MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Soft Nudging for Tax Compliance: Systematic Review on the Effectiveness of Non-Deterrence Nudging Interventions on Tax Compliance

verfasst von / submitted by

Alexander Truzka, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2021 / Vienna 2021

Contents

Abstract

Many recent field experiments investigated the effectiveness of non-deterrence nudging interventions on tax compliance as an alternative to traditional audits and sanctions. However, previous research to date has not yet systematically investigated how the effectiveness of non-deterrence interventions is influenced by individual and contextual factors. The current systematic review therefore integrated findings from individual field experiments into a comprehensive framework. The study effects were investigated for different subpopulations of taxpayers, taking into account their occupation and compliance history. Additionally, the effects were analysed for various regions in which the field experiments were conducted. The body of literature was analysed according to the PRISMA guidelines. The results showed that non-deterrence nudging interventions were more frequently effective in increasing tax compliance in a previously non-compliant population compared to a general population of taxpayers. Additionally, nudging interventions were found to be more frequently effective in European than in North American field studies. When looking at the interventions in detail, it was found that most interventions including a simplification element were effective, however those effects mainly stem from a population of individual taxpayers. Social norm interventions also showed frequent positive effects, mainly in the population of previously non-compliant taxpayers. Reciprocity and information treatments did not display frequent positive effects in any subpopulation or regional context. Further evidence is necessary, especially for a corporate sample of taxpayers as well as for countries outside of Europe, North, and South America.

Keywords: tax compliance, non-deterrence, nudging, compliance history, regional variation

Introduction

What motivates tax compliance? This is a question that researchers, tax authorities, and governmental institutions in general, are increasingly eager to solve. The body of literature on this issue is constantly growing and has seen a rapid expansion in publications on various levels, especially in the last five to ten years. From laboratory experiments to surveys, as well as natural field experiments, and cross-country analyses (for reviews of existing literature see, for example, Alm et al., 2020b; Mascagni, 2018; Slemrod, 2019), various approaches tried to explain the mechanisms behind tax compliance. The current systematic review aims to add to previous literature on tax compliance by focusing on individual and contextual factors that influence decisions for tax compliance. Before diving into this topic in detail, however, it is important to discuss why tax compliance is of fundamental importance in the first place.

As previous research already noted, "the collection of taxation is a crucial function for governments worldwide" (Hallsworth et al., 2017, p. 14) to provide them with the funds necessary to invest in country-wide public goods and services like infrastructure, healthcare and social security, as well as long-term development, and to accelerate growth (Organization of Economic Development and Growth (OECD), 2013). If governments understand the needs of their taxpayers and respond and invest accordingly, the taxpayers will benefit from their compliance while governments, in turn, benefit from higher tax revenues (OECD, 2019), creating a perpetual virtuous cycle profitable for both sides. However, there are still many, especially developing, countries that raise relatively low tax revenues compared to most OECD countries (OECD, 2013). Given a relatively strong correlation between tax revenues and the level of a country's development (OECD, 2013), this further highlights the importance of tax collection and, therefore, tax compliance. Interestingly though, according to cross-country analyses conducted by the OECD (2019), there are also differences in the amount of tax revenue collection between countries with similar development statuses. These findings show that the amount of tax revenues collected cannot be solely explained by differences between countries' developmental statuses. The variation raises questions regarding other influences on the collection of tax revenues, thereby shifting away from cross-country comparisons to including an individual level. Considering individual taxpayers' motivations to pay or evade their taxes within different settings in countries, thus, leads to the context of the current review.

As mentioned at the beginning, there is a large body of literature as well as reviews attempting to explain mechanisms behind tax compliance that came to varying conclusions (for a review, see Alm, 2019). Most research, however, agreed that tax compliance is influenced

by two main aspects: tax enforcement and individual tax morale (Kirchler et al., 2008; Luttmer & Singhal, 2014; Mascagni, 2018). The tax enforcement approach is based on a rational choice model proposed by Allingham and Sandmo (1972). It is assumed that compliance is driven by financial considerations and, besides tax rate and income, mainly dependent on the level of enforcement within a country (i.e., the probability of detection and sanctions that are imposed when tax evasion is detected). The tax morale approach, on the other hand, focuses on voluntary tax compliance and its determinants, which takes into account context-dependent factors like intrinsic motivation, reciprocity, fairness, peer effects, social influences, and information imperfections (Luttmer & Singhal, 2014). Voluntary tax compliance, compared to enforced compliance, implies that the taxpayer will be compliant without proactive action by the government (Plumley, 2002). Both aspects are assumed to interact, meaning that individual tax morale is affected by the legitimacy of enforced tax policies (i.e., how taxpayers perceive tax enforcement strategies; Tyran & Feld, 2001).

It is, however, fundamentally difficult to empirically establish a link between a policy intervention and subsequent changes in compliance levels, especially when factoring in individual tax morale. Taxpayers, for once, have a strong incentive to conceal their illegal activities due to potential penalties imposed on them when their activities are discovered (Alm, 2019). Conversely, there exist various contextual factors (e.g., culture or geographic region; Hallsworth, 2014) as well as different individual preferences (e.g., preference for simplicity vs. complexity; Gould, 1997) that influence individual tax compliance decisions, which limit the generalizability of the results to other contexts. Even though laboratory experiments are considered quite useful for empirically assessing tax compliance (Hallsworth, 2014), they do face stronger constraints regarding generalizability due to a general lack of external validity (i.e., the transfer of findings to real-world settings; Camerer, 2012). Consequently, researchers have begun to cooperate more with tax authorities to implement controlled field experiments (see Section 1.7; Hallsworth, 2018). In those field experiments, the typical approach was to send out messages (e.g., a letter or electronic notification) to real taxpayers via the authority and investigate whether the message (i.e., an intervention or treatment) influences tax compliance. With experimental and control groups in real-world environments, field studies have some advantages in identifying potential causal determinants for compliance decisions (Alm, 2019).

The meta-analysis conducted by Antinyan and Asatryan (2020) examined which types of policy interventions in field experiments work particularly well on enhancing tax compliance. They compared the effectiveness of deterrence treatments (i.e., an information or threat

highlighting tax enforcement policies) with non-deterrence treatments (i.e., an information highlighting factors associated with tax morale) and showed that deterrence treatments on average increase compliance, while the findings for non-deterrence treatments were less conclusive. However, Antinyan and Asatryan (2020) neither accounted for environmental factors (e.g., regional differences) nor for individual differences (e.g., different attitudes towards taxes within the population) in their framework. Still, given the differing circumstances across individuals, firms, and countries, there is a "need to move from analysing 'what works' to 'what works, for whom, when, and why'", as "for any set of results, we need to understand if we are dealing with findings that will not be obtained in different circumstances" (Hallsworth, 2018, p. 446).

