the compliance decision.” *Id.* at 676.

135. James Alm et al., *Do Individuals Comply on Income Not Reported by Their Employer?*, 37 PUB. FIN. REV. 120, 121 (2009).

136. *Id.* at 127.

137. *Id.*

138. *Id.* “Increasing the audit probability from 10 to 30 percent increases compliance by 4.9 percentage points, all else equal.” *Id.* at 136. That variable was significant at  $p < 0.01$ . *See id.* at 135 tbl.3.

54.2%, while for nonmatched income it was 41.4%, and the difference was statistically significant.<sup>139</sup>

Opportunity to evade is relevant for firms, too. For example, Dina Pomeranz found a positive effect of an audit threat on compliance by Chilean firms with a Value-Added Tax (VAT), which was submitted monthly.<sup>140</sup> The Chilean tax authority sent 102,000 randomly selected companies a letter stating, in part, “your firm has been selected for analysis. In the event that any irregularities are detected, you could be summoned for an audit.”<sup>141</sup> The result was an immediate increase in VAT remittances,<sup>142</sup> and it was primarily driven by the transactions for which there is no paper trail—sales to customers—as opposed to between-firm transactions in the supply chain, where firms have an incentive to ask for receipts from their suppliers.<sup>143</sup> Thus, it appears that the existence of a paper trail, like an increase in the audit rate, raises perceived detection risk.<sup>144</sup>

Similarly, James Alm et al. found that the average compliance rate across the New Mexico firms in their study was 43%, but it varied widely across industries.<sup>145</sup> They found that firms that have a greater opportunity to evade taxes—as evidenced by providing services, having an out-of-state mailing address, and exhibiting greater variation in claimed deductions—are less compliant with a sales tax.<sup>146</sup> However, firms’ compliance increased when the (perceived) audit rate increased.<sup>147</sup> In fact, if an individual monitors firm compliance (contrary to the usual arrangement, where firms

---

139. *Id.* at 133 ( $p < 0.05$ ).

140. Dina Pomeranz, *No Taxation Without Information: Deterrence and Self-Enforcement in the Value Added Tax*, 105 AM. ECON. REV. 2539, 2543 (2015).

141. *Id.* at 2547 & n.16.

142. “The median declared VAT increases by about 12 percent and then slowly decreases to reach the same levels as the control group after about 15 months.” *Id.* at 2557.

143. *Id.* at 2540, 2543–44.

144. *Id.* at 2541.

145. James Alm et al., *Audit Selection and Firm Compliance with a Broad-Based Sales Tax*, 57 NAT’L TAX J. 209, 217 (2004).

146. *Id.* at 225. Missing filing deadlines was also a statistically significant predictor of noncompliance. *Id.* at 223.

147. *Id.* at 211, 223–25.

generally provide information reports on individuals<sup>148</sup>), studies suggest that firms' compliance increases.<sup>149</sup>

## 2. Other factors affecting tax compliance

These studies, which highlight the importance of the opportunity to evade, support the idea that deterrence spurs compliance. Of course, lack of opportunity to evade does not explain all tax compliance.<sup>150</sup> For one thing, U.S. taxpayers pay an estimated 37% of tax dollars due on income not subject to information reporting, such as self-employment income.<sup>151</sup> That figure is quite low in comparison to the other voluntary compliance figures, but it is not zero.<sup>152</sup>

---

148. See Leandra Lederman, *Reducing Information Gaps to Reduce the Tax Gap: When Is Information Reporting Warranted?*, 78 *FORDHAM L. REV.* 1733, 1739–40 (2010) (explaining that information reporting is most efficient when done by centralized parties acting at arms' length and with a bookkeeping infrastructure, which typically means businesses).

149. See generally Todd Kumler et al., *Enlisting Employees in Improving Payroll-Tax Compliance: Evidence from Mexico* (Nat'l Bureau of Econ. Research, Working Paper No. 19385, Apr. 2015), <http://www.columbia.edu/~ev2124/research/KumlerVerhoogen&FriasApril2015.pdf> (finding that pension reform that gave younger workers an incentive to monitor their employers' wage reporting, and the information to do so, resulted in reduced underreporting of wages for younger workers by firms (and thus reduced payroll tax evasion)); Joana Naritomi, *Consumers as Tax Auditors* (May 2016) (working paper), [https://www.dropbox.com/s/1e0bctgji4s01c/naritomi\\_enforcement\\_May2016.pdf?dl=0](https://www.dropbox.com/s/1e0bctgji4s01c/naritomi_enforcement_May2016.pdf?dl=0) (finding that a São Paulo, Brazil, program that provided consumers with incentives in the form of tax rebates and lottery participation for requesting receipts, as well as for checking online the retailers' reports of their transactions, increased retailers' reported revenues for tax purposes by at least 22% over a four-year period).

150. See Andreoni et al., *supra* note 126, at 822 (explaining that, for 1992, 91.7% of all reportable income was reported, although only three-quarters of income was subject to information reporting). Taxpayers who receive most of their income in forms subject to information reporting could cheat by inventing or exaggerating deductions or credits not subject to information reporting. However, IRS data show that much more noncompliance occurs on the income side of the ledger. *New IRS Study Provides Preliminary Tax Gap Estimate*, I.R.S. News Release IR-2005-38 (Mar. 29, 2005), <https://www.irs.gov/uac/New-IRS-Study-Provides-Preliminary-Tax-Gap-Estimate> ("For individual underreporting, more than 80 percent comes from understated income, not overstated deductions."). One reason is that some taxpayers who are willing to skip an item may not be willing to affirmatively lie. M. Bernard Aidinoff et al., *Report and Recommendations on Taxpayer Compliance*, 41 *TAX LAW.* 329, 376 (1988). In addition, inflated amounts that appear on the return are easier for the IRS to audit than wholly unreported items are.

151. See INTERNAL REVENUE SVC., *TAX GAP FOR TAX YEAR 2006 OVERVIEW* (Jan. 6, 2012), [https://www.irs.gov/pub/newsroom/overview\\_tax\\_gap\\_2006.pdf](https://www.irs.gov/pub/newsroom/overview_tax_gap_2006.pdf).

152. Of course, much of that reporting could be deterrence-motivated, including reporting amounts received with a paper trail, such as via credit card.

Part of that level of compliance may be due to deterrence aspects that are missing from the very basic model described above, such as criminal penalties. First, individuals may be risk averse about the prospect of going to prison, as well as the stigma associated with criminal tax evasion.<sup>153</sup> They may also overestimate the likelihood of prosecution, which is very small.<sup>154</sup> Two empirical studies using IRS data found that criminal tax enforcement increases tax compliance.<sup>155</sup>

Second, the simple economic model applied above treated the audit rate as exogenously determined, which is not realistic. The likelihood that a taxpayer will be audited is endogenous, in that it depends on what the taxpayer reports.<sup>156</sup> The IRS does some random audits for research purposes and selects some taxpayers for audit for other reasons, but its primary approach is to score tax returns using a secret formula designed to detect which returns are most noncompliant.<sup>157</sup> Thus, an average audit rate may be misleadingly low as applied to someone engaging in tax evasion.

---

153. See Steven Klepper & Daniel Nagin, *The Anatomy of Tax Evasion*, 5 J.L. ECON. & ORG. 1, 21 (1989) (“For tax evasion, fear of stigmatization may be a particularly powerful deterrent because most taxpayers have no criminal record.”); see also Jeffrey A. Dubin, *Criminal Investigation Enforcement Activities and Taxpayer Noncompliance*, 35 PUB. FIN. REV. 500, 523 (2007) (finding that incarceration and probation have a greater influence than fines on tax compliance).

154. Between fiscal years 2007 and 2016, the IRS initiated an average of 2621 tax investigations each year, peaking at 3325 in 2013. *Enforcement Statistics*, IRS, <https://www.irs.gov/compliance/criminal-investigation/enforcement-statistics-criminal-investigation-ci-enforcement-strategy> (last updated Sept. 30, 2017).

155. See Dubin, *supra* note 153, at 502 (finding statistically significant positive results from IRS Criminal Investigation prosecutions resulting in sentences on tax compliance); Alan H. Plumley, *The Determinants of Individual Income Tax Compliance: Estimating the Impacts of Tax Policy, Enforcement, and IRS Responsiveness*, I.R.S. Publication 1916, at 36 (Nov. 1996), <https://www.irs.gov/pub/irs-soi/pub1916b.pdf> (finding that U.S. states with a higher proportion of criminal tax convictions have higher compliance with tax-reporting obligations, and also finding that criminal convictions have “an equally significant, but smaller, positive impact on offsets reporting”).

156. See, e.g., James Alm et al., *Tax Compliance with Endogenous Audit Selection Rules*, 46 KYKLOS 27, 27 (1993) (“[T]he government tax agency does not select tax returns randomly for audit but instead uses information from the returns to determine strategically whom to audit.”).

157. Leigh Osofsky, *Concentrated Enforcement*, 16 FLA. TAX REV. 325, 335–36 (2014) (“[V]arious government authorities have indicated that the IRS . . . employs a ‘worst-first’ approach by focusing on taxpayers who are likely to be the most noncompliant, as determined by deviation from others.”).

Sophisticated models of tax compliance treat the audit rate as endogenous.<sup>158</sup>

Third, that very basic economic model ignores the possibility of “non-pecuniary costs of evading taxes such as embarrassment, loss of self-esteem, and social status.”<sup>159</sup> A taxpayer may also increase compliance to try to avoid the time, hassle, and expense involved in an audit.<sup>160</sup>

However, other factors undoubtedly affect tax compliance. For example, some studies have shown a limited positive effect of procedural fairness by the tax collector.<sup>161</sup> A cross-country study in Latin America found a positive effect of better governance institutions, particularly citizen “voice” in government, supporting Swiss scholar Bruno Frey’s argument that higher tax morale increases tax compliance.<sup>162</sup> Yet studies generally do not show a positive effect of moral suasion letters that focus on the public goods that the government provides to taxpayers, such as

---

158. See, e.g., Allingham & Sandmo, *supra* note 103, at 331–32 (including a model with audit rate endogenously determined); Andreoni et al., *supra* note 126, at 824–31 (discussing two variations on a model, one in which the tax authority pre-commits to an audit rule and one in which it does not).

159. Sharmila King & Steven M. Sheffrin, *Tax Evasion and Equity Theory: An Investigative Approach*, 9 INT’L TAX & PUB. FIN. 505, 507 (2002).

160. See Helen V. Tauchen, *Tax Compliance: An Investigation Using Individual Taxpayer Compliance Measurement Program (TCMP) Data*, 9 J. QUANTITATIVE CRIMINOLOGY 177, 194 (1993) (“Fears of bureaucracy, time costs, and other factors associated with an audit may play as important a role in securing compliance as do the monetary penalties that might result from an audit.”); cf. Stephan Muehlbacher et al., *Uncertainty Resolution in Tax Experiments: Why Waiting for an Audit Increases Compliance*, 41 J. SOCIO-ECON. 289, 290 (2012) (finding higher tax compliance in a lab experiment if the random audits were conducted three weeks later instead of immediately).

161. See Lederman, *supra* note 128, at 1000–04 (discussing studies of this issue).

162. See Richard M. Bird et al., *Societal Institutions and Tax Effort in Developing Countries*, 15 ANNALS OF ECON. & FIN. 185, 209 (2014). The voice variables are statistically significant, generally at  $p < 0.01$ . See *id.* at 211 tbl.2. That study also found a positive effect of a “Tax Morale” variable that was based on the per-country level of self-reported belief in the justifiability of tax cheating. See *id.* at 201, 207, 216 tbl.7 (defining the terms and the variable; reporting effects, which were significant at  $p < 0.05$  for “Revenue Effort,” defined as current revenues/gross domestic product, but not significant for “Tax Effort”).

education and snow removal,<sup>163</sup> even in Switzerland,<sup>164</sup> where tax morale is stated to be high.<sup>165</sup>

By contrast, several field studies have shown that appealing to compliance norms may have a positive effect.<sup>166</sup> In fact, norms-based appeals may bolster enforcement efforts.<sup>167</sup> A 2009 study of individuals from the neighboring countries of Botswana and South Africa<sup>168</sup> supports the idea that enforcement efforts are more effective where compliance norms are stronger. The study first compared the self-reported commitment to honest taxpaying among individuals in the two countries and found self-reported honesty to be higher in Botswana than in South Africa.<sup>169</sup> That suggests a stronger taxpaying norm in Botswana.

The study then compared the simulated tax compliance decisions of individuals from the two countries because they have

163. See, e.g., Arian, *supra* note 73, at 27, 41–45, 58 (finding that a letter to Israeli corporations, explaining how tax dollars were allocated, providing reasons for paying taxes, and highlighting the societal harm from not paying, resulted in a small but statistically significant effect in the direction of *non*compliance with a Value Added Tax (VAT)); STEPHEN COLEMAN, MINN. DEP'T OF REVENUE, THE MINNESOTA INCOME TAX COMPLIANCE EXPERIMENT: STATE TAX RESULTS 5 (1996), [http://www.revenue.state.mn.us/research\\_stats/research\\_reports/19xx/research\\_reports\\_content\\_complnce.pdf](http://www.revenue.state.mn.us/research_stats/research_reports/19xx/research_reports_content_complnce.pdf) (finding no effect of letter stating “your income tax dollars are spent on services that we Minnesotans depend on” and listing items). *But cf.* Michael Chirico et al., *An Experimental Evaluation of Notification Strategies to Increase Property Tax Compliance: Free-Riding in the City of Brotherly Love*, in 30 TAX POL'Y & ECON. 129, 146 tbl.5 (Jeffrey R. Brown ed.) (2016), <http://www.nber.org/chapters/c13690.pdf> (finding that a Public Service letter had a statistically significant effect at  $p < 0.05$  on one subgroup, single-property owners). A U.K. study combined a norms-based appeal and moral suasion into one letter, confounding the analysis. See John Hasseldine et al., *Persuasive Communications: Tax Compliance Enforcement Strategies for Sole Proprietors*, 24 CONTEMP. ACCT. RES. 171, 178 (2007). That letter was significant only for self-preparers (at  $p < 0.01$ ). *Id.* at 181 tbl.1, 184.

