

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Effectiveness of Deterrence Nudging Interventions on Tax Compliance – A
Systematic Literature Review

verfasst von | submitted by

Olivia Engelbrecht BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on the
student record sheet:

UA 066 840

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Psychologie

Betreut von | Supervisor:

Univ.-Prof. i.R. Dr. Erico Kirchler

Author Note

This systematic review was conducted as part of a collaborative research project developed with my colleague, Alexander Truzka, who examined a separate aspect of the research data. The systematic literature search and development of the coding scheme was conducted jointly, while the coding of the studies and subsequent data analyses were performed separately, with each researcher analyzing a distinct subset of the complete dataset. The collaborative aspects of this research have been comprehensively documented throughout the thesis to ensure transparency and clarity regarding the division of work.

Table of Contents

Abstract.....	4
Introduction	5
1.1 The Deterrence Approach.....	6
1.2 Slippery Slope Framework.....	8
1.3 Responsive Regulation	9
1.4 Investigating Tax Compliance	11
1.5 The Rise of Natural Field Experiments	12
1.6 The Present Study.....	13
Methods.....	14
2.1 PRISMA Statement	14
2.2 Literature Research Strategy	14
2.3 Selection Criteria	15
2.4 Coding Procedure	15
Results.....	16
3.1 Study Selection.....	16
3.2 Interrater Reliability	17
3.3 Categorization of Studies and Overall Frequencies	18
3.3.1 Outcome Measures	18
3.3.2 Type of Intervention.....	26
3.3.3 Population	26
3.3.4 Region.....	27
3.3.5 Type of Tax.....	28
3.3.6 Effectiveness	29
3.3.7 Control Group	29
3.4 Empirical Analysis	30
3.4.1 Overall Frequency Distribution of Effects	30
3.4.2 Effectiveness and Compliance History.....	34
3.4.3 Effectiveness and Employment Context.....	40
3.4.4 Effectiveness and Regional Context	41
3.4.5 Correspondence Analysis	42
Discussion	44
Acknowledgments.....	52
References	53
Appendix A.....	60
Appendix B.....	63
Appendix C	75
Appendix D	77
Zusammenfassung.....	79

Abstract

Not all nudges aim to persuade gently: this systematic review investigates how deterrence-based strategies shape taxpayer behavior across diverse contexts. Drawing on existing empirical literature within the framework of natural field experiments, studies were coded and categorized according to compliance history, employment context, regional setting, and the specific type of intervention employed. The analysis revealed that deterrence interventions were more frequently effective than ineffective among individuals with a history of non-compliance. At the same time, studies involving general population samples more often reported no significant effects. However, when measured at the extensive margin, focusing on the binary decision of whether to comply or not, positive effects were observed in both subgroups, with already evident improvements among non-compliant individuals becoming even more pronounced. Regionally, the effectiveness of deterrence interventions varied, highlighting the influence of differing geographical contexts and tax systems. Interventions conducted in Europe demonstrated powerful results. In South America, the predominant share of outcomes was also positive, although the effects were less pronounced than in Europe. In North America, ineffective outcomes were slightly more prevalent. Evidence from Africa, Asia and the Middle East remains limited, restricting broader generalizations. Overall, interventions involving threats of audit and detection were found to be slightly more effective than penalty threats in enhancing tax compliance. Although concerns have been raised in previous literature regarding the potential for deterrence-based strategies to backfire, this review found little evidence of negative effects; most interventions increased compliance or had no measurable impact, with positive effects slightly outweighing zero effects overall. These findings support prior meta-analytic evidence, suggesting that deterrence remains a key factor in shaping tax behavior, and underscore the importance of contextual sensitivity in the design of effective compliance interventions.

Keywords: tax compliance, nudging, deterrence, tax audit, penalty threat

Introduction

In recent years, tax authorities have increasingly focused on taxpayer behavior and the factors that increase tax compliance. Since effective tax collection is essential to a well-functioning state and the provision of public goods, efforts are made to shed light on which determinants affect tax compliance (see, for example, Hallsworth et al., 2017). It is a key challenge to curb tax evasion, as it not only leads to substantial losses in government revenues but also contributes to significant social inequities. Contrary to the well-known and frequently quoted statement by Benjamin Franklin, taxes are not at all inevitable. Tax administrations must rely on self-reported data, so individuals often do not comply, since they seemingly have an economic incentive to underreport income (Alm & Torgler, 2011). Intentional and illegal actions are taken to reduce tax obligations, and frequently in a successful way. This takes place by underreporting incomes, sales, or wealth, overstating deductions, or simply failing to file appropriate tax returns. Additionally, tax authorities face financial constraints in implementing deterrence measures, such as increasing audit rates. The study also found that less than 1% of all individual income tax returns are audited, and subsequent penalties on evasions constitute merely a fraction of the amount of those unpaid taxes. Assessing how much economic activity is concealed presents a significant challenge, as individuals who take part in evasion are highly motivated to conceal their illegal actions. Estimates provided by official sources and academic researchers consistently indicate that tax evasion represents an economically significant activity (Hashimzade et al., 2013). For instance, it has been reported that the loss of government revenue due to tax evasion in the United States amounts to \$500 billion (Bott et al., 2017).

To increase tax compliance, the concept of nudging has become a widely adopted approach, now utilized by policymakers and researchers globally. As defined by Thaler and Sunstein (2008), a nudge is an “aspect of the choices architecture that alters people’s behavior predictably without forbidding any options or significantly changing their economic incentives” (p. 6). Compared to the minimal costs of implementing nudges (e.g., sending out letters), the potential benefits regarding the rise in tax collection can be remarkably substantial. In their respective studies, Hallsworth et al. (2017) estimate a £9 million increase in gathered taxes, while Bott et al. (2017) report an approximate amount of \$25 million rise in tax revenues, almost doubling the average self-reported income, all due to such cost-efficient interventions.

In a meta-analysis conducted by Antinyan and Asatryan (2019), results show that deterrence interventions are generally more potent in increasing tax compliance levels than non-deterrence interventions. Deterrence interventions are nudges that inform taxpayers about audit probabilities and potential penalties, whereas their softer, non-deterrent counterparts highlight elements of tax morale, focusing on voluntary compliance and intrinsic motivation, e.g., the provision of public goods. Antinyan and Asatryan (2019) found that, on average, only deterrence interventions were successful in increasing compliance levels, while non-deterrence interventions were on average ineffective in doing so. Additionally, individuals who were previously non-compliant were found to be more likely to be affected by nudges than compliers. These findings are relatively minor in scale, however, this evidence could lead to the interpretation that for one's individual compliance decision, personal financial motives, and thus the avoidance of detection and penalties tend to be a more compelling motivator than moral duty.

The current systematic review aims to expand on the individual and geographical contexts in which deterrence interventions influence tax compliance by evaluating, synthesizing, and further investigating the findings of existing studies on this topic. Section 1.1 introduces the standard utility model on tax compliance developed by Allingham and Sandmo (1972). Section 1.2 will examine the limitations of this model and further explore the 'Slippery Slope' Framework proposed by Kirchler et al. (2008), along with the responsive regulation model (Braithwaite, 2007) in Section 1.3. In the subsequent Sections 1.4 and 1.5, the various methods for investigating tax compliance, including the more recent approach of natural field experiments, will be discussed. This thesis is based on a cooperative research design, developed together with a colleague, who investigated a distinct aspect of the complete dataset. Further details on the collaborative research process and the objectives of the present systematic review are provided in Section 1.6.

1.1 The Deterrence Approach

According to the model by Allingham and Sandmo (1972), taxpayers make decisions based on expected utility, weighing the financial benefits of tax evasion against the risks of detection and penalties. Since the decision to declare taxes does not inherently trigger a response in the form of a penalty, taxpayers are presented with two options: They can either declare their income correctly, or they can declare less than that. By choosing the latter strategy, the outcome depends on whether the taxpayer is audited; if not, they are in a more advantageous position than under the first strategy. Instead, if they are audited without

declaring their taxes truthfully, the consequence would be a penalty. Hence, decision-making in this context is non-trivial and, according to this model, depends on four main factors: actual income level, tax rates, audit probability, and the magnitude of fines. Assuming taxpayers base their decisions on tax declaration in a way that maximizes their expected utility, they would not pay more taxes than they are obliged to. Therefore, reducing non-compliance requires implementing vigilant monitoring, the threat of sanctions and penalties, and restrictions on opportunities for tax fraud. Without these measures, the standard utility model suggests that taxpayers will not comply.

The deterrence approach tracing back to the pioneering work of Allingham and Sandmo (1972) has been dominant in the field of tax compliance research. However, this prediction is in contrast with the real-world observation of individuals tax-paying behavior. It substantially overestimates non-compliance, considering the very low likelihood of an audit, the fairly modest penalties, and the high number of taxpayers that comply (Hashimzade et al., 2013; Kleven et al., 2011). This compliance paradox led to the suggestion that the more intriguing question is not why individuals evade taxes, but why they pay them. (Alm, 2012).

The predictions of the standard utility theory, which holds the assumption that individuals always act perfectly rational, were further challenged by Thaler (1988). He unveiled a variety of anomalies in human behavior, such as the *Endowment Effect*, which states that individuals tend to overvalue items they own, regardless of their objective market value. Regarding tax policy changes, they might show more resistance towards tax increases, since the psychological impact of perceived loss is commonly greater than the positive feelings obtained from potential gains associated with tax reductions (see also Yang, 2023).

Another heuristic used in decision-making under uncertainty is described within the *Framing Effect*. The way tax information is presented can significantly influence compliance. For example, emphasizing the benefits of tax payments may increase compliance, while pointing out penalties may result in resistance (Casey & Scholz, 1991).

These mentioned heuristics are only a fraction of the numerous mental shortcuts individuals take when making decisions, especially under uncertainty. This demonstrates that agents do not have stable, well-defined preferences, and rational choices are not always made (McCaffery et al., 2003; Thaler, 1988). Such evidence suggests deviations from utility maximization theories centered around the so-called homo economicus and that compliance cannot be solely explained by the benefit-cost analysis of amoral individuals. When making decisions, taxpayers might additionally be influenced by factors such as tax morale (Graetz & Wilde, 1985).

1.2 Slippery Slope Framework

The above-mentioned limitations of the expected utility theory indicate that the compliance decisions made by individuals is highly influenced by additional factors not accounted for by the elemental ‘economics-of-crime’ framework (Alm & Torgler, 2011). An extensively referenced concept, the Slippery Slope Framework by Kirchler, Hoelzl, and Wahl (2008), seeks to explain the factors influencing tax compliance. This model distinguishes between enforced and voluntary compliance, emphasizing the essential roles of both deterrence and non-deterrence factors. According to this theory, the tax climate in a society varies on a continuum between antagonistic and synergistic. Antagonistic climates describe societies where taxpayers and authorities work against each other, resulting in a ‘cops and robbers’ mentality, where compliance can merely be reached by enforcement. In contrast, in a synergistic climate, both parties cooperate and collaborate towards common goals, and authorities prioritize transparency and respectful interactions with taxpayers. This can be described as a ‘service and client’ mindset, where voluntary compliance is likely due to taxpayers’ sense of responsibility.

Furthermore, this framework states that compliance is influenced by two principal and interacting dimensions: the power of authorities and the trust in authorities. The power dimension refers to the authorities’ ability to enforce compliance through deterrent measures, such as audits and penalties, as perceived by the taxpayer. In contrast, trust in this context relates to individuals’ perceptions of the authorities regarding their trustworthiness and benevolence. Hence, this model emphasizes the importance of the interaction between both parties, suggesting that the regulations set by the authorities and taxpayers’ attitudes toward these authorities play key factors in influencing compliance behavior.

Additionally, the dimensions of trust and power interact dynamically. If, according to this framework, trust in authorities is low, tax compliance tends to increase when the power of authorities increases, such as through higher audit probabilities and severe penalties. Taxpayers may comply primarily out of fear of detection or punishment, resulting in enforced compliance. Conversely, high-trust environments where taxpayers perceive authorities as benevolent and fair, they are more likely to comply voluntarily, without the need for coercion. Trust variations are most critical when power is low, although they become irrelevant when power is at its maximum since authorities can enforce full compliance. In contrast, variations in power are most relevant when trust is low, however, they are inconsequential when trust is at its highest since taxpayers will contribute their share regardless.

The dynamic nature of this model is further evident in the way changes in one dimension can also affect the other. For example, increased power in the form of more frequent auditing and more severe fines can be perceived as a lack of trust towards the honest taxpayer, especially if they are conducted in harsh or seemingly unfair ways. Taxpayers may feel that they are being treated as potential criminals rather than responsible citizens, which, in turn, may result in reducing their overall trust in authorities. This may lead to resentment, reducing the likelihood of voluntary compliance in the future. On the other hand, if authorities are perceived as effective in detecting fraud and penalizing tax evaders in a fair manner, taxpayers may develop greater trust. Witnessing fair consequences for wrongdoing can foster a more positive perception of tax authorities, leading to increased compliance. Additionally, taxpayers' attitudes toward authorities are strongly influenced by how these authorities communicate their actions.

In summary, the relationship between power and trust is intricate and dynamic, and variations in one dimension can profoundly affect the other, influencing taxpayer behavior in numerous ways.

1.3 Responsive Regulation

While the Slippery Slope Framework emphasizes the dynamic relationship between power and trust, as well as underlying mechanisms and psychological aspects thereof, the responsive regulation model by Braithwaite (2007) focuses on regulatory strategies authorities can apply depending on the behavior of taxpayers. To encourage compliance, the model suggests various levels of regulatory interventions that tax authorities can implement, which are illustrated in the form of a pyramid. It is structured from the least intrusive methods at the bottom to the most intrusive strategies at the top. At a base level, education and service lay the grounds for ensuring compliance in a non-intrusive way, providing taxpayers with information, guidance, and support, to help them understand the tax system and their responsibilities. Slightly more intrusive tactics, as shown in the middle of the pyramid, include reminders, warnings, and audits to signal to taxpayers that their non-compliant behavior is disapproved. Finally, at the top of the pyramid and its highest level of intrusion, tax authorities may resort to more severe measures, such as penalties or audits. The idea behind the regulatory pyramid is that authorities can achieve compliance using lower levels, relying on education and persuasion rather than coercion. The most intrusive level is used only as a last resort, after lower levels have failed to ensure compliance. The pyramid further emphasizes the importance of escalating interventions only when necessary, followed by de-

escalation once compliance is achieved. The objective of this approach lies in cultivating a cooperative relationship between tax authorities and taxpayers, ultimately resulting in a more effective and sustainable tax system. Responsive regulation, according to Braithwaite (2007), stands in contrast to regulatory formalism, which relies on predefined rules and procedures, often applying them universally without considering individual contexts. This rigid approach relies on punishment and enforcement as the primary tools for achieving compliance, using automated processes that may alienate taxpayers and overlook the underlying causes of certain behaviors. Braithwaite also argues that deterrence interventions may have counterproductive effects, especially if taxpayers perceive them as unjust, potentially resulting in adverse outcomes that undermine the overall objective of a well-functioning tax system.

To summarize, responsive regulation strives for active engagement with taxpayers through dialogue and understanding, allowing for more flexible responses in regulatory methods and focusing on building a culture of cooperation and shared responsibility between taxpayers and tax authorities.

In addition to highlighting the importance of measures set by authorities, Braithwaite (2003) defines five motivational postures individuals can hold towards regulatory systems and how they affect their compliance decisions. Two of these postures indicate a generally positive attitude towards authorities, termed *commitment* and *capitulation*. Committed taxpayers genuinely feel morally obliged to act in the collective's interest and contribute their share with good will. The second positive posture, capitulation, involves accepting the tax office as a legitimate power as long as one complies with its demands. In contrast, the three distinct postures of defiance, which indicate a negative attitude towards authorities, are labelled *resistance*, *disengagement*, and *game playing*. Resistance reflects doubts about the tax office's intentions, promotes vigilance against perceived overreach and a willingness to curb its power. Disengagement represents another posture of resistance, however, in this case, it is more pervasive and signifies disenchantment with authorities, judging the system as irredeemable. The primary objective for disengaged individuals is not to challenge authorities but to maintain a significant distance while neglecting their tax responsibilities. Finally, individuals who adopt a game-playing attitude perceive laws as tools to be manipulated for personal gain, indicating a strategic approach to engaging with the tax system. These five postures capture the range of attitudes and behaviors taxpayers exhibit towards authorities, reflecting their varying levels of acceptance, defiance, and strategic engagement with the tax system.

Regarding the relations of these different types of motivational postures to compliance, Kirchler and Wahl (2010) found that voluntary tax compliance is positively correlated with the positive postures of commitment and capitulation, and as expected, negatively correlated with resistance, disengagement, and game playing. The negative attitude of resistance is positively correlated with enforced compliance. Resistance is also positively associated with tax evasion, as are disengagement and game-playing attitudes, while the latter is also additionally correlated with tax avoidance. While tax avoidance as well as tax evasion are forms of non-compliance, the first describes a way to reduce taxes by legal means by taking advantage of loopholes in the law (Sandmo, 2005). Tax evasion, however, is illegal, as it involves the intentional act of non-compliance by understating income or overstating deductions, thereby reducing tax owed (Elffers et al., 1987; Sandmo, 2005).

It is important to note that, while both voluntary and enforced compliance lead to the fulfilment of tax obligations, there is no correlation between these two forms of compliance, since the individual's motives to comply are inherently different. Voluntary compliance stems from the taxpayers' trust in authorities, whereas enforced compliance results from the authorities' ability to conduct audits and impose penalties effectively (Kirchler et al., 2008).

1.4 Investigating Tax Compliance

While audits are the most direct method for detecting tax fraud, they are certainly not the most efficient (Hallsworth, 2014). Audits are expensive, officials might not detect undeclared income, and it is difficult to distinguish between fraud and an actual error. Furthermore, there is a lack of consistency among auditors, and tax authorities are unable to detect fraudulent individuals who simply fail to file tax returns. These limitations have contributed to the increased use of tax experiments in controlled laboratory settings to analyze tax compliance (Hallsworth, 2014). The experiments take place over several rounds, where at the beginning, subjects earn or are given an income and then must decide the amount they declare. As in real-life situations, taxes are paid only on reported income, not on any underreported amounts. Taxpayers are informed beforehand that failing to declare their full income carries a certain probability of detection and may result in a fine. At the end of the experiment, each subject is paid the amount they successfully accumulated (Alm & Torgler, 2011).

Experiments within these controlled laboratory settings hold several advantages (Hallsworth, 2014). Most importantly, they make evasion visible, since it can be easily measured within these controlled conditions. Secondly, they allow for greater manipulation

and isolation of variables, facilitating the measurement of their causal effect without external interference. Another advantage lies within the cost-effectiveness of those experiments, and the opportunity for these conditions to be configured and reconfigured quite easily (Webley et al., 1991). However, the use and interpretation of tax compliance experiments in controlled settings should be approached with caution (Alm & Torgler, 2011). Like lab experiments in general, the findings of tax compliance experiments often do not apply to human behavior in the real world. Additionally, in these types of studies, the subjects most frequently participating are students, which may not be entirely representative of real-world taxpayers. Subjects in an experiment may also be prone to modify their behavior since they are aware that they are being observed while taking part in an experiment. Therefore, the external validity of tax experiments in laboratory settings is limited, and results should be interpreted cautiously.

1.5 The Rise of Natural Field Experiments

Several limitations of laboratory tax experiments have been addressed with the emergence of field tax experiments (Mascagni, 2018). Instead of isolating subjects in artificial laboratory settings, field experiments take place under real-life conditions. The participants are real taxpayers in their usual environment and data is collected from actual returns.