Therefore, the aim of the present review on field experiments is to address these issues by taking into account not only different forms of interventions (i.e., deterrence and non-deterrence interventions), but also the population of taxpayers on which the intervention was performed, as well as regional variations in which compliance was assessed. This allows for shedding further light on the 'whydunit' of tax compliance. In Section 1.1 of this review, the traditional approach on tax evasion proposed by Allingham and Sandmo (1972), which focuses on the use of deterrent methods, will be discussed. In Section 1.2, some limitations of this approach will be pointed out and subsequently, the 'Slippery Slope'-model proposed by Kirchler et al. (2008), that aimed to integrate deterrence and non-deterrence considerations, will be discussed (see Section 1.3). Section 1.4 will then focus on the heterogeneous mechanisms behind non-deterrent methods. Section 1.5 and Section 1.6 will discuss taxpayer heterogeneity as well as regional variation, both of which are important to consider when implementing interventions and policies in the context of tax compliance. Finally, in Section 1.7, the use of and focus on field experiments will be explained, leading to the objectives and the rationale of the current systematic review in Section 1.8.

1.1 The Deterrence Approach on Tax Evasion

As discussed at the beginning, the compliance of taxpayers is supposed to be strongly affected by tax enforcement policies, which are largely influenced by the traditional approach to tax evasion proposed by Allingham and Sandmo (1972). The traditional model was based on the standard economic assumption of human behaviour, arguing that individuals act rationally, have unlimited willpower, and are solely interested in maximizing their utility and advancing their private goals. Therefore, they proposed that individuals' decision to pay or evade their taxes is derived from their expected utility after weighing the risk that arises from:

1) the possibility of evasion being detected and 2) the severity of punishment when caught against potential gains when evasion goes unnoticed. Besides these factors, taxpayers also consider the tax rate and their income within this decision process (Allingham & Sandmo, 1972). Based on their model, non-compliance can be reduced if the tax authority implements effective tax enforcement policies as, for example, increasing the frequency of audits or heightening the salience and severity of penalties (Hallsworth, 2018). Conversely, if there are not any or insufficient enforcement policies in place that remove opportunities to evade, tax-payers will have no incentive to voluntarily cooperate with tax authorities (Hallsworth, 2014).

The assumptions underlying the traditional model by Allingham and Sandmo (1972) gave rise to the enactment of various enforcement treatments in laboratory (see Alm et al., 2020b) and field experiments (see Hallsworth, 2014) to investigate deterrent effects on tax compliance. Deterrence interventions mostly aim at highlighting audit rates and penalties due to the assumption that taxpayers are afraid of both, detection, and punishment (Slemrod et al., 2001). Laboratory experiments typically implemented these interventions in the form of vary-ing audit probabilities and penalties within the framework of tax games (for a review, see Alm et al., 2020b). One of the assumptions being that the higher, for example, the audit probabil-ity, the fewer taxpayers should be willing to evade their taxes during the game. Deterring field experiments, on the other hand, focused mainly on increasing the salience of audit probabili-ties and penalties in messages that researchers sent out to actual taxpayers in cooperation with tax authorities (e.g., Boning et al., 2020; Brockmeyer et al., 2019; Harju et al., 2018). Meta-analyses on both, laboratory and field experiments, could show that interventions emphasiz-ing traditional determinants of compliance, such as audit probabilities and penalty rates, in-deed have a potential deterrent effect on tax compliance (e.g., Alm et al., 2020b; Antinyan & Asatryan, 2020; Blackwell, 2007).

1.2 Limitations of the Deterrence Approach

Allingham and Sandmo's (1972) model, however, significantly underestimated true compliance rates within the population, when considering the relatively low probability of actually being subjected to an audit (Alm, 2012). Evidence suggests that most individuals correctly declare their income and pay their taxes most, if not all of the time (Alm, 2019). The World Values Survey, which has been conducted regularly since 1981 in almost 100 coun-tries, aimed to study changing values and their impact on social and political life. The survey also included questions regarding the justifiability of "cheating on taxes if you have a chance". Over 80 percent of participants in the survey stated a value of 8 or higher on a 10-

point scale, where 10 indicated that cheating on taxes is "never justifiable".[1] These responses further highlighted a general tendency of taxpayers towards compliance in many countries (Inglehart et al., 2014). It was also assumed that the sole implementation of enforcement strategies might even crowd out the intrinsic motivation of compliant taxpayers because they would feel obliged to pay instead of wanting to pay their taxes (Frey, 1997).

Additionally, the perspective of an individual deciding to pay or evade taxes based on a single motivation (i.e., financial considerations; Alm et al., 2020a), as proposed by the Allingham and Sandmo (1972) model, is perceived as restrictive and ineffective (Ayres & Braithwaite, 1992). The traditional approach is neither applicable to every context (Braithwaite, 2003; Kirchler et al., 2008) nor does it take into account additional influencing factors such as individual tax morale and its determinants, like perceptions of fairness and reciprocity (Kirchler, 2007).

Besides motivational considerations of taxpayers, it should also be noted that tax authorities find themselves with limited financial resources to implement deterrence measures like increasing audit rates. For example, a threat of 100 percent audit rates (see Kleven et al., 2011) is highly unlikely because of resource constraints. Therefore, taxpayers who doubt the credibility of such threats will eventually be less susceptible to them (Slemrod, 2019), which, in turn, means that the effectiveness of deterrence interventions and policies also depends upon the validity of the threat.

1.3 The Importance of Non-Deterrence

Due to the rather one-sided perspective and the above-mentioned limitations of the deterrence approach, research increasingly began to include non-financial considerations in their assumptions regarding the compliance decision process, which will be subsequently labelled as 'non-deterrence' factors. The *Slippery Slope Framework* (SSF) proposed by Kirchler et al. (2008), for example, highlighted the importance of considering non-deterrence factors like trust in the context of tax compliance and showed how deterrence and non-deterrence factors interact.

In this model it was argued that tax compliance is, on one hand, affected by the 'power of authorities', which is the ability of authorities to enforce compliance. On the other hand, 'trust in authorities' plays a crucial role, thereby incorporating both considerations in the model (Kirchler et al., 2008). The SSF presumed that tax compliance will be low, and taxpay-

---

[1] Data from the 2005-2009 wave of the World Values Survey from 53 participating countries, see https://www.worldvaluessurvey.org/WVSDocumentationWV5.jsp for more information

ers would act according to maximising their utility if neither power of nor trust in authorities is high within a given context. However, authorities could enhance tax compliance by either increasing voluntary compliance through non-deterrent methods or enforcing compliance by exerting power. Finally, the model proposed that both, the power of and trust in authorities, need to be high to achieve a maximum level of compliance among the taxpaying population (Kirchler et al., 2008). The assumptions behind the SSF were examined in a large-scale study in 44 countries and according to the findings, the framework seems to be applicable to various tax contexts (Batrancea et al., 2019). It was found that trust, indeed, increases voluntary compliance, whereas power triggers enforced compliance. Additionally, the power of the tax authority seems to positively affect voluntary compliance, given that taxpayers already exhibited trust in authorities, thereby confirming the interaction of both aspects. This shows that building trust in authorities might be an important pre-condition for the effective enforcement of tax compliance.

Furthermore, it was argued that the relationship between trust and power can be explained by how the power is exerted (Leonidou et al., 2008). In the context of tax compliance, for example, this means that when coercive power (i.e., fear or threats as an incentive) is used to target tax evaders or taxpayers classified as 'at risk', the use of power is perceived as legitimate and in turn, positively affects taxpayers' trust in authorities and, subsequently, voluntary tax compliance. However, if coercive power is directed randomly at a general taxpaying population, the use of power may be perceived as illegitimate and trust in authorities is negatively affected as a result.