164. See Benno Torgler, *Moral-Suasion: An Alternative Tax Policy Strategy? Evidence from a Controlled Field Experiment in Switzerland*, 5 ECON. GOVERNANCE 235, 239–40 (2004). Torgler found an insignificant effect on timely filing or payment in Trimbach, Switzerland, of this moral appeal: “If the taxpayers did not contribute their share, our commune with its 6226 inhabitants would suffer greatly. With your taxes you help keep Trimbach attractive for its inhabitants.” *Id.* at 240, 249–50 (emphasis omitted).

165. See *supra* text accompanying notes 118–120.

166. See Lederman, *supra* note 70, at 1471–75 (describing these studies).

167. See *id.* at 1497–99.

168. Cummings et al., *supra* note 116, at 447.

169. *Id.* at 451–52. “The marginal effects indicate that being a resident of Botswana rather than of South Africa increases the probability of reporting the highest tax honesty by around 6 percentage points, and this result is robust . . .” *Id.* at 452. This is significant at  $p < 0.01$ . *Id.* at 452 tbl.3.



similar tax systems, but South Africa is much less politically stable.<sup>170</sup> The authors framed the experiment as a tax-reporting decision where reporting could be audited and a fine imposed on undisclosed income.<sup>171</sup> The audit rate was endogenously determined based on the amount of income reported.<sup>172</sup>

The study found greater compliance by individuals in Botswana than those in South Africa, as the authors expected.<sup>173</sup> They attributed that result to higher tax morale in Botswana, which has better governance.<sup>174</sup> However, what the study actually tested was the effect of differing compliance norms and cultures on a simulated taxpaying decision.<sup>175</sup> Because the two countries are similar apart from their quality of governance, it is logical to hypothesize that the better government resulted in stronger taxpaying norms.<sup>176</sup> However, the authors did not test that, and it is theoretically possible that stronger taxpaying norms (resulting from some other cause) and the resulting higher tax payments produced better governance structures.<sup>177</sup>

Regardless, what is most important from a deterrence perspective is that the Botswana/South Africa study also found that, for both countries, compliance rates increase with an increase in the audit or penalty rate.<sup>178</sup> This suggests that deterrence does not crowd out intrinsic motivations to pay taxes. In fact, the authors found that “while compliance does increase with enforcement effort, the effect is less in the country for which governance is less good.”<sup>179</sup> This suggests that enforcement and tax morale—if that is

---

170. *Id.* at 449–50.

171. *Id.* at 453.

172. *Id.* at 454.

173. *Id.* at 455–56. This effect is significant at  $p < 0.05$ . *Id.*

174. *Id.* at 450 (“Botswana . . . refers to itself as the ‘gem of Africa’ in many . . . publications.” (citation omitted)).

175. *Id.* at 449, 452 (“[T]he objective of our research is to examine the effect of cultural factors and social norms on tax compliance behavior . . .”).

176. The authors state that “[i]f these norms evolve from perceptions that the quality of governance is high, we predict we will observe differences in tax compliance that are correlated with these perceptions.” *Id.* at 449.

177. *Cf.* Bird et al., *supra* note 162, at 209 (pointing out these two possibilities in their study).

178. Cummings et al., *supra* note 116, at 454. The effect of the audit rate is significant at  $p < 0.01$ . *Id.* at 455 tbl.6. The significance of the penalty rate is not reported.

179. *Id.* at 448. This study thus supports the argument that enforcement complements compliance norms. *See* Lederman, *supra* note 70.

what is driving the differing compliance rates across the two countries – may be *complements*, not substitutes.

And even if higher tax morale stemming from good government increases compliance, as Frey argues, this section has shown that that does not by any means counter the effectiveness of deterrence. The next section adds to the force of the deterrence model by showing the positive effect on taxpayer compliance of audits, which are a key component of the deterrence model, and the smaller, but still generally positive, effect of sanctions.

### B. The Effectiveness of Particular Deterrence Tools

#### 1. Audits and audit threats

Generally speaking, the existence of audits has been found to have a positive effect on tax compliance. Several studies have found that increasing the audit rate increases compliance.<sup>180</sup> In the United States, the IRS has found both a direct effect of audits on tax collections<sup>181</sup> and an indirect effect of audits of approximately six dollars for each dollar collected directly through enforcement.<sup>182</sup>

---

180. See James Alm, *Tax Compliance and Administration*, in HANDBOOK ON TAXATION 741, 756 (W. Bartley Hildreth & James A. Richardson eds., 1999) (“Nearly all studies have found that a higher (random) audit rate leads to more compliance . . . ; however, . . . this impact appears to be small and nonlinear, so that the deterrent effect of a higher audit rate eventually diminishes.” (citation omitted)); see also, e.g., Alm et al., *supra* note 135, at 135–36 (finding in lab experiment that “[t]he compliance rate increases by 3.8 percentage points for an increase in the audit success rate of 25 percentage points”); Nipoli Kamdar, *Corporate Income Tax Compliance: A Time Series Analysis*, 25 ATLANTIC ECON. J. 37, 45–46 (1997) (finding for the sample period of 1961 to 1987 that “audits play a statistically significant role in promoting corporate [tax] compliance”); Plumley, *supra* note 155, at 35–36 (finding an 11.7 to 1 indirect effect of audits started in 1991). Note that laboratory experiments, which use computer simulations, can determine the participant’s actual compliance. Field experiments cannot, so they generally compare changes in reported income (or some other reported amount).

181. The National Taxpayer Advocate reported that the direct effect of IRS audits of 1.5% of self-employed taxpayers in fiscal year 2014 was more than \$3 billion in recommended additional taxes. NAT’L TAXPAYER ADVOCATE, *Audit Impact Study*, in 2 TAS RESEARCH AND RELATED STUDIES 67, 68 (2015), [http://www.taxpayeradvocate.irs.gov/Media/Default/Documents/2015ARC/ARC15\\_Volume2\\_3-AuditImpact.pdf](http://www.taxpayeradvocate.irs.gov/Media/Default/Documents/2015ARC/ARC15_Volume2_3-AuditImpact.pdf). For fiscal years 2008–10, the IRS anticipated collecting an average of \$52 billion via late payments and enforcement actions. See INTERNAL REVENUE SVC., *supra* note 113, at 3.

182. See *Department of the Treasury: Internal Revenue Service*, in APPENDIX, BUDGET OF THE UNITED STATES GOVERNMENT, FISCAL YEAR 2016, at 1035, 1035, <https://www.gpo.gov/fdsys/pkg/BUDGET-2016-APP/pdf/BUDGET-2016-APP-1-19.pdf>; see also Jeffrey A. Dubin et al., *The Effect of Audit Rate on the Federal Individual Income Tax, 1977–1986*, 43 NAT’L

Moreover, a recent study found that an increase in the IRS's face-to-face audit rate has a positive spillover effect on individuals' compliance with *state* income taxes.<sup>183</sup> This makes sense, given that state income tax regimes generally use the federal income tax base as a starting point.<sup>184</sup>

The positive effect of audits is not limited to the United States. For example, Kleven et al. conducted an experiment in Denmark involving both actual audits and letters threatening audit. They thoroughly audited half of a representative sample of more than 42,000 individual income taxpayers, some employees and some self-employed.<sup>185</sup> They found a statistically significant positive effect of the audits on income reported in the following year. The difference was entirely due to income not subject to third-party reporting,<sup>186</sup> which is the income that is less visible absent an audit.

The year after the audits, Kleven et al. subdivided the audited and unaudited taxpayers who were employees into three groups each.<sup>187</sup> The subdivisions were for a group that got no letter, a group that was sent a letter informing its recipients that 50% of the group would be audited, and a group sent a letter stating that the recipient's return definitely would be audited.<sup>188</sup> The study found that "audit threats have a positive impact on self-reported income"<sup>189</sup> and that the effect was about twice as strong for the guaranteed audit than for the 50% audit threat.<sup>190</sup> The effect of

---

TAX J. 395, 405 (1990) (finding that the indirect effect of audits is responsible for six out of every seven dollars of tax revenue); cf. Tauchen, *supra* note 160, at 179 (finding a two-to-one indirect effect in a 1979 dataset).

183. Liucija Birskyte, *Effects of Tax Auditing: Does the Deterrent Deter?*, 8 RES. J. ECON., BUS. & ICT 1, 6 (2013). The study used aggregate data from all forty-one states that levied a broad-based income tax. *Id.* at 3. The study found that "[a] 1% increase in federal audit rate, on the average, increases individual income state tax collected per return by 1.74 dollars, holding other variables constant." *Id.* at 5. This was significant at the  $p < 0.01$  level. *Id.*

184. See Heather M. Field, *Binding Choices: Tax Elections & Federal/State Conformity*, 32 VA. TAX REV. 527, 532 (2013).

185. Kleven et al., *supra* note 126, at 660.

186. *Id.* at 654. "For the full sample, the effect on total net income is 2557 kroner or 30.1 cents per additional kroner of audit adjustment. The effect on tax liability is 1375 kroner, corresponding to 41.7 cents per dollar of audit adjustment. These estimates are strongly significant." *Id.* at 685.

187. *Id.* at 653, 662.

188. *Id.* at 653, 662-63.

189. *Id.* at 654.

190. *Id.* at 688.

receiving an audit-threat letter was similar for the group that had been audited the prior year and the group that had not.<sup>191</sup> Kleven et al. explained:

Because 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. We may think of the detection probability as a product of two probabilities: the probability of audit and the probability of detection conditional on audit. Audits may have an effect through both channels.<sup>192</sup>

The Chilean VAT study mentioned above included an audit-with-prior-notice treatment. The tax authority sent half of a sample of companies with reporting patterns suggesting tax evasion<sup>193</sup> prior notice of an audit, then summoned the whole group for an audit.<sup>194</sup> The study found that the deterrent effect of the audit announcement had valuable spillover effects on companies in the supply chain: “The randomly administered audit announcement leads to strong spillover effects that increase VAT payments by the suppliers of the treated firms. In line with the asymmetric incentives between clients and suppliers in the self-enforcing mechanism of the VAT, tax payments of client firms do not increase.”<sup>195</sup> The pro-compliance spillover effects are consistent with both IRS data on the U.S. federal income tax<sup>196</sup> and the study that found a positive spillover of IRS audits onto state tax collections.<sup>197</sup>

John Hasseldine and his coauthors studied sales and net profit reporting in the United Kingdom by sole proprietors who reported sales under £15,000 for two consecutive tax years.<sup>198</sup> That amount reflects an important reporting threshold: individuals and partnerships with sales less than that amount were required to

---

191. *Id.* (“The broad conclusion from these estimates is that letter effects are roughly the same in the 0% and 100% audit groups.”).

192. *Id.* at 681.

193. These patterns included things such as continually reporting sales less than costs, without going out of business. Pomeranz, *supra* note 140, at 2549.

194. *Id.* at 2541, 2550.

195. *Id.* at 2541.

196. *See supra* note 182 and accompanying text.

197. *See supra* note 183 and accompanying text.

198. Hasseldine et al., *supra* note 163, at 171–72.

provide the tax authority only “turnover (sales), allowable deductions, and net profit,” whereas other businesses had to itemize all their income and expenses—a more burdensome reporting obligation.<sup>199</sup>

Hasseldine et al. studied five types of letters from Inland Revenue, including three sanction letters: (1) an audit threat, (2) a “threat of audit with possible penalties,” and (3) an audit with prior notice (described as a “virtual guarantee of forthcoming audit once the return is filed”).<sup>200</sup> They found that each of the three sanction letters increased the sole proprietors’ reported net profits, compared to the control group.<sup>201</sup> The three sanction letters also all had a statistically significant effect on turnover reported over the threshold.<sup>202</sup>

As discussed above, in a recent church tax study in Germany, the auditors knew the amount of tax due and thus could determine if a taxpayer under complied, fully complied, or over complied.<sup>203</sup> One set of treatments varied the audit probabilities announced in a letter between 10, 20, and 50%, as well as including notched probabilities stating that taxpayers faced an audit probability of 50% if they pay €10 or less, and zero chance of audit otherwise.<sup>204</sup>

The study found that this deterrence intervention had “strong effects on compliance for baseline evaders, but small and mostly insignificant effects for baseline donors”<sup>205</sup>—those who had overpaid the tax.<sup>206</sup> In addition, the notched audit letter was very effective at increasing tax payments to more than €10.<sup>207</sup> It is possible that taxpayers treat a church tax differently from other types of taxes. However, it is worth noting that this study not only

---

199. *Id.* at 177.

200. *Id.* at 172.

201. The “respective mean changes (*F*-statistics) of [Letters 3, 4, and 5 were] . . . £438 (4.303), £731 (19.431), and £976 (31.156).” *Id.* at 184. These increases were statistically significant at  $p < 0.05$ . *Id.*

202. *Id.* at 181 ( $p < 0.05$ ). While 39.7% of the control group reported turnover above the threshold, the sanction groups were all above 50%. *Id.* (“The respective percentages ( $\chi^2$ -statistics) for those four groups were . . . 53.8 percent ( $\chi^2 = 60.929$ ), 56.4 percent ( $\chi^2 = 82.075$ ), and 53.7 percent ( $\chi^2 = 42.575$ ).”).

203. See Dwenger et al., *supra* note 42, at 204.

204. *Id.* at 223.

205. *Id.* at 205.

206. See *id.* at 204.

207. *Id.* at 223.

supports the effectiveness of audit threats but also the possibility of a slight “crowding in” of intrinsic motivations.<sup>208</sup> This effect is similar to the study comparing Botswana and South Africa, which found that enforcement increased compliance more in Botswana, the country with stronger compliance norms.<sup>209</sup>

The Minnesota Tax Compliance Experiment, which examined the effects of several different letters on individual taxpayers, found a slightly more mixed effect of an audit with prior notice (a letter to a group of taxpayers stating that his or her return would be closely examined, which it subsequently was).<sup>210</sup> That letter increased reported income and taxes paid for about 96.7% of the population – all but high-income taxpayers.<sup>211</sup> Moreover, the audit with prior notice, which had a pro-compliance effect on taxpayers in the low- and middle-income strata, had a greater effect on low- and middle-income taxpayers who were “high-risk”<sup>212</sup> – those filing a federal Schedule C for self-employment income or Schedule F for farm income.<sup>213</sup>

However, the audit letter’s results were mixed and possibly negative for the 3.3% of the population defined as “high income” (those who have federal adjusted gross income of more than \$100,000).<sup>214</sup> A recent study by Mazzolini et al. found a similar result in Italy on the effect of actual audits: “[T]he average positive audit effect that we detected . . . is driven by low and middle reported-income taxpayers, while the effect is even negative at the highest decile.”<sup>215</sup>

---

208. *Id.*

209. *See supra* text accompanying note 179.