According to Roth's (1986) categorization, experiments aimed at testing and refining formal economic theories are described as 'speaking to theorists', while large-scale field experiments offer the opportunity to directly influence policy-making strategies, which he refers to as 'whispering into the ears of princes'. In the latter category, experimenters can engage in a dialogue with policymakers, not just with each other. Substituting lab settings with field experiments in real-life situations can provide a more nuanced understanding of tax compliance. While lab experiments may highlight theoretical foundations as a proof of concept, field experiments are better suited to test concrete policy options, which greatly increases their relevance (Mascagni, 2018).

According to Harrison and List (2004), field experiments can be classified into the following categories:

- *artefactual field experiments* are the same as conventional lab experiments but with a nonstandard subject pool (i.e. non-students);
- *framed field experiments* provide an additional field context in either the commodity, task, or information set;

- and, within *natural field experiments*, subjects naturally undertake tasks and are not aware that they are taking part in an experiment.

Natural field experiments (NFEs) are based on randomized treatment and control groups and the subsequent comparison of those (Mascagni, 2018). Subjects are provided with information intended to influence their tax compliance behavior, such as perceptions of audit probability, fairness, and tax morale. The treatments are frequently conducted by sending letters directly to taxpayers. These letters can contain deterrent messages stating that audits are conducted in a routine manner, and evading taxes can lead to facing severe fines. Another approach is to send letters including non-deterrent messages that highlight the importance of paying taxes, e.g., in terms of the provision of public goods, such as financing schools and hospitals. The latter caters to subjects' perceptions of tax morale and fairness, whereas the first has more of a threatening undertone and compliance is interpreted as a mere gamble with the risk of punishment (Hallsworth, 2014). Treatment groups are then compared to control groups, whereof subjects have not received a letter, or received a letter without any informational treatment. The variations in taxpayer behavior across groups, as well as before and after receiving the letters, can be examined, allowing researchers to determine the causal effect of various factors on tax compliance (Mascagni, 2018).

1.6 The Present Study

The current systematic review examined the effectiveness of deterrence nudging interventions on tax compliance across various population subgroups, considering individuals' compliance history and employment context. Building on the theoretical foundations outlined earlier, from the classical economic deterrence model (Allingham & Sandmo, 1972) to the more nuanced concepts of the Slippery Slope Framework (Kirchler et al., 2008) and the responsive regulation model (Braithwaite, 2007), this review seeks to empirically evaluate how deterrence strategies operate across different individual contexts. The review also incorporates methodological insights from the growing use of natural field experiments, which offer a valuable enhancement to traditional lab-based approaches by capturing real-world taxpayer behavior in natural settings. These experiments provide the foundation for the evidence analyzed in this review, enabling a comprehensive examination of the causal impact of deterrence messages on compliance outcomes. In addition to examining contextual variation, deterrence interventions were categorized based on the type of deterrence message used in each study to assess which strategies were most effective in promoting tax

compliance. Following the synthesis and coding process, the complete dataset was split into two separate analyses. This review focuses exclusively on deterrence interventions, while the effectiveness on non-deterrence interventions was analyzed separately by a colleague (Truzka, 2021). The coding scheme was jointly developed, while the coding of the studies and the subsequent data analyses were carried out separately by each researcher. Finally, beyond individual-level factors, deterrence interventions were also examined in relation to the geographical region where the studies were conducted, in order to explore the role of broader systemic and cultural differences in shaping taxpayer behavior.

Methods

2.1 PRISMA Statement

To ensure the reporting quality as well as the transparency of this systematic review, the criteria of the PRISMA Statement (“preferred reporting items for systematic reports and meta-analyses”) were followed. Developed by Moher et al. (2009), the PRISMA Statement consists of two integral parts: A checklist containing 27 items, which serves as a guideline for the writing process of systematic reviews and meta-analyses, as well as a four-phase flow diagram to guide through the process of literature research and study selection. Following the PRISMA Statement strengthens the quality of systematic reviews and meta-analyses by facilitating standardization, enhancing transparency, and reducing the risk of biases (Moher et al., 2009).

2.2 Literature Research Strategy

To conduct the literature research, multiple sources were used to accumulate relevant studies for this systematic review. As a starting point and to gain an overview of existing studies, an informal literature search was conducted using Google Scholar. Subsequently, the databases PsycInfo, PSYINDEX, Web of Science, and Scopus were employed to conduct a comprehensive literature review.

The keywords used for screening the databases were split into different categories regarding the methodology (“field experiment”), the tax compliance context (“tax compliance”, “tax evasion”), deterrence interventions (“deterrence”, “audit”, “penalty”, “detection”, “threat”), and non-deterrence interventions (“non-deterrence”, “social norm”, “moral appeal”, “public good”, “simplification”, “reward”, “information”).

2.3 Selection Criteria

To be included in this systematic review, studies were required to meet a set of clearly defined criteria to ensure the relevance and comparability of the findings: First, only natural field experiments were accepted, ensuring that real taxpayers behaved in their natural environment. Second, studies had to employ nudging interventions that targeted tax compliance behavior. These interventions could be based on deterrence principles (e.g., emphasizing penalties or audits), non-deterrence strategies (e.g., appealing to social norms, trust, or moral obligations), or a combination of both. Third, only direct effects on taxpaying behavior were of interest within this systematic review, while effects regarding mediation or moderation, spill-over or network effects were not included. Finally, in order to assess and compare the effectiveness of the interventions across studies, it was necessary for the study to report statistical significance levels to classify them as positive, zero, or negative. This classification served as the foundation for synthesizing the evidence and drawing conclusions about the overall effectiveness of different nudging strategies.

2.4 Coding Procedure

Following the study selection process, the most relevant information about each study was gathered in a first coding sheet as a general overview. The data was extracted and assorted into the following categories:

- title, author(s), year of publication
- key results
- message content (i.e., deterrence, non-deterrence, or both)
- independent variable (i.e., type of intervention(s))
- dependent variable (i.e., outcomes measured)
- type of tax (e.g., income tax, property tax, etc.)
- participants (e.g., individuals, companies, self-employed), including the sample size
- country
- effectiveness

In an additional column, the exact wording of each intervention was transcribed. This initial data extraction was the basis for synthesizing the information into the final coding table, which consists of the following categories:

- title, author(s), year of publication
- study outcome
- type of intervention (i.e., deterrence, non-deterrence, or both)

- type of nudge element
- population (e.g., individuals, companies, self-employed)
- region
- type of tax (e.g., income tax, property tax, etc.)
- effectiveness
- control group (e.g., no letter, simple letter)
- sample size

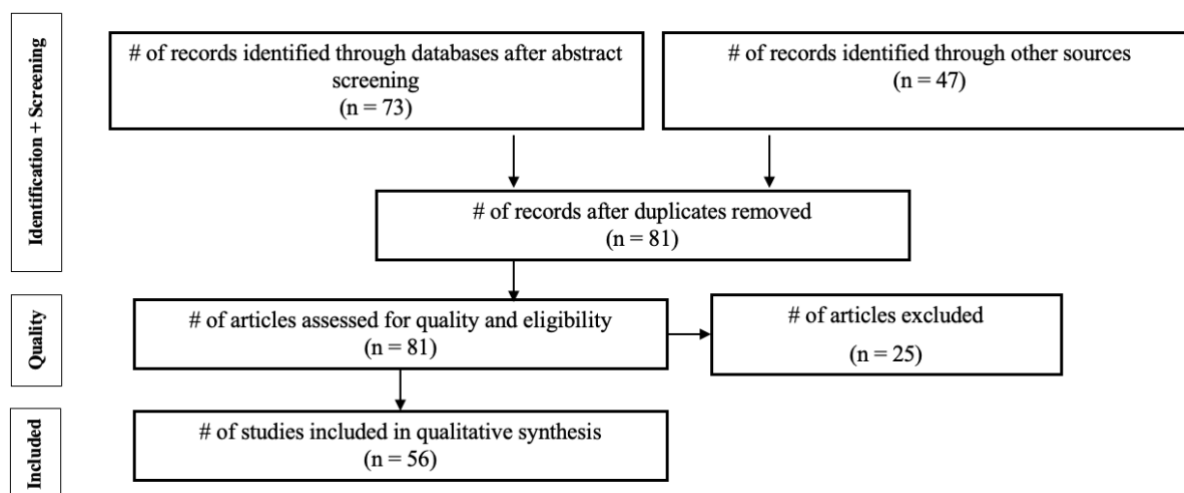
According to this coding pattern, each intervention used in the reviewed studies was systematically categorized. For each of these broad categories, more specific subcategories were defined, to which the individual interventions were further allocated. The subcategories will be described in detail in Section 3.

Due to the abstract nature of the categorization process and the potential loss of specific contextual information, an additional column was included in the coding table to provide more detailed descriptions of each intervention. The coding procedure for each intervention was conducted by both reviewers.

Results

3.1 Study Selection

The study selection process was guided by the PRISMA flow diagram, depicted in Figure 1. The studies identified through the databases mentioned in Section 2 ($n = 120$) were further reduced by removing duplicates ($n = 81$). The selection of studies was then revised in detail, causing some of the studies to be excluded from the sample for several reasons. After screening the studies in greater detail, it was found that in some of the studies, the target population was no real taxpayers, the intervention was not delivered via letter, e-mail, or SMS, or the outcome was measured qualitatively and not quantitatively, which, as mentioned in Section 2 were all critical for the selection of the studies. A small number of studies was further excluded for the following reasons: the outcome was measured in follow-up, network, or spillover effects only (i.e., no direct effects), the outcome was not of interest for the present study (e.g., demographical effects), or the interventions used in the study did not qualify as a deterrence or non-deterrence nudge. Consequential to further exclusion of these studies, a total of 56 studies were included in this systematic review (see Table 1), which were subsequently categorized into the coding table by both raters.

Figure 1*PRISMA Flow Diagram*

3.2 Interrater Reliability

To assess the degree of agreement between the raters, Fleiss' kappa (Fleiss & Cohen, 1973) was calculated for the categories that showed the greatest variation between studies. *Type of Intervention*, *Population*, and *Effectiveness* were the three main categories for which interrater reliability was evaluated, as well as the overall interrater reliability. Fleiss' kappa quantifies the degree to which observed agreement between raters exceeds the level expected by chance, assuming random categorization. Following the characterization by Landis and Koch (1977), Fleiss et al. (J. Fleiss et al., 2003) defined specific value ranges for kappa to indicate the degree of agreement. Values greater than approximately .75 are considered to represent excellent agreement beyond chance, values below .40 may be taken to reflect poor agreement beyond chance, and values between .40 and .75 are interpreted as indicating fair to good agreement beyond chance. For the category *Type of Intervention*, the κ coefficient reached .692; for the category *Population* .777; and .925 for the category *Effectivity*. Across all three categories, this resulted in an interrater reliability of $\kappa = .825$, indicating an excellent level of agreement between both raters.

Following the process of categorization, any variations in ratings between the two raters were discussed and with an experienced third rater. Additionally, categories with inexact definitions were precisely redefined with his support and the guidance of the supervisor of this systematic review. Figures displaying Fleiss' kappa for the combined and separate categories are provided in Appendix D.

3.3 Categorization of Studies and Overall Frequencies

The initial sample consisting of 56 studies included in the coding table was further narrowed down, as for the data exploration of this systematic review, exclusively deterrence interventions were of interest. Consequentially, this resulted in a sample of 42 studies containing at least one deterrence intervention, with 216 deterrence interventions in total. The overall frequencies will be displayed below each category using bar charts.

3.3.1 Outcome Measures

To measure each studies direct effects, the following categories were defined in the coding table: The *Extensive Margin* refers to a binary outcome indicating the probability of compliance (i.e., whether taxes were paid, or filing occurred after the intervention had been conducted). The *Intensive Margin*, on the other hand, measures the extent or intensity of compliance compared to the control group, such as the amount of taxes paid after receiving the intervention. The category *Timely Payment* is defined as another binary outcome indicating the probability of compliance within two months following the intervention. The fourth outcome category *Tax Deductions/Claims* measures if the intervention had any effects on tax deductions or credit claims. The outcome was classified based on whether there was a statistically significant impact on these variables. As shown in Figure 2, the category intensive margin was coded 114 times, representing the most frequent outcome category.

Figure 2

Overall Frequency Distribution of Outcomes

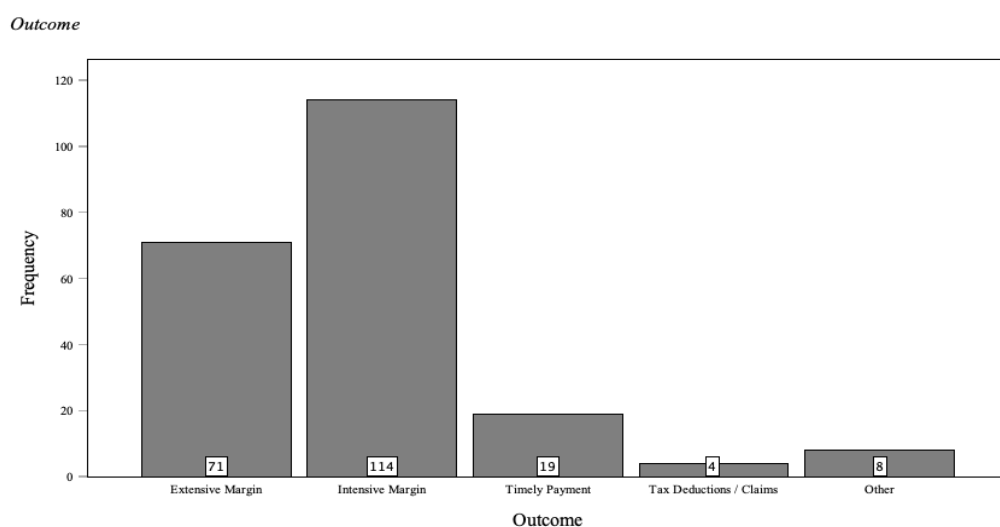


Table 1

Overview of Studies Containing Deterrence Interventions in this Systematic Review

Author(s), Year of Publication	Country	Participants	Deterrence Intervention(s)	Dependent Variable(s) of Interest
Alm, Cifuentes, Niño, and Rocha (2019)	Colombia	3,606 individual social protection contribution taxpayers	Possible/likely/certain audit message, likely penalty message	Reported health care payments, reported pension payments
Appelgren (2008)	Sweden	768 sole proprietors	Random and rational audit threat	Changes in declared income
Ariel (2012)	Israel	4,935 corporations	Penalty threat	Gross sales values reported, VAT payments, VAT deductions
Bérgolo, Ceni, Cruces, Giacobasso, and Perez-Truglia (2019)	Uruguay	16,392 small- and medium-sized firms (primary sample) 4048 small- and medium sized firms (secondary sample)	Audit-statistics message (providing information about audit and penalty rates), Audit-endogeneity message (quantitative information about audit probabilities and penalty rates) (primary sample) Audit threat (25%) message, Audit threat (50%) message (secondary sample)	Perceived audit probability, Taxes paid based on administrative data
Boning, Guyton, Hodge, Slemrod, and Troiano (2018)	USA	12,172 businesses with Alert C (higher risk of falling behind in required deposits than the general population, yet not as high as groups A and B)	Information about firms' responsibilities, penalties etc., in-person visit from IRS officer	Income and payroll tax remittance, probability of tax remittance

Bott, Cappelen, Sørensen, and Tungodden (2017)	Norway	15,708 individual taxpayers	Detection letter (tax administration has received information that individuals had income and/or assets abroad)	Amount of self-reported foreign income
Brockmeyer, Smith, Hernandez, and Kettle (2019)	Costa Rica	12515 firms that were covered by at least one third-party report (Experiment 1) 37242 firms that were not covered by a third-party report (Experiment 2)	Threat message including possible "third-party information", Threat message including specific examples of "third party" reports (Experiment 1) Threat without including "third party information", threat including weaker "third party information" (Experiment 2)	Income tax compliance rate, Tax payment, other compliance outcomes (e.g., sales tax declarations, sales tax payments etc.)
Carrillo, Pomeranz, and Singhal (2017)	Ecuador	7993 firms	Missing revenue reminders, missing revenue reminders detailing the missing amount	Income Tax response rate and reported revenue
Castro and Scartascini (2015)	Argentina	23,195 individual property owners	Penalty threat message	Rate of property tax payment
Chirico, Inman, Loeffler, MacDonald, and Sieg (2016)	USA	4,927 individual property owners	Reminder letter about detection and penalty threat	Rate of property tax payment, amount of revenue collection
Chirico, Inman, Loeffler, MacDonald, and Sieg (2019)	USA	19,039 individual property owners	Reminder letter about detection and penalty threat	Tax payment rate, amount of taxes paid

Coleman (1996)	USA	47,000 individual taxpayers	Message containing threat about higher audit probability	Changes in reported federal taxable income, taxes paid
Cranor, Goldin, Homonoff, and Moore (2020)	USA	90,349 individual taxpayers	Detailed penalty reminder, general penalty reminder	Full tax payment or payment plan creation
De Neve, Imbert, Spinnewijn, Tsankova, and Luts (2019)	Belgium	1,216,317 individual taxpayers (Experiment 1); 229,751 individual taxpayers (Experiment 2); 188,180 individual taxpayers (Experiment 3); 1,500,000 online tax filers (Experiment 4); 148,925 individual taxpayers (Experiment 5)	Deterrence message (penalty; penalty + immediacy; penalty + enforcement)	Extensive margin and intensive margin outcomes
Del Carpio (2013)	Peru	22,318 individual property owners	Information about the average level of enforcement, payment reminder threat	Rate of property tax payment
Dunning, Monestier, Pineiro, Rosenblatt, and Tunón (2015)	Uruguay	28,646 individual taxpayers	Penalty threat reminder	Intended compliance measured through access to a web account, actual tax payments
Dwenger, Kleven, Rasul, and Rincke (2016)	Germany	39,782 individual church taxpayers	Letter informing about audit probability	Rate of church tax payment vs. the amount of tax due
Eerola, Kosonen, Kotakorpi, Lyytikäinen, and Tuimala (2019)	Finland	45,000 individual property owners	Notification about the general increase of enforcement, notification about the increase in enforcement based on the use	Rate of property tax payments, amount of reported property taxes

			of third-party information on ownership of dwellings	
Fellner, Sausgruber, and Traxler (2013)	Austria	50,498 individual TV license fee payers	Detection threat message	GIS registrations or contract updates
Hallsworth, List, Metcalfe, and Vlaev (2017)	UK	119,527 individual taxpayers	Letter informing about charging with interest	Rate of tax payments
Harju, Kosonen, and Ropponen (2014)	Finland	4800 small businesses	Letter informing about audit probability of 5%, letter informing about audit probability of 33%	VAT tax reporting
Hasseldine, Hite, James, and Toumi (2007)	UK	7,307 sole proprietors	Increased audit treatment letter, certain audit treatment letter, financial penalty treatment letter	Reported turnover, Change in net profit
Hernandez, Jamison, Korczyk, Mazar, and Sormani (2017)	Poland	149,925 individual taxpayers	Penalty threat and execution order treatment, omission information	Rate of tax payments, amount of taxes paid
Hernandez, Karver, Negre, and Perng (2019)	Kosovo	11,603 individuals (letter sample), 23,622 firms (e-mail sample), 22,828 firms (SMS sample)	Simple reminder combined with a deterrence message	Rate of tax declaration, amount of taxes paid
Hiscox et al. (2018)	Australia	1,279 existing businesses (Experiment 1); 328 new businesses (Experiment 2)	Penalty threat reminder e-mail (Experiment 1), Information stressing obligations necessary for businesses (Experiment 2)	Tax compliance rate within the DGST scheme, tax payments