Levels of trust within the population seem to be a rather important factor that should be taken into account by authorities. Trust has been repeatedly linked not only to tax compliance but to economic development and growth in general (Algan & Cahuc, 2010; Guiso et al., 2006). Ortiz-Ospina and Roser (2016) correlated trust-related questions (e.g., "most people can be trusted") in the data from the World Values Survey from 2014 with the GDP per capita of various countries and found that trust, indeed, seems to be linked to a country's prosperity. To account for this, the non-deterrence approach suggests that instead of controlling taxpayers' behaviour by solely threatening with deterrence measures, potentially undermining trust in authorities (Batrancea et al., 2019; Leonidou et al., 2008) and intrinsic motivation to pay taxes (Frey, 1997), taxpayers should be treated fairly and with respect. In addition to that, it was argued that they should be given clear information, help, and provided with services to make compliance with tax authorities as easy as possible (Hallsworth, 2014). Therefore, in the

following section, the importance of various non-deterrence factors and their mechanisms for understanding and increasing individual tax compliance will be discussed.

1.4 The Inclusion of Non-Deterrence Factors

Individuals are assumed to deviate from their expected utility not only due to the trust-worthiness of authorities, but also because of cognitive biases related to limited self-control and cognitive resources (e.g., Dwenger et al., 2016), incomplete information available to them (e.g., Del Carpio, 2014), framing effects (e.g., Boyer et al., 2016), as well as considerations of morality, peer behaviour, social contexts, fairness, and reciprocity (see Luttmer & Singhal, 2014). The behavioural science approach incorporates these deviations from the traditional assumptions to present a more realistic view of how individual taxpayers make compliance decisions. As non-deterrence factors involve a greater variety and more abstraction in concepts (e.g., in terms of the wording of interventions; Hallsworth, 2014), evidence regarding their effect on tax compliance is quite inconsistent (Antinyan & Asatryan, 2020; Slemrod, 2019). Consequently, some assumptions and mechanisms behind non-deterrence factors that are especially relevant for the current systematic review will be outlined in detail. For once, to highlight the importance of non-financial considerations in the context of tax compliance, but also to shed some light on potential reasons for inconclusive findings in recent literature.

1.4.1 Psychological Costs

Non-deterrent factors like morality, peer behaviour, and fairness were suggested to represent some form of *psychological cost* to the individual that is considered when making compliance decisions (Gordon, 1989). Individuals will comply if they think that compliance is the 'right thing to do', suggesting that the decision depends upon some form of social norm pervasive within the population (Alm & Torgler, 2011). Social norms, in general, can be defined as shared expectations of acceptable behaviour within groups (Lapinski & Rimal, 2005). Therefore, if peer behaviour suggests that tax compliance is the norm, deviating from this norm could potentially increase the psychological cost of non-compliance (Traxler, 2010) due to, for example, feelings of guilt or shame (Erard & Feinstein, 1994) that affect the individuals' self-image. It was also found that social norms have an even stronger positive impact on tax compliance if the taxpayer in question identified with the peer group exhibiting the behaviour (Wenzel, 2005). However, the behaviour of peers might as well influence taxpayers' beliefs about the norm of compliance in a negative way (Fellner et al., 2013). If the environment suggests that most people do not pay their taxes, the individual might not perceive any psy-

chological cost when evading taxes themselves. Differences in prior beliefs about social norms might thus explain some of the mixed findings in regard to non-deterrence interventions that attempted to influence behaviour through incurring psychological costs via the inclusion of a social norm (e.g., Blumenthal et al., 2001; Fellner et al., 2013).

1.4.2 Reciprocity

Reciprocity, in a positive sense, can be broadly defined as a mutual exchange, where two parties give each other help and advantages (Falk & Fischbacher, 2006). In the context of tax compliance, reciprocity concerns the mutual relationship between the taxpayer and the government. This relationship is influenced by trust in authorities (Kirchler et al., 2008), which was discussed in Section 1.3, as well as general attitudes towards the government and its institutions (Braithwaite, 2003), which will be further detailed in Section 1.5. For now, when talking about reciprocity, the focus lies on the direct tax-benefit mechanism that influences the compliance decision process.

It was argued that taxpayers' compliance decisions are affected by how their tax payments are perceived in relation to the public goods and services they receive (Bordignon, 1993). Various field experiments, therefore, introduced non-deterrence interventions that informed taxpayers about the public goods and services that the government provides through tax revenues. However, the effects were inconclusive, with some studies finding positive effects (Hallsworth et al., 2017; Hernandez et al., 2017), while others found no effects on tax compliance (Blumenthal et al., 2001; Castro & Scartascini, 2015). This might be due to the fact that individual tax compliance is strongly associated with taxpayers' *satisfaction* about how their tax money is spent (OECD, 2013). Laboratory experiments could also show that compliance of participants is greater when they are given the possibility to vote on how their tax money is used (Casal et al., 2016; Wahl et al., 2010). Informing individuals about the public goods and services provided and, therefore, enhancing their awareness on how their tax money is spent, might only increase compliance if the public good or service is also perceived as a benefit to the individual. Additionally, the findings that information about the provision of public goods and services did not affect tax compliance could also be explained by the argument that the information itself was not powerful enough to update taxpayers' beliefs. It was, for example, argued that beliefs about the perception of public goods might be formed through a lifetime of experiences (Luttmer & Singhal, 2014).

The mechanism of reciprocity, however, does not influence taxpayers' compliance only in terms of the provision of public goods and services. Laboratory experiments could also

show that directly rewarding individuals for their honesty increased tax compliance, for example, when offering eligibility for a lottery (Alm et al., 1992; Bazart & Pickhardt, 2011; Feld et al., 2006). However, field experiments found relatively little evidence for the reciprocal effects of direct rewards on tax compliance (Dunning et al., 2015; Dwenger et al., 2016). In their field experiment, Koessler et al., (2019) argued that the effectiveness of direct rewards was strongly dependent on its attractiveness for the target population, similarly to the arguments mentioned in regard to satisfaction with the provision of public goods (OECD, 2013). Overall, the reciprocal mechanism in the context of tax compliance seems to be very reliant on how the public goods or rewards are perceived, as well as how beneficial they are for the taxpayers.

1.4.3 Cognitive Deviations

Multiple researchers argued that (cognitive) deviations from the expected utility model also play an important role in the tax compliance decision process (Alm, 2019; Luttmer & Singhal, 2014). First, individual decisions tend to be influenced by how information is presented to them (i.e., how information is framed). A classic example would be the so-called 'loss aversion', which leads to individuals being more inclined to take a risk (Kahneman & Tversky, 1979) or behave in a dishonest way (Schindler & Pfattheicher, 2017) to avoid losses compared to acquiring gains of equal value. Field experiments in the context of tax compliance have already shown that framing information in letters sent to taxpayers (e.g., presenting information as a 'loss' vs. a 'gain') have affected tax compliance in different ways (e.g., Boyer et al., 2016; Dunning et al., 2015; Hallsworth et al., 2017).