210. COLEMAN, *supra* note 163, at 10–12.

211. *Id.* at 10–12, 22. Once the results were weighted to make the sample proportional to the population, the additional taxes, when computed for the population of Minnesota, would amount to \$73 million for 1993. Stephen Coleman, *Income Tax Compliance: A Unique Experiment in Minnesota*, GOV’T FIN. REV., Apr. 1997, at 5.

212. COLEMAN, *supra* note 163, at 12 (finding high-risk low- and middle-income taxpayers increased reported taxes by \$186 more than the controls, on average, compared to \$36 for the low-risk low- and middle-income taxpayers).

213. *Id.* 2.

214. *Id.* at 11.

215. Gabriele Mazzolini et al., *The Deterrence Effect of Real-World Operational Tax Audits* 21, 45 tbl.A1 (Dep’t of Econ. Mgmt. & Statistics, Univ. of Milan–Bicocca, Working Paper No. 359, Jan. 2017), <http://dems.unimib.it/repec/pdf/mibwpaper359.pdf>.

A recent study by Jason DeBacker et al. of U.S. C corporations found an analogous result: corporations receiving a positive adjustment (increase in taxes) in an IRS audit “subsequently become more tax aggressive than when they receive a zero or negative adjustment.”<sup>216</sup> By contrast, their study of individuals found that a positive adjustment increased compliance in the first years after the audit.<sup>217</sup> Taken together, the DeBacker, Mazzolini, and Minnesota findings suggest that an audit may increase tax aggressiveness in taxpayers with the most resources (who presumably can afford to have a tax adviser and may be more likely to treat tax reporting as a negotiation<sup>218</sup>).

In addition, although Minnesota’s audit letter had a positive effect on taxpayers who had had their prior year’s taxes adjusted, it had a negative effect on the subgroup who had paid a penalty the year before. Moreover, the higher the penalty had been, the larger the effect was.<sup>219</sup> This could suggest an attempt at loss recoupment, an effect discussed in Part III of this Article.

A recent study using country-level data hypothesized and found a “U-shaped association between auditing<sup>220</sup> and tax evasion,”<sup>221</sup> meaning “that compliance increases until a certain auditing level is reached, and decreases beyond that level (i.e., an elevated auditing level backfires).”<sup>222</sup> However, it is important to note that that study used as its measure of tax evasion “business executives’ perception of the extent to which tax evasion is common

---

216. Jason DeBacker et al., *Legal Enforcement and Corporate Behavior: An Analysis of Tax Aggressiveness After an Audit*, 58 J. LAW & ECON. 291, 311 (2015).

217. See *infra* text accompanying note 343.

218. 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, 480–81 (2001).

219. COLEMAN, *supra* note 163, at 11. Results like this suggest that in studies along the lines of the Israeli Day-care Center Study, it is worth examining separately the effects of the fine regime on those who have been fined and those who have not been fined. See *supra* notes 61 and 68.

220. The overall audit rate is “the number of verification actions” divided by “the total number of registered taxpayers for personal and corporate income taxes.” Juan P. Mendoza et al., *The Backfiring Effect of Auditing on Tax Compliance*, 62 J. ECON. PSYCHOL. 284, 287 (2017). Thus, it is a blend of individual and corporate audit rates. “[Eighty-five percent] of the untransformed auditing level observations lie between 0.2 and 20 verification actions per 100 taxpayers.” *Id.* at 288.

221. *Id.* at 284.

222. *Id.* at 285.

practice in their country,"<sup>223</sup> not actual government estimates of tax evasion rates. Thus, what the study actually found is a correlation between an overall audit rate that is higher than the average for the countries studied<sup>224</sup> and a higher perception of tax evasion on the part of business executives. This outcome could suggest that the correlation runs the other way: a comparatively high audit rate causes businesspeople to perceive that noncompliance is high.<sup>225</sup> The authors try to address that issue but state that "[a]lthough [various] tests provide support for our hypothesis, endogeneity cannot be fully ruled out with the available data . . . ."<sup>226</sup>

223. *Id.* at 286. "Respondents are top and middle management executives." *Id.* at 287.

224. The study includes fifty countries, including many countries that belong to the Organisation for Economic Development. *See id.* at 289 tbl.2 (listing countries ranging from Argentina to United States).

225. That would presumably be because a high audit rate is a signal that the government has determined that the enforcement rate needs to be high. A comparatively low audit rate could also cause businesspeople to perceive that noncompliance is high because of the risk that enforcement levels are too low to provide significant deterrence.

226. Mendoza et al., *supra* note 220, at 292. To deal with endogeneity, the Mendoza et al. study used "a two-step system GMM estimator (Roodman, 2006), which uses lags as instruments of possibly endogenous regressors." *Id.* at 291. They note that "the number of instruments does not exceed the number of countries (i.e.,  $47 < 50$ ) . . ." *Id.* This comports with the "rule of thumb . . . to worry if the instrument count exceeds  $N$ ." David Roodman, *How to Do xtabond2: An Introduction to "Difference" and "System" GMM in Stata 43* (Ctr. for Glob. Dev., Working Paper No. 103, Dec. 2006), <http://dx.doi.org/10.2139/ssrn.982943>. However, in a companion paper that Mendoza et al. also cite, Roodman explains that that limit is too high when the equation has endogenous regressors, and that the two-step system GMM estimates become particularly problematic as the number of instruments approaches  $N$  (in this case, the number of countries). *See* David Roodman, *A Note on the Theme of Too Many Instruments*, 71 OXFORD BULL. ECON. & STAT. 135, 140 (2009); *see also id.* at 142 (citing other studies that "suggest that merely keeping the instrument count below  $N$  does not safeguard the [Hansen]  $J$ -test"). Roodman also provides examples of this problem using replications of published empirical work. One study Roodman replicates "includes 75 instruments, compared with 77 countries and 353 observations." *Id.* at 153. He runs tests that reveal downward biases of standard errors caused by the large number of the instruments in that study and finds that, if the number of instruments is reduced, the Hansen  $J$  statistic declines dramatically, suggesting invalidity of system GMM instruments. *Id.* at 155. And if the problematic system GMM instruments are dropped, several explanatory variables lose their statistical significance. Roodman concludes that the "facts suggest that instrument invalidity is the source of the . . . results." *Id.* at 156. The author thanks Michael Alexeev for pointing out this issue.

It would be helpful to examine Mendoza et al.'s analysis of the data. The article states that "[i]nstructions to access the dataset and conduct the analysis are presented in the Appendix." Mendoza et al., *supra* note 220, at 286. However, the appendix does not seem to accompany the article.



Overall, these results suggest that an audit regime and audit threats generally are effective deterrents. The next section focuses on the effects of penalties, and Part III discusses the direct effect of audits on the audited taxpayer's subsequent reporting.

## 2. Penalties

*a. Public goods games.* The effects of penalties or penalty threats may not be the same as the effects of audits or audit threats. However, laboratory experiments involving contributions to a public good provide support for the beneficial effects of sanctions.<sup>227</sup> In these games, each participant is given an initial sum and the opportunity to invest some or all of it in a group fund. The experimenter will multiply the pot by a positive number and divide the resulting amount equally among all participants regardless of who contributed to the pot. Typically, the game is played for a set number of rounds.<sup>228</sup> The game is structured so that the best result for the group is if everyone contributes the full stake provided by the experimenter, while each individual's maximizing strategy is to contribute nothing, free riding on the others' contributions.<sup>229</sup>

These experiments have found that many participants do contribute, and on average, players start by contributing 40 to 60% of their stake.<sup>230</sup> In later rounds, players usually mirror others' contributions.<sup>231</sup> Because some players free ride from the outset,

---

227. The experimental games discussed in this section all involve public goods, with players contributing to a common pool. Some of the initial experiments, however, such as ultimatum games, are two-party, reciprocity games. Houser et al. found a negative effect of punishment in a reciprocity context. Daniel Houser et al., *When Punishment Fails: Research on Sanctions, Intentions and Non-Cooperation*, 62 *GAMES & ECON. BEHAV.* 509, 523 (2008) (concluding in a gift-exchange game, where an "investor" sends an amount to a "trustee," which the experimenter triples, that trustees responded negatively to monetary incentives, thus crowding out pro-social behavior).

228. See Lederman, *supra* note 70, at 1482.

229. See Elinor Ostrom, *Collective Action and the Evolution of Social Norms*, 14 *J. ECON. PERSP.* 137, 139 (2000) (explaining that a rational player rationally should contribute zero in the last round, and thus the penultimate round, and so on).

230. *Id.* at 140.

231. *Id.* at 142. One study found that approximately half of the players increased their contributions in response to increases in others' contributions, 14% did the same up to certain contribution levels and then decreased their contributions, and approximately one-third of players were free riders. Urs Fischbacher et al., *Are People Conditionally Cooperative? Evidence from a Public Goods Experiment*, 71 *ECON. LETTERS* 397, 398 (2001).

and many mirror others' behavior, the typical basic game results in a progressive decline in contributions over the ten rounds.<sup>232</sup>

By contrast, when players can pay to punish defectors, that possibility generally dramatically increases cooperation.<sup>233</sup> A study of the same public goods game in sixteen different countries found that a punishment option resulted in differing levels of cooperation in different countries, but in all of them, the availability of punishment prevented a breakdown in cooperation.<sup>234</sup> Another experiment found that the least-trusting players were the most likely to become strong contributors where there was a punishment mechanism.<sup>235</sup> One study found that punishment by a third party – someone who had previously played the game – had a weaker effect than punishment by other players, but it still significantly increased contributions.<sup>236</sup>

Mizuho Shinada and Toshio Yamagishi examined the mechanism by which punishment works. Their study of two one-shot prisoner's dilemma games suggested that punishment has both a *direct* effect on cooperation by altering its payoffs and an *indirect* effect because it increases the expectation that others will cooperate.<sup>237</sup> Moreover, this was true whether punishments were administered by other players or by a third party (the

232. Ostrom, *supra* note 229, at 140 (reporting that typically 70% of players contribute nothing in the pre-announced final round of a public goods game).

To test whether players are learning the optimal strategy through successive rounds, some experiments included a surprise "restart." See James Andreoni, *Why Free Ride? Strategies and Learning in Public Goods Experiments*, 37 J. PUB. ECON. 291, 300 (1988). They found that players returned to high levels of giving in round eleven, behaving similarly to round one, ruling out the possible explanation that it took players multiple rounds simply to learn the incentives in the game. *Id.* at 298.

233. Ernst Fehr & Simon Gächter, *Cooperation and Punishment in Public Goods Experiments*, 90 AM. ECON. REV. 980 (2000); Colin F. Camerer & Ernst Fehr, *Measuring Social Norms and Preferences Using Experimental Games: A Guide for Social Scientists* 11–12 (Inst. for Empirical Research in Econ., Univ. of Zurich, Working Paper No. 97, Jan. 2002), <https://www.zora.uzh.ch/id/eprint/51997/1/iewwp097.pdf>.

234. Benedikt Herrmann et al., *Antisocial Punishment Across Societies*, 319 SCI. 1362 (2008); Benedikt Herrmann et al., *Supporting Material for Antisocial Punishment Across Societies* 33 (Feb. 12, 2008), <http://science.sciencemag.org/content/sci/suppl/2008/03/06/319.5868.1362.DC1/Herrmann.SOM.pdf>

235. Fehr & Gächter, *supra* note 233, at 984–85; see also Ostrom, *supra* note 229, at 141.

236. Camerer & Fehr, *supra* note 233, at 13.

237. Mizuho Shinada & Toshio Yamagishi, *Punishing Free Riders: Direct and Indirect Promotion of Cooperation*, 28 EVOLUTION & HUM. BEHAV. 330, 331–34 (2007).

experimenter).<sup>238</sup> This is consistent with economic modeling of conditional cooperators<sup>239</sup> and with IRS data suggesting that enforcement actions have a positive indirect (or “shadow”) effect on other taxpayers.<sup>240</sup>

In sum, these studies suggest that the availability of costly punishment is very effective at increasing cooperation in an experimental setting.<sup>241</sup> The next section looks at whether this extends to the tax context in field experiments.

*b. Tax-focused experiments.* A number of experiments have examined the effects of sanctions on tax compliance. As mentioned above, an IRS study found that the proportion of criminal tax convictions positively influences taxpayer reporting of income, although that effect is somewhat reduced by increased reporting of offsets, such as credits.<sup>242</sup>

With respect to monetary sanctions, studies sometimes find a positive effect,<sup>243</sup> but they generally do not find as strong an effect of fines as they do of audit threats,<sup>244</sup> or they do not get statistically

---

238. *Id.* at 334–37.

239. See Jon S. Davis et al., *Social Behaviors, Enforcement, and Tax Compliance Dynamics*, 78 ACCT. REV. 39 (2003) (modeling this approach); Lederman, *supra* note 70 (arguing that this is what happens in the tax compliance context).

240. See *supra* note 182 and accompanying text.

241. Because punishment is costly, the cost of punishment could outweigh the gains to the group from increased cooperation. Simon Gächter et al., *The Long-Run Benefits of Punishment*, 322 SCI. 1510, 1510 (2008). However, punishment generally increases cooperation over time and thus reduces the use of punishment. Gächter et al. conducted experiments involving games with punishment available and without a punishment option. They found higher average contributions in the punishment-option games, and they also found greater benefits of punishment in a fifty-round game than in a ten-round game. *Id.*

242. See Plumley, *supra* note 155, at 36.

243. See Alm, *supra* note 180, at 756 (“A higher fine rate leads to marginally more compliance, with an estimated reported income-fine rate elasticity less than 0.1.”); cf. Robert Mason & Lyle D. Calvin, *Public Confidence and Admitted Tax Evasion*, 37 NAT’L TAX J. 489, 493 (2001) (finding from a survey of Oregon taxpayers that “[s]anction fear . . . is strongly related to compliance and suggests that the dissatisfied honest do not engage in tax fraud because they are afraid of getting caught”); Michael W. Spicer, *Civilization at a Discount: The Problem of Tax Evasion*, 39 NAT’L TAX J. 13, 15 (1986) (“Survey studies indicate that the incidence of tax evasion is inversely related to the fear of sanctions . . .”).