Iyer, Reckers, and Sanders (2010a)	USA	800 state retail firms	Letters enhancing perceived detection risk, letters raising penalty awareness, letters raising both	Reported Use and B&O tax
Iyer, Reckers, and Sanders (2010b)	USA	1000 construction firms	Letters enhancing perceived detection risk, letters raising penalty awareness, letters raising both	Reported Use and B&O tax
Kettle, Hernandez, Ruda, and Sanders (2016)	Guatemala	43,387 individual taxpayers	Information about potential penalties, audit/detection threat, information about social norms combined with detection threat	Tax declaration rate, log amount conditional on payment, amount unconditional on payment
Kettle, Hernandez, Sanders, Hauser, and Ruda (2017)	Guatemala	627,242 individual taxpayers	Information about penalties for dishonesty, information about enforcement, information about social norms combined with detection threat	Amount of online tax declaration
Kleven, Knudsen, Thustrup, Kreiner, Pedersen, and Saez (2011)	Denmark	42784 taxpayers (partly self-reported, partly third-party reported) (17764 self-employed)	Deterrence letter threatening with certain audit, deterrence letter threatening with 50% audit	Reported income tax levels
Koumpias (2017)	Greece	4,316 taxpayers	E-Mail reminder and phone call (including sanction threats)	Extensive margin, intensive margin
Larkin, Sanders, Andresen, and Algate (2019)	UK	9,130 individual taxpayers	Enforcement salience letter	Rate of tax payments, amount of taxes paid

Mascagni, Nell, and Monkam (2018)	Rwanda	10,800 business and self-employed taxpayers	Penalty threat treatment (letter, e-mail, SMS)	Rate of reduced income tax liabilities
Meiselman (2018)	USA	9,523 individual taxpayers	Penalty salience letter, Punishment probability letter	Rate of remitted tax payments
Ortega and Sanguinetti (2013)	Venezuela	6,000 firms	Detection threat letter	Tax balance (differences between tax accruals and tax payments)
Perez-Truglia and Troiano (2018)	USA	34,334 individual taxpayers	Penalty information (with interest) treatment	Probability of leaving a tax delinquent list (participants can get off the list if they fully pay the difference owed or agree to a repayment plan)
Pomeranz (2015)	Chile	102,000 small and medium-sized firms	Audit threat letter	Monthly VAT payments
Robitaille, House, and Mazar (2018)	Canada	12,505 Organizations	Penalty threat including deadline	Extensive and intensive margin
Sanders, Reckers, and Iyer (2008)	USA	277 use tax paying firms; 1,278 B&O tax paying firms	Penalty threat, penalty threat including nondeterrent accountability manipulation (signing affidavit that no tax was due)	Reported use and B&O tax base
Schwartz and Orleans (1967)	USA	173 individual taxpayers	Penalty threat letter treatment	Changes in levels of reported income, tax deductions, income tax after credits

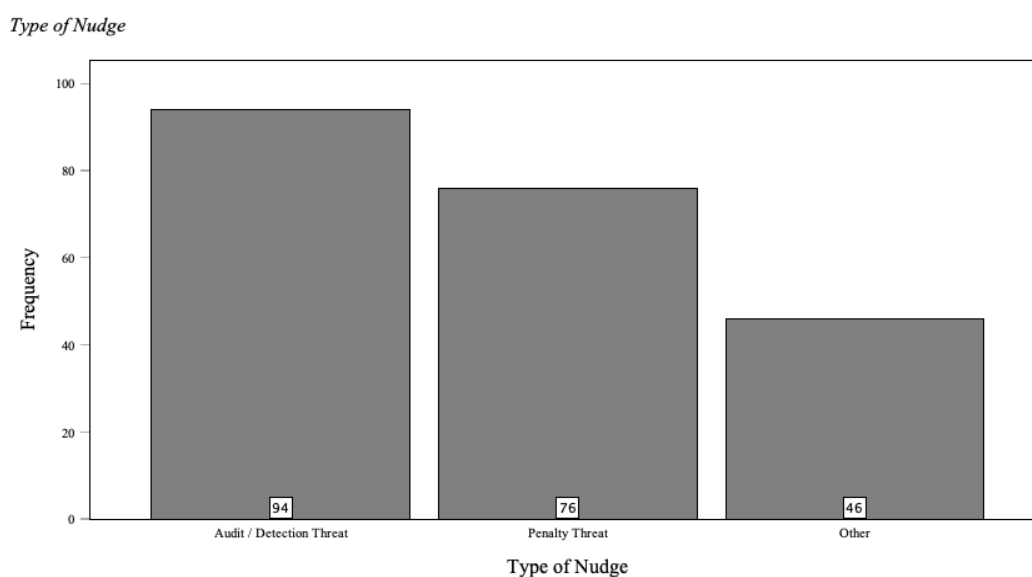
Shimeles, Gurara, and Woldeyes (2017)	Ethiopia	3,730 businesses	Audit threat and penalty threat via hand-delivered letter	Profit tax payment
Slemrod, Blumenthal, and Christian (2001)	USA	1724 individual taxpayers	Audit threat letter	Average reported federal taxable income, average reported federal tax after credits average reported Minnesota tax liability

3.3.2 Type of Intervention

For the subsample of studies that implemented deterrence-based interventions, subcategories were defined to distinguish between the different types of nudges. The first category, *Audit/Detection Threat*, refers to interventions that employed explicit warnings or reminders about the possibility of a tax audit or detection. The second category, *Penalty Threat*, includes interventions that primarily focused on threatening sanctions or punishments for non-compliance. The third category, *Other*, was used for interventions that did not fit into one of the two previous categories or that employed a combination thereof. This category also included non-deterrence nudges (e.g., additional information or simplification) that contained an additional distinct deterrent element (e.g., in the letter header). Overall, interventions categorized as *Audit/Detection Threat* appeared most frequently, coded 94 times, as depicted in Figure 3.

Figure 3

Overall Frequency Distribution of Types of Interventions



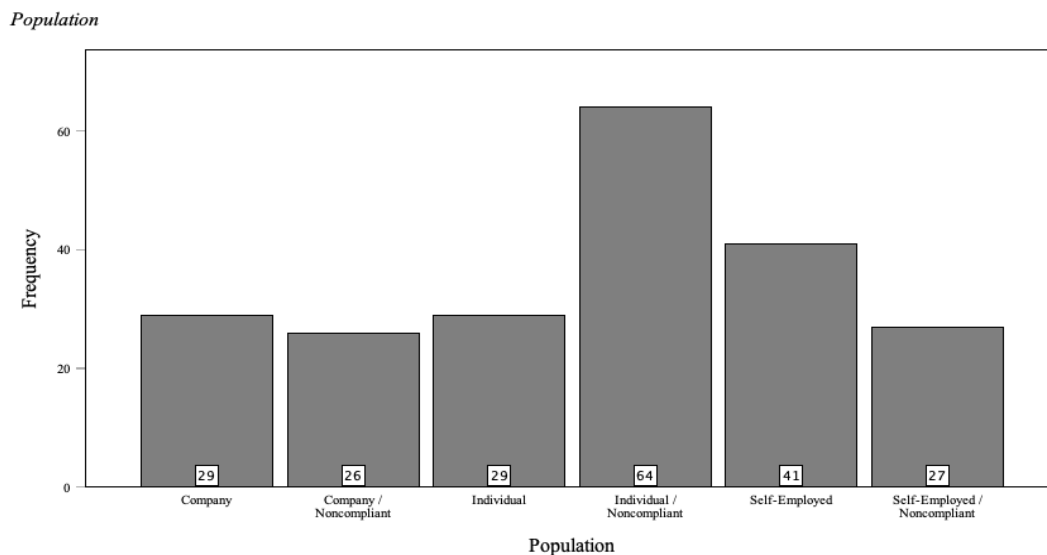
3.3.3 Population

Within the population category, three distinct subcategories were further defined. The first of these subcategories, *Company*, refers to businesses that employ a relatively large number of employees. The second subcategory, *Self-Employed*, encompasses sole proprietors as well as businesses that typically have fewer employees and may operate with less formal administrative structures. The third subcategory under the population category was labeled *Individuals*, which refers to individual taxpayers who are personally obliged to contribute

their share of taxes. To allow for a comparison of the population based on their compliance history, the participants' employment status was combined with an additional subcategory representing their previous compliance. Thus, the label *non-compliant* was added to each of the employment status subcategories if there was a record of previously failing to pay taxes accordingly. If previous non-compliance was not explicitly mentioned in the underlying study, participants were categorized as a mixed or general taxpayer group, entailing both those with and without a history of non-compliance. The subsample *Individual/non-compliant* was coded most frequently, accounting for 64 cases. The overall frequencies are provided in Figure 4.

Figure 4

Overall Frequency Distribution of Population Subcategories

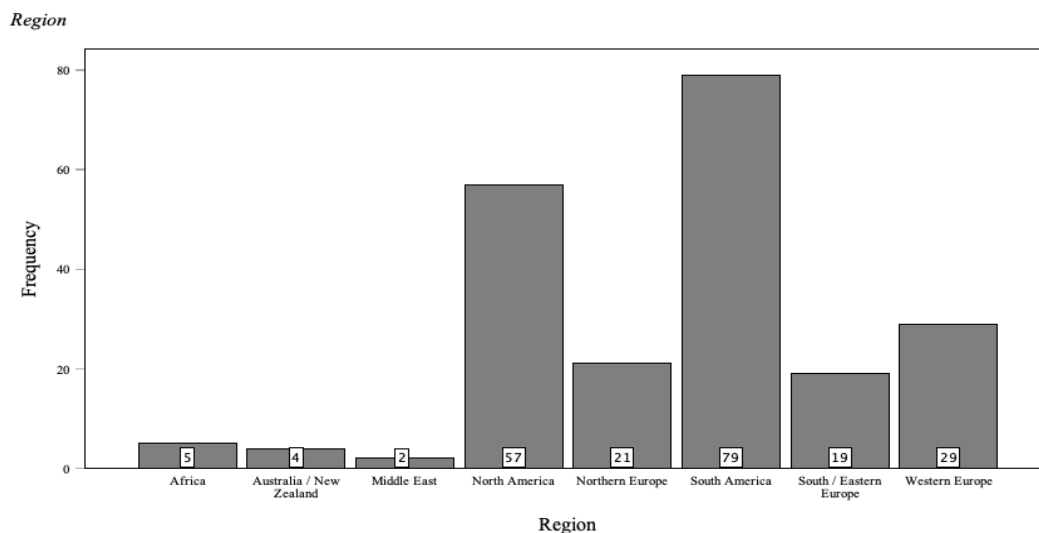


3.3.4 Region

The geographical context in which the included studies were implemented was systematically classified into the following predefined subcategories: *Africa*, *Australia / New Zealand*, *Middle East*, *North America*, *South America* (including Central America), *Northern Europe*, *Southeastern Europe*, and *Western Europe*, which includes Central European countries. Figure 5 provides a visual representation of frequencies, with the largest share of regional subcategories from *South America*, coded 79 times overall.

Figure 5

Overall Frequency Distribution of Regional Subcategories

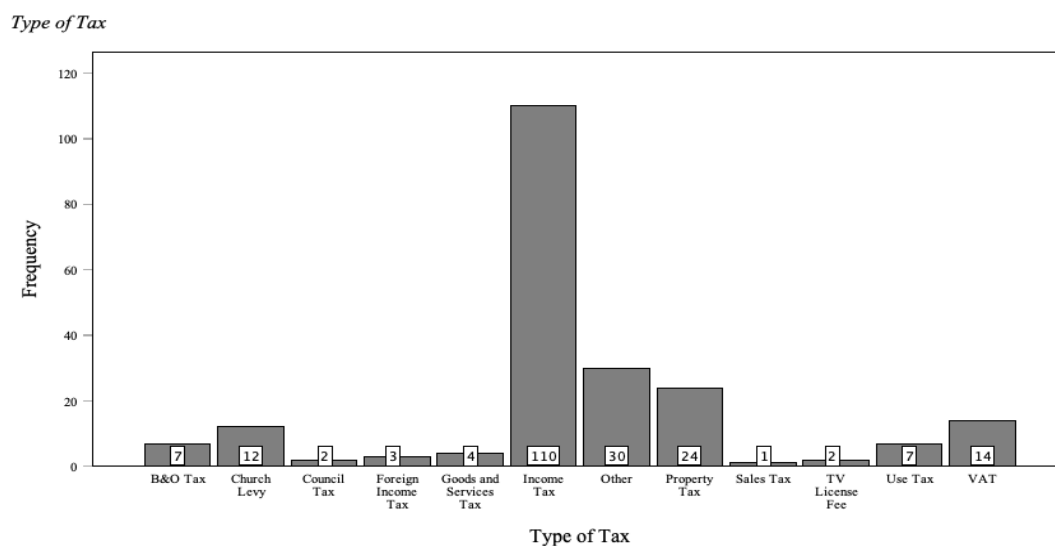


3.3.5 Type of Tax

The individual taxation types were divided into the following subcategories: *Income Tax*, *VAT*, *Property Tax*, *Council Tax*, *Foreign Income Tax*, *Goods and Services Tax*, *B&O Tax* (Washington State Business & Occupation Tax), *Sales Tax*, *TV License Fee*, *Use Tax* and *Church Levy*. An additional subcategory labeled *Other* was included for taxes that did not align with the predefined categories or were not explicitly addressed in the study. The most frequently observed type of tax was *Income Tax*, coded 110 times, as shown in Figure 6.

Figure 6

Overall Frequency Distribution of Types of Taxes

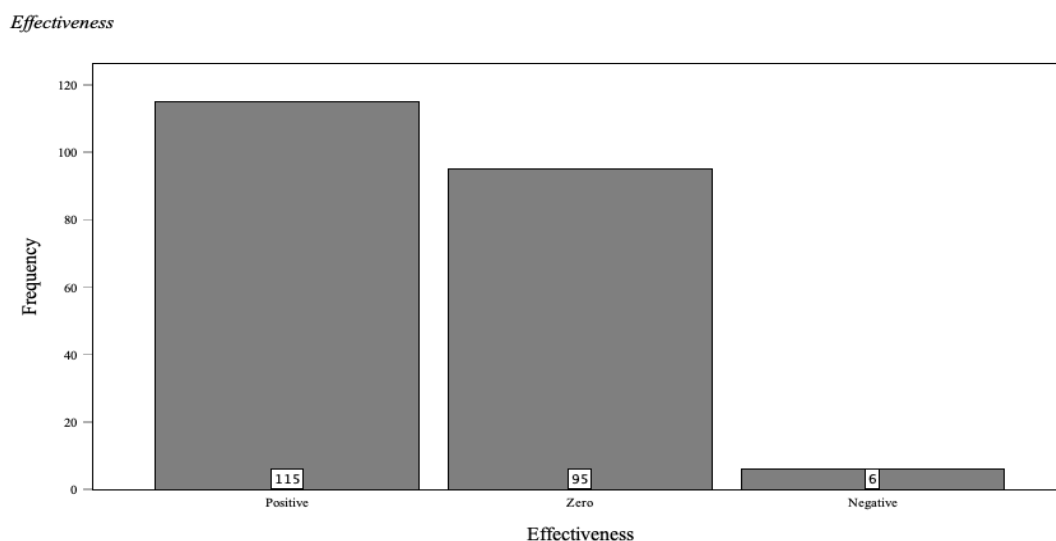


3.3.6 Effectiveness

To measure the effectiveness of each intervention, the results were classified as either *Positive Effect*, *Zero Effect*, or *Negative Effect* on the measured outcome. An effect was considered positive if the intervention had significantly increased tax compliance or tax deductions, and negative if the intervention led to a decrease in tax compliance or tax deductions, causing a negative or backfiring effect. If there was no significant effect of an intervention, it was classified as *Zero Effect*. Significance levels in the context of this systematic review were set at $p < .05$. In the deterrence sample, most interventions were coded as effective, particularly 115 cases. A negative or backfiring effect could be observed in six cases, as illustrated in Figure 7.

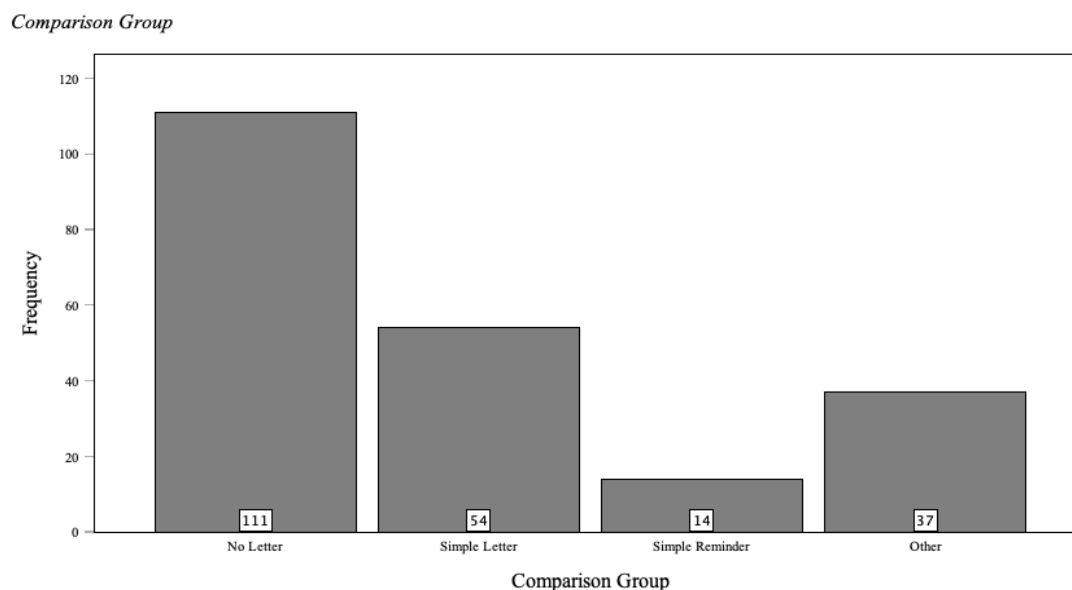
Figure 7

Overall Frequency Distributions of Effectiveness



3.3.7 Control Group

Two relevant subcategories were defined to classify which type of control group the intervention was compared against. The intervention was either compared to a *No Letter* control group, which was not exposed to any intervention or message, or a *Simple Letter / Simple Reminder* control group, in which subjects received a message without nudging elements. The subcategory *Other* contains treatments that were not clearly defined or could not be categorized into one of the described subcategories. The frequency distributions of control groups are presented in Figure 8. Interventions were most frequently compared against a *No Letter* control group, coded 111 times.

Figure 8*Overall Frequency Distributions of Control Groups*

3.4 Empirical Analysis

The results of this systematic review will be presented in the following section, with a primary focus on the frequency distributions observed within the dataset. Deterrence nudging interventions and their respective effects will be displayed across various subcategories, which include different taxpayer populations, such as individuals, self-employed taxpayers or companies. Additionally, the results will be analyzed in relation to the specific type of intervention applied, enabling comparisons between the effectiveness of distinct deterrence mechanisms. Furthermore, the analysis will incorporate geographical variation, examining how deterrence interventions and their effects differ across regions.

3.4.1 Overall Frequency Distribution of Effects

To provide a comprehensive understanding of the distribution of intervention effectiveness across various dimensions, the following section presents a detailed overview of each subcategory, specifically distinguishing between positive, zero, or negative effects as reported in the studies. To systematically compare these effects, Table 2 summarizes the frequencies of interventions that utilized deterrence nudges and presents their effectiveness within each predefined subcategory.