Second, information imperfections (i.e., when not all information is available to the taxpayers) can also lead their behaviour to deviate from expected utility, for example, when misperceiving probabilities. Taxpayers tend to overestimate the small probability of tax evasion being detected, as they usually do not possess sufficient information about actual audit rates and therefore assume that the probability of being audited is much higher than it actually is (Scholz & Pinney, 1995). However, taxpayers also tend to underestimate actual compliance rates within the population (Fellner et al., 2013). One field experiment in Peru, for example, disclosed information on both, true audit, and true compliance rates to taxpayers. While disclosing information on compliance rates had a large positive effect on tax compliance, disclosing information on actual enforcement probabilities did not raise compliance (Del Carpio, 2014). Similarly to Fellner et al. (2013), it was argued that the effectiveness of such interventions depends upon current beliefs about compliance and audit rates based on available infor-

mation and whether the intervention is able to update those beliefs in the desired direction, which has been the case with non-deterrent but not with deterrent information in her field experiment.

Finally, limited cognitive resources might incline taxpayers to deviate from a rational expected utility decision. Individuals might simply forget to remit their taxes or feel overwhelmed with the amount of information and subsequently procrastinate dealing with their tax payments (Luttmer & Singhal, 2014). It was already shown to help taxpayers to improve their tax compliance if they are provided with simple payment reminders (e.g., Del Carpio, 2014; Hallsworth et al., 2017) to tackle the problem of limited attention. Another approach is to present the information in a simplified way, for example, through visual stimuli or a reduction in content, both of which have been shown to increase tax compliance (De Neve et al., 2019; Dwenger et al., 2016).

Taking into account the above-mentioned deviations from the expected utility model, some of the heterogeneous treatment effects of non-deterrent interventions on tax compliance can be explained. Therefore, the non-deterrence approach suggests that contextual and environmental aspects (e.g., social norms or the provision of public goods; Dolan et al., 2012) together with individual preferences (Gould, 1997) need to be considered when implementing such policies and interventions. It was additionally argued that their effectiveness is determined by what kind of taxpaying population the interventions are communicated to (Hallsworth, 2014). Accordingly, in the following two sections, the importance of considering individual beliefs, motivations, and preferences, as well as regional differences within the non-deterrence approach will be discussed, thereby highlighting some of the advantages of the current systematic review in the process.

1.5 Taxpayer Heterogeneity

Taxpayers exhibit great diversity in their motivation to pay their taxes, and governments and tax authorities need to assess these individual motivations as well as current beliefs in the taxpaying population before deciding which approach to use to communicate (Hallsworth, 2014). This was called a 'full house' of behaviours that authorities need to recognize to appropriately respond to their population (Gould, 1997). As described above, some individuals might indeed be solely motivated by financial considerations, whereas others might be very inclined to adjust their behaviour to that of the peer group and are, thus, stronger motivated by social considerations (Alm, 2019). However, not only deterrent and non-deterrent factors in-

fluence the behaviour of taxpayers. The relationship between the taxpayer and the government or authority also plays a crucial role in this context.

In Section 1.3, it was shown how trust in general and the trustworthiness of authorities influence taxpaying behaviour (see Batrancea et al., 2019; Kirchler et al., 2008). However, it is not only of importance how the taxpayer perceives the authority based on their actions but also how the individual taxpayer is disposed towards the authority and taxes in general (e.g., through a political alignment; Cullen et al., 2018). In her line of work, Braithwaite (2003) identified five different *motivational postures* that describe attitudes toward tax authorities. Unlike the deviations from expected utility, which were argued to be rather context-dependent (see Section 1.4), motivational postures were assumed to be less context-sensitive and represent an overall set of beliefs about taxes that are "consciously held and openly shared" (Braithwaite, 2003, p. 4). The motivational postures range from *commitment* and *capitulation*, which are described as postures of 'deference', to *resistance*, *disengagement, and game playing*, which are labelled as postures of 'defiance' regarding the orientation towards the authority. Deference postures represent a moral obligation to act in the interest of society and, therefore, reflect an overall positive orientation. On the contrary, resistant taxpayers doubt the good intentions of authorities to cooperate with them, and disengaged taxpayers see no point in cooperating with or challenging the authorities in the first place. Both of these postures reflect an overall negative orientation towards the authority. Finally, game players try to find ways to circumvent tax laws by exploiting grey areas in legislation (Braithwaite, 2003). The motivational postures proposed by Braithwaite (2003) were not the only attempts at capturing general attitudes within the population of taxpayers. In 2010, the Internal Revenue Service (IRS) also suggested several 'motivational postures', partly overlapping with those proposed in the Braithwaite model. Their postures range from 'pathologically honest', 'conflicted', 'fearful', 'surprised', to 'careless/negligent/procrastinator', 'strategic', and 'pathologically defiant' (for details, see Alm, 2019).

These attempts at segmenting taxpayers have in common that they suggest authorities need to respond differently to different kinds of taxpayers (Braithwaite, 2009). For example, taxpayers with an overall positive orientation towards the authority should not be subjected to harsh controls and severe punishments, but instead should be provided with help and information (Kirchler, 2007). This form of responsive regulation (Ayres & Braithwaite, 1992), in turn, could improve taxpayers' trust in authorities, further improving tax compliance in the process (Alm et al., 2020a). Taxpayers' motivations and attitudes towards the authority should therefore be kept in mind when designing interventions. As mentioned at the begin-

ning, researchers have increasingly pointed out that it is important to understand not only what way of communication works best, but also understand for what reasons specific interventions work (Hallsworth, 2018).

Besides individual motivations of taxpayers, authorities also need to recognize differences between taxpayers in terms of their 'occupation' (i.e., whether they are employed, self-employed, or operate in a corporate context). It was pointed out that research still largely focuses on individual taxpayers (Hallsworth, 2018), even though businesses make up most of the tax revenue collected by countries (Slemrod & Velayudhan, 2017). Firms or self-employed individuals were argued to differ from employed individuals in terms of their tax compliance (Alm et al., 2020a). Self-employed individuals, for example, were classified as a high-risk group, because they pay most of their taxes themselves and therefore have a higher opportunity to cheat (e.g., Kleven et al., 2011) compared to employed individuals, who in many countries have their taxes deducted from their gross income by their employer. Thus, it was argued that the higher opportunity to cheat is the result of self-employed individuals paying their taxes 'out-of-pocket' and, therefore, perceiving their tax payments as a loss (Kirchler et al., 2009). Accordingly, self-employed individuals were supposed to be more prone towards risk-taking to avoid these losses (Kahneman & Tversky, 1979). One study, for example, found that countries with a high self-employment rate collected a smaller amount of taxes than countries with higher employment rates (Kleven, 2014).