244. See Boris Maciejovsky et al., *Misperception of Chance and Loss Repair: On the Dynamics of Tax Compliance*, 28 J. ECON. PSYCHOL. 678, 684 (2007) (finding, in a lab experiment in Vienna, that “[e]ffect sizes suggest that audits have a stronger impact on compliance than fines”). The details of this experiment are discussed further below. See *infra* notes 270–280 and accompanying text.

significant results.<sup>245</sup> For example, a laboratory experiment conducted in the United States found that the severity of the fine (either twice or six times the evaded amount) had no effect on compliance.<sup>246</sup> “[I]n fact, tax evasion is less with a two-times-fine rate with a three in six audit probability than a six-times-fine rate with a one in six audit probability.”<sup>247</sup> Another U.S. lab experiment found that “[c]ompliance increases with an increase in the fine rate; however, the coefficient on FINERATE is so small that the fine rate elasticity is virtually zero, and the coefficient is also not highly significant.”<sup>248</sup>

Why might sanctions have a small effect? The authors of the latter study, Jim Alm, Betty R. Jackson, and Michael McKee, explain that “[s]ince the probability of detection is small, large responses to changes in the fine rate would require extreme degrees of risk aversion.”<sup>249</sup> Thus, a large fine may be little deterrent if there is little likelihood it will be imposed.

Alm, Jackson, and McKee note that “[a] policy implication is that increasing penalties may not have a noticeable effect on compliance, unless the probability of detection is increased significantly.”<sup>250</sup> A study in Washington State on reporting

245. See, e.g., Arial, *supra* note 163, at 43–52 (finding no statistically significant effect on corporate taxpayers of a letter “inform[ing] taxpayers that filing a false report will result in harsh sanctions”). The text of the letter is not included but is described. See *id.* at 43 (“The fact that tax noncompliance is a grave matter was emphasized more than once. . . . The types of available modes of punishment were highlighted. . . . Moreover, the letter informed the participant that the tax authority now uses new methods of auditing taxpayers. . . . [which] was intended to increase the perceived probability of detection and apprehension.”).

246. Paul Webley, *Audit Probabilities and Tax Evasion in a Business Simulation*, 25 ECON. LETTERS 267, 270 (1987).

247. *Id.*; see also Maciejovsky et al., *supra* 244, at 684 (“Effect sizes suggest that audits have a stronger impact on compliance than fines.”). *But cf.* Nehemiah Friedland et al., *A Simulation Study of Income Tax Evasion*, 10 J. PUB. ECON. 107, 110–11 (1978) (finding in study of fifteen Israeli undergraduates that increasing the fine rate from three times to fifteen times the amount evaded was a more effective deterrent than increasing the audit rate from one out of fifteen to five out of fifteen).

248. James Alm et al., *Estimating the Determinants of Taxpayer Compliance with Experimental Data*, 45 NAT’L TAX J. 107, 110 (1992).

249. *Id.* See also Chirico et al., *supra* note 163, at 150, 154 fig.2 (finding that a threat of seizure or sale of property written in large type resulted in positive effects, but they were small and statistically insignificant).

250. Alm et al., *supra* note 248, at 110.

compliance supports this suggestion.<sup>251</sup> The study examined the effects of “increasing sanction awareness and threats of enhanced detection risk” on payments by construction businesses of (1) an excise tax on gross revenues called the Business and Occupation (B&O) tax and (2) “use taxes” due on items purchased out of state but used in state.<sup>252</sup> The study found that receipt of either a “sanction-awareness” or “detection” letter<sup>253</sup> was associated with increased reporting of the base subject to use tax,<sup>254</sup> the effect they studied because of differences in applicable tax rates.<sup>255</sup> However, for both taxes, the penalty letter resulted in only marginally significant effects without a manipulation of detection likelihood.<sup>256</sup>

Overall, these studies suggest that audit threats are much more effective than sanction threats at increasing compliance. The general effectiveness of audit threats supports the notion that deterrence has positive rather than negative effects, as does the IRS’s data on the indirect effects of audits and the deterrent effect of criminal tax prosecutions.<sup>257</sup> However, although these overall effects are important, they do not isolate the effects of audit on the

---

251. Govind Iyer et al., *Increasing Tax Compliance in Washington State: A Field Experiment*, 63 NAT’L TAX J. 7 (2010).

252. *Id.* at 7–10 (parentheticals omitted) (“The use tax noncompliance rate is estimated at 18.1 percent while the B&O tax noncompliance rate is estimated at 2.2 percent.”).

253. The detection letter “explained that the DOR was trying to improve compliance with tax laws by examining tax returns to assess use tax underpayments. The detection manipulation also informed firms that the DOR required them to provide a detailed report of assets purchased and that these reports should be consistent with their Federal Depreciation Schedule . . .” *Id.* at 14.

254. *Id.* at 21. “[W]hen penalty information is communicated to the taxpayers (penalty/no detection scenario), the reported use tax base increases to 8.97 (significant at . . .  $p = 0.07$ .” *Id.* at 22. They also found “a significant effect for the Penalty\*Detection variable (. . .  $p = 0.01$ .” *Id.* at 21.

255. *Id.* at 16. The study also found a modest spillover of the detection letter onto compliance with B&O taxes. *Id.* at 29.

256. *Id.* at 29. More specifically:

When the taxpayers are asked to provide federal tax information (detection/no penalty scenario), the reported [logarithmic] use tax base variable increases [from 8.42] to 9.28 (significant at . . .  $p = 0.004$ ). Likewise, when penalty information is communicated to the taxpayers (penalty/no detection scenario), the reported use tax base increases to 8.97 (significant at . . .  $p = 0.07$ ).

*Id.* at 21–22; see also *id.* at 29 (“Penalty information [for B&O tax] is, as in the case of use tax, marginally significant in the absence of detection manipulation.”); *id.* at 22 (in use tax penalty/no detection manipulation,  $p$  was 0.07).

257. See Plumley, *supra* note 155; see also *supra* note 155 and accompanying text; *supra* notes 180–182 and accompanying text.

audited taxpayers themselves. That issue is discussed in the next Part.

### III. POST-AUDIT REDUCTION OF COMPLIANT BEHAVIOR

It is possible that the audit itself affects taxpayer reporting behavior in a way that audit threats or audits of others do not. Audited taxpayers' subsequent reporting behavior may also differ depending on aspects of the audit experience, such as whether the audit was random or not; whether the taxpayer was found undercompliant, fully compliant, or overcompliant in the audit; whether the taxpayer was required to pay a fine; and whether the taxpayer was subject to repeated audits. Several studies have examined the effect of audits on the future reporting behavior of those audited.<sup>258</sup> The next section discusses laboratory experiments, and the following section discusses field experiments.

#### A. The "Bomb-Crater Effect" in Laboratory Experiments

Several laboratory experiments have examined the effects of audits on post-audit compliance behavior. These experiments are structured such that participants retaining more lab dollars net of "taxes" are paid more for participating.<sup>259</sup> One of the earliest tested eight different conditions, each with sixty rounds.<sup>260</sup> Each participant, an economics student from the University of Trento,<sup>261</sup>

---

258. See, e.g., Brian Erard, *The Influence of Tax Audits on Reporting Behavior*, in WHY PEOPLE PAY TAXES: TAX COMPLIANCE AND ENFORCEMENT 95, 113 (Joel Slemrod ed., 1992) (studying the effects of IRS audits on subsequent-year reporting and reaching inconclusive results); *id.* at 97-98 (reporting the results of a study finding that taxpayers experiencing a random research audit of their 1969 returns were slightly less likely to be found noncompliant on their 1971 returns, but the magnitude of noncompliance was unaffected) (citing Susan B. Long & Richard D. Schwartz, *The Impact of IRS Audits on Taxpayer Compliance: A Field Experiment in Specific Deterrence*, Paper Presented at the Annual Law and Society Association Meeting (1987)); *infra* Part III.B (discussing other studies); see also *supra* text accompanying notes 211-214 (noting Minnesota Tax Compliance Experiment's finding that the audit letter had a positive effect on individuals who had had their prior year's taxes adjusted but a negative effect on those who had paid a penalty that prior year).

259. See, e.g., Luigi Mittone, *Dynamic Behavior in Tax Evasion: An Experimental Approach*, 35 J. SOCIO-ECON. 813, 814 (2006).

260. *Id.* at 816-18. One can think of those sixty rounds as corresponding to sixty years' worth of tax filing in the life of an individual.

261. Students typically have little taxpaying experience. A recent study in the United Kingdom conducted a taxpaying experiment on 200 students, 200 individuals who work as

was assigned an amount of gross income in a round, told the amount of taxes due, and notified with a pop-up window each time the audit probability for the next three rounds changed.<sup>262</sup> The experimenter, Luigi Mittone, randomly determined the audited rounds in advance.<sup>263</sup> However, each audit entailed an audit of that round plus the previous three rounds.<sup>264</sup> Mittone tested the effect of several factors, such as informing the participants that the taxes collected would be contributed to a scholarship fund.<sup>265</sup>

Mittone found, in part, that “the number of evaders was noticeably lower in the experiments with redistribution . . . and with a public good . . . than it was in those with no moral constraint . . . .”<sup>266</sup> In addition, he found the striking pattern that “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.”<sup>267</sup> He termed this the “bomb crater effect.”<sup>268</sup> This

---

employees and have their income fully reported by the employer and do not have to file a tax return, and 120 self-employed individuals who are required to file returns. C.Y. Lawrence Choo et al., *Do Students Behave Like Real Taxpayers in the Lab? Evidence from a Real Effort Tax Compliance Experiment*, 214 J. ECON. BEHAV. & ORG. 102, 103 (2016). It found that, in the low-fine treatments, students were the least compliant ( $p = 0.069$ ). *Id.* at 108.

262. Mittone, *supra* note 259, at 819. The income assigned to participants ranged from 0.36 Euro cents to 0.51 Euro cents in different rounds. The tax rate was 20, 30, or 40% in each round. The penalty for evasion was 4.5 times the evaded taxes. *Id.* at 818.

263. *Id.* at 819 (“rounds 13, 31, 34, 48, 54, 58 for the first group, and rounds 3, 24, 27, 40, 46, 50 for the second group. . .”).

264. *Id.* at 818.

265. *Id.* at 820.

266. *Id.* at 821.

267. *Id.* at 823.

268. *Id.* at 824. The term comes from the reported tendency of soldiers in World War I to shelter in bomb craters because they believed a bomb was unlikely to fall twice in the same place. *Id.* at 824 n.9. This 2006 article seems to use the term “bomb crater effect” as referring to a decline in compliance after an audit that occurs because participants think it’s unlikely they’ll experience another audit right away; it states, “This shall be called the ‘bomb crater effect’ (BCE): the subjects decide to evade immediately after a fiscal audit because they believe that it cannot happen twice in the same place (time).” *Id.* at 823–24 (footnote omitted).

Subsequently, Maciejovsky explored the possibility that “loss repair” explains the observed decline in compliance. See Maciejovsky et al., *supra* note 244, at 679. In a 2017 article Mittone coauthored, the Abstract states, “The experimental literature has identified the Bomb Crater Effect (BoCE), i.e., the fact that tax compliance drops immediately after a taxpayer is audited. From a theoretical perspective, BoCE has been explained either by the misperception of chance, also known as the gambler’s fallacy, or by the loss repair effect.” Luigi Mittone et al., *The Bomb-Crater Effect of Tax Audits: Beyond the Misperception of Chance*, 61 J. ECON. PSYCH. 225, 225 (2017). This Article uses the term in that later sense, as simply describing an observed decline in compliance, without referring to a possible cause.

effect, in terms of the median percentage of tax paid, was highly statistically significant.<sup>269</sup>

A study published the following year attempted to distinguish between two possible explanations of the bomb-crater effect: misperceived chance and recouping of audit losses.<sup>270</sup> Maciejovsky et al. recruited students in Vienna for two lab experiments. In the first experiment, participants earned income by buying or selling assets in a competitive market. "After each trading period, participants were asked to declare their earnings, separately for sales revenues, resulting from selling assets, and for dividends paid for holding assets."<sup>271</sup> Audit probabilities were 15% in one treatment and 30% in the other, and the penalty for evasion was 50% or 100% of the omitted income.<sup>272</sup> Maciejovsky et al. found that tax compliance was generally higher in the high-audit treatment than in the low-audit treatment.<sup>273</sup> In addition:

[C]ompliance decreased sharply after an audit and increased slowly in the next three consecutive trading periods in the high-audit condition. In the low-audit condition compliance decreased after an audit steadily. . . . It might well be that observing more than four subsequent filing periods would reveal an increase of compliance rates also in the low-audit condition. However, our design did not allow for such an analysis. These results suggest that participants' compliance decisions are driven by misconception of the audit probability, as shown in the literature on gambler's fallacy.<sup>274</sup>

In their second experiment, Maciejovsky et al. focused on whether the participants truly believed the audits were random.<sup>275</sup> Accordingly, one of the participants rolled a six-sided die, and all participants were audited if the roll was a one or a two, yielding a one-third chance of audit.<sup>276</sup> Participants were also assigned an income amount; the tax rate was 40%, and the penalty was 50% of

---

269. Mittone, *supra* note 259, at 825 ( $p = 0.0000$ ).

270. Maciejovsky et al., *supra* note 244, at 679.

271. *Id.* at 682.

272. *Id.*

273. *Id.* at 684.

274. *Id.*

275. *Id.* at 685–86.