Table 2*Frequencies of Effects within Subcategories*

	Positive Effect (N=115)	Zero Effect (N=95)	Negative Effect (N=6)	Total (N=216)
Outcome				
Extensive Margin	55 (77.5%)	14 (19.7%)	2 (2.8%)	71 (100.0%)
Intensive Margin	42 (36.8%)	68 (59.6%)	4 (3.5%)	114 (100.0%)
Other	7 (87.5%)	1 (12.5%)	0 (0.0%)	8 (100.0%)
Tax Deductions/Claims	0 (0.0%)	4 (100.0%)	0 (0.0%)	4 (100.0%)
Timely Payment	11 (57.9%)	8 (42.1%)	0 (0.0%)	19 (100.0%)
Type of Intervention				
Audit/Detection Threat	49 (52.1%)	42 (44.7%)	3 (3.2%)	94 (100.0%)
Penalty Threat	35 (46.1%)	38 (50.0%)	3 (3.9%)	76 (100.0%)
Other	31 (67.4%)	15 (32.6%)	0 (0.0%)	46 (100.0%)
Population				
Company	7 (24.1%)	19 (65.5%)	3 (10.3%)	29 (100.0%)
Company / non- compliant	15 (57.7%)	11 (42.3%)	0 (0.0%)	26 (100.0%)
Individual	16 (55.2%)	13 (44.8%)	0 (0.0%)	29 (100.0%)
Individual / non- compliant	44 (68.8%)	17 (26.6%)	3 (4.7%)	64 (100.0%)
Self-Employed	14 (34.1%)	27 (65.9%)	0 (0.0%)	41 (100.0%)
Self-Employed / non- compliant	19 (70.4%)	8 (29.6%)	0 (0.0%)	27 (100.0%)
Region				
Africa	1 (20.4%)	4 (80.0%)	0 (0.0%)	5 (100.0%)
Australia / New Zealand	4 (100.0%)	0 (0.0%)	0 (0.0%)	4 (100.0%)
Middle East	0 (0.0%)	2 (100.0%)	0 (0.0%)	2 (100.0%)
North America	27 (47.4%)	30 (52.6%)	0 (0.0%)	57 (100.0%)
Northern Europe	11 (52.4%)	10 (47.6%)	0 (0.0%)	21 (100.0%)
South America	41 (51.9%)	35 (44.3%)	3 (3.8%)	79 (100.0%)
South/Eastern Europe	11 (57.9%)	5 (26.3%)	3 (15.8%)	19 (100.0%)
Western Europe	20 (69.0%)	9 (31.0%)	0 (0.0%)	29 (100.0%)

Tax				
B&O Tax	0 (0.0%)	7 (100.0%)	0 (0.0%)	7 (100.0%)
Church Levy	6 (50.0%)	6 (50.0%)	0 (0.0%)	12 (100.0%)
Council Tax	2 (100.0%)	0 (0.0%)	0 (0.0%)	2 (100.0%)
Foreign Income Tax	2 (66.7%)	1 (33.3%)	0 (0.0%)	3 (100.0%)
Goods and Services Tax	4 (100.0%)	0 (0.0%)	0 (0.0%)	4 (100.0%)
Income Tax	64 (58.2%)	43 (39.1%)	3 (2.7%)	110 (100.0%)
Other	11 (36.7%)	16 (53.3%)	3 (10.0%)	30 (100.0%)
Property Tax	15 (62.5%)	9 (37.5%)	0 (0.0%)	24 (100.0%)
Sales Tax	1 (100%)	0 (0.0%)	0 (0.0%)	1 (100.0%)
TV License Fee	2 (100.0%)	0 (0.0%)	0 (0.0%)	2 (100.0%)
Use Tax	2 (28.6%)	5 (71.4%)	0 (0.0%)	7 (100.0%)
VAT	6 (42.9%)	8 (57.1%)	0 (0.0%)	14 (100.0%)
Comparison				
No Letter	68 (61.3%)	41 (36.9%)	2 (1.8%)	111 (100.0%)
Other	16 (43.2%)	18 (48.6%)	3 (8.1%)	37 (100.0%)
Simple Letter	19 (35.2%)	34 (63.0%)	1 (1.9%)	54 (100.0%)
Simple Reminder	12 (85.7%)	2 (14.3%)	0 (0.0%)	14 (100.0%)

An analysis of frequency distributions across outcome subcategories revealed that within the category Extensive Margin, 77.5% of the interventions were effective. In contrast, within in the outcome category Intensive Margin a majority of interventions showed no significant effect (59.6%), while only 36.8% demonstrated a positive effect.

Considering the type of intervention, in the Audit/Detection Threat category more than half of the interventions showed positive effects (52.1%), while merely 3.2% backfired or had negative effects. In the subcategory Penalty Threat, more interventions had no significant effect (50.0%) than a positive effect (46.1%).

Examining the frequency distributions of deterrence nudging interventions regarding the population they targeted, the subsample Company displayed more ineffective interventions (65.5%) than positive effects (24.1%). Similarly, in the subsample Self-Employed, the majority of interventions (65.9%) showed no significant effect, while only 34.1% of resulted in positive effects. However, in the Individual subsample, more than half of the interventions were effective (55.2%). Interestingly, in each of the non-compliant subsamples the deterrence interventions were more successful, showing a higher frequency of positive effects than no effects. In the Company/non-compliant category, 57.7% of interventions demonstrated positive effects. The category Individual/non-compliant showed

68.8% effectiveness, while in the Self-Employed/non-compliant subsample, 70.4% of interventions had a positive effect. Backfiring or negative effects were rarely observed, occurring in three interventions within the Company subsample and three interventions within the Individual/non-compliant subsample.

To provide a more comprehensive overview, the frequency distribution of effectiveness in the Population category was further analyzed with respect to participants' compliance history (i.e., general or mixed vs. non-compliant subsample), excluding employment status, in addition to the frequencies shown in Table 2. This analysis revealed that interventions in the non-compliant subsample were more frequently effective than ineffective, with 66.7% positive effects and 30.8% zero effects. In the mixed sample, ineffective interventions outnumbered effective ones, with 59.6% showing no effect and 37.4% demonstrating positive effects. Focusing exclusively on employment status (i.e., Company vs. Individual vs. Self-Employed) and disregarding participants' compliance history, interventions in the Company subsample appeared to be more frequently ineffective (54.5%) than effective (40.0%). In the Individual subsample, deterrence interventions showed greater effectiveness (64.5%) than no effectiveness (32.3%), while in the Self-Employed subsample, zero effects were marginally more frequent (51.5%) than positive effects (48.5%).

Regarding regional subcategories, most of the interventions took place in North and South America. In South America, more than half of the deterrence interventions (51.9%) showed a positive effect, while 44.3% had no significant effect. In contrast, in North America, a greater proportion of interventions were ineffective in increasing compliance (52.6%), while positive effects were observed in 47.4% of interventions. Throughout all European subcategories, deterrence interventions increased compliance more frequently than not. For example, 69% of interventions in Western Europe showed positive effects. Negative or backfiring effects were observed in three interventions in South America (3.8%), as well as in three interventions in South/Eastern Europe, constituting 15.8% within the respective subcategory.

Regarding the Comparison subcategories, deterrence interventions were most frequently compared to a No Letter control group, as noted in the previous section. Most of these interventions compared to the No Letter comparison subgroup showed positive effects (61.3%). In contrast, in the second most frequent control group, Simple Letter, the interventions were mostly ineffective (63.0%), with 35.2% resulting in positive effects.

Among different tax types, Income Tax was the most frequently coded, appearing 110 times. In this subcategory, the majority of interventions demonstrated positive effects

(58.2%), while 39.1% showed no significant impact on enhancing compliance in the respective outcome. However, the category Type of Tax will not be further examined in this systematic review, as it was included solely to provide additional context.

3.4.2 Effectiveness and Compliance History

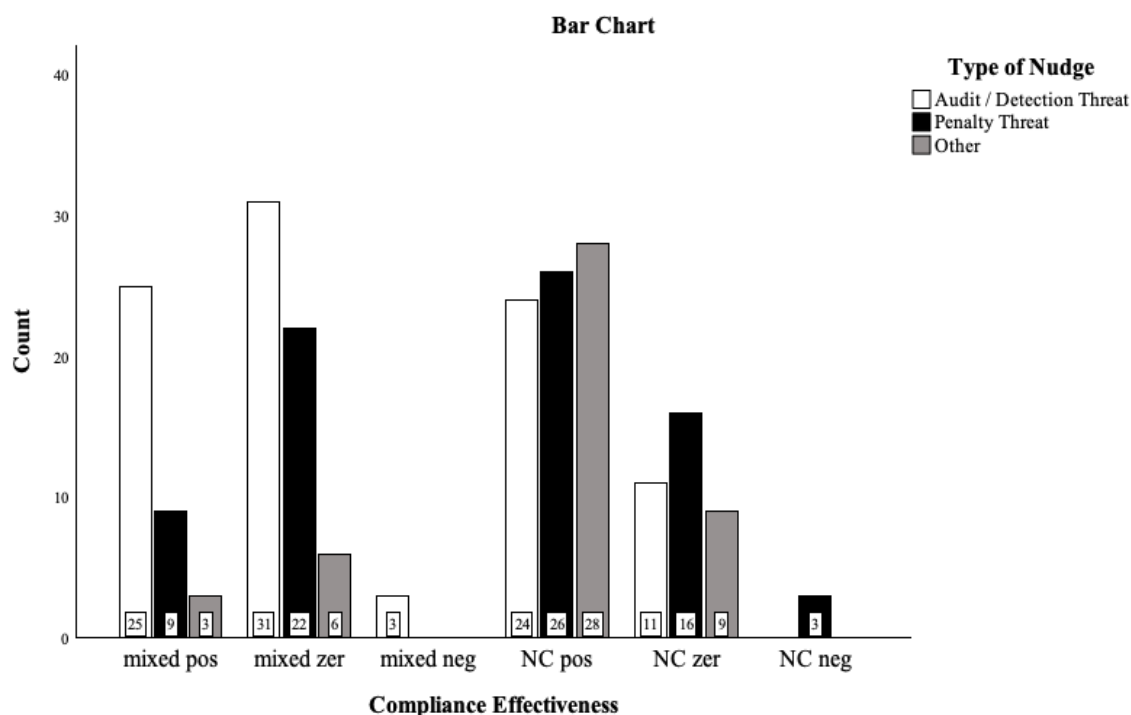
To gain a deeper understanding of which specific deterrence interventions had positive, zero, or negative effects in different population subsamples, and to provide a more detailed overview of the data, the effectiveness of the intervention subgroups was further investigated in terms of participants' compliance history. In the following section, the general or mixed taxpayer subsample (with no stated record of previous non-compliance) and the non-compliant subsample (known to have previously failed to fulfil their tax obligations) will be compared regarding the effectiveness of interventions. To further assess the effects in this context, the subcategories Individual, Company, and Self-Employed were combined and recoded into two groups: mixed and non-compliant. Subsequently, the two categories combined with effectiveness were compared to the subcategories within Type of Nudge (Audit/Detection Threat, Penalty Threat, and Other). The frequency distributions in this context are depicted in Figure 9.

An analysis of the frequency distributions revealed that negative or backfiring effects were rarely observed in both the mixed and non-compliant subsamples. In the mixed subsample, Audit/Detection Threat was more frequently effective in promoting tax compliance (26.6%) compared to Penalty Threat (11.8%). In contrast, in the non-compliant subsample, the difference in effectiveness between these types of interventions was less pronounced. Yet, in the intervention subcategory Audit/Detection Threat in the mixed sample, non-significant effects were observed more frequently than positive effects.

The intervention subcategory Other, which proved more effective in the non-compliant subsample compared to the previous subcategories, is also depicted below; however, it does not provide additional insights into the effectiveness of compliance interventions. Intervention subtypes categorized as Other did not fit in the existing subcategories, either because they were fundamentally different, or because they combined elements of Audit/Detection Threat and Penalty Threat. Furthermore, they did not occur with sufficient frequency to establish a new subcategory. Nevertheless, to preserve a broader perspective, interventions of this subcategory were not excluded from the frequency analysis.

Figure 9

Frequency Distributions of Deterrence Interventions in the Mixed and Non-Compliant Subsamples



Note. NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

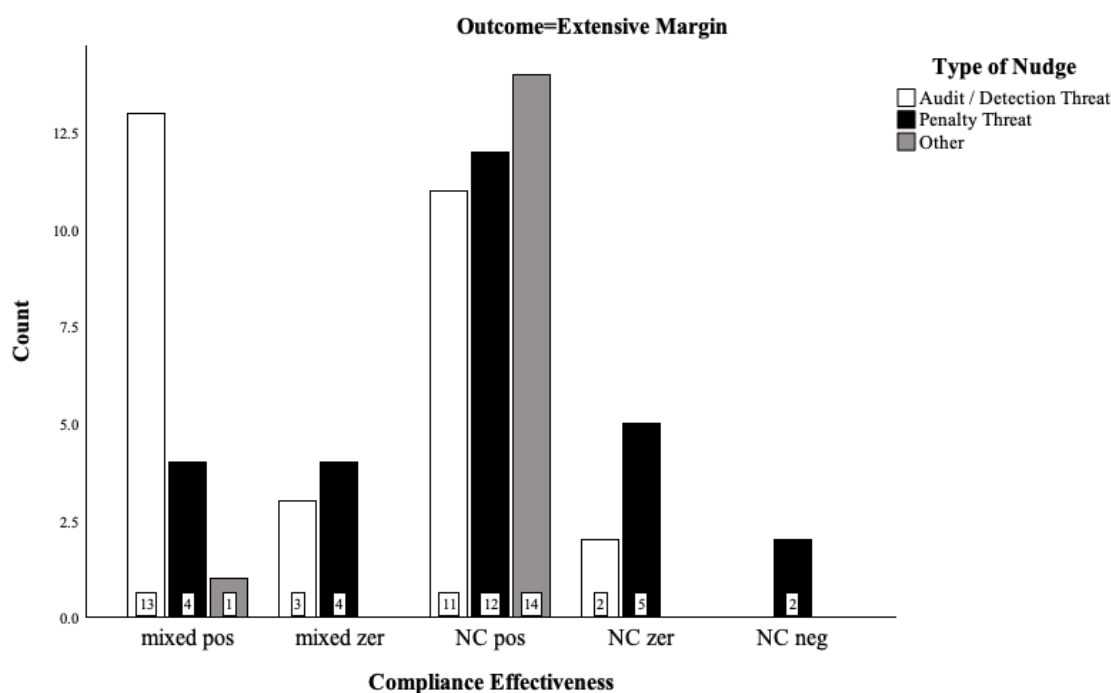
Frequency Distributions within Extensive Margin and Intensive Margin Outcomes. In the previous section, the effectiveness of the interventions was presented for all outcomes, including the Extensive Margin, Intensive Margin, Timely Payment, Tax Deductions/Claims, and Other. The subsequent analysis focuses on the two most frequently observed outcomes, extensive and intensive margin.

The outcome category Extensive Margin, as previously mentioned, describes a binary outcome of the probability of compliance, i.e., whether to fulfill tax liabilities or not after the intervention had been conducted. Since the outcome subcategory Timely Payment is defined as a binary outcome that also includes information on the probability of compliance within two months after the intervention, it was incorporated into the Extensive Margin subcategory. Figure 10 displays frequency distributions for the mixed and non-compliant subsample, with effectiveness measured at the extensive margin. In this context, the mixed subsample demonstrated greater effectiveness with Audit/Detection Threat interventions (44.8%) compared to Penalty Threat interventions (14.8%). In contrast, within the non-compliant subsample, Penalty Threat interventions appeared slightly more effective (44.7%) than

Audit/Detection Threat interventions (37.9%). In the non-compliant sample, interventions in the subcategory Other had more frequently positive effects than in the mixed sample. It is also noteworthy that, when measured at the extensive margin outcome, the effectiveness of deterrence interventions generally increased compared to results across combined outcome categories, with more positive than zero effects in the mixed sample as well. In the non-compliant sample, positive effects were already the most frequent across combined outcome categories; however, an even stronger effect emerged when measured at the extensive margin, with more than five times as many positive effects as zero effects.

Figure 10

Frequency Distribution of Deterrence Interventions in the Mixed and Non-Compliant Subsamples Measured at the Extensive Margin



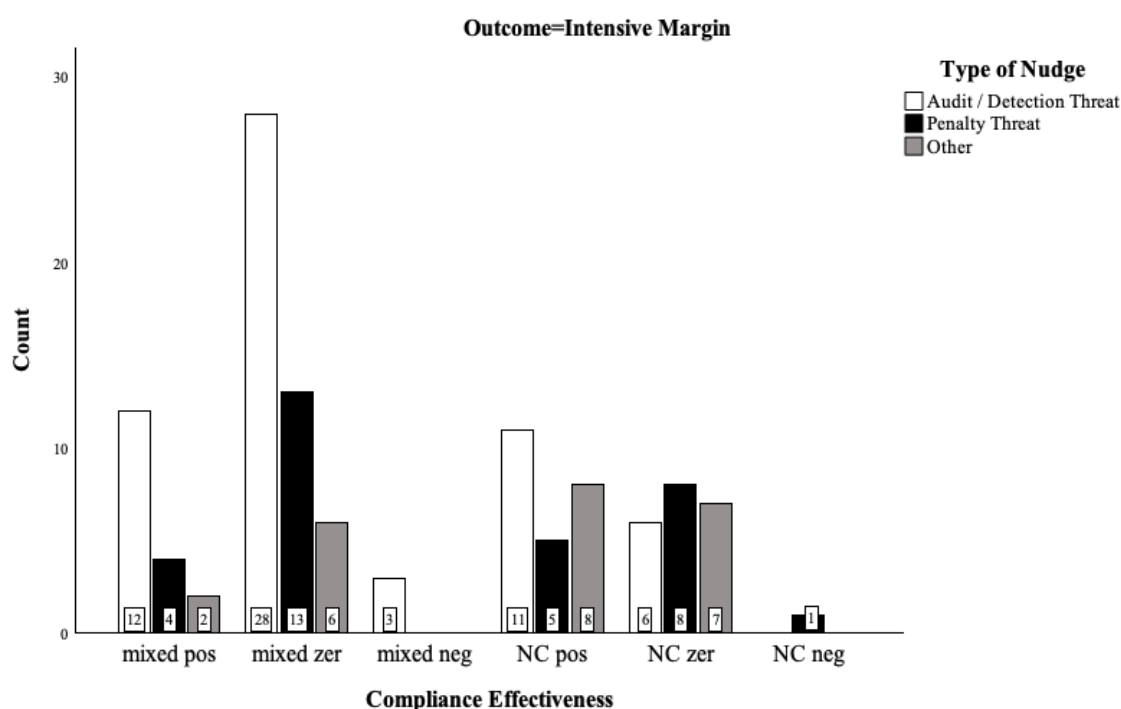
Note. NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

The following analysis will focus on the frequency distribution of effectiveness across different types of interventions measured at the intensive margin, for both the mixed and non-compliant subsamples. The outcome category Intensive Margin, as described previously, measures the extent or intensity of compliance compared to the control group, e.g., the amount of taxes paid after receiving the intervention.

The frequency analysis for both subsamples with their respective effectiveness measured at the intensive margin are depicted in Figure 11. In the intensive margin outcome, the frequency distribution of intervention subcategories is relatively parallel to the frequency distributions of the extensive margin outcome. The mixed subsample demonstrated greater effectiveness in the intervention subcategory Audit/Detection Threat (20.0%) compared to Penalty Threat (12.9%). Also, interventions categorized as Other in the non-compliant sample appeared to have a more substantial impact on increasing tax compliance (34.8%) than in the mixed subsample (8.7%). Compared to the effects observed at the extensive margin outcome, effectiveness at the intensive margin was generally lower, with a higher prevalence of zero effects.