In terms of larger companies, many prefer to use professional tax preparers or have their own accounting department to do their tax returns for them. In a study in the US, it was found that 97 percent of companies use professional preparers (Klepper et al., 1991). Even though this is neither current data nor is it applicable to other countries, it does, however, give an indicator for the large number of companies that seem to depend on tax professionals. However, relying on professional preparers does not necessarily have a positive impact on tax compliance. Tax professionals' job is making sure that firms correctly declare their taxes. Still, tax compliance was found to be lower for professional preparers (Erard, 1997), possibly due to them being more likely to take advantage of legal grey areas (Klepper & Nagin, 1989).

The arguments presented in this section show how important it is for tax authorities to take differences in motivation and attitudes, as well as opportunities to cheat or bypass regulations within their taxpaying population into account when designing policies and interventions to improve tax compliance. Interventions should be tailored to the population to achieve the most promising results, however, there is still only isolated evidence on what kinds of interventions work best (Alm, 2019; Hallsworth, 2018) regarding, for example, different dis-

positions in the population of taxpayers. The current review, therefore, coded the population of taxpayers on which field experiments were performed, not only in terms of their occupational context, but also in terms of their compliance history (i.e., their past taxpaying behaviour) as a potential indicator for disposition towards the authority. This should provide further insight and possible explanations for the above-mentioned inconclusive findings, taking into account general attitudes and taxes in various contexts. The next section focuses on social and institutional environments, in which field experiments were performed, to outline why it is important to also take environmental circumstances into account in the context of tax compliance decision processes.

1.6 Regional Variation

An OECD report from 2013 examined factors that influence tax morale in different geographic regions. Similarly to Batrancea et al. (2019; see Section 1.3), they could link trust in governments and institutions to higher voluntary tax compliance (i.e., tax morale), which was found to be relatively high in Western Europe, North America, Latin America, and Australia. In contrast, Eastern Europe, Africa, and Asia exhibited relatively low trust in governments and subsequently also lower tax morale. Further, they could show that satisfaction with public goods and services correlates with trust in governments, both of which are factors that were assumed to positively affect tax morale (OECD, 2013). Considering that trust seems to be a driving factor for tax compliance especially in less developed countries (Batrancea et al., 2019), improving the quality of public goods and services, as well as enhancing taxpayers' awareness about the use of their taxes through interventions might have a higher positive impact on tax compliance in these countries than, for example, in wealthier countries.

Countries with higher levels of taxation as a percentage of GDP (e.g., OECD countries) appear to have higher levels of tax morale in general. A report from the OECD (2019) argued that this might be indicative of a reinforcing cycle, which arises from higher tax revenue and propels the provision of public goods and services, subsequently enhancing voluntary tax compliance. These arguments could explain some variation in the evidence from field experiments examining tax compliance in different countries that would benefit from further investigation.

With differences in the provision of public goods and related trust in governments between countries being one factor that potentially explains regional variation in tax compliance, laboratory experiments with students across countries could also demonstrate significant differences in compliance behaviour in similar regional settings. Studies, for example, com-

18

pared students from the US and Spain (Alm et al., 1995), as well as Botswana and South Africa (Cummings et al., 2009), and observed their behaviour in tax games. They found that even though the setting was similar, students showed differences in their levels of tax compliance, which they attributed to different social norms and orientations towards compliance between the countries. It was also found that greater diversity in the taxpaying population, especially due to heterogeneity by race or religious membership, reduced tax compliance (Alm et al., 2016). This could explain potential differences between economically similar countries with varying degrees of heterogeneity in their population.

Evidence listed in this section adds to the literature previously discussed in terms of deviations from expected utility as well as taxpayer heterogeneity, thereby further illustrating the importance of factoring in situational, dispositional, social, and environmental aspects when interpreting the effectiveness of non-deterrence interventions and policies. This is something researchers are increasingly aware of, as has been discussed in previous reviews (e.g., Alm, 2019; Hallsworth, 2018; Slemrod, 2019), however, comprehensive evidence is still lacking. Unlike previous reviews and meta-analyses, the current systematic review, therefore, in addition to context-dependent deviations from expected utility, incorporates the population on which the intervention was performed and the region where the field experiment was conducted into the analysis. This allows for standardizing the findings and presenting a more comprehensive picture on 'what motivates tax compliance', leading to further explanations on previously inconclusive findings for non-deterrence interventions. Finally, it should be highlighted why the focus of this review lies on field experiments, as well as described how most field experiments centred on tax compliance were implemented.

1.7 The Use of Field Experiments

Field experiments increasingly overcome problems of translating research into practice (Grimshaw et al., 2013). The lack of generalizability, especially in laboratory experiments in the context of tax compliance, was briefly discussed at the beginning. Even though in laboratory experiments various changes can be introduced to measure how these changes subsequently affect tax compliance (e.g., varying audit probabilities or the size of fines), the environment in which they are performed differ from the real world. For example, taxpayers usually do not have very detailed information on actual audit probabilities in the real world (Scholz & Pinney, 1995). Additionally, participants in laboratory experiments, usually tax games, neither use their own money (unless incentivized) nor do they face severe consequences when they decide to evade taxes or cheat in the game. This leads to risk-taking being

a more entertaining option than compliance (Webley et al., 1991). However, there is another difference from real-world environments. In a laboratory setting, participants are observed by experimenters, whereas tax evasion in a natural setting usually takes place 'out of sight'. Therefore, participants may comply more often due to the effects of social desirability (Webley et al., 1991). Either way, a bias towards compliance or non-compliance may be created during laboratory experiments (Kirchler et al., 2010).

In recent years, research on tax compliance, therefore, shifted towards conducting field experiments with a population of real taxpayers to avoid some of the issues concerning laboratory experiments. Field experiments typically adopt an approach often found in behavioural sciences, where real taxpayers are randomly allocated to receive or not to receive a message (e.g., a letter or electronic notification) that aims to positively influence their tax compliance. This approach relies on the idea of 'nudging' taxpayers to improve their individual behaviour. In the context of this work, 'nudges' are defined as interventions that attempt to influence individuals' choices or behaviours in a predictable way while at the same time respecting their freedom of choice (Benartzi et al., 2017), by taking advantage of cognitive biases in individual decision-making (e.g., highlighting information to reduce cognitive effort due to limited resources; Hansen, 2016). Furthermore, nudges are seen as relatively cost-effective (Thaler & Sunstein, 2008) compared to, for example, increasing audit rates to improve tax compliance. As such, potential payoffs can be relatively high (e.g., Hallsworth et al., 2017), making it an attractive tool for authorities to use. Therefore, researchers increasingly collaborate with tax authorities to send out messages containing some form of policy-relevant information aimed to nudge taxpayers to increase their compliance (e.g., "nine out of ten people pay their tax on time", "your taxes contribute to the provision of public goods"). To examine, whether the intervention was effective, the treatment group compliance is then compared with control group compliance, whereas the control group either receives no message or a message that does not contain a nudging element.

At this point, it should be mentioned that field experiments also have their drawbacks. External validity may be limited in field experiments as well because interventions are typically carried out in a particular setting, raising doubts as to whether the same intervention would work in a different setting as well (Cartwright & Hardie, 2013). Furthermore, the timing when the intervention was performed also matters, as the intervention might have a different impact if the message is sent out at the beginning of a tax year compared to the end of a tax year (Burtless, 1995). This could be concerning, given that the behavioural science approach is very context-dependent in general (Kahneman & Tversky, 1984). Finally, field ex-

periments might also fail to capture spill-over effects (i.e., changes of one behaviour that spill over into another behaviour) or unintended consequences elicited by the intervention (Hallsworth, 2018), for example, taxpayers may find ways to balance out losses incurred by complying (e.g., Carrillo et al., 2017).