276. *Id.* at 686.



undeclared income.<sup>277</sup> This experiment also found a bomb-crater effect.<sup>278</sup>

To test whether the bomb-crater effect was driven by loss recoupment, the authors examined the effects audits had on compliant versus noncompliant participants, presumably because noncompliant/penalized participants have greater losses to recoup after audit.<sup>279</sup> They found no differences.<sup>280</sup> Accordingly, because they had posited only one alternative explanation—misperceived chance—they concluded that misperceived chance drove the bomb-crater effect.

A subsequent lab experiment by Barbara Kastlunger et al. investigated the effects on compliance of various patterns of repeated audits.<sup>281</sup> One of the effects they examined was the bomb-crater effect, looking at whether it could better be explained by misperceived chance or loss recoupment.<sup>282</sup> Participants were given 1000 Experimental Currency Units in each of sixty rounds and told that the tax rate was 20%, the audit rate was 15%, and the fine for evasion was three times the unpaid tax (in addition to the unpaid tax). Each participant was audited nine times. In the control condition, the audits were randomly distributed over the sixty rounds. Experimental condition 1 (E1) experienced audits in all three of the first three rounds, with the remaining six audits in the rest of the first twenty rounds.<sup>283</sup> Experimental condition 2 (E2) had all of the audits concentrated between rounds 20 through 39, so

---

277. *Id.* at 686–87. “The exact amount [of income assigned] was randomly drawn from a uniform distribution ranging from 80 ECU [Experimental Currency Units] to 120 ECU, whereby 100 ECU was equivalent to 10 Euro.” *Id.* at 686.

278. *Id.* at 688. (“The average compliance rate dropped from 0.63 in the period of an audit to 0.56 immediately after an audit ( $t(49)=2.51; p<0.05; d=0.72$ ).”). Maciejovsky et al. also found that the return to baseline compliance after an audit was a quadratic function. *Id.*

279. *Id.* (“Loss repair can be detected if the drop of compliance for the dishonest taxpayers is significantly more pronounced than the corresponding drop for the honest taxpayers.”).

280. *Id.*

281. Barbara Kastlunger et al., *Sequences of Audits, Tax Compliance, and Taxpaying Strategies*, 30 J. ECON. PSYCHOL. 405, 405, 408 (2009) (studying economics students at the University of Trento).

282. *Id.* at 409.

283. *Id.* at 408. The only consecutive audits after that were in rounds 9 and 10. *Id.*

there were no audits at the beginning or end of the sixty rounds.<sup>284</sup> Experimental condition 3 (E3) had all of its audits in the first twenty rounds, but all audits were in even rounds (round 2, round 4, round 6, etc., through round 18).<sup>285</sup>

Overall, Kastlunger et al. found that both previous tax payments and fines had a slight positive effect on compliance in the next round, and audits had a substantial negative effect, with all of these effects significant at  $p < 0.01$ .<sup>286</sup> Although a bomb-crater effect existed for the control group (random audits), and for condition E3 (the fairly transparent alternating audits), conditions E1 and E2 lacked that effect. Instead, audits positively predicted compliance.<sup>287</sup> Accordingly, Kastlunger et al. determined that consecutive audits counter the bomb-crater effect.<sup>288</sup>

This suggested to them that misperceived chance causes bomb craters. Accordingly, like Maciejovsky, they disaggregated the effects of audits on compliant and noncompliant taxpayers in the control group.<sup>289</sup> They found,

[f]ocusing on compliant cases in  $t_0$  [the audit round], in 52.7% of compliant audited cases tax payments were reduced to some extent in  $t_1$  [the post-audit round] (21.8%) or participants evaded completely (30.9%) in  $t_1$ . By contrast, only in 36.9% of non-compliant cases at  $t_0$  participants reduced their tax payments (9.4%) or evaded completely (27.5%) at  $t_1$ ; whereas, 19.4% of the non-compliant cases showed increased or total compliance after the audit.<sup>290</sup>

---

284. *Id.* (“[T]here were three audits right at the beginning, whereas the remaining audits were positioned between rounds 23 and 39, with two further continuous audits after rounds 28 and 29.”).

285. *Id.*

286. *Id.* at 409, 412 tbl.2. For an explanation of  $p$ -values, see Darrow, *supra* note 48.

287. Kastlunger et al., *supra* note 281, at 409, 410 fig.1.

288. *Id.* at 409.

289. *Id.* (“If loss-repair accounts for the bomb crater effect, only participants who were fined in the audit round ( $t_0$ ) should increase evasion in the following round ( $t_1$ ), whereas compliant participants should tend to remain compliant.”). Compliant taxpayers could actually try to recoup the loss of the tax paid, but they would have less to recoup from that round than taxpayers who paid the tax plus a penalty.

290. *Id.*; see also *id.* at 412 tbl.3.

They concluded from this that their results “do not confirm loss-repair tendencies but suggest misperception of chance.”<sup>291</sup>

It is true that those results do not confirm loss repair; audited noncompliant taxpayers did not increase their evasion more than the audited compliant. However, if the bomb-crater mechanism is misperception of the chance of audit, one would expect differences in compliant and noncompliant taxpayers’ post-audit behavior to result only by chance. Here, the *compliant* taxpayers showed the greater bomb-crater effect. Kastlunger et al. do not report whether the differences are statistically significant.<sup>292</sup> If they are, it would support a distinct effect of audit on compliant taxpayers, an issue discussed in the next section.

Also, what is the actual chance of audit that participants may have misperceived? They were told that the audit rate was 15%. And in each condition, nine out of sixty rounds were audited, so each participant was audited nine times. This is not the same as rolling a die, however, where each roll is independent and has the same odds. When rolling a die that has a 15% chance of prompting an audit, participants might perceive that the likelihood of audit diminishes with each audit and even reaches zero after nine audits. That would be incorrect. But here, participants who hold those beliefs would actually *correctly* perceive the audit likelihood.

Kastlunger et al. also found that tax payments decreased after participants had not been audited for ten rounds and continued to decrease until the experiment ended.<sup>293</sup> In order to investigate whether a single later audit after a longer period without audits would restore compliance to the previous levels, they conducted a second experiment, with two conditions. The first condition (E1\*) was like E1 except that the audit in round 3 was moved to round 31. The second condition (E2\*) followed the same pattern of audits

---

291. *Id.* at 409. Kastlunger et al. also found for noncompliant participants who decreased their tax payments the round after audit, “[t]he correlation between experienced fines and compliance at  $t_1$  was highly significant with  $r = -.47, p < .01, n = 59$ , indicating that detection and punishment of non-compliance led to lower compliance in the following round.” *Id.* at 411-12.

292. *See id.* at 409, 412 tbl.3.

293. *Id.* at 413.

(two consecutive audits, an audit five rounds later, etc.), except with the first audit in round 20, so the last audit was in round 51.<sup>294</sup>

The authors found that “[t]he average payments per round, as well as the sum of honest tax filings, did not differ between the two conditions and were similar to Study 1.”<sup>295</sup> Audits had no significant effect on compliance in E1\* but had a positive influence in E2\*. The authors concluded that two consecutive audits may suppress bomb craters and that one audit during a compliance decrease may suffice to increase it.<sup>296</sup> Thus, Kastlunger et al.’s focus was on the timing of audits, rather than on the post-audit behavior of compliant and noncompliant taxpayers.

Those studies used random or predetermined audits. By contrast, a lab experiment by James Alm and Jim McKee examined the effect on compliance of an endogenous audit rule that compared the individual’s compliance level to that of others.<sup>297</sup> Although that study’s focus was on whether participants could coordinate on a zero-compliance equilibrium, it included a dummy variable for whether the participant was audited in the previous round, and found a negative correlation between that and compliance.<sup>298</sup> The authors explained that because the audit rule called for auditing the least-compliant participant in that round, decreasing compliance “is a best response strategy if the individual expects others to lower their compliance because they were not audited in the previous round.”<sup>299</sup> That study did not disaggregate compliant participants, so it is unclear if a particular subset of participants was driving that effect.

A more recent study used a different endogenous-audit rule to test whether such an audit rule eliminated bomb craters.<sup>300</sup> Emily Satterthwaite used Amazon’s Mechanical Turk (an online

---

294. *Id.* at 413–14.

295. *Id.* at 415.

296. *Id.* at 415, 416.

297. James Alm & Jim McKee, *Tax Compliance as a Coordination Game*, 54 J. ECON. BEHAV. & ORG. 297 (2004).

298. *See id.* at 309 tbl.3, 310 (not stating whether that effect was statistically significant).

299. *Id.* at 310.

300. *See* Emily Ann Satterthwaite, *Can Audits Encourage Tax Evasion?: An Experimental Assessment*, 20 FLA. TAX REV. 1 (2016).

marketplace for workers)<sup>301</sup> to recruit mostly non-student U.S. participants. The study involved a random audit (the control) and an endogenous audit in which likelihood of audit was determined by whether the previous audit found the taxpayer noncompliant (the treatment).<sup>302</sup>

Each condition had sixty rounds, although participants were told the study would last at least sixty-two rounds, so as to avoid a “last period” effect. In each round, participants were endowed with \$800 to \$1200 of “extra income” on which taxes were due. The tax rate was 30%, the average audit probability (over all participants) was 10%, and the penalty was 100% of unpaid taxes (plus the unpaid taxes).<sup>303</sup> Participants in the treatment group were told, “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.”<sup>304</sup>

Satterthwaite found that average compliance rates in the treatment group were higher, in every period, than under the random-audit rule.<sup>305</sup> In both conditions, compliance decreased

---

301. See AMAZON MECHANICAL TURK, <https://www.mturk.com/mturk/welcome> (last visited Oct. 16, 2018).

302. Satterthwaite, *supra* note 300, at 9. Random audits were scattered throughout the experiment. *Id.* In the treatment group, if the random audit showed full reporting, the participant remained subject to random audit. If the random audit showed cheating, that participant was audited again in the next period. *Id.* A participant ceased to be subject to random audits after being audited six times. *Id.*

303. *Id.* at 45.

304. *Id.* at 47. The details of the “flagging” were not disclosed to participants but were as follows:

[1.] [I]f the initial audit reveals that she reported all her income truthfully, she will be returned to the random IID [independent and identically distributed] audit lottery.

[2.] If the initial audit reveals underreported income, she will be audited again in the immediately succeeding period.

[3.] 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 . . . .

[4.] 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.

*Id.* at 48.

305. *Id.* at 54.

over time, although somewhat more so under random audits.<sup>306</sup> Satterthwaite found evidence of a bomb-crater effect for the control group, although in the first half of the experiment there were eight periods in which average post-audit compliance was higher than average pre-audit compliance.<sup>307</sup> For the treatment group, after about round 10, average post-audit compliance exceeded pre-audit compliance and in many periods approached or attained perfect compliance.<sup>308</sup> However, in the first half of the experiment, there were eight periods in which the treatment group, on average, exhibited a bomb-crater effect.<sup>309</sup> She concluded that a known endogenous-audit rule eliminated a bomb-crater effect following the first audit.<sup>310</sup>

Satterthwaite also investigated the effect of consecutive audits.<sup>311</sup> In the control condition, after one random audit, compliance dropped by almost 10 percentage points. After a second random audit in a row, compliance increased by approximately 7 percentage points. After a third, compliance increased by 12 percentage points, and after four random audits in a row, it increased by about 2 percentage points and began to plateau.<sup>312</sup> By contrast, the treatment condition, in which participants had been told that detected cheating would be a flag for audit, did not begin with a bomb crater.<sup>313</sup> Instead, after the first audit, compliance rose by 5 percentage points. After a second consecutive audit, compliance decreased by approximately 2 percentage points.<sup>314</sup> After a third consecutive audit, compliance increased by almost 7 percentage points, and after the fourth audit in a row, it decreased slightly.<sup>315</sup> Thus, in the endogenous audit treatment, Satterthwaite

---

306. *Id.*

307. *Id.* at 55.

308. *Id.* at 56.

309. *Id.*

310. *Id.* at 63.

311. In the control group, even a compliant taxpayer could experience consecutive audits. In the treatment group, a compliant taxpayer could also be selected at random for a consecutive audit.

312. Satterthwaite, *supra* note 300, at 61.

313. *Id.*

314. *Id.*

315. *Id.*

found an *increase* in compliance after the first audit but alternating effects for subsequent audits.<sup>316</sup>

Satterthwaite's results provide support for presenting audits as non-random. They may provide some support for repeated audits, as does the Kastlunger et al. study discussed above<sup>317</sup> and the DeBacker et al. study of IRS (non-random) audits of corporate taxpayers.<sup>318</sup> However, it is hard to know what is driving the alternating increases and decreases in compliance in Satterthwaite's endogenous audit group.<sup>319</sup>

Professor Satterthwaite's experiment initially did not separate out compliant from noncompliant taxpayers, and her published article does not report any results relating to those subgroups. However, she conducted a basic subgroup analysis in response to my inquiry. Specifically, for each of the audit conditions, she calculated average compliance in round  $t_1$  (the post-audit round) of four groups: (1) those audited in round  $t_0$  who were compliant in that round, (2) those audited in round  $t_0$  who were *not* compliant in that round, (3) those *not* audited in round  $t_0$  who were compliant in that round, and (4) those *not* audited in round  $t_0$  who were *not* compliant in that round.<sup>320</sup>

When Professor Satterthwaite compared the two subgroups in the random audit condition who were compliant in round  $t_0$ , she found less than 100% compliance in round  $t_1$  for both of those subgroups. The subgroup that had not been audited in round  $t_0$  had

---

316. *Id.* Repeated audits may have various effects. For example, a lab experiment found that "the number of audits had a significant and negative effect on tax evasion in the last round." Michael W. Spicer & Rodney E. Hero, *Tax Evasion and Heuristics: A Research Note*, 26 J. PUB. ECON. 263, 266 (1985) ( $p < 0.02$ ).

317. See *supra* text accompanying note 296.

318. See DeBacker et al., *supra* note 216, at 311 ("These results imply that firms are much less tax aggressive after a second audit than they are after the first audit, which suggests that they have revised their audit probability upward.").