Figure 11

Frequency Distribution of Deterrence Interventions in the Mixed and Non-Compliant Subsamples Measured at the Intensive Margin



Note. NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

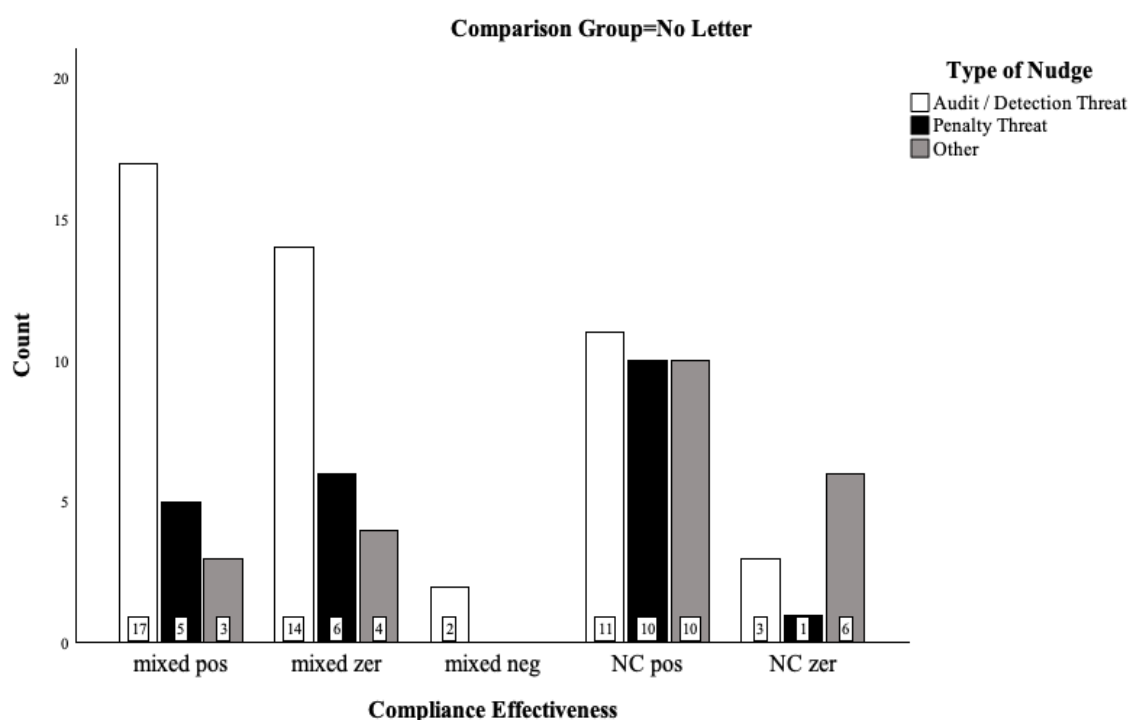
Frequency Distributions compared to No Letter and Simple Letter Control Groups. In the following section, frequency distributions of effectiveness were compared to assess how the control group influenced outcomes within the mixed and non-compliant subsamples.

Specifically, whether they were compared to a control group with no intervention or one that received a message without deterrent nudging elements. Control group subcategories include No Letter, Simple Letter, Simple Reminder and Other. The Simple Letter and Simple Reminder subcategories were combined for this analysis, as both lacked deterrent nudging elements. The subsequent frequency analyses compare effectiveness relative to each control group for both subsamples.

Within the No Letter control group, the mixed subsample showed greater positive effects from Audit/Detection Threats (36.2%) than Penalty Threats (22.7%). In contrast, the non-compliant subsample responded more positively to Penalty Threats (45.5%) than Audit/Detection Threats (23.4%). Negative or backfiring effects were rarely observed in both the mixed and non-compliant subsamples, consistent with the overall frequencies. Figure 12 presents the frequency distributions of the subsamples regarding compliance history compared to No Letter control groups.

Figure 12

Frequency Distribution of Deterrence Interventions in the Mixed and Non-Compliant Subsamples Compared to a No Letter Control Group

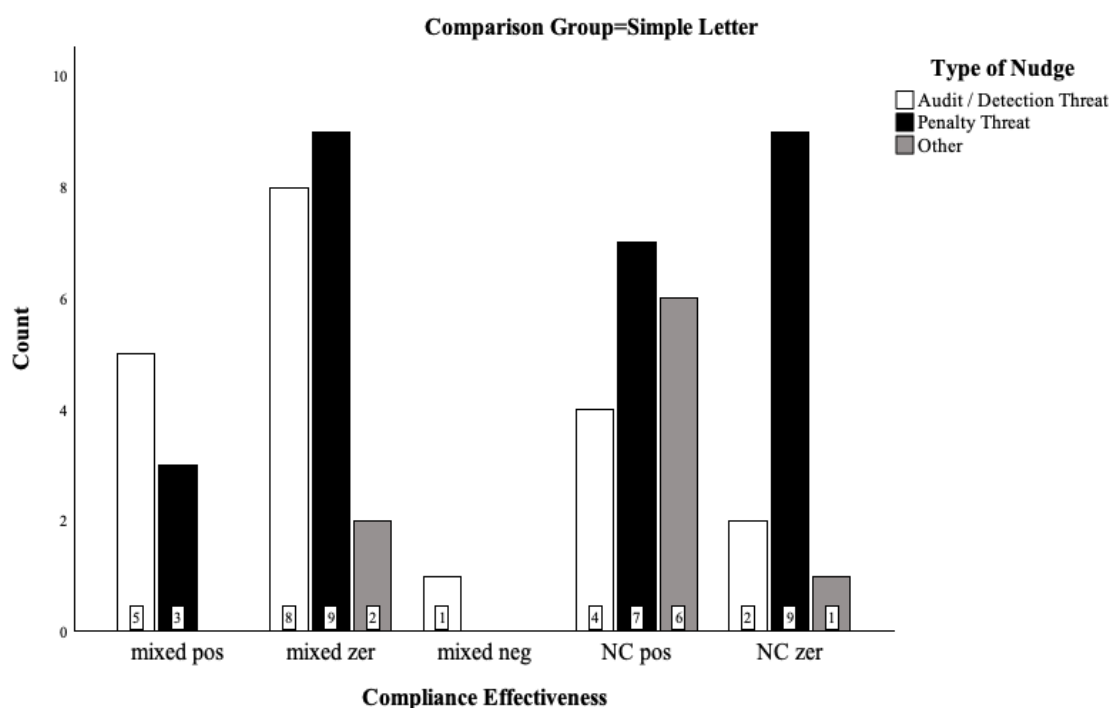


Note. NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

When examining the frequency distributions of effectiveness relative to a Simple Letter control group, tax compliance of the mixed subsample appeared to be more substantially positively impacted by Audit/Detection Threats (25.0%) than by Penalty Threats (10.7%). Within the non-compliant subsample, interventions classified under the subcategory Other appeared to have the greatest effect on compliance (66.7%), followed by the intervention subcategory Penalty Threat (25.0%). Audit/Detection Threats were found to be the least successful intervention subtype at 20.0%. However, when exploring the frequency distributions of the mixed and non-compliant subsamples in terms of their effectiveness against Simple Letter control groups, the positive effects appear to be weakened compared to No Letter control groups, exhibiting a higher frequency of non-significant effects than positive effects. Figure 13 presents the frequency distributions of the subsamples regarding compliance history compared against Simple Letter control groups.

Figure 12

Frequency Distribution of Deterrence Interventions in the Mixed and Non-Compliant Subsamples Compared to a Simple Letter Control Group



Note. NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

3.4.3 Effectiveness and Employment Context

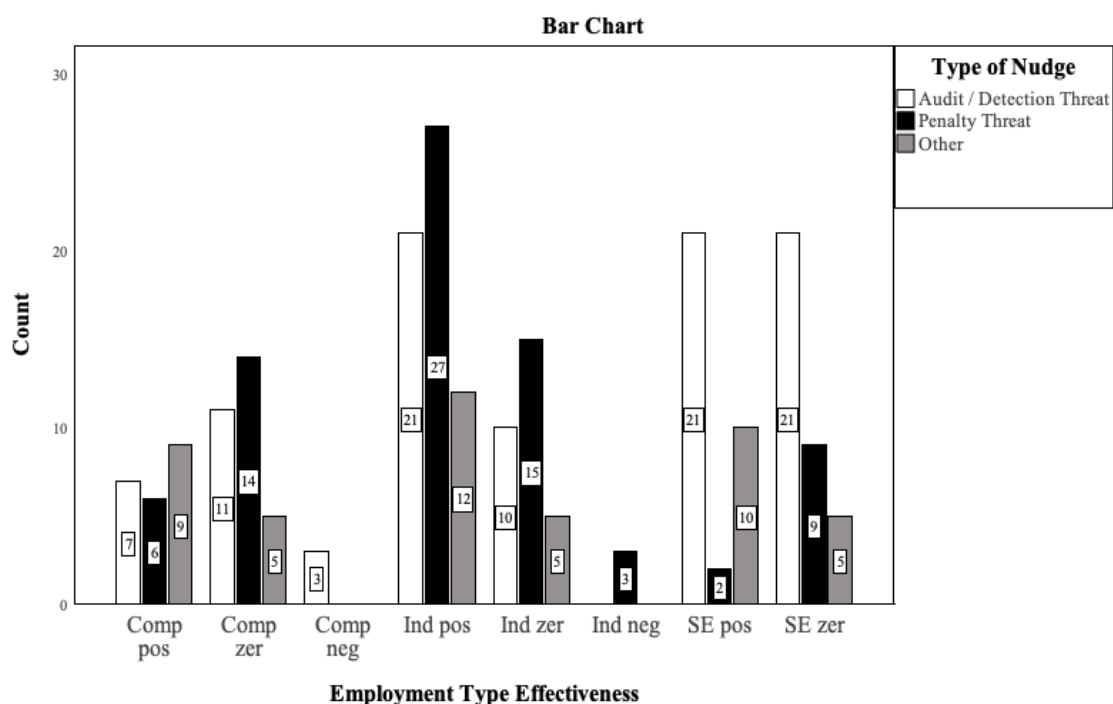
For the subsequent analysis of this systematic review, the influence of employment context on the effectiveness of deterrence interventions was examined. To evaluate these effects, the Population category was divided into the subcategories Company, Individual, and Self-Employed, assessing the effectiveness of each regardless of participants' compliance history.

The frequency distributions of the subsamples regarding employment status will be presented in Figure 14. Within the Self-Employed subgroup, Audit/Detection Threats appeared to be more effective (22.3%) compared to Penalty Threats (2.6%). There was no observation of negative or backfiring effects in the Individual or Self-Employed subsamples, and three observations in the Company subsample. In individuals, Penalty Threats were found to be slightly more effective (35.5%) than Audit/Detection Threats (22.3%), while in the Company category, these two intervention types were of similar effectiveness.

Further analyses of frequency distributions by employment type across distinct outcome categories and control groups will be presented in Appendix B.

Figure 14

Frequency Distribution of Deterrence Interventions in the Subsamples of Company, Individual and Self-Employed



Note. Comp = Company. Ind = Individual. SE = Self-Employed. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

The frequency distributions across all subpopulations, considering both compliance history and employment type, will not be discussed in detail within the main results section. This is due to the division of the overall population into 18 distinct subcategories, which, while allowing for a more nuanced classification, resulted in very low or zero frequencies in several cases. The small number of observations within many of these subgroups limits the potential for meaningful interpretation or reliable comparison across categories. Nevertheless, the full frequency distributions for all subpopulations are provided in Appendix B.

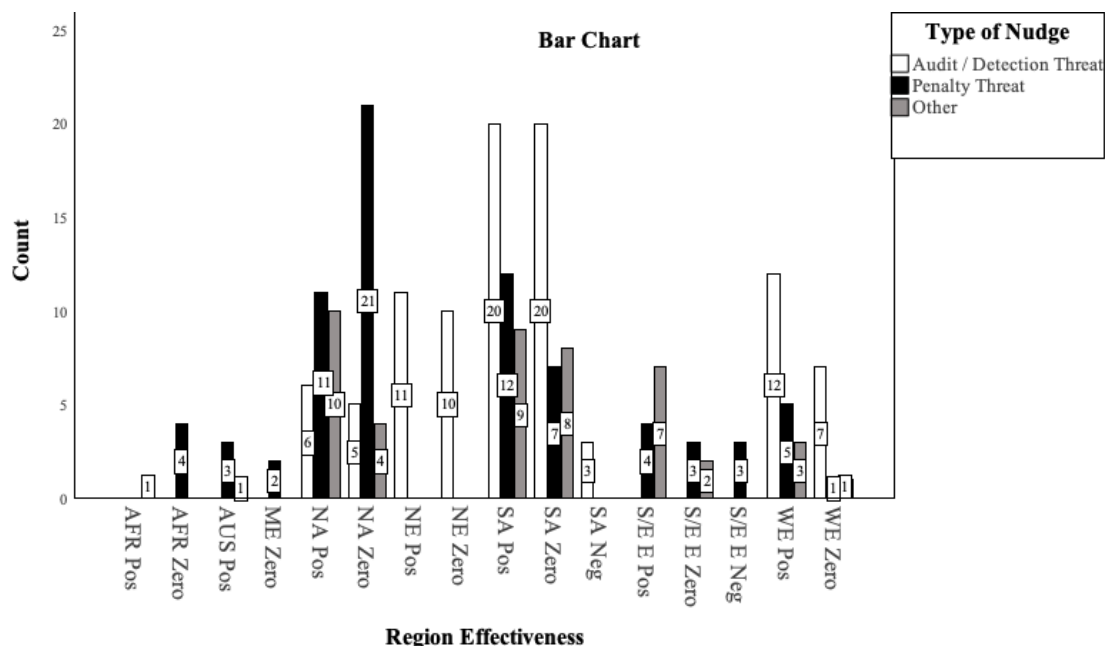
3.4.4 Effectiveness and Regional Context

The effectiveness of different intervention types across geographical regions was assessed in the following section to explore potential regional differences in how deterrence-based nudges influence taxpayer behavior. However, it is important to note that several regions, such as Africa or the Middle East, were represented by a limited number of studies, resulting in low or zero frequencies for certain intervention subtypes. These limitations should be considered when interpreting the results, as they restrict the external validity of the findings for underrepresented areas.

The frequency distributions of regional subsamples, along with corresponding effectiveness outcomes for each intervention subtype are presented in Figure 15. In North America, Penalty Threat interventions were observed to be more effective (14.5%) than Audit/Detection Threat interventions (6.4%), although non-significant interventions marginally outweighed the effective ones. In contrast, in South America, Audit/Detection Threats led to higher compliance rates (21.3%) compared to Penalty Threats (15.8%), with positive effects more frequently observed than non-significant outcomes. In Europe, and particularly in Western Europe, deterrence-based interventions were generally more effective overall, with Audit/Detection Threat interventions (12.8%) leading to higher compliance than Penalty Threat interventions (6.6%). Negative or backfiring effects were rarely observed across all regions, with only three interventions in South America and one in Southern Europe. Further frequency distributions for geographical regions across separate outcome categories and compared against specific control groups are presented in Appendix B.

Figure 15

Frequency Distribution of Deterrence Interventions in the Regional Subcategories



Note. AFR = Africa, AUS = Australia / New Zealand. ME = Middle East. NA = North America. NE = Northern Europe. SA = South America. S/E E = South / Eastern Europe. WE = Western Europe. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

3.4.5 Correspondence Analysis

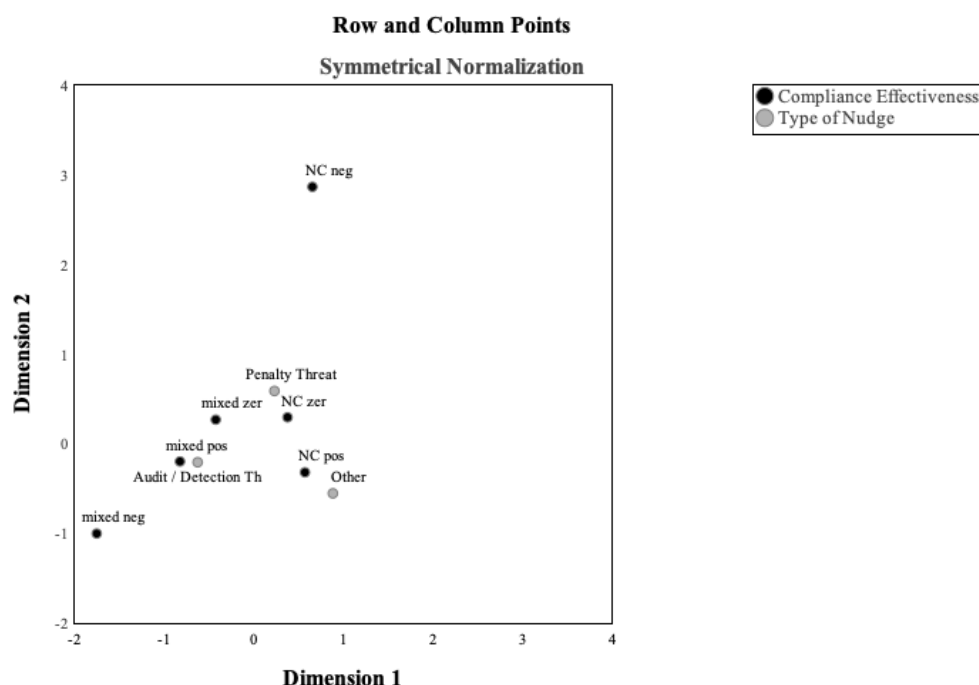
In addition to the previously presented frequency distributions derived from crosstabulations, Correspondence Analysis was conducted to further examine the effectiveness of deterrence interventions within both the mixed and non-compliant subsamples. The interpretation of the crosstabulations provided a general overview of how the frequency of different intervention types and their effectiveness occurred in various subsamples (e.g. mixed vs. non-compliant).

To gain a deeper understanding of these patterns, Correspondence Analysis was applied subsequently, since it allowed for a more nuanced analysis by reducing the dimensionality of the data and visualizing the relationships between different subcategories. In this context, the relationship between intervention types and effectiveness across compliance history subgroups is presented graphically in Figure 16, revealing clusters of intervention types with varying degrees of effectiveness for each subgroup. The row points (i.e., black points) in the biplot represent participants' compliance history and the effectiveness of the intervention, while the column points (i.e., grey points) represent different

intervention subtypes. Proximity between points indicates the strength of their association, with closer points representing a stronger relationship between categories. For example, the proximity of the intervention type Audit/Detection Threat and mixed positive could indicate that this intervention type is an effective strategy for the mixed subsample. The significant distance between the non-compliant subsample, which showed negative effects, and the other categories suggests a weaker relationship, indicating that previously non-compliant individuals do not share strong associations with the other groups in the analysis and may exhibit distinct underlying patterns. The distance between the non-compliant subsample resulting in positive effects is relatively equivalent to all three intervention subtypes, Audit/Detection Threat, Penalty Threat, and Other, suggesting that for previously non-compliant individuals to act compliant it has minimal impact which specific deterrence intervention type is employed. Further correspondence analyses regarding the frequency distributions of effectiveness in different employment types (Company vs. Individual vs. Self-Employed) and geographical regions are included in Appendix C.

Figure 16

Correspondence Analysis for Frequency Distributions of Deterrence Interventions in the Mixed and Non-Compliant Subsamples



Note. Black points represent the rows (i.e., the population subcategories). Grey points represent the columns (i.e., the deterrence interventions). NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Discussion

This systematic review investigated the effectiveness of deterrence nudging interventions on tax compliance across different population subcategories, considering compliance history and employment context. The effectiveness of these interventions was also analyzed across different geographical regions. Furthermore, the effectiveness of distinct subcategories of deterrence interventions was assessed by examining the frequencies of positive, zero, and negative effects reported in the studies included in this review.