Still, most field experiments so far adopted the message approach (for a review, see Antinyan & Asatryan, 2020). The process of randomization allows for attributing changes in compliance mostly to receiving the message containing the intervention, due to matching levels of baseline tax compliance in their samples (Hallsworth, 2018). This method provides insights into tax compliance, beneficial for both researchers and tax authorities. Hallsworth (2014) concluded that "a good natural field experiment will allow policy-makers to evaluate the effectiveness and cost-effectiveness of specific interventions" (p.17). However, both Hallsworth (2014) and Muehlbacher and Kirchler (2016) noted that, due to a lack of systematic research comparing tax compliance results in field settings, researchers are still drawing conclusions from isolated studies, making firm assumptions difficult, which leads to the objective of this thesis.

1.8 The Present Study

The current systematic review aimed to identify how different non-deterrence nudging interventions affect tax compliance. At the same time, a colleague investigated the effects of deterrence nudging interventions on tax compliance. Therefore, the data from this systematic review was divided into two separate analyses. The interventions were assigned to different deterrence and non-deterrence sub-categories to be able to standardize the heterogeneous approaches used in the underlying field experiments. Additionally, the population of taxpayers was considered in terms of their occupation and compliance history, as well as the region in which the field experiment was conducted to present the factors influencing tax compliance in a more standardized and comprehensive way. This allowed for tackling some of the above-mentioned issues in terms of isolated and inconclusive evidence regarding the effectiveness of deterrence and non-deterrence interventions and policies used in the field (e.g., Alm, 2019; Hallsworth, 2018; Slemrod, 2019). Further, it sets this review apart from previous reviews and meta-analyses (e.g., Antinyan & Asatryan, 2020) that focused on analysing what interventions work while putting little emphasis on the context in which those interventions were implemented. Consequently, this work tries to answer why some non-deterrence nudging interventions work on tax compliance while others do not. Additionally, the review investigates

how certain contextual factors (e.g., individual preferences and regional variation) influence decisions regarding tax compliance.

Methods

2.1 Design

This systematic review followed the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement (PRISMA; Moher et al., 2009) as closely as possible within the current framework. As effect estimates were not analysed in this study, the presentation of results deviates from PRISMA guidelines (e.g., no forest plot was included). Additionally, risks of bias across studies were not coded. The PRISMA flow diagram was used to display the search strategy through different phases of the systematic review (see Moher et al., 2009). The method section of this thesis initially describes the procedure for the whole dataset of field experiments that used deterrence and / or non-deterrence interventions, and subsequently focuses on the characteristics of field experiments that involve non-deterrence interventions.

2.2 Search Strategy

The studies included in the review were accumulated through six different sources and were screened based on pre-defined selection criteria, which will be explained in detail in the section below. The following databases were selected to identify important research: PsycInfo, Web of Science, PSYNDEX, and Scopus. Additionally, studies were identified based on references of already existing research in the context of tax compliance, including previous systematic reviews, meta-analyses, and experiments conducted in this field. Articles collected by an independent reviewer in a previously done literature search via Google Scholar in an early stage of this project were also assessed. There was no set limit in the search regarding the publication period, nor was there a condition concerning the impact factor of studies of interest. Furthermore, due to the relatively novel approach of sending out nudging interventions to real taxpayers, a limited pool of available studies was expected. Therefore, further restrictions based on the above-mentioned criteria were not feasible. The search terms (i.e., keywords) used were divided into terms regarding the methodology (i.e., 'field experiment', 'letter', 'reminder'), the context ('tax compliance', 'tax evasion'), and the interventions used ('deterrence', 'audit', 'detection', 'penalty', 'threat', 'non-deterrence', 'social norm', 'moral appeal', 'persuasion', 'public goods', 'simplification', 'information', 'reward'

and 'framing'), and the same search terms were used for all four databases. The keywords were determined based on set methodological criteria, existing literature in the context of tax compliance (e.g., Antinyan & Asatryan, 2020; Hallsworth, 2014), and forms of nudging interventions used in different contexts (e.g., Carlsson et al., 2018). For the literature search, keywords concerning the methodology and the context were combined with keywords concerning the type of intervention used to narrow down the pool of relevant studies. The literature search was concluded on August 12, 2020, meaning that studies published since then, with potential relevance for this systematic review, were not included.

2.3 Selection Criteria

The following criteria had to be met for a study to be included in this systematic review: i) the study was conducted in a natural environment (i.e., a field experiment), where taxpayers (i.e., individuals, self-employed, or companies) were making real decisions regarding their taxpaying behaviour; ii) the interventions used in the study were nudging interventions following the definition in Section 1.7; iii) the interventions could be classified as using deterrence or non-deterrence elements, or both as described in Sections 1.1 and 1.4 (which will be further detailed in Section 3.3.3); iv) the outcome variable of interest was measured as a direct effect on the taxpaying behaviour of the target population (i.e., no mediator or moderator effects, spill-over effects, or network effects were of interest); v) the outcomes reported included significance levels which allowed for categorizing the effects of the studies into negative, zero, and positive effects (see Section 3.3.2).

2.4 Study Selection

The literature search and study selection were conducted by two independent reviewers. The specified selection criteria were already applied during the literature search in the selected databases, that is, abstracts were already screened in the process of collating relevant studies and articles not meeting the specified criteria were not further assessed. However, this process was not documented in detail, meaning that there is no record of all articles identified through database search before the screening, as should be the case according to the PRISMA flow diagram (Moher et al., 2009). Therefore, the PRISMA flow diagram was slightly adapted for it to match the selection process in this study, which can be observed in Figure 1.

2.5 Data Extraction

After the assessment and reasonable exclusions of articles, the following information was extracted and roughly summarized (see Data Extraction Sheet in Appendix E): i) title, authors, and year of publication; ii) key results described in the study; iii) message content (i.e., deterrence, non-deterrence, or both); iv) characteristics of the intervention(s); v) short description of the outcome(s) measured; vi) type of tax (e.g., income tax, VAT, etc.); vii) target group (e.g., individuals, self-employed, companies) and sample size; viii) country of origin; ix) effectiveness of the intervention. This data extraction was a first attempt to assess similarities and heterogeneous information in the various field experiments. Based on this information, relevant criteria were determined, which were then used to synthesize the data in a coding table (see Section 2.6). In the process of creating the coding table, extracted information from the field studies was revisited and revised. As such, the information in the data extraction sheet does not entirely match the information categorized in the coding table. The coding table contains updated information and should, therefore, be referenced to.