319. In part, that may be regression to the mean. See Susan B. Long, *Commentary on Brian Erard, The Influence of Tax Audits on Reporting Behavior*, in *WHY PEOPLE PAY TAXES: TAX COMPLIANCE AND ENFORCEMENT* 115, 119 (Joel Slemrod ed., 1992) (stating with respect to a 1992 study, "[h]ad Erard selected taxpayers whose initial audit results were 'extreme' in the opposite direction (taxpayers initially highly compliant), he would have doubtless found a similar regression toward the mean with these taxpayers becoming less compliant after their initial audit"); cf. Erard, *supra* note 258, at 100–01 (acknowledging that point).

320. See Email from Emily Satterthwaite, Professor, Univ. of Toronto, to author (Dec. 19, 2017) (on file with author).

an average compliance rate of 0.929 in round  $t_1$ . This decrease from 100% compliance despite the lack of audit suggests that regression to the mean explains at least part of the post-audit decline in compliance.<sup>321</sup> The average compliance among these two subgroups of compliant taxpayers was lower for those who were audited, however, with the audited subgroup having an average compliance rate of 0.820, suggesting that the audit had a negative effect.<sup>322</sup>

In Satterthwaite's endogenous audit condition, the same pattern occurred, but the difference between the round  $t_1$  compliance rates was smaller: an average compliance rate of 0.940 for the not-audited who were compliant in round  $t_0$  and an average of 0.915 for the audited who were compliant in round  $t_0$ .<sup>323</sup> The greater decline in compliance in round  $t_1$  by those audited in round  $t_0$  compared to those not audited in round  $t_0$  also suggests that experiencing an audit has a negative effect on next-round compliance. Thus, Satterthwaite's findings in this regard are consistent with other laboratory experiments that found a bomb-crater effect immediately after audit.

### B. Field Experiments on Post-Audit Compliance

The laboratory experiments discussed above generally focus on the effects of audits on all audited participants. By contrast, field experiments have raised the prospect of a post-audit decline in compliance only of the subgroup of individual taxpayers found compliant on audit.<sup>324</sup> First, a U.K. field experiment examined the effects of audits on three categories of unincorporated taxpayers: medium businesses, small businesses, and personal taxpayers.<sup>325</sup>

321. See *supra* note 319; *infra* text accompanying notes 396–403.

322. See Email from Emily Satterthwaite, *supra* note 320.

323. *Id.*

324. See, e.g., Marcelo Bergman & Armando Nevarez, *Do Audits Enhance Compliance? An Empirical Assessment of VAT Enforcement*, 59 NAT'L TAX J. 817, 821, 825 (2006) (examining the effects of audits on VAT compliance in Chile and Argentina).

325. Norman Gemmill & Marisa Ratto, *Behavioral Responses to Taxpayer Audits: Evidence from Random Taxpayer Inquiries*, 65 NAT'L TAX J. 33, 41 (2012) ("[T]hrough the M- and S-segments may be thought to be similar, many small businesses—those with a turnover of less than £15,000—are akin to personal taxpayers with some self-employment income."). "Each segment consists of around 2,500–3,000 individuals, of which around 20–25 percent were audited." *Id.* at 43.



The study's authors, Norman Gemmell and Marisa Ratto, compared taxpayers randomly audited in 2000 with taxpayers who were not and looked at tax reported by these taxpayers for the three years before and after 2000.<sup>326</sup> Overall, the study found no statistically significant effects of the audit, before separating out subgroups of taxpayers.<sup>327</sup> Gemmell and Ratto then looked separately at the medium businesses, small businesses, and personal taxpayers and divided them into "compliant" and "noncompliant" based on the audit results.<sup>328</sup>

The study found a positive effect of audits on the tax reported by taxpayers found noncompliant but a negative effect on the tax reported by taxpayers found compliant.<sup>329</sup> As far as the magnitude of a possible negative effect of audit on subsequent compliance, Gemmell and Ratto explain that "for Personal taxpayers identified as compliant, declared tax fell due to audit by about 7 percent on average over 2001–2003."<sup>330</sup> The effects on small and medium businesses were larger (11% and 17%, respectively), but with larger margins of error.<sup>331</sup> The positive effects of audits on taxpayers whose returns *were* adjusted were generally of greater magnitude: 5%, 24%, and 18% for personal taxpayers, small businesses, and medium business respectively.<sup>332</sup> Gemmell and Ratto hypothesized that the stronger compliance response by businesses than by personal taxpayers reflects their greater opportunity to evade taxes.<sup>333</sup>

---

326. *Id.* at 42. They state: "To minimize the likelihood that previous or future audits influence observed reactions to the 2000 audit, we excluded all taxpayers who were audited in these six years (1997–1999, 2001–2003)." *Id.* at 41. Looking at years prior to the audit should help give a more complete picture of a taxpayer's overall compliance level.

327. *Id.* at 47; *cf.* Bergman & Nevarez, *supra* note 324, at 821 (finding in Chile and Argentina that, overall, audits had no effect on compliance).

328. Gemmell & Ratto, *supra* note 325, at 42. Of the 2006 audited taxpayers, they found 1342 (66.9%) compliant and 664 (33.1%) noncompliant. *See id.* at 42 tbl.2. The control group contained 6420 unaudited taxpayers. *See id.*

329. *Id.* at 48 ("In all cases positive parameters are obtained for the noncompliant group and negative parameters are obtained for the compliant group."). The authors recognize and control for the fact that although the audited group and control group are selected randomly, the subgroups of audited compliant and audited noncompliant are not. *Id.* at 45–46. They do not have subgroups of unaudited compliant and noncompliant. *See id.*

330. *Id.* at 50.

331. *Id.* at 49 tbl.5.

332. *Id.*

333. *Id.* at 55.

In this study, only some of the effects were statistically significant, however. For example, in the “Before/after 2000” regression, for those found compliant, while the coefficient for personal taxpayers was significant at  $p < 0.05$ , the coefficient for medium businesses was not significant, and the one for small businesses was significant only at  $p < 0.10$ .<sup>334</sup> Gemmell and Ratto explain that “[u]nsurprisingly perhaps, the negative preventive effects for the compliant are less robustly estimated for medium businesses where they are a smaller fraction of the audited.”<sup>335</sup> Overall, they concluded that any negative effect of random audits on noncompliant taxpayers should be weighed against the beneficial indirect effects of these audits.<sup>336</sup>

In the United States, DeBacker et al. used IRS data to study the effects of random audits on individuals’ tax compliance.<sup>337</sup> They examined data from years 2006, 2008, and 2009 of the IRS National Research Program,<sup>338</sup> which runs the random audits the IRS conducts to update the formula it uses to score returns on their likelihood of noncompliance. This was a much larger study than Gemmell and Ratto’s, with billions of observations.<sup>339</sup> Overall, DeBacker et al. found that audits had positive effects: they

---

334. *Id.* at 47 tbl.4. The other specification was “Before 2000/after inquiry closure.” *Id.* In that regression, the coefficients for compliant personal taxpayers and medium businesses were significant at  $p < 0.01$ , and the coefficient for compliant small businesses was not statistically significant. *Id.*

335. *Id.* at 48.

336. *Id.* at 53, 55.

337. Jason DeBacker et al., *Once Bitten, Twice Shy? The Lasting Impact of IRS Audits on Individual Tax Reporting 1* (Jan. 16, 2016) (unpublished manuscript) (on file with the author).

338. *Id.* at 10.

339. *Id.* at 35 tbl.1 (reporting 4.36 billion observations for the NRP sample and 4.33 billion observations for the random sample). For the NRP sample, DeBacker et al. report 404,252,738 observations in the base (NRP-audit) year, of which 164,338,287 had a positive adjustment to tax liability and 29,778,483 had a negative adjustment. *Id.* There were 210,135,968 no-change returns, comprising 51.98% of the total. *Id.* This contrasts with general data reported by the IRS, which for 2015 shows an overall no-change rate for individual income tax returns of 9% for field audits and 12% for correspondence audits. DEP’T OF THE TREASURY, INTERNAL REVENUE SVC., INTERNAL REVENUE SERVICE DATA BOOK, 2015, at 1, 24 tbl.9a (Mar. 2016), <https://www.irs.gov/pub/irs-soi/15databk.pdf>.

increased reported taxable income, a statistically significant effect,<sup>340</sup> and that increase persisted over time.<sup>341</sup>

DeBacker et al. also separately examined the effect on reported taxable income on taxpayers whose random audit resulted in a positive adjustment, a negative one, or no adjustment.<sup>342</sup> They found that for taxpayers with a positive adjustment, taxable income increased in the first and second years post-audit, and it stayed elevated through the final year they reported, year six.<sup>343</sup> The results for that group were statistically significant.<sup>344</sup>

For taxpayers whose audit resulted in no adjustment either way, DeBacker et al. found that taxable income declined the first year post-audit—which is consistent with a bomb-crater effect—increased to the audit-year level in the second year post-audit, gradually increased through year four, then declined slightly in the fifth and sixth years, ending a bit above where it started.<sup>345</sup> However, in contrast to the results for taxpayers whose returns had been adjusted, these results generally were not statistically significant.<sup>346</sup>

---

340. DeBacker et al., *supra* note 337, at 18–19 (“[A]udits increase reported taxable income by \$1,109, which is statistically significant at the 1% level.”); *see also id.* at 30–31 (“The results from a simple difference-in-differences specification indicate that an audit increases reported taxable income by more than \$1,100 per year, equivalent to 2.9% of the average income.”).

341. *Id.* at 17 (“Reported taxable income increases in the first and second years after an audit and remains elevated even after six years. Adjusted gross income (AGI) and wage income follow a similar pattern to that of taxable income.”).

342. *Id.* at 45 fig.4.

343. *Id.* (showing progressive increases in the first and second years post-audit, a gradual decline in years three through five to about the level of the first post-audit year, and a leveling off in year six).

344. *Id.* at 19 (“We see the strongest and most statistically significant response from the group with positive adjustments. Almost all the coefficients for the other two groups are statistically insignificant and have much lower point estimates than for the positive adjustment group.”).

345. *Id.* at 45 fig.4. The figure shows “Change in Reported Taxable Income,” which is thus 0 in the year of audit. In the first post-audit year, the change in this group relative to the audit year is about -650; in the second, close to 0; in the third, about 450; in the fourth, about 650; in the fifth, about 600; and in the sixth, about 450. *See id.*

The results for the *negative* adjustment group (those who were found on audit to have overcomplied) essentially zigzagged from increase to decrease and back: the change in this group relative to the year of audit was about 1750 in year 1, a slightly negative figure in year 2, about 800 in year 3, about 1900 in year 4, about -600 in year 5, and about 1100 in year 6. *See id.*

346. *Id.* at 19.

Another U.S. study using IRS data, and similar in design to the U.K. study, found a negative effect of audit on taxpayers found compliant on audit, as the U.K. study did. This U.S. study, conducted by Sebastian Beer et al., was commissioned by the National Taxpayer Advocate (NTA), Nina Olson, and analyzed the forward-looking effects of IRS audits of self-employed taxpayers' 2007 returns.<sup>347</sup> The main design difference between the NTA's study and the U.K. study was that the audits included in the NTA's study were not random; the returns had been selected for audit in the ordinary course, which generally is based on likelihood of noncompliance.<sup>348</sup>

Overall, like Gemmell and Ratto's UK study, the NTA's study found a positive effect of audit on subsequent reporting. First, for all taxpayers in the sample (compliant or not), taxable income was 20% higher than the control (unaudited) group three years after audit.<sup>349</sup> Second, that effect was more pronounced for the subgroup of taxpayers who were assessed additional taxes on audit: "[O]ne year after having undergone enforcement activity, [they] . . . report approximately 250 percent more in taxable income than taxpayers in the control group. Three years after the audit, the estimated differential remains quite high at 120 percent."<sup>350</sup> These results initially suggest that audits may be most beneficial where

347. NAT'L TAXPAYER ADVOCATE, *supra* note 181, at 67 n.1, 78. The study was conducted by Sebastian Beer, Matthias Kasper, Erich Kirchler, and Brian Erard. *Id.* at 67. It compared "a random sample of 2,204 Schedule C filers with under \$200,000 in total positive income who were audited subsequent to filing their TY 2007 returns with data for a control sample of 4,705 who were not audited . . ." *Id.* at 71.

348. The study notes that "the focus of this study is on taxpayers selected through an ordinary operational audit process. . . . Operational audits tend to be targeted towards tax returns with a high potential for noncompliance." *Id.*

349. *Id.* at 73. Similarly, a study in Denmark found that audits deter evasion the following year: "For the full sample, the effect on total net income is 2557 kroner or 30.1 cents per additional kroner of audit adjustment. The effect on tax liability is 1375 kroner, corresponding to 41.7 cents per dollar of audit adjustment. These estimates are strongly significant." Kleven et al., *supra* note 126, at 685. That study did not look separately at taxpayers found compliant on audit. *See generally id.*

350. NAT'L TAXPAYER ADVOCATE, *supra* note 181, at 72. Like the Gemmell and Ratto U.K. study, this study does not separate the control group of unaudited taxpayers by compliant and noncompliant. *Id.* at 71. The NTA's study uses "a Matched Difference-in-Differences estimator" to address this problem. *Id.* at 76.

they are most needed, at least with respect to U.S. self-employed taxpayers.<sup>351</sup>

With respect to taxpayers found compliant (or over-compliant<sup>352</sup>) on audit, however, like Gemmell and Ratto, the NTA found a negative effect of audit on subsequent reported taxable income.<sup>353</sup> In only one of the six specifications, the baseline differences-in-differences regression, was the magnitude of the decline statistically significant.<sup>354</sup> There, three years after audit, those taxpayers reported 35.8% less in taxable income than the control group, a difference significant at  $p < 0.05$ .<sup>355</sup>

A lab experiment conducted in the United Kingdom with three groups of participants (students, employees, and the self-employed) also found that compliance decreased after audit, including among compliant participants, although the primary driver of this effect was the student sample.<sup>356</sup> In contrast to these studies, a study of the effect of audits on VAT compliance in Chile and Argentina found a negative *overall* effect of audit and a *positive* effect on the audited taxpayers who did not receive an assessment. In particular, Marcelo Bergman and Armando Nevarez found that, “[o]n average, the non-audited have better compliance than audited tax-payers after enforcement.”<sup>357</sup> However, in Chile, 38% of the taxpayers audited had no assessment, and the audit had a

---

351. The study notes that “we cannot rule out that our estimates are influenced by the economic downturn in 2008.” *Id.* at 70.