To ensure transparency and quality, this systematic review adhered to the PRISMA Statement (Moher et al., 2009), which provides a 27-item reporting checklist and a four-phase flow diagram to guide the literature search and study selection process. This approach supports standardization and minimizes potential bias in systematic reviews and meta-analyses.

The intervention subtypes analyzed within this framework were threats of audits or detection, penalty threats, and a third subcategory containing deterrence interventions that could not be classified under the main subcategories. This was either due to their fundamentally distinct nature and insufficient frequency to justify the formation of separate category, or because they combined elements of both primary subtypes. To ensure a comprehensive overview, interventions of this subcategory were preserved in the frequency analysis despite their limited informational value regarding which specific intervention subtypes produce the most significant outcomes. To assess the effectiveness of each intervention, the results were classified as positive, zero, or negative, based on their impact on the measured outcome. An effect was classified as positive if the intervention significantly increased tax compliance; conversely, a negative effect was defined as a decrease in tax compliance, indicating a backfiring result. If no significant effect was observed, the intervention was categorized as zero effect, with significance levels set at $p < .05$.

Effectiveness was evaluated based on the outcome category, the dependent variable in this context, with the main subcategories either as a binary outcome of the probability of compliance, as well as the extent or intensity of compliance compared to the control group.

As previously noted, all studies included in the analysis were conducted within the framework of natural field experiments, where participants, who are actual taxpayers in their usual environment, are unaware of their participation in an experiment, and data are collected from actual returns. The objective of this systematic review was to examine the influence of

mentioned factors regarding populational subcategories, regional differences, and varying types of deterrence nudging interventions on tax compliance in a comprehensive synthesis.

The following section presents a detailed analysis of the key findings and their implications for tax compliance behavior. The results showed that deterrence interventions, implemented within the context of natural field experiments, are an effective method for influencing individuals' taxpaying behavior, especially among those with a history of non-compliance. The results revealed several important patterns that not only contribute to our understanding of tax compliance but also reinforces the theoretical foundations included in the introduction. The theoretical model of Allingham and Sandmo (1972) suggests that tax compliance is driven by a rational decision-making process where individuals weigh the potential costs of non-compliance against the probability of detection and severity of punishment. The findings of this review support this theory by demonstrating that deterrence nudges, such as audit threats or penalty threats, primarily operate on individual's financial motives, especially when targeting non-compliant individual populations. These interventions increase the perceived likelihood of detection or the severity of penalties, shifting the cost-benefit assessment in favor of compliance. As expected under Allingham and Sandmo's model, such deterrence-based interventions proved more effective in previously non-compliant populations, where individuals had already calculated the benefit of tax evasion outweighing potential costs.

In the subsequent section, frequency distributions of effectiveness in terms of population characteristics will be addressed in depth, beginning with participants' compliance history. The analysis of effectiveness was investigated distinguishing between the non-compliant subsample, i.e., participants with a documented record of failure to comply with tax obligations in the past, and a general or mixed subsample, consisting of taxpayers with no explicit mention of their previous compliance behavior in the underlying studies. Investigating the effectiveness regarding compliance history indicated that interventions targeting the non-compliant subsample were more frequently effective than ineffective, with positive effects occurring more than twice as often as zero effects. This observation also aligns with findings from the meta-analysis by Antinyan and Asatryan (2019), which indicated that groups with a higher proportion of non-compliers are generally more susceptible to nudges, with particularly strong effects observed for deterrence nudges. Bott et al. (2017) also found that sending letters designed to increase the perceived probability of detection among previously non-compliant participants resulted in a statistically significant increase in the proportion of individuals self-reporting foreign income.

In contrast, within the mixed subsample, ineffective interventions appeared to outnumber those with positive effects when analyzed across all outcome categories, suggesting that deterrence interventions are more successful in non-compliant populations. However, when focusing on the extensive margin outcome, the majority of deterrence interventions were also found to be effective in the mixed sample. When comparing intervention effectiveness within the non-compliant subsample across different outcome categories, similar patterns emerged. Positive effects were already the most frequent across combined outcomes; however, this tendency became more pronounced at the extensive margin outcome, where positive effects occurred over five times as often as zero effects. The findings on increased effectiveness at the extensive margin aligned with previous research (see Alm & Malézieux, 2021; Antinyan & Asatryan, 2019).

Within the mixed subsample, deterrence interventions were more frequently found to be ineffective than effective. Within the positive effects in this subgroup, the intervention subcategory Audit/Detection Threat was significantly more effective in promoting tax compliance than the subcategory Penalty Threat. In contrast, within the non-compliant subsample, the difference in the effectiveness between intervention subcategories was less pronounced. Negative or backfiring results from deterrence interventions were rarely observed, both in the mixed and non-compliant subsamples.

It is also noteworthy that, when comparing deterrence interventions with different control groups, positive effects were found to be higher when the intervention was compared to a no-letter control group, rather than to a control group that received simple letters without deterrence nudging elements. This suggests that the mere receipt of a message from tax authorities, even in the absence of deterrence nudges, may have a significant impact on taxpayer behavior. The presence of any official communication could create a perceived threat, thereby introducing considerable bias and influencing compliance behavior. These findings are consistent with those observed by Alm et al. (2019).

To examine the influence of employment context on the effectiveness of deterrence interventions, frequency distributions across companies, individuals, and the self-employed were compared within this systematic review. Overall, interventions were most effective for individuals, with 64.5% showing positive outcomes. In contrast, both Company and Self-Employed subsamples exhibited slightly more ineffective interventions than positive ones. This could be attributed to the increased opportunities for tax evasion arising from their responsibility for self-reporting, as highlighted by Alm et al. (2022).

Among the intervention subtypes, Audit/Detection Threats had the greatest impact on tax compliance within the Self-Employed subgroup, surpassing other types of interventions. This suggests that, within this specific group of taxpayers, Audit/Detection Threats are notably more effective than Penalty Threats. Conversely, within the Individual subgroup, Penalty Threats were found to be slightly more effective than Audit/Detection Threats, which could imply that, for individuals, the immediate financial consequences of non-compliance may outweigh the perceived risk of being detected, aligning with the idea that economic incentives play a significant role in compliance decisions. In the Company subsample, the specific type of intervention did not significantly affect success in increasing tax compliance. Notably, the subgroup Company was the only one where interventions led to backfiring effects.

Overall, the analysis of the effectiveness of different types of deterrence nudges indicated that Audit/Detection Threats were more frequently associated with positive effects, though the difference compared to other intervention types was not substantial. As previously noted, the mixed subsample showed a stronger positive response to Audit/Detection Threats compared to Penalty Threats. In contrast, within the non-compliant subsample, Penalty Threats yielded more positive effects than Audit/Detection Threats. The remaining intervention subcategory, Other, had a notable impact on the non-compliant sample. This might suggest that deterrence nudges incorporating additional non-deterrence elements (e.g., supplementary informational content, simplification, or social norm messages) or even combinations of Audit/Detection Threats and Penalty Threats, may serve as particularly effective strategies for enhancing tax compliance, particularly in previously non-compliant populations.

Regionally, deterrence interventions showed varying levels of effectiveness, highlighting the importance of contextual factors. In Europe, specifically Western Europe, the considerably high success rate of interventions (69%) underscores a more favorable environment for deterrence-based approaches. In South America, the frequencies of effectiveness were less distinct, with just over half resulting in positive effects, while in North America, a slightly higher proportion of ineffective interventions (52.9%) could be observed. Negative or backfiring effects were rare overall, observed only in a few interventions in South America and South/Eastern Europe, indicating that while some risk exists, it is relatively limited. Analyzing the effectiveness of specific intervention subtypes in different geographical regions showed that Audit/Detection Threats were more effective than Penalty Threats in South America and Europe, while in North America, Penalty Threats were found to be

particularly effective in raising tax compliance. These regional variations in the effectiveness of deterrence interventions underscore the critical role of contextual factors in influencing outcomes, particularly institutional trust. The higher rate of successful deterrence interventions in Western Europe could be attributed to a higher level of trust in the tax system and perceived legitimacy of tax authorities, as discussed by Kirchler et al. (2008). In contrast, regions like South and North America displaying mixed results could suggest that trust in authorities and varying tax cultures may have influenced the effectiveness of interventions. These results align with the Slippery Slope Framework's premise by Kirchler et al. (2008) that trust plays a key role in taxpayers' compliance decisions, and interventions are more successful in environments where tax authorities are perceived as legitimate and trusted.

While this systematic review provides valuable insights, several limitations must be considered when interpreting the findings. Methodologically, the assessment of the effectiveness of deterrence interventions in the contexts discussed was based on absolute frequencies. Since the overall sample size within the framework of field studies was relatively limited, the underrepresentation or absence of data in certain distributions, particularly in regions such as Africa and the Middle East, restricted further statistical analysis. This gap should also be acknowledged when interpreting regional patterns and, more broadly, the overall conclusions of this review.

Concerning the distinction of the different subtypes of interventions, the decision was made to include the intervention subcategory "Other" in the analysis. As already noted, interventions of this type did not fit into existing subcategories, either due to their unique nature or due to a combination of Audit/Detection and Penalty Threats. In a few cases, an intervention was classified as Other, if there was a prominent deterrence aspect including an additional non-deterrence element, such as informational content, simplification, or social norm messages. Although the inclusion of this subcategory did not offer further insight into the effectiveness of specific types of interventions, it was retained to provide a more comprehensive perspective in the frequency analysis. This decision may be considered a limitation when interpreting the findings.

One notable limitation arising from the underlying studies included in this review concerns the limited information due to unreported data. Similar difficulties were noted in meta-analyses in this context (see, e.g., Antinyan & Asatryan, 2019). Specifically, the differentiation of participants based on their compliance history was somewhat unclear. In contrast to the non-compliant subsample, which consisted of individuals with a well-documented record of non-compliant behavior, the general or mixed subsample lacked such a

clear classification. Due to the absence of relevant data on past compliance in some of the included studies, it cannot be assumed that this group consisted solely of compliant individuals. Therefore, the general subsample should be interpreted as a heterogeneous group, comprising both compliant and non-compliant individuals, rather than a population of consistently law-abiding taxpayers.

Additionally, deterrence interventions of the studies included in this review were implemented across various contexts, with countries applying distinct tax systems and individuals demonstrating varying levels of tax compliance culture. Therefore, the effectiveness of interventions is context-dependent, with their success varying according to specific tax systems, taxpayer attitudes, and employment types.

The findings of this systematic review suggest several opportunities for future research, particularly in further exploring the contextual factors that influence the effectiveness of deterrence nudging interventions on tax compliance across diverse populations and regions. Most importantly, natural field experiments should be conducted more frequently with larger population samples to enable more generalizable conclusions about the effectiveness of deterrence interventions. Although recent years have seen significant progress in moving “from the lab to the field” (Mascagni, 2018), which is now considered the gold standard in tax compliance research, there remains considerable potential for further studies, especially in regions such as Africa, Asia, and the Middle East. Compared to experiments in laboratory settings, they offer several key advantages. Fully embedded in real taxpaying contexts, they provide a stronger foundation for policy recommendations, benefit from greater statistical power due to the possibility of larger sample sizes, and offer deeper insights into the diversity of taxpayer responses to various compliance factors.

Additionally, this systematic review reinforces the importance of using a no-message control group in the design of field interventions, rather than a simple message without deterrence elements. The analysis revealed that positive effects were more pronounced when deterrence-based interventions were evaluated against a no-letter control group, suggesting that the mere receipt of a message from tax authorities can influence taxpayer behavior. This finding suggests that the mere receipt of official communication, regardless of its specific content, may create a perceived threat, potentially biasing taxpayer responses and influencing compliance. This insight aligns with the observations made by Alm et al. (2019) and is crucial for evaluating the effectiveness of various deterrence interventions.

Furthermore, in future research, it may be valuable to explore the potential benefits of combining elements from both Audit/Detection Threat and Penalty Threat intervention types.

As highlighted in this review, some interventions were classified as “Other” due to a combination of deterrence intervention subtypes or, in some cases, the inclusion of additional non-deterrence elements. Frequent positive effects emerging from this type of nudge suggest that hybrid interventions could potentially yield better results, particularly within the non-compliant population. Investigating such combinations could offer a more comprehensive understanding of how different intervention types interact to influence taxpayer behavior.

Moreover, future research should prioritize the collection of more detailed and systematic data on participants characteristics, particularly regarding compliance history and employment context. In several of the studies included in this review, such information was either insufficiently reported or entirely absent. Addressing this gap in future research would not only enable a more accurate differentiation of target populations, but also significantly enhance the robustness of statistical analyses. Furthermore, access to more comprehensive participant data would allow for a greater precision in identifying patterns, such as differing responsiveness to deterrence messages based on past compliance behavior. More detailed data would also provide a stronger empirical foundation for evaluating the effectiveness of deterrence-based interventions and would support the development of more tailored and targeted policy recommendations.

The current systematic review aimed to further explore the individual and regional contexts in which deterrence nudging interventions can improve tax compliance. It demonstrated that these interventions, when applied within the framework of natural field experiments, are a potent tool for nudging individuals’ tax behavior towards adherence, particularly among previously non-compliant populations. The effects were significantly amplified when measured at the extensive margin, which refers to the binary outcome of whether individuals choose to fulfill their tax obligations after the intervention. The findings of this review align with those of the meta-analysis conducted by Antinyan & Asatryan (2019), which conclude that deterrence interventions are a powerful driver in enhancing tax compliance for previously non-compliant individuals. This suggests that individual financial motives play a significant role in tax behavior, potentially outweighing factors related to tax morale, such as social norms or reciprocity.

Researchers in the field of tax compliance have raised concerns that deterrence interventions may backfire, especially if taxpayers perceive them as overregulating or unfairly targeting them, potentially leading to unintended negative effects; outcomes that counter the broader societal goal of a well-functioning tax system (see, for example, Braithwaite, 2007).

However, the findings of the present review indicate that such negative outcomes do not pose a meaningful threat.

Overall, the results of this systematic review demonstrated that deterrence interventions were more likely to promote compliance than to yield no measurable effect. When measured at the extensive margin, deterrence interventions were found to be effective for both the general and non-compliant samples. The positive effects were particularly pronounced among taxpayers with a history of non-compliance, as deterrence measures appeared to help establish clearer behavioral expectations and contribute to reducing tax evasion.

Acknowledgments

I want to thank Dr. Erich Kirchler for his guidance and support throughout the process of this thesis. His expertise, thoughtful feedback, and patience have been greatly appreciated.

References

- Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1, 323–338. [https://doi.org/10.1016/0047-2727\(72\)90010-2](https://doi.org/10.1016/0047-2727(72)90010-2)
- Alm, J. (2012). Measuring, explaining, and controlling tax evasion: Lessons from theory, experiments, and field studies. *International Tax and Public Finance*, 19(1), 54–77. <https://doi.org/10.1007/s10797-011-9171-2>
- Alm, J., Cifuentes, L. R., Niño, C. M. O., & Rocha, D. (2019). Can behavioral “nudges” improve compliance? The case of Colombia social protection contributions. *Games*, 10(4), 43. <https://doi.org/10.3390/g10040043>
- Alm, J., Gerbrands, P., & Kirchler, E. (2022). Using “responsive regulation” to reduce tax base erosion. *Regulation & Governance*, 16(3), 738–759. <https://doi.org/10.1111/rego.12359>
- Alm, J., & Malézieux, A. (2021). 40 years of tax evasion games: A meta-analysis. *Experimental Economics*, 24(3), 699–750. <https://doi.org/10.1007/s10683-020-09679-3>
- Alm, J., & Torgler, B. (2011). Do ethics matter? Tax compliance and morality. *Journal of Business Ethics*, 101(4), 635–651. <https://doi.org/10.1007/s10551-011-0761-9>
- Antinyan, A., & Asatryan, Z. (2019). Nudging for tax compliance: A meta-analysis. *ZEW - Centre for European Economic Research*. Discussion Paper 19-055. <http://dx.doi.org/10.2139/ssrn.3500744>
- Appelgren, L. (2008). The effect of audit strategy information on tax compliance—an empirical study. *eJTR*, 6, 67. Retrieved from <https://classic.austlii.edu.au/au/journals/eJITaxR/2008/4.html>
- Ariel, B. (2012). Deterrence and moral persuasion effects on corporate tax compliance: findings from a randomized controlled trial. *Criminology*, 50(1), 27–69. <https://doi.org/10.1111/j.1745-9125.2011.00256.x>
- Bérgolo, M., Ceni, R., Cruces, G., Giacobasso, M., & Perez-Truglia, R. (2019). Tax audits as scarecrows: Evidence from a large-scale field experiment. *National Bureau of Economic Research*. <https://doi.org/10.3386/w23631>
- Boning, W. C., Guyton, J., Hodge, R., & Slemrod, J. (2018). Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms. *National Bureau of Economic Research*. <https://doi.org/10.3386/w24305>

- Bott, K. M., Cappelen, A. W., Sorensen, E., & Tungodden, B. (2017). You've got mail: A randomised field experiment on tax evasion. *Management science*, 66(7), 2801-2819. <https://doi.org/10.2139/ssrn.3033775>
- Braithwaite, V. (2003). Dancing with tax authorities: Motivational postures and non-compliant actions. *Taxing democracy*, 3, 15-39. <https://doi.org/10.4324/9781315241746>
- Braithwaite, V. (2007). Responsive regulation and taxation: Introduction. *Law & Policy*, 29(1), 3-10. <https://doi.org/10.1111/j.1467-9930.2007.00242.x>
- Brockmeyer, A., Smith, S., Hernandez, M., & Kettle, S. (2019). Casting a wider tax net: Experimental evidence from Costa Rica. *American Economic Journal: Economic Policy*, 11(3), 55-87. <https://doi.org/10.1257/pol.20160589>
- Carrillo, P., Pomeranz, D., & Singhal, M. (2017). Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics*, 9(2), 144-164. <https://doi.org/10.1257/app.20140495>
- Casey, J. T., & Scholz, J. T. (1991). Beyond deterrence: Behavioral decision theory and tax compliance. *Law & Society Review*, 25(4), 821-843. <https://doi.org/10.2307/3053871>
- Castro, L., & Scartascini, C. (2015). Tax compliance and enforcement in the pampas: evidence from a field experiment. *Journal of Economic Behavior & Organization*, 116, 65-82. <https://doi.org/10.1016/j.jebo.2015.04.002>
- Chirico, M., Inman, R. P., Loeffler, C., MacDonald, J., & Sieg, H. (2016). An experimental evaluation of notification strategies to increase property tax compliance: free-riding in the city of brotherly love. *Tax Policy and the Economy*, 30(1), 129-161. <https://doi.org/10.1086/685595>
- Chirico, M., Inman, R., Loeffler, C., MacDonald, J., & Sieg, H. (2019). Deterring property tax delinquency in Philadelphia: An experimental evaluation of nudge strategies. *National Tax Journal*, 72(3), 479-506. <https://doi.org/10.17310/ntj.2019.3.01>
- Coleman, S. (1996). The Minnesota income tax compliance experiment: State tax results. *MPRA Paper*, 4, 528-541. Retrieved from <https://mpra.ub.uni-muenchen.de/4827/>
- Cranor, T., Goldin, J., Homonoff, T., & Moore, L. (2020). Communicating tax penalties to delinquent taxpayers. *National Tax Journal*, 73(2), 331-360. <https://doi.org/10.17310/ntj.2020.2.02>
- De Neve, J.-E., Imbert, C., Spinnewijn, J., Tsankova, T., & Luts, M. (2019). How to improve tax compliance? Evidence from population-wide experiments in Belgium. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3389405>