2.6 Data Synthesis and Coding Table

The coding table was created to standardize information extracted from the included field studies (see Coding Table in Appendix E). The following main categories were determined based on the data extraction sheet: i) study outcome; ii) type of intervention used in the study; iii) type of nudge element used in the intervention; iv) target population on which the intervention was performed; v) region in which the study was conducted; vi) type of tax that the outcome concerned; vii) effectiveness of the intervention; viii) sample size per intervention. Additionally, a main category was added that specified which type of control group the intervention was compared to (i.e., control groups that either received no letter or a letter containing no nudging element), to assess whether effects were less pronounced if there are already existing letter interventions implemented. For each of those main categories, various sub-categories were constructed, according to which extracted information was categorized by two independent raters. Characteristics of the sub-categories will be explained in detail in Section 3.3. During the process of creating the sub-categories, the study characteristics were differentiated carefully, with a focus on forming not too many categories that might contain only a few or individual nominations. Due to the heterogeneity of the data in the underlying field studies, it was important not to lose important information in the process of categorization. Therefore, a column was added with details about the intervention to the coding table to trace back how the interventions were defined in the respective field studies. Finally, another column was added to include additional information (e.g., details about the population or the

outcomes, modality with which an intervention was sent etc.) that might be relevant for the exploration of effects in the dataset. It should be noted, however, that the columns with additional information were not included in the analysis itself due to the heterogeneous nature of its content.
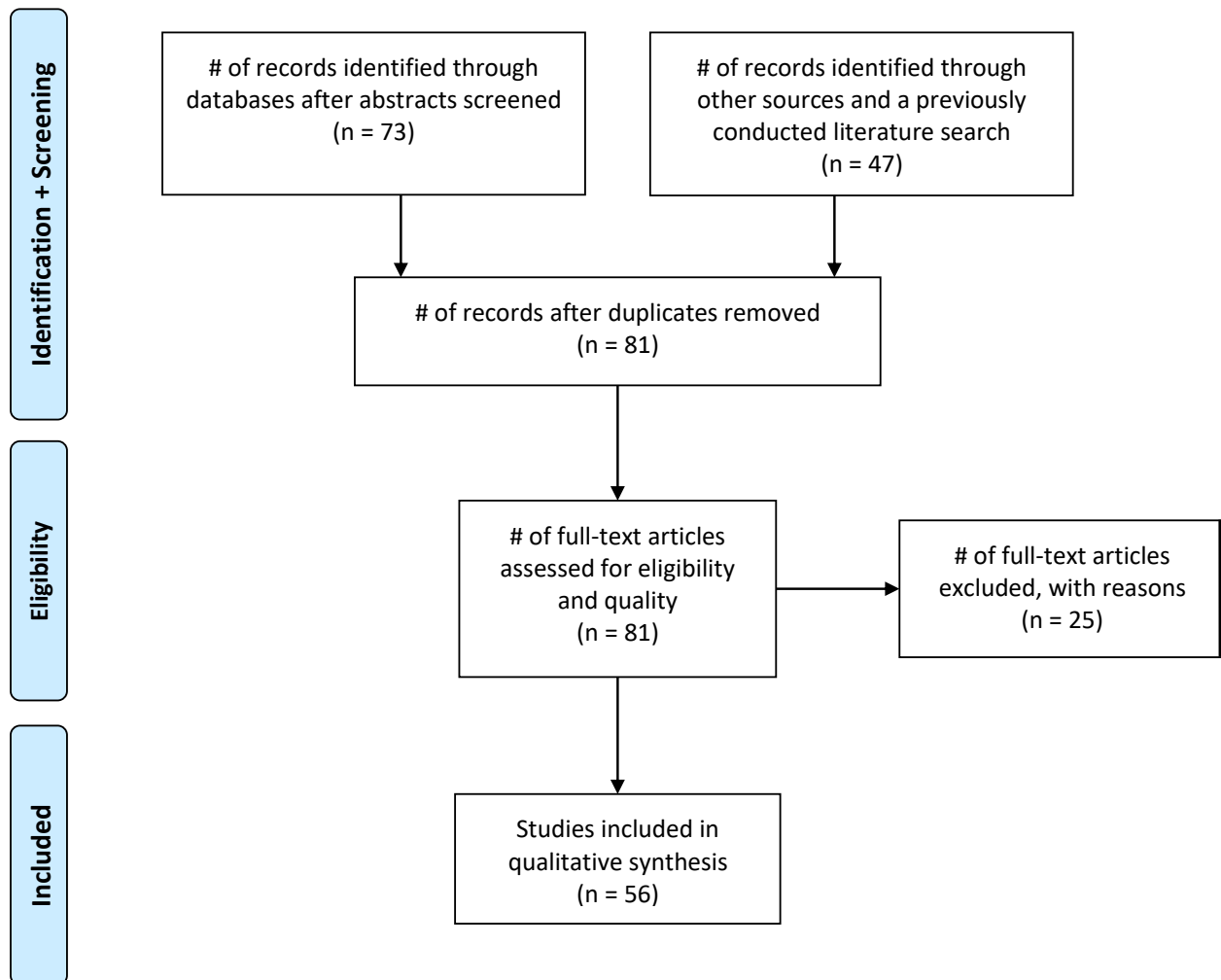
Figure 1. PRISMA flow diagram of the study selection process.

Results

3.1 Study Selection

After finishing the literature search in the selected databases, studies meeting the specified criteria were integrated based on the initial screening ($n = 73$) and articles identified through other sources and a previously conducted literature search, as described above ($n = 47$), removing duplicates ($n = 39$) in the process. The remaining full-text articles ($n = 81$) were thoroughly assessed and discussed by both reviewers with the support of an independent

third reviewer that supported this work. Various articles (*n* = 25) were excluded for the following reasons: i) the interventions used in the study were either not comparable to the majority of the studies; ii) the key driving factor for the effectiveness of the intervention was not the nudging element itself (e.g., training programmes involving forms of nudging); iii) the effect examined was not due to a direct effect, but rather a form of moderation or mediation; iv) the assessed study was a follow-up study examining network or spill-over-effects; v) the focus of the outcome was not changes in the actual taxpaying behaviour (e.g., demographical effects, effects in the context of ICT-usage). After the assessment and reasonable exclusions of articles, a sample of 56 studies remained (see Figure 1), which were then categorized into the coding table by the two raters. An overview of the sample of studies, including independent and dependent variables of the underlying field studies, can be found in Table 1.

3.2 Interrater Reliability

To ensure that the data was mapped appropriately within the predefined sub-categories, the interrater reliability was calculated for those main categories in the coding table that showed the greatest variation in the wording or the definition within the underlying studies, that is, categories that left scope for different interpretations between the raters. The main categories for which interrater reliability was calculated were: i) type of nudge element used in the intervention; ii) target population on which the intervention was performed; and iii) effectiveness of the intervention. Interrater reliability was calculated using Fleiss' kappa, as it was deemed the most suitable instrument for measuring interrater reliability when classifying items into categories (Fleiss & Cohen, 1973). Fleiss' kappa expresses the degree to which the observed proportion of agreement among both raters exceeds the degree that would be expected if the raters categorized randomly (Fleiss et al., 2004). A Fleiss' kappa of less than .40 can be considered a poor agreement beyond chance, whereas a kappa of more than .75 can be considered an excellent agreement beyond chance. Values between .40 and .75 represent a relatively good agreement beyond chance (Fleiss et al., 2004). The interrater reliability was calculated individually for each of the three main categories as well as across all three categories, based on the rating tables (see Rating Tables in Appendix E). The κ coefficient amounted to .692 for the main category *Type of Nudge Element*; .777 for the main category *Target Population*; and .925 for the main category *Effectivity* of the intervention. Overall, this resulted in an interrater reliability of κ = .825 across all three main categories, which can be considered an excellent agreement beyond chance between the two raters (Fleiss et al., 2004). The interrater reliability for the sub-categories can be found in Appendix D. Categorizations on which both raters dis-

agreed were further discussed with the help of an additional third rater. The final categorization was set following an agreement between all three raters. At this point, the categorization was further discussed with the supervisor of this thesis. This resulted in a rework on the definition of the *Company* and *Self-Employed* target population sub-categories, due to the fact that these two sub-categories were not defined precisely enough to differentiate between them. All raters agreed upon the suggested changes.