352. The study does not mention overcompliant taxpayers. Taxpayers with no positive adjustment (which could include a refund) were labelled as having no tax change. *See id.* at 71, 72 (“To avoid confusion . . . , we will refer to . . . the subsample that does not receive an additional recommended tax assessment as the no-tax-change experimental group ‘E-NC,’ rather than as . . . ‘compliant[]’ . . .”).

353. *See id.* at 86, tbl.5.

354. *See id.* (reporting six regressions and specifying statistically significant results at the 0.01, 0.05, and 0.10 levels).

355. *Id.* at 86. The Executive Summary reports that “[t]he [thirty-five percent] difference is significant at the one percent level.” *Id.* at 69. This seems to be an error.

356. Choo et al., *supra* note 261, at 112 ( $p < 0.001$ ) (“Starting with the case where the subject was not evading in period  $t$ , we observe a negative and highly significant coefficient on  $\text{Not Evade}_{it} \times \text{Audited}_{it-1}$ , indicating that expected compliance goes down in the period subsequent to an audit taking place. . . . In short, we find evidence for the bomb-crater effect in our experiment, but that effect is driven primarily by the student sample.” (citations omitted)).

357. Bergman & Nevarez, *supra* note 324, at 821.

positive effect on that group.<sup>358</sup> (The authors were unable to conduct the same analysis for Argentina, where only 2% of the taxpayers did not receive assessments on audit.<sup>359</sup>)

It is possible that a country in which noncompliance is high may result in different taxpayer behavior.<sup>360</sup> However, the recent study in Italy<sup>361</sup> by Mazzolini et al. also found similar results to those of the NTA and Gemmel and Ratto. Mazzolini et al. studied audits of self-employed taxpayers in the ordinary course. They found that “[r]eported income increases on average by approximately 8.2 percent after audits”<sup>362</sup> and remains higher for three years after audit.<sup>363</sup> Like the NTA, they found a negative effect on the subgroup of compliant taxpayers for the year after audit.<sup>364</sup> However, that result was statistically insignificant.<sup>365</sup>

Mazzolini et al. find the result with respect to compliant taxpayers expected<sup>366</sup> and “consistent with the predictions of the

358. *Id.* at 826, 829. At the beginning of the article, the authors state that in Chile, “approximately 22 percent of VAT fails to be collected, while in neighboring Argentina tax evasion is roughly double.” *Id.* at 817–18. However, they do not consider taxpayers who were not assessed on audit necessarily compliant; they state, “We assume that there is no perfect compliance. It appears that most taxpayers cheat; however, the difference is in the size of noncompliance. Those who were not sanctioned appear to be small-scale cheaters.” *Id.* at 824 n.14.

359. *Id.* at 826. The percentage was so high in part because audits were not random. *See id.* at 820.

360. *Id.* at 829–30 (“[U]nder the noncompliance equilibrium (as in Argentina), enforcement fails to enhance specific deterrence and very likely fails to achieve general deterrence as well.”).

361. “Italy is known to be a country with low tax morale, where tax amnesties are frequent and institutions (including the tax administration) are relatively less efficient than those operating in other developed countries.” Mazzolini et al., *supra* note 215, at 30.

362. *Id.* at 7.

363. *Id.* at 22. Similarly, a study of Armenia found that “[f]ollowing an audit, firms respond by increasing reported income (sub-figure 13a). The magnitude of this response amounts to an average increase of 9, 17 and 7 percent in the first three years of audit, respectively, compared to the pre-audit year.” Zareh Asatryan & Andreas Peichl, *Responses of Firms to Tax, Administrative and Accounting Rules: Evidence from Armenia* 30–31 (CESifo, Working Paper No. 6754, Nov. 2017), [https://ideas.repec.org/p/ces/ceswps/\\_6754.html](https://ideas.repec.org/p/ces/ceswps/_6754.html). However, “every additional dollar of reported income in years  $t$  to  $t+1$  is, on average, matched by a 0.90 dollar increase in reported deductions.” *Id.* at 31. This result is consistent with Carrillo et al., *Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement*, 9 AM. ECON. J. APPLIED ECON. 144 (2016). *See id.* at 146 (“[F]irms increased reported costs by 96 cents for every dollar of revenue adjustment.”).

364. Mazzolini et al., *supra* note 215, at 23.

365. *Id.*

366. *Id.* at 6–7.

standard [(deterrence)] model of tax compliance.”<sup>367</sup> They explain that, in a no-change audit, “we expect a downward revision of the probability assigned by the cheating taxpayer to the success of tax inspectors in discovering evasion.”<sup>368</sup> In other words, before the audit, the average taxpayer likely expected to owe some positive amount after audit. After experiencing an audit outcome of owing nothing (or even receiving a refund), that taxpayer should be expected to revise that expectation downward. Thus, “[a]ccording to the rational taxpayer model, on average the effect on future reported income should be lower than in the positive-adjustment case.”<sup>369</sup> That is, if deterrence works, an audited taxpayer who is found to owe more tax should be more deterred from noncompliance than an audited taxpayer who is found to owe nothing.

Note that the focus here is on *specific deterrence*—that is, the effect on the audited taxpayer. The regime-level effect of audit rates is the focus of many studies, and many studies do not separate out the effects on those actually experiencing an enforcement action.<sup>370</sup> But for a taxpayer who actually experiences an audit, if the audit results in no additional tax (let alone a penalty), it is intuitive that that reduces the deterrent effect of audits *to that taxpayer* (that is, the specific-deterrent effect), at least for some period of time.

This analysis reflects a learning effect—at a minimum learning from the audit outcome. It may also reflect learning from the audit experience ways in which noncompliance was effective or likely would not be detected.<sup>371</sup> As Kleven et al. noted, audits provide information to taxpayers, particularly with respect to the perceived likelihood of detection.<sup>372</sup>

In addition, taxpayers who were audited and found not to owe more tax may infer that the likelihood of audit for the next few

---

367. *Id.* at 23.

368. *Id.* at 6.

369. *Id.*

370. See *supra* notes 61, 68, and 219 (pointing this out).

371. NAT'L TAXPAYER ADVOCATE, *supra* note 181, at 69–70 (“This newfound awareness of opportunities for reporting and paying lower taxes combined with a low perceived future audit risk could drive some taxpayers to report less income on subsequent returns.”); cf. Mazzolini et al., *supra* note 215, at 24 (“[I]n the case of cheating taxpayers, we expect a downward revision of the perceived probability that tax inspectors are successful in discovering evasion.”).

372. Kleven et al., *supra* note 126, at 681; see also Gemmell & Ratto, *supra* note 325, at 55.

years is low.<sup>373</sup> Endogenous audits—where audit selection is not random—suggest a decreased likelihood of a quick second audit of a taxpayer who was found compliant, leading to a seemingly “safe” period for noncompliance.<sup>374</sup> In theory, that should not be true for random audits, but people may misperceive the likelihood of a random event recurring.<sup>375</sup> In other words, after an audit where the taxpayer was found compliant, the taxpayer may reduce the perceived likelihood of an audit for the next few years, reducing the perceived probability of detection. Under the deterrence model, this could reduce the taxpayer’s expected cost of evasion.<sup>376</sup>

Note that the explanations regarding learning and a “safe period” do not need to disaggregate the subgroup of audited no-change taxpayers into those who were actually compliant and those who were in fact noncompliant but simply did not get caught. It is possible that any of the taxpayers in this subgroup expect that the likelihood of audit in the future is low (whether they were actually compliant or not).<sup>377</sup> It also is possible that any of these taxpayers learned from the audit a “safe” method of cheating, although this likely is most true of taxpayers who were actually noncompliant, in that they directly observed noncompliance going undetected.

The NTA offered three possible explanations for her finding that audited taxpayers who were not found to owe more tax reduced their subsequent tax payments. One combines the “safe

373. Mazzolini et al., *supra* note 215, at 23–24 (“[T]he taxpayer . . . [may] think[] that, after a null-outcome [(no-change)] audit, she is removed, for some time at least, from the set of taxpayers targeted by the [tax agency].”); NAT’L TAXPAYER ADVOCATE, *supra* note 181, at 70 (referring to a “low perceived future audit risk”).

374. This approach is supported by the pattern observed by Kastlunger et al. in their first lab experiment: “The behavior of the participants . . . reminds [sic] of a ‘cops and robbers’ situation. When participants expected audits, they were compliant and, to a great extent, they evaded when they felt ‘safe.’” Kastlunger et al., *supra* note 281, at 413 (citation omitted).

375. This concept underlies the term “bomb-crater effect” and is similar to the belief that lightning doesn’t strike twice in the same place. *See supra* note 269.

376. *See supra* text accompanying notes 103–104.

377. Mazzolini et al., *supra* note 215, at 23 (“A null-outcome [(no-change)] audit may occur either if the taxpayer was actually compliant and AE [(Agenzia delle Entrate, the Italian tax agency)] acknowledges this or if the taxpayer evaded taxes and AE was not able to uncover evasion. In either case, the taxpayer has no reason to increase her subsequent income reports . . . .”); NAT’L TAXPAYER ADVOCATE, *supra* note 181, at 70 (referring to a “low perceived future audit risk”).



period” and “safe methods” hypotheses described above.<sup>378</sup> Another one focuses on (presumably) compliant taxpayers and hypothesizes that they experienced a reduction in tax morale.<sup>379</sup> This might be termed the “crowding-out hypothesis,” as it focuses primarily on compliant taxpayers and suggests that extrinsic enforcement crowded out intrinsic motivations (a crowding-out effect). The third explanation is that cheaters were emboldened by not having been caught on audit.<sup>380</sup>

Of course, there may have been a mix of fully compliant taxpayers and noncompliant taxpayers. A field experiment such as this one cannot distinguish between actual compliance and noncompliance that was not detected on audit because the auditors do not know the actual amount of tax due. In this respect, laboratory experiments are particularly helpful because the experimenter perceives actual compliance levels. While Maciejovsky et al. found no difference in the post-audit compliance rates of those compliant or noncompliant in the audit round,<sup>381</sup> the results of the Kastlunger et al. lab experiment testing repeated audits, discussed above,<sup>382</sup> could suggest that taxpayers who were actually compliant when audited tend to decrease their compliance post-audit. However, that study involved audits at predetermined times, rather than audit selection based on the taxpayer’s likelihood of evasion. Satterthwaite’s data, by contrast, found a lower compliance rate in the round after audit by the noncompliant than by the fully compliant.<sup>383</sup>

---

378. NAT’L TAXPAYER ADVOCATE, *supra* note 181, at 70.

379. *Id.* at 69 (“[A]n experience of coercive enforcement activity could reduce tax morale among (seemingly) compliant taxpayers . . .”).

380. *Id.* at 69–70 (“[T]he observed reduction in reported income may be attributable to noncompliant taxpayers whose misreporting was not detected during the audit. Experiencing an audit that results in no recommended additional tax assessment may embolden such taxpayers to become even more noncompliant in the future.”).

381. Maciejovsky et al., *supra* note 244, at 688; *see also supra* text accompanying notes 271–273 (summarizing this study).

382. *See* Kastlunger et al., *supra* note 281, at 409, 412 tbl.3.

383. That was true in both the random audit and endogenous audit conditions, but the compliance rates in both the audit round and the next round were higher in the endogenous audit condition. *See* Attachment to Email from Emily Satterthwaite, Professor, Univ. of Toronto, to author (May 12, 2016) (on file with author). The details are as follows:

Given the fact that the overall misreporting percentage for the group of taxpayers involved in the NTA's study (self-employed taxpayers) is much higher than average,<sup>384</sup> it seems likely that at least some of the taxpayers labelled compliant were actually noncompliant.<sup>385</sup> In addition, the fact that the majority of taxpayers audited by the IRS, at least according to the DeBacker study,<sup>386</sup> are found compliant supports that idea, assuming that the IRS's selection criteria are at all effective. To the extent that some of the taxpayers whose returns were not adjusted to owe more tax were actually noncompliant, failing to detect that noncompliance means that the audit not only left money on the table for that tax year, it may have emboldened these taxpayers to increase their evasion in subsequent years.<sup>387</sup>

Note that there is another possible category of taxpayers in this subgroup, which the three explanations the NTA offers do not address: overcompliers. In the NTA's study, all taxpayers with no "additional recommended tax assessment" were grouped together.<sup>388</sup> Taxpayers who correctly received a refund are

Compliance in Round $t_1$		
	Audited in round $t_0$ & found compliant	Audited in round $t_0$ & found noncompliant
Random audit	0.820 ( $n=340$ )	0.400 ( $n=290$ )
Endogenous audit	0.915 ( $n=421$ )	0.805 ( $n=114$ )

*Id.*

384. Income subject to no information reporting has an estimated compliance rate of only 37%, compared to an average rate of almost 82% and a high of 99% for income subject to both withholding and substantial information reporting. See INTERNAL REVENUE SVC., *supra* note 113, at 3; *supra* text accompanying note 113; *supra* text accompanying notes 129–130. Self-employed taxpayers have a high opportunity to evade taxes. Mazzolini et al., *supra* note 215, at 10.

385. Cf. Morse et al., *supra* note 126, at 37 ("According to government reports, most individuals with business income fail to pay all their taxes, although some appear to cheat more than others."); *id.* at 67 (concluding that "[c]ash business owners rely on parallel cash economies to underreport receipts and thereby evade income, employment and sales taxes").

386. See *supra* note 339. The Gemmill and Ratto study found a supermajority compliant. See *supra* note 328. However, that was in random audits. See *supra* text accompanying note 326.