- Del Carpio, L. (2013). Are the neighbors cheating? Evidence from a social norm experiment on property taxes in Peru. *Princeton University*. Retrieved from <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=82a72f3a4d270a0d0a4c431a05705fb15b8e4cdf>
- Dunning, T., Monestier, F., Pineiro, R., Rosenblatt, F., & Tunón, G. (2015). Positive vs. negative incentives for compliance: Evaluating a randomized tax holiday in Uruguay. *SSRN Electronic Journal*, (August). <https://doi.org/10.2139/ssrn.2650105>
- Dwenger, N., Kleven, H., Rasul, I., & Rincke, J. (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy*, 8(3), 203–232. <https://doi.org/10.1257/pol.20150083>
- Eerola, E., Kosonen, T., Kotakorpi, K., Lyytikäinen, T., & Tuimala, J. (2019). Tax compliance in the rental housing market: Evidence from a field experiment. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3464295>
- Elffers, H., Weigel, R. H., & Hessing, D. J. (1987). The consequences of different strategies for measuring tax evasion behavior. *Journal of Economic Psychology*, 8(3), 311–337. [https://doi.org/10.1016/0167-4870\(87\)90026-2](https://doi.org/10.1016/0167-4870(87)90026-2)
- Fellner, G., Sausgruber, R., & Traxler, C. (2013). Testing enforcement strategies in the field: Threat, moral appeal, and social information. *Journal of the European Economic Association*, 11(3), 634–660. <https://doi.org/10.1111/jeea.12013>
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33(3), 613–619. <https://doi.org/10.1177/001316447303300309>
- Fleiss, J., Paik, M., & Levin, B. (2003). *Statistical methods for rates and proportions*. Wiley Series in Probability and Statistics. <https://doi.org/10.1002/0471445428>
- Graetz, M. J., & Wilde, L. L. (1985). The economics of tax compliance: Fact and fantasy. *National Tax Journal*, 38(3), 355–363. <https://doi.org/10.1086/ntj41792032>
- Hallsworth, M. (2014). The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy*, 30(4), 658–679. <https://doi.org/10.1093/oxrep/gru034>
- Hallsworth, M. (2018). New ways of understanding tax compliance: From the laboratory to the real world. In A. Lewis (Ed.), *The Cambridge Handbook of Psychology and Economic Behaviour* (2nd ed., pp. 430–451). Cambridge University Press. <https://doi.org/10.1017/9781316676349.015>

- Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148, 14–31. <https://doi.org/10.1016/j.jpubeco.2017.02.003>
- Harju, J., Kosonen, T., & Ropponen, O. (2014). Do honest hairdressers get a haircut? On tax rate and tax evasion. *ZEW 2014 Public Finance Conference*. Retrieved from <https://ntanet.org/wp-content/uploads/proceedings/2014/143-kosonen-harju-ropponen-honest-hairdressers-get-haircut-tax.pdf>
- Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4), 1009–1055. <https://doi.org/10.1257/0022051043004577>
- Hashimzade, N., Myles, G. D., & Tran-Nam, B. (2013). Applications of behavioural economics to tax evasion. *Journal of Economic Surveys*, 27(5), 941–977. <https://doi.org/10.1111/j.1467-6419.2012.00733.x>
- Hasseldine, J., Hite, P., James, S., & Toumi, M. (2007). Persuasive communications: Tax compliance enforcement strategies for sole proprietors. *Contemporary Accounting Research*, 24(1), 171–194. <https://doi.org/10.1506/P207-004L-4205-7NX0>
- Hernandez, M., Jamison, J., Korczyk, E., Mazar, N., & Sormani, R. (2017). *Applying behavioral insights to improve tax collection: Experimental evidence from Poland*. World Bank Policy Report. <https://doi.org/10.1596/27528>
- Hernandez, M., Karver, J., Negre, M., & Perng, J. (2019). *Promoting Tax Compliance in Kosovo with Behavioral Insights*. World Bank Policy Report. <https://doi.org/10.1596/31357>
- Hiscox, M. J., Bialecki, J., Cotching, H., Daffey, M., Greenwell, H., Su, M. K.-M., ... Weeks, B. (2018). Improved compliance with the deferred GST scheme. *Commonwealth of Australia, Department of the Prime Minister and Cabinet*. Retrieved from <https://creativecommons.org/licenses/by/4.0/deed.enhttp://www.itsanhonour.gov.au/co-at-arms>
- Iyer, G. S., Reckers, P. M. J., & Sanders, D. L. (2010a). A field experiment to explore the effects of detection and penalties communications and framing among Washington State retail firms. *Advances in Accounting*, 26(2), 236–245. <https://doi.org/10.1016/j.adiac.2010.05.004>
- Iyer, G. S., Reckers, P. M., & Sanders, D. L. (2010b). Increasing tax compliance in Washington state: A field experiment. *National Tax Journal*, 63(1), 7-32. <https://doi.org/10.17310/ntj.2010.1.01>

- Kettle, S., Hernandez, M., Ruda, S., & Sanders, M. (2016). *Behavioral interventions in tax compliance - Evidence from Guatemala*. World Bank, Washington, DC.
<https://doi.org/10.1596/1813-9450-7690>
- Kettle, S., Hernandez, M., Sanders, M., Hauser, O., & Ruda, S. (2017). Failure to CAPTCHA attention: Null results from an honesty priming experiment in Guatemala. *Behavioral Sciences*, 7(2). <https://doi.org/10.3390/bs7020028>
- Kirchler, E., Hoelzl, E., & Wahl, I. (2008). Enforced versus voluntary tax compliance: The “slippery slope” framework. *Journal of Economic Psychology*, 29(2), 210–225.
<https://doi.org/10.1016/j.joep.2007.05.004>
- Kirchler, E., & Wahl, I. (2010). Tax compliance inventory: TAX-I voluntary tax compliance, enforced tax compliance, tax avoidance, and tax evasion. *Journal of Economic Psychology*, 31(3), 331–346. <https://doi.org/10.1016/j.joep.2010.01.002>
- Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica*, 79(3), 651–692. <https://doi.org/10.3982/ECTA9113>
- Koumpias, A. M. (2017). The effects of compliance reminders on personal income tax payments in Greece; Evidence from a regression discontinuity design. *National Tax Journal*, 110. Retrieved from <https://aysps.gsu.edu/files/2016/12/Koumpias-JMP.pdf>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Larkin, C., Sanders, M., Andresen, I., & Algate, F. (2019). Testing local descriptive norms and salience of enforcement action: A field experiment to increase tax collection. *Journal of Behavioral Public Administration*, 2(1), 1–11.
<https://doi.org/10.30636/jbpa.21.54>
- Mascagni, G. (2018). From the lab to the field: A review of tax experiments. *Journal of Economic Surveys*, 32(2), 273–301. <https://doi.org/10.1111/joes.12201>
- Mascagni, G., Nell, C., & Monkam, N. (2018). One size does not fit all: A field experiment on the drivers of tax compliance and delivery methods in Rwanda. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3120363>
- Meiselman, B. S. (2018). Ghostbusting in Detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics*, 158(March 2017), 180–193.
<https://doi.org/10.1016/j.jpubeco.2018.01.005>

- McCaffery, E. J., Baron, J., & Baron, J. (2003). Heuristics and biases in thinking about tax. In *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association* (Vol. 96, pp. 434-443).
<https://doi.org/10.2139/ssrn.467440>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine* 6(7). <https://doi.org/doi:10.1371/journal.pmed.1000097>
- Ortega, D., & Sanguinetti, P. (2013). Deterrence and reciprocity effects on tax compliance: Experimental evidence from Venezuela. In *AEA Randomized Controlled Trials*. American Economic Association. <https://doi.org/10.1257/rct.491-3.0>
- Perez-Truglia, R., & Troiano, U. (2018). Shaming tax delinquents. *Journal of Public Economics*, 167, 120–137. <https://doi.org/10.1016/j.jpubeco.2018.09.008>
- Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 105(8), 2539–2569.
<https://doi.org/10.1257/aer.20130393>
- Robitaille, N., House, J., & Mazar, N. (2018). Effectiveness of planning prompts on organizations' likelihood to file their overdue taxes: A multi-wave field experiment. *Management Science*, 67(7), 4327-4340.
<https://doi.org/10.1287/mnsc.2020.3744>
- Sanders, D. L., Reckers, P. M. J., & Iyer, G. S. (2008). Influence of accountability and penalty awareness on tax compliance. *The Journal of the American Taxation Association*, 30(2), 1–20. <https://doi.org/10.2308/jata.2008.30.2.1>
- Sandmo, A. (2005). The theory of tax evasion: A retrospective view. *National Tax Journal*, 58(4), 643–663. <https://doi.org/10.17310/ntj.2005.4.02>
- Schwartz, R. D., & Orleans, S. (1967). On legal sanctions. *The University of Chicago Law Review*, 34(2), 274. <https://doi.org/10.2307/1598934>
- Shimeles, A., Gurara, D. Z., & Woldeyes, F. (2017). Taxman's dilemma: coercion or persuasion? Evidence from a randomized field experiment in Ethiopia. *American Economic Review*, 107(5), 420-424. <https://doi.org/10.1257/aer.p20171141>
- Slemrod, J., Blumenthal, M., & Christian, C. (2001). Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 79(3), 455–483. [https://doi.org/10.1016/S0047-2727\(99\)00107-3](https://doi.org/10.1016/S0047-2727(99)00107-3)
- Thaler, R. H. (1988). Anomalies: The winner's curse. *Journal of Economic Perspectives*, 2(1), 191–202. <https://doi.org/10.1257/jep.2.1.191>

- Thaler, R. H. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale Univ. Press.
- Truzka, A. (2021). *Soft nudging for tax compliance: Systematic review on the effectiveness of non-deterrence nudging interventions on tax compliance* [Master's thesis, University of Vienna]. <https://doi.org/10.25365/thesis.70482>
- Webley, P., Robben, H., Elffers, H., & Hessing, D. (1991). *Tax evasion: An experimental approach*. Cambridge University Press.
- Yang, W. (2023). The endowment effect: A behavioural economic perspective on land use policy, environmental protection and taxation. *Highlights in Business, Economics and Management*, 21, 62–68. <https://doi.org/10.54097/hbem.v21i.13605>

Appendix A

Table A1

Description of Subcategories

Category	Nr. of Studies	Nr. of Observations	Description
Outcome			
Extensive Margin	24	71	Probability to pay, file, or report taxes
Intensive Margin	30	114	Amount of reported taxes or reported income
Tax Deductions/Claims	3	4	Deductions from gross revenues
Timely Payment	8	19	Probability to pay, file, or report taxes within a timeframe of two months
Other	3	8	Other types of compliance indicators (e.g., tax compliance indicator obtained from administrative data; see DelCarpio, 2014)
Type of Intervention			
Audit/Detection Threat	20	94	Interventions containing threats regarding an upcoming audit
Penalty Threat	20	76	Interventions containing threats of punishment, e.g., the obligatory payment of a fine
Other	16	46	Interventions containing elements of both prior intervention types and interventions containing additional non-deterrence elements
Population			
Company	7	29	Small to large companies with several employees
Company / non-compliant	5	26	Small to large companies with several employees and a record of previously missed tax payments
Individual	9	29	Individual taxpayers with no business affiliation
Individual / non-compliant	15	64	Individual taxpayers with no business affiliation and a record of previously missed tax payments
Self-Employed	9	41	Sole proprietors or small-scale businesses with few employees

Self-Employed / non-compliant	5	27	Sole proprietors or small-scale businesses with few employees and a record of previously missed tax payments
Region			
Africa	2	5	Studies conducted in African countries
Australia / New Zealand	1	4	Studies conducted in Australia or New Zealand
Middle East	1	2	Studies conducted in Middle Eastern countries
North America	13	57	Studies conducted in the US or Canada
Northern Europe	5	21	Studies conducted in Scandinavian or Baltic countries
South America	11	79	Studies conducted in Middle or South American countries
South/Eastern Europe	3	19	Studies conducted in Eastern Europe and the Balkans
Western Europe	6	29	Studies conducted in Middle or Western European countries
Tax			
B&O Tax (Business and Occupation Tax)	3	7	Gross revenue tax in the state of Washington (e.g., see Sanders et al., 2008)
Church Levy	1	12	Tax obligation for all members of the Catholic and Protestant churches (e.g., in Germany; see Dwenger et al., 2016)
Council Tax	1	2	Tax administered at the subnational level (in the UK; see Larkin et al., 2019)
Foreign Income Tax	1	3	Self-reported foreign income (e.g., in Norway; see Bott et al., 2017)
Goods and Services Tax	1	4	Tax deferral on imported goods (e.g., in Australia; see Hiscox et al., 2018)
Income Tax	21	110	Tax of income or profits earned (personal and corporate)
Property Tax	5	24	Tax on the value of a property (e.g., see Castro & Scartascini, 2015)
Sales Tax	1	1	Business income tax applied to gross sales of firms belonging to industry, retail, or service sectors (e.g., in Venezuela; see Ortega & Sanguinetti, 2013)
TV License Fee	1	2	Broadcasting fee for the reception of public television and radio services (e.g., in Austria; see Fellner et al., 2013)

Use Tax	3	7	Tax on goods purchased out-of-state (e.g., in the US, see Iyer et al., 2010)
VAT	5	14	Consumption tax levied on the value at each stage of production or distribution of goods and services (e.g., see Ariel, 2012)
Other	3	30	Other forms of taxes (e.g., healthcare or pension payments; see Alm, 2019)
Effectiveness			
Positive Effect	34	115	Significant positive effect on outcome following the intervention ($p < .05$)
Zero Effect	28	95	No significant effect on outcome following the intervention ($p < .05$)
Negative Effect	3	6	Significant negative or backfiring effect on outcome following the intervention ($p < .05$)
Comparison			
No Letter	25	111	The control group received no intervention
Simple Letter	14	54	The control group received an intervention without nudging elements
Simple Reminder	4	14	The control group received a reminder without nudging elements
Other	10	37	The control group received an intervention with nudging elements

Note. Column 2 displays the frequency of studies including each subcategory, column 3 displays the frequency of observations of each subcategory.

Appendix B

Table B1

Frequency Distribution of Deterrence Interventions in the Mixed and Non-Compliant Samples

Nudge	mixed pos	mixed zer	mixed neg	NC pos	NC zer	NC neg
Audit / Detection Threat	25	31	3	24	11	0
Penalty Threat	9	22	0	26	16	3
Other	3	6	0	28	9	0

Note. NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B2

Frequency Distribution of Deterrence Interventions in the Mixed and Non-Compliant Samples Measured at the Extensive Margin

Nudge	mixed pos	mixed zer	mixed neg	NC pos	NC zer	NC neg
Audit / Detection Threat	13	3	0	11	2	0
Penalty Threat	4	4	0	12	5	2
Other	1	0	0	14	0	0

Note. NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B3

Frequency Distribution of Deterrence Interventions in the Mixed and Non-Compliant Samples Measured at the Intensive Margin

Nudge	mixed pos	mixed zer	mixed neg	NC pos	NC zer	NC neg
Audit / Detection Threat	12	28	3	11	6	0
Penalty Threat	4	13	0	5	8	1
Other	2	6	0	8	7	0

Note. NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B4

Frequency Distribution of Deterrence Interventions in the Mixed and Non-Compliant Samples Compared to a No Letter Control Group

Nudge	mixed pos	mixed zer	mixed neg	NC pos	NC zer	NC neg
Audit / Detection Threat	17	14	2	13	5	0
Penalty Threat	5	9	0	15	2	0
Other	3	4	0	15	7	0

Note. NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B5

Frequency Distribution of Deterrence Interventions in the Mixed and Non-Compliant Samples Compared to a Simple Letter Control Group

Nudge	mixed pos	mixed zer	mixed neg	NC pos	NC zer	NC neg
Audit / Detection Threat	5	8	1	4	3	0
Penalty Threat	4	10	0	3	10	0
Other	0	2	0	3	1	0

Note. NC = non-compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B6

Frequency Distribution of Deterrence Interventions in the Samples of Company, Individual, and Self-Employed

Nudge	C pos	C zer	C neg	I pos	I zer	I neg	SE pos	SE zer	SE neg
Audit / Detection Threat	7	11	3	21	10	0	21	21	0
Penalty Threat	6	14	0	27	15	3	2	9	0
Other	9	5	0	12	5	0	10	5	0

Note. C = Company. I = Individual. SE = Self-Employed. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B7

Frequency Distribution of Deterrence Interventions in the Samples of Company, Individual and Self-Employed measured at the Extensive Margin

Nudge	C pos	C zer	C neg	I pos	I zer	I neg	SE pos	SE zer	SE neg
Audit / Detection Threat	4	0	0	14	3	0	6	2	0
Penalty Threat	3	3	0	13	5	2	0	1	0
Other	2	0	0	9	0	0	4	0	0

Note. C = Company. I = Individual. SE = Self-Employed. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B8

Frequency Distribution of Deterrence Interventions in the Samples of Company, Individual, and Self-Employed measured at the Intensive Margin

Nudge	C pos	C zer	C neg	I pos	I zer	I neg	SE pos	SE zer	SE neg
Audit / Detection Threat	2	10	3	7	6	0	14	18	0
Penalty Threat	2	10	0	5	5	1	2	6	0
Other	5	4	0	2	4	0	3	5	0

Note. C = Company. I = Individual. SE = Self-Employed. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B9

Frequency Distribution of Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Compared to a No Letter Control Group

Nudge	C pos	C zer	C neg	I pos	I zer	I neg	SE pos	SE zer	SE neg
Audit / Detection Threat	4	5	2	8	3	0	18	11	0
Penalty Threat	5	7	0	13	3	0	2	1	0
Other	4	2	0	5	4	0	9	5	0

Note. C = Company. I = Individual. SE = Self-Employed. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B10

Frequency Distribution of Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Compared to a Simple Letter Control Group

Nudge	C pos	C zer	C neg	I pos	I zer	I neg	SE pos	SE zer	SE neg
Audit / Detection Threat	2	6	1	5	1	0	2	4	0
Penalty Threat	1	7	0	6	8	0	0	5	0
Other	1	3	0	2	0	0	0	0	0

Note. C = Company. I = Individual. SE = Self-Employed. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B11

Frequency Distribution of Deterrence Interventions in the Samples of Company, Individual, and Self-Employed, Including Compliance History

Nudge	C pos	C zer	C neg	CN pos	CN zer	CN neg	I pos	I zer	I neg	IN pos	IN zer	IN neg	SE pos	SE zer	SE neg	SEN pos	SEN zer	SEN neg
Audit / Detection Threat	4	6	3	3	5	0	11	9	0	10	1	0	10	16	0	11	5	0
Penalty Threat	2	11	0	4	3	0	5	2	0	22	13	3	2	9	0	0	0	0
Other	1	2	0	8	3	0	0	2	0	12	3	0	2	2	0	8	3	0

Note. C = Company. CN = Company Non-Compliant. I = Individual. IN = Individual Non-Compliant. SE = Self-Employed. SEN = Self-Employed Non-Compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B12

Frequency Distribution of Deterrence Interventions in the Samples of Company, Individual, and Self-Employed, Including Compliance History Measured at the Extensive Margin

Nudge	C pos	C zer	C neg	CN pos	CN zer	CN neg	I pos	I zer	I neg	IN pos	IN zer	IN neg	SE pos	SE zer	SE neg	SEN pos	SEN zer	SEN neg
Audit / Detection Threat	3	0	0	1	1	0	8	3	0	6	0	0	2	0	0	4	3	0
Penalty Threat	0	3	0	3	0	0	5	2	0	17	7	2	0	1	0	0	0	0
Other	0	0	0	2	1	0	0	0	0	9	1	0	1	0	0	5	0	0

Note. C = Company. CN = Company Non-Compliant. I = Individual. IN = Individual Non-Compliant. SE = Self-Employed. SEN = Self-Employed Non-Compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B13