387. NAT'L TAXPAYER ADVOCATE, *supra* note 181, at 91.

388. See *supra* note 352. The Italian study by Mazzolini et al. does not mention the possibility of taxpayers receiving a refund. See Mazzolini et al., *supra* note 215, at 9, 48 fig.1 (audit outcomes).

overcompliers. Overcompliant taxpayers are situated differently from others in that a reduction from overcompliance does not necessarily produce undercompliance. Moreover, those taxpayers' motivations may be different, so disaggregating taxpayers found overcompliant would be helpful. The DeBacker et al. study, which is also of IRS audits and for a similar time period, though not limited to self-employed taxpayers,<sup>389</sup> observed a different post-audit pattern between audited taxpayers who experienced no adjustment and taxpayers who received a negative adjustment (refund). While the reported taxable income of the former declined one year post-audit, then increased for the next few years, the reported taxable income of the latter group increased in the first year after audit and then generally zigzagged.<sup>390</sup>

The NTA's second proposed explanation, that the audit may have reduced tax morale, crowding out intrinsic motivation to pay taxes, deserves detailed examination.<sup>391</sup> The report does not explain why an audit might reduce compliant taxpayers' intrinsic motivation. One could posit that a compliant taxpayer who is audited may feel like a "chump" for complying.<sup>392</sup> That is, a compliant taxpayer may feel penalized despite (or for) doing the right thing or may begin to feel that the government is hapless—wasting resources auditing the compliant.<sup>393</sup>

---

389. The NTA studied the effect of audits of the 2007 tax year. See *supra* note 347 and accompanying text. DeBacker et al. studied the IRS's random (research) audits of the 2006 through 2009 years. See *supra* text accompanying notes 337–338.

390. See DeBacker et al., *supra* note 337, at 45 fig.4; see also *supra* note 345 (reporting the numbers).

391. The crowding-out hypothesis focuses on taxpayers who have intrinsic motivations to comply and were actually compliant in the audit year. Field experiments cannot readily distinguish between true and faked compliance, but laboratory experiments can. A lab experiment could examine the effect of an audit (whether random or endogenous) that labels some participants' reports "compliant" even when they are not, to see the extent to which that has the effect of reducing subsequent compliance (as the deterrence model would predict), or whether only audits of truly compliant participants have that effect (as the crowding-out theory would predict).

392. See Lederman, *supra* note 70, at 1497–98 ("[S]anctioning people who do not contribute to public goods can reassure others that they will not be 'chumps' if they contribute.").

393. Deterrence can remove that feeling—as long as it is applied to the noncompliant. See *id.* Thus, this hypothesis would not be inconsistent with the fact that audit threats have been found to increase compliance. See *supra* Part I.B.1. Audit threats differ from actual audits in that a compliant taxpayer could interpret an audit threat to mean that *noncompliant* taxpayers will be investigated and burdened. The actual audit is a burden even to a

However, a follow-up study on taxpayer attitudes commissioned by the NTA<sup>394</sup> found that “audited self-employed taxpayers perceive higher degrees of procedural justice, information justice, interpersonal justice, and distributive justice than the unaudited control group.”<sup>395</sup> In addition, audited taxpayers who were not found to owe more tax reported higher levels on all four categories of justice (procedural, informational, interpersonal, and distributive) than audited taxpayers who had a positive tax adjustment.<sup>396</sup> The same was true of those taxpayers who received a refund.<sup>397</sup> Although these were not the same self-employed taxpayers as in the earlier study,<sup>398</sup> the survey results suggest that the audit experience did not produce negative feelings for self-employed taxpayers who were found not to owe more tax, undermining the crowding-out hypothesis.

Moreover, a tax-morale explanation is not needed to explain a post-audit decline in tax payments by taxpayers found not to owe more tax, once the effect on specific deterrence of a no-change audit is considered.<sup>399</sup> That is, the absence of a penalty or even tax due is not much of a deterrent because it likely reduces the perceived sanction. This in itself would reduce deterrence under the

---

compliant taxpayer because of the time and stress involved, as well as possible financial costs of engaging tax advisors. A study in Austria suggests that too much audit supervision can backfire by delaying payments. See Katharina Gangl et al., *Effects of Supervision on Tax Compliance: Evidence from a Field Experiment in Austria*, 123 *ECON. LETTERS* 378, 379–80 (2014) (explaining that intervention involving tax authority visits to new firms plus monthly audits during the firm’s first year statistically significantly *decreased* timely payment, but for firms paying late, supervision reduced taxes due by an estimated 27 to 50%).

394. The study was conducted by Matthias Kasper, Sebastian Beer, Erich Kirchler, and Brian Erard. MATTHIAS KASPER ET AL., NAT’L TAXPAYER ADVOCATE & TAXPAYER ADVOCATE SERV., *Audits, Identity Theft Investigations, and Taxpayer Attitudes: Evidence from a National Survey*, in 2 *TAS RESEARCH AND RELATED STUDIES* 148 (2017), [https://taxpayeradvocate.irs.gov/Media/Default/Documents/2017-ARC/ARC17\\_Volume2\\_05\\_AuditsIDtp.pdf](https://taxpayeradvocate.irs.gov/Media/Default/Documents/2017-ARC/ARC17_Volume2_05_AuditsIDtp.pdf).

395. *Id.* at 163.

396. *Id.* at 166 fig.5.8; *see also id.* at 180 app. A (relevant questions and numerical scale).

397. *Id.*

398. The study included “2,729 self-employed taxpayers (1,363 audited and 1,366 non-audited) . . .” *Id.* at 159. Among other selection criteria, the audited taxpayers “[h]ad an operational audit between tax years 2010 and 2015.” *Id.* at 158. (The earlier study compared a “random sample of 2,204 Schedule C filers . . . who were audited subsequent to filing their TY 2007 returns with data for a control sample of 4,705 who were not audited.” NAT’L TAXPAYER ADVOCATE, *supra* note 181, at 71.)

399. *See supra* text accompanying notes 361–365.

deterrence model,<sup>400</sup> but the effect is compounded when it is combined with the “safe period” notion discussed above; a taxpayer who was audited and not found to owe more tax might very well reduce the perceived likelihood of audit in the near future.<sup>401</sup>

It is also worth noting that some compliant taxpayers’ subsequent evasion may be regression to the mean.<sup>402</sup> That is, an audit is a snapshot of one particular year. Such a snapshot may not entirely capture a taxpayer’s behavior over time. For example, a taxpayer may report 90% of income, on average, but report somewhat more in some years—perhaps even 100%—and somewhat less in others. In general, at the extreme ends of the distribution, the taxpayer’s subsequent behavior is most likely to move closer to the mean. And taxpayers audited and found fully compliant approach one extreme on the distribution of audit results.<sup>403</sup>

A 1988 study conducted by Hessing et al. on Dutch individuals’ tax returns supports this idea.<sup>404</sup> Although a number of the details in the brief report of the study are not fully clear,<sup>405</sup> the article explains that the Dutch tax authorities checked every tax return every year and simply corrected inaccurate returns,<sup>406</sup> without imposing a penalty.<sup>407</sup> Hessing et al. generally found that noncompliance in one year predicts noncompliance in a future

---

400. See *supra* text accompanying notes 103–104.

401. See *supra* text accompanying notes 373–374.

402. This effect is easier to isolate with a matched group of unaudited fully compliant taxpayers, which the U.K. and U.S. studies did not have. However, those studies did use statistical techniques to control for the fact that the subgroups of audited compliant and audited noncompliant are not randomly assigned. See *supra* note 329 (U.K. Gemmill and Ratto study); *supra* note 350 (NTA’s U.S. study).

403. See Long, *supra* note 319, at 118.

404. See Hessing et al., *supra* note 112, at 298–304.

405. See Richard Bird, *Commentary on Dick J. Hessing et al., Does Deterrence Deter? Measuring the Effect of Deterrence on Tax Compliance in Field Studies and Experimental Studies*, in *WHY PEOPLE PAY TAXES: TAX COMPLIANCE AND ENFORCEMENT* 306, 309 (Joel Slemrod ed., 1992) (“[W]ith respect to the second field study reported by the authors, unfortunately once again some crucial pieces of information are missing from the highly condensed account.”); see also *infra* note 408.

406. Hessing et al., *supra* note 112, at 298.

407. The article explains that the tax authority imposed no penalty unless, after the return had been processed, subsequent information that was not available upon initial inspection indicated tax evasion. In that case, the tax authority could impose a penalty of 25 to 100% of the evaded taxes. *Id.*

year.<sup>408</sup> Interestingly, the taxpayers whose prior returns were not corrected – and thus presumably were found largely compliant by the Dutch tax authorities – were found to have in 1986 a noncompliance percentage of about 11% and an evasion percentage of about 3%.<sup>409</sup> Yet these taxpayers apparently did not experience any contact from the tax authority.<sup>410</sup> Accordingly, it would seem that the reduced compliance by those taxpayers had nothing to do with an audit. Thus, this study also lends at least limited support to the idea that some taxpayers who are found fully compliant in one year may not be the next year, regardless of whether they are audited.

Gemmell and Ratto's U.K. study examined whether having no adjustment at all on audit was the key to the result. The authors looked at whether taxpayers with a small adjustment behaved similarly to taxpayers with no adjustment.<sup>411</sup> They found limited support for that notion, but the results were not statistically significant.<sup>412</sup> Thus, that inquiry does not help resolve the question of how much of the post-audit effect is specific to fully compliant taxpayers.

Overall, the results of the studies on self-employed or small business taxpayers in the United Kingdom, United States, and Italy suggest that while undergoing an audit generally increases subsequent tax payments, it does not do so for taxpayers found not to owe additional tax. In fact, taxpayers in that subgroup may

408. *See id.* at 302, 303 fig.2 (also finding that result for tax evasion). Based on the figure, the percentages of noncompliance for 1986 in the three groups seem to be about 11% (presumably of Dutch Guilders) for taxpayers whose prior returns had not been corrected, 28% for taxpayers whose returns had contained errors that were corrected, and 36% for taxpayers whose prior returns were corrected for tax evasion. The exact numbers are not given, and there appears to be a typo on the *y*-axis, such that 0% should be 10%. Also, the discussion and figure are not clear as to which prior year(s) are being compared. *See id.*

409. *Id.* at 303 fig.2.

410. Hessing et al. describe the process as follows:

In the Dutch tax system, every single tax form is checked every year. . . . When a tax form is found to be incorrect, the tax inspector simply corrects the tax form and collects the tax associated. . . . If the (positive) correction takes place during the normal processing of the tax form, there is no possibility of administering a fine.

*Id.* at 298.

411. Gemmell & Ratto, *supra* note 325, at 50.

412. *Id.* at 52.

reduce their tax payments. While one could argue that the audit experience crowded out intrinsic motivations to pay tax, there is little evidence to support that argument. Rather, the result is consistent with the deterrence model: a positive outcome is not much of a deterrent.

### CONCLUSION

Given the importance to the government of tax compliance, it is important to know whether and when deterrence works. Some have suggested that audits or audit threats may backfire.<sup>413</sup> However, not only has an IRS researcher found in a detailed study that audits increase tax compliance,<sup>414</sup> empirical studies generally find that enforcement fosters compliance.<sup>415</sup> In part, that may be because failing to pursue back taxes simply allows taxpayers to ignore the delinquency.<sup>416</sup> In addition, enforcement may signal not that many taxpayers are evading taxes but rather that *few* taxpayers successfully evade.<sup>417</sup> In fact, it is non-enforcement that “sends a signal . . . that others do not wish to enforce the tax laws and that tax evasion is in some sense socially acceptable, and the social norm

---

413. See, e.g., Kahan, *supra* note 3, at 380–81 (“When the IRS engages in dramatic gestures to make individuals aware that it is redoubling its efforts to catch and punish tax evaders, it also causes individuals to infer that more taxpayers than they thought are choosing to cheat. This inference in turn triggers a reciprocal motive to evade, which dominates the greater material incentive to comply associated with the higher than expected penalty.” (footnotes omitted)); Mendoza et al., *supra* note 220 (finding a correlation between a higher audit rate in a country and a higher perceived rate of tax evasion); cf. Frey, *supra* note 50, at 387 (reporting that pooled data for Switzerland finds that a higher probability of detection correlates with higher tax evasion).

414. See Plumley, *supra* note 155.

415. See *supra* Part II.B.

416. See Ryan Briggs, *How to Fix Philly's Poor Tax Collection*, PHILLY.COM (July 2, 2015) (connecting high delinquency rates in Philadelphia property taxes and business taxes to lack of enforcement and difficulties in making payment, and adding, “The city’s liquor tax, hotel tax and school income tax . . . similarly rely on self-reporting, and often go uncollected with few warnings or penalties.”); see also Chirico et al., *supra* note 163, at 138 (explaining, regarding Philadelphia’s property tax, “In the past, the [Department of Revenue’s] efforts at collection from these very tardy taxpayers have been limited to simply remailing the usual reminder letter.”).

417. See Lederman, *supra* note 70, at 1497–98; see also Davis et al., *supra* note 239.

of compliance disappears. Such an outcome is common in many countries, such as the Philippines and Italy . . . ."<sup>418</sup>

The notion that deterrence, perversely, will reduce tax compliance generally is not consistent with the ample empirical evidence in the United States and elsewhere. For example, the IRS finds a positive "shadow" effect on tax collections from each dollar collected via enforcement,<sup>419</sup> and field studies generally find positive effects of audit threats and of audits with prior notice.<sup>420</sup>

Several field studies have found evidence in the context of unincorporated small business taxpayers of a reduction in reported tax liabilities following an audit that did not yield additional tax due. This suggests that audits that do not result in a tax payment have an opportunity cost not only with respect to that tax year but also with respect to several subsequent years for that taxpayer. This outcome, while perhaps initially surprising, is consistent with the deterrence model; a favorable outcome after audit may lower that taxpayer's perceived likelihood of subsequent audit and the perceived magnitude of the sanction.

In addition, it is important to remember that the overall direct effect of audits is increased tax payments by those audited, and the positive indirect effect of audits has been estimated to be much larger than the direct positive effect of audits. Thus, it would be unwise to conclude that reduced enforcement would have a positive effect on compliance. In fact, at low audit rates such as those in the United States, the evidence suggests that *increasing* the audit rate would increase overall tax compliance.

---

418. James Alm & Jorge Martinez-Vazquez, *Institutions, Paradigms, and Tax Evasion in Developing and Transition Countries*, in PUBLIC FINANCE IN DEVELOPING AND TRANSITIONAL COUNTRIES 146, 151 (Jorge Martinez-Vazquez & James Alm eds., 2003).

419. See *supra* note 182 and accompanying text.

420. See *supra* Part II.B.1.