Frequency Distribution of Deterrence Interventions in the Samples of Company, Individual, and Self-Employed, Including Compliance History Measured at the Intensive Margin

Nudge	C pos	C zer	C neg	CN pos	CN zer	CN neg	I pos	I zer	I neg	IN pos	IN zer	IN neg	SE pos	SE zer	SE neg	SEN pos	SEN zer	SEN neg
Audit / Detection Threat	1	6	3	1	4	0	3	6	0	4	0	0	8	16	0	6	2	0
Penalty Threat	2	7	0	0	3	0	0	0	0	5	5	1	2	6	0	0	0	0
Other	1	2	0	4	2	0	0	2	0	2	2	0	1	2	0	2	3	0

Note. C = Company. CN = Company Non-Compliant. I = Individual. IN = Individual Non-Compliant. SE = Self-Employed. SEN = Self-Employed Non-Compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B14

Frequency Distribution of Deterrence Interventions in the Samples of Company, Individual, and Self-Employed, including Compliance History Compared to a No Letter Control Group

Nudge	C pos	C zer	C neg	CN pos	CN zer	CN neg	I pos	I zer	I neg	IN pos	IN zer	IN neg	SE pos	SE zer	SE neg	SEN pos	SEN zer	SEN neg
Audit / Detection Threat	1	2	2	3	3	0	8	3	0	0	0	0	8	9	0	10	2	0
Penalty Threat	1	6	0	4	1	0	2	2	0	11	1	0	2	1	0	0	0	0
Other	1	0	0	3	2	0	0	2	0	5	2	0	2	2	0	7	3	0

Note. C = Company. CN = Company Non-Compliant. I = Individual. IN = Individual Non-Compliant. SE = Self-Employed. SEN = Self-Employed Non-Compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B15

Frequency Distribution of Deterrence Interventions in the Samples of Company, Individual, and Self-Employed, Including Compliance History Compared to a Simple Letter Control Group

Nudge	C pos	C zer	C neg	CN pos	CN zer	CN neg	I pos	I zer	I neg	IN pos	IN zer	IN neg	SE pos	SE zer	SE neg	SEN pos	SEN zer	SEN neg
Audit / Detection Threat	2	4	1	0	2	0	1	0	0	4	1	0	2	4	0	0	0	0
Penalty Threat	1	5	0	0	2	0	3	0	0	11	9	0	0	5	0	0	0	0
Other	0	2	0	1	1	0	0	0	0	5	1	0	0	0	0	1	0	0

Note. C = Company. CN = Company Non-Compliant. I = Individual. IN = Individual Non-Compliant. SE = Self-Employed. SEN = Self-Employed Non-Compliant. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B16

Frequency Distribution of Deterrence Interventions in the Regional Subcategories (Part 1)

Nudge	AFR pos	AFR zer	AFR neg	AUS pos	AUS zer	AUS neg	ME pos	ME zer	ME neg	NA pos	NA zer	NA neg
Audit / Detection Threat	0	0	0	0	0	0	0	0	0	6	5	0
Penalty Threat	0	4	0	3	0	0	0	2	0	11	21	0
Other	1	0	0	1	0	0	0	0	0	10	4	0

Note. AFR = Africa. AUS = Australia/New Zealand. ME = Middle East. NA = North America. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B17*Frequency Distribution of Deterrence Interventions in the Regional Subcategories (Part 2)*

Nudge	NE pos	NE zer	NE neg	SA pos	SA zer	SA neg	S/E E pos	S/E E zer	S/E E neg	WE pos	WE zer	WE neg
Audit / Detection Threat	11	10	0	20	20	3	0	0	0	12	7	0
Penalty Threat	0	0	0	12	7	0	4	3	3	5	1	0
Other	0	0	0	9	8	0	7	2	0	3	1	0

Note. NE = Northern Europe. SA = South America. S/E E = Southeastern Europe. WE = Western Europe. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B18*Frequency Distribution of Deterrence Interventions in the Regional Subcategories Measured at the Extensive Margin (Part 1)*

Nudge	AFR pos	AFR zer	AFR neg	AUS pos	AUS zer	AUS neg	ME pos	ME zer	ME neg	NA pos	NA zer	NA neg
Audit / Detection Threat	0	0	0	0	0	0	0	0	0	4	0	0
Penalty Threat	0	4	0	2	0	0	0	0	0	6	4	0
Other	0	0	0	0	0	0	0	0	0	7	0	0

Note. AFR = Africa. AUS = Australia/New Zealand. ME = Middle East. NA = North America. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B19

Frequency Distribution of Deterrence Interventions in the Regional Subcategories Measured at the Extensive Margin (Part 2)

Nudge	NE pos	NE zer	NE neg	SA pos	SA zer	SA neg	S/E E pos	S/E E zer	S/E E neg	WE pos	WE zer	WE neg
Audit / Detection Threat	5	0	0	7	2	0	0	0	0	8	3	0
Penalty Threat	0	0	0	5	0	0	2	1	2	1	0	0
Other	0	0	0	2	0	0	5	0	0	1	0	0

Note. NE = Northern Europe. SA = South America. S/E E = Southeastern Europe. WE = Western Europe. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B20

Frequency Distribution of Deterrence Interventions in the Regional Subcategories Measured at the Intensive Margin (Part 1)

Nudge	AFR pos	AFR zer	AFR neg	AUS pos	AUS zer	AUS neg	ME pos	ME zer	ME neg	NA pos	NA zer	NA neg
Audit / Detection Threat	0	0	0	0	0	0	0	0	0	2	5	0
Penalty Threat	0	0	0	0	0	0	0	1	0	2	14	0
Other	1	0	0	0	0	0	0	0	0	3	4	0

Note. AFR = Africa. AUS = Australia/New Zealand. ME = Middle East. NA = North America. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B21

Frequency Distribution of Deterrence Interventions in the Regional Subcategories Measured at the Intensive Margin (Part 2)

Nudge	NE pos	NE zer	NE neg	SA pos	SA zer	SA neg	S/E E pos	S/E E zer	S/E E neg	WE pos	WE zer	WE neg
Audit / Detection Threat	6	9	0	11	16	3	0	0	0	4	4	0
Penalty Threat	0	0	0	4	4	0	2	2	1	1	0	0
Other	0	0	0	3	7	0	2	2	0	1	0	0

Note. NE = Northern Europe. SA = South America. S/E E = Southeastern Europe. WE = Western Europe. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B22

Frequency Distribution of Deterrence Interventions in the Regional Subcategories Compared to a No Letter Control Group (Part 1)

Nudge	AFR pos	AFR zer	AFR neg	AUS pos	AUS zer	AUS neg	ME pos	ME zer	ME neg	NA pos	NA zer	NA neg
Audit / Detection Threat	0	0	0	0	0	0	0	0	0	5	2	0
Penalty Threat	0	4	0	3	0	0	0	2	0	8	1	0
Other	1	0	0	1	0	0	0	0	0	3	1	0

Note. AFR = Africa. AUS = Australia/New Zealand. ME = Middle East. NA = North America. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B23

Frequency Distribution of Deterrence Interventions in the Regional Subcategories Compared to a No Letter Control Group (Part 2)

Nudge	NE pos	NE zer	NE neg	SA pos	SA zer	SA neg	S/E E pos	S/E E zer	S/E E neg	WE pos	WE zer	WE neg
Audit / Detection Threat	7	7	0	15	9	2	0	0	0	3	0	0
Penalty Threat	0	0	0	9	4	0	0	0	0	1	0	0
Other	0	0	0	9	8	0	2	2	0	2	0	0

Note. NE = Northern Europe. SA = South America. S/E E = Southeastern Europe. WE = Western Europe. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B24

Frequency Distribution of Deterrence Interventions in the Regional Subcategories Compared to a Simple Letter Control Group (Part 1)

Nudge	AFR pos	AFR zer	AFR neg	AUS pos	AUS zer	AUS neg	ME pos	ME zer	ME neg	NA pos	NA zer	NA neg
Audit / Detection Threat	0	0	0	0	0	0	0	0	0	1	3	0
Penalty Threat	0	0	0	0	0	0	0	0	0	3	17	0
Other	0	0	0	0	0	0	0	0	0	2	1	0

Note. AFR = Africa. AUS = Australia/New Zealand. ME = Middle East. NA = North America. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Table B25

Frequency Distribution of Deterrence Interventions in the Regional Subcategories Compared to a Simple Letter Control Group (Part 2)

Nudge	NE pos	NE zer	NE neg	SA pos	SA zer	SA neg	S/E E pos	S/E E zer	S/E E neg	WE pos	WE zer	WE neg
Audit / Detection Threat	2	1	0	3	9	1	0	0	0	5	7	0
Penalty Threat	0	0	0	1	2	0	2	1	0	1	1	0
Other	0	0	0	0	0	0	1	1	0	0	1	0

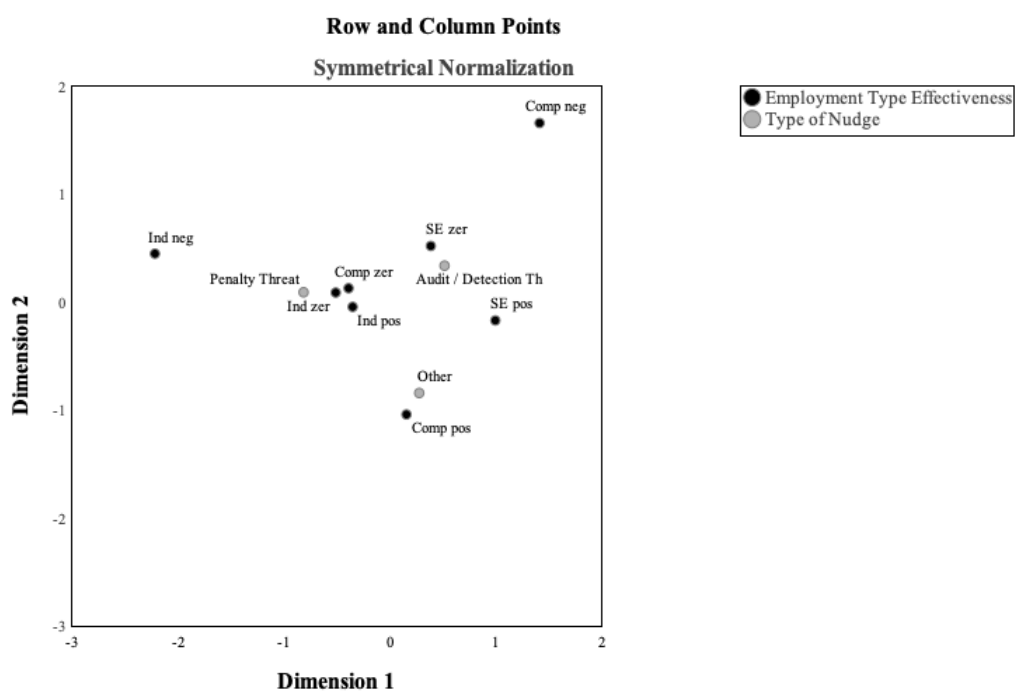
Note. NE = Northern Europe. SA = South America. S/E E = Southeastern Europe. WE = Western Europe. Pos = Frequency of positive effects. Zer = Frequency of zero effects.

Neg = Frequency of negative effects.

Appendix C

Figure C1

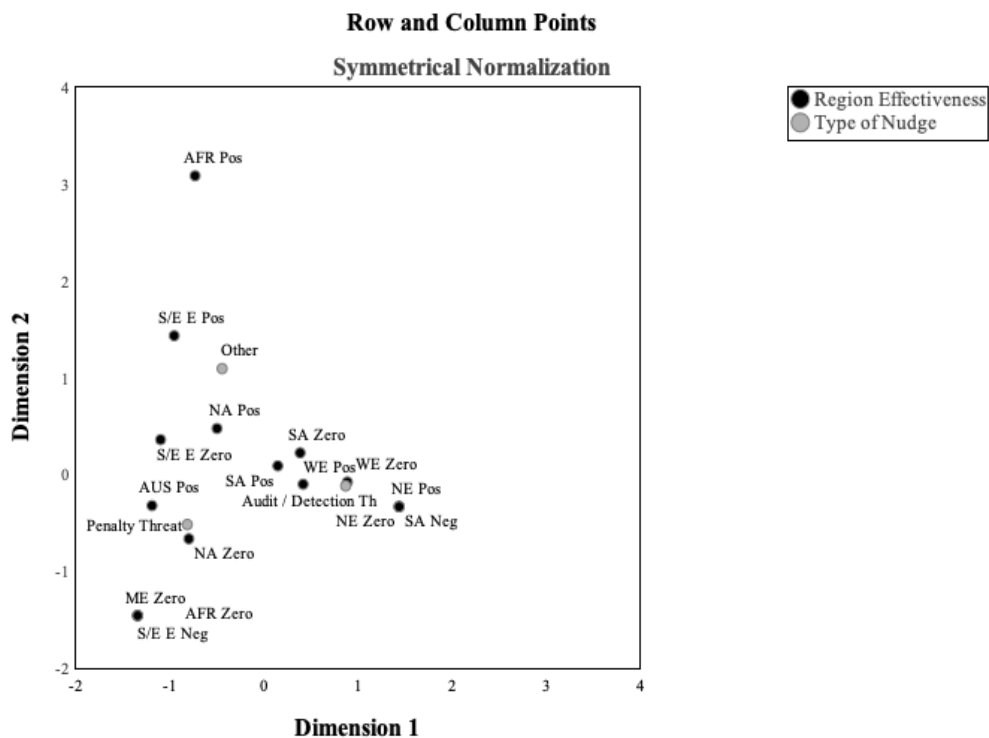
Correspondence Analysis for Frequency Distributions of Deterrence Interventions in the Subsamples of Company, Individual, and Self-Employed



Note. Black points represent the rows (i.e., the population subcategories). Grey points represent the columns (i.e., the deterrence interventions). Comp = Company. Ind = Individual. SE = Self-Employed. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Figure C2

Correspondence Analysis for Frequency Distributions of Deterrence Interventions in the Regional Subsamples



Note. Black points represent the rows (i.e., the regional subcategories). Grey points represent the columns (i.e., the deterrence interventions). AFR = Africa. AUS = Australia/New Zealand. ME = Middle East. NA = North America. NE = Northern Europe. SA = South America. S/E E = Southeastern Europe. WE = Western Europe. Pos = Frequency of positive effects. Zer = Frequency of zero effects. Neg = Frequency of negative effects.

Appendix D

Figure D1

Fleiss' Kappa for Two Independent Raters Across All Three Main Categories

Subjects = 1611
Raters = 2
Kappa = 0.825

z = 107
p-value = 0

	Kappa	z	p.value
Audit / Detection Threat	0.879	35.294	0.000
Company	0.776	31.164	0.000
Company / Noncompliant	0.743	29.814	0.000
Individual	0.798	32.025	0.000
Individual / Noncompliant	0.928	37.245	0.000
Information	0.602	24.162	0.000
Interaction	0.483	19.369	0.000
Moral Appeal	0.397	15.947	0.000
Negative	0.855	34.302	0.000
Negative Framing	1.000	40.137	0.000
Other	0.190	7.642	0.000
Penalty Threat	0.782	31.386	0.000
Positive	0.949	38.079	0.000
Positive Framing	1.000	40.137	0.000
Public Goods	0.821	32.934	0.000
Punishment	-0.001	-0.050	0.960
Reward	0.950	38.128	0.000
Self-employed	0.545	21.861	0.000
Simple Letter	0.845	33.912	0.000
Simple Reminder	0.912	36.621	0.000
Simplification	0.724	29.077	0.000
Social Norm	0.824	33.076	0.000
Zero	0.966	38.757	0.000

Figure D2

Fleiss' Kappa for Two Independent Raters for the Category 'Population'

Subjects = 537
Raters = 2
Kappa = 0.777

z = 31.6
p-value = 0

	Kappa	z	p.value
Company	0.748	17.343	0.000
Company / Noncompliant	0.722	16.727	0.000
Individual	0.753	17.441	0.000
Individual / Noncompliant	0.893	20.695	0.000
Self-employed	0.525	12.168	0.000

Figure D3

Fleiss' Kappa for Two Independent Raters for the Category 'Type of Intervention'

Subjects = 537
Raters = 2
Kappa = 0.692

z = 47.8
p-value = 0

	Kappa	z	p.value
Audit / Detection Threat	0.862	19.966	0.000
Information	0.594	13.755	0.000
Interaction	0.427	9.885	0.000
Moral Appeal	0.376	8.714	0.000
Negative Framing	1.000	23.173	0.000
Other	0.144	3.337	0.001
Penalty Threat	0.762	17.661	0.000
Positive Framing	1.000	23.173	0.000
Public Goods	0.811	18.792	0.000
Punishment	-0.004	-0.087	0.931
Reward	0.947	21.949	0.000
Simple Letter	0.842	19.520	0.000
Simple Reminder	0.908	21.048	0.000
Simplification	0.719	16.653	0.000
Social Norm	0.810	18.761	0.000

Figure D4

Fleiss' Kappa for Two Independent Raters for the Category 'Effectiveness'

Subjects = 537
Raters = 2
Kappa = 0.925

z = 24.5
p-value = 0

	Kappa	z	p.value
Negative	0.849	19.681	0.000
Positive	0.924	21.405	0.000
Zero	0.940	21.789	0.000

Zusammenfassung

Verhaltensinterventionen wie Nudges werden häufig mit sanfter Einflussnahme assoziiert, doch nicht alle Strategien folgen diesem Ansatz. Die vorliegende Arbeit untersucht, wie abschreckungsbasierte Maßnahmen das Steuerverhalten in unterschiedlichen Kontexten beeinflussen. Ergebnisse aus Studien im Rahmen natürlicher Feldexperimente wurden systematisch im Hinblick auf vergangenes Steuerverhalten, den Beschäftigungskontext, das regionale Setting sowie die Art der Intervention kategorisiert und analysiert. Die Analyse zeigt, dass Abschreckungsmaßnahmen bei Personen mit bereits dokumentierter Steuerpflichtsmisachtung überwiegend wirksam waren. In Studien mit Stichproben aus der Allgemeinbevölkerung hingegen zeigten sich häufiger keine signifikanten Effekte. Zusätzlich wurden die Ergebnisse in Bezug auf verschiedene abhängige Variablen ausgewertet. Die binäre Messung der Entscheidung, ob Steuern gezahlt wurden oder nicht, ergab mehrheitlich positive Effekte in beiden Gruppen, wobei diese für Personen mit vorheriger Steuerpflichtsmisachtung noch ausgeprägter ausfielen. Regional variierte die Wirksamkeit abschreckender Maßnahmen erheblich: Besonders starke positive Effekte wurden in Europa festgestellt, in Südamerika überwogen ebenfalls positive Ergebnisse, jedoch weniger deutlich. Insgesamt erwiesen sich Interventionen mit der Androhung einer Steuerprüfung als etwas wirksamer als solche mit Strafandrohungen. Obwohl die bestehende empirischen Literatur auf mögliche negative oder gegenteilige Effekte durch Abschreckung hinweist, fanden sich in dieser Analyse kaum Hinweise darauf. Die meisten Interventionen führten entweder zu einer höheren Steuercompliance oder hatten keinen signifikanten Einfluss, wobei insgesamt die positiven Effekte überwogen. Diese Ergebnisse stützen metaanalytische Befunde und unterstreichen sowohl die Bedeutung abschreckender Maßnahmen als zentralen Einflussfaktor auf das Steuerverhalten als auch die Notwendigkeit, Compliance-Strategien kontextspezifisch zu gestalten.

Stichworte: Steuercompliance, Nudging, Deterrence, Steuerprüfung, Strafandrohung