3.3 Characteristics of Included Studies and Categorization

As the data exploration for this thesis was focused solely on the effectiveness of non-deterrence interventions, the sample of 56 studies in the dataset was further reduced to a sample of 40 studies that include at least one non-deterrence intervention according to the categorization. Therefore, from this point onwards, only the non-deterrence subsample is concerned. All included studies are so-called 'letter studies', that is, each of the included studies contained a written text delivered in the form of a letter, an e-mail, an SMS, or was handed out in person to real taxpayers.

3.3.1 Outcome measures

Based on the included studies, four different main outcome categories were created for the coding table to capture the effects of the interventions on tax compliance. The outcome categories and the frequency of their observation will be detailed in this section. It is important to note that more than one outcome could be observed within the same study (e.g., a study that included extensive and intensive margin outcomes).

3.3.1.1 Extensive Margin. The extensive margin of tax compliance was defined as the probability to pay, file, or report taxes within a given period of time. The probability was based on the binary decision of the taxpayers to pay or not to pay (or file, or report) taxes after receiving the intervention in the respective study. In the non-deterrence sample, this outcome was observed 92 times in 21 studies.

3.3.1.2 Timely Payment. The outcome for extensive margin was subdivided into another outcome that measured the probability of paying (or filing, or reporting) taxes within a certain timeframe of fewer than two months after receiving the intervention according to the field experiments. This outcome was intended to assess whether the intervention has a stronger impact on the likelihood of paying taxes shortly after the receipt. The outcome was coded 54 times in 10 studies.

Table 1

Overview of Studies containing Non-Deterrence Nudging Elements in this Systematic Review

Author(s), Year of Publication	Country	Participants	Non-deterrence Intervention(s)	Dependent Variable(s) of Interest
Alm, Cifuentes, Niño, and Rocha (2019)	Colombia	3,606 individual social protection contribution taxpayers	Individual benefit message, Group benefit message	Reported health care payments, Reported pension payments
Ariel (2012)	Israel	4,935 corporations	Public goods message	Gross sales values reported, VAT payments, VAT deductions
Bérgolo, Ceni, Cruces, Giaccobasso, and Perez-Truglia (2017)	Uruguay	16,392 small- and medium-sized firms (primary sample)	Public goods message	Perceived audit probability, Taxes paid based on administrative data
Bhargava and Manoli (2015)	USA	35,050 respondent tax filers	Complex notice, Complex worksheet, Benefit display notice, Information about transactional costs, Indemnification information, Informational flyer, Envelope message, Personal stigma reduction message, Social stigma reduction message, Simple notice	Tax credit claims
Biddle, Fels, and Sinning (2018)	Australia	2,938 businesses	Changing due date treatment, Social norm letter, Changing colours treatment, Information about donation treatment	Rate of tax payments, Amount of payments of liabilities
Blumenthal, Christian, and Slemrod (2001)	USA	60,061 individual taxpayers	Public goods message, Social norm message	Change in federal taxable income reported and Minnesota tax liability paid
Bott, Cappelen, Sørensen, and Tungodden (2019)	Norway	15,708 individual taxpayers	Fairness treatment, Social benefits treatment, Simple letter treatment	Amount of self-reported foreign income
Boyer, Dwenger, and Rincke (2016)	Germany	39,788 church taxpayers	Compulsory tax payment letter, Voluntary tax payment letter	Extensive margin outcome, Intensive margin outcome
Castro and Scartascini (2015)	Argentina	23,195 individual property owners	Public goods message, Social norms message	Rate of property tax payment
Chirico, Inman, Loeffler, MacDonald, and Sieg (2016)	USA	4,927 individual property owners	Public service appeal, Civic duty appeal	Rate of tax payment, Amount of revenue collection
Chirico, Inman, Loeffler, MacDonald, and Sieg (2019)	USA	19,039 individual property owners	Public goods provided in neighbourhood, Public goods provided city-wide, Appeal to peer behaviour, Appeal to civic duty, Standard reminder	Tax payment rate, Amount of taxes paid
Coleman (1996)	USA	47,000 individual taxpayers	Information about enhanced customer service, Public goods message, Social norm message	Changes in reported federal taxable income, Taxes paid
Cranor, Goldin, Homonoff, and Moore (2020)	USA	90,349 individual taxpayers	Social norm 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)	Simplification treatment, Social norm treatment, Public goods (positive and negative framing) treatment, Social norm and public goods treatment, Pop-up pie chart treatment	Extensive margin and intensive margin outcomes
Del Carpio (2014)	Peru	22,318 individual property owners	Social norms message, Simple reminder	Rate of property tax payment
Dunning et al. (2015)	Uruguay	28,646 individual taxpayers	Individual and social reward treatments, Standard reminder	Intended compliance measured through access of a web account, Actual tax payments
Dwenger, Kleven, Rasul, and Rincke (2016)	Germany	39,782 individual church taxpayers	Reducing misperception of audit, Social and private rewards	Rate of church tax payment vs. amount of tax due
Eerola, Kosonen, Kotakorpi, Lyytikäinen, and Tuimala (2019)	Finland	45,000 individual property owners	Letter providing simplified information, Standard reminder	Rate of property tax payments, Amount of reported property taxes
Fellner, Sausgruber, and Traxler (2013)	Austria	50,498 individual TV license fee payers	Moral appeal letter, Social norm letter, Simple letter	GIS registrations or contract updates
Gillitzer and Sinning (2020)	Australia	4,787 businesses	Standard reminder letter	Tax payments made within 7 weeks of due date
Hallsworth, List, Metcalfe, and Vlaev (2017)	UK	101,471 individual taxpayers (Experiment 1); 119,527 individual taxpayers (Experiment 2)	Various descriptive and injunctive social norm treatments, Gain-framed and loss-framed public service treatments, Appeal to moral duty treatment, Additional information treatment	Rate of tax payments
Hasseldine, Hite, James, and Toumi (2007)	UK	7,307 sole proprietors	Enabling letter, Public goods letter	Reported turnover, Change in net profit
Hernandez, Jamison, Korczyc, Mazar, and Sormani (2017)	Poland	149,925 individual taxpayers	Gain-framed and loss-framed public goods treatment, Social norm treatment, Omission treatments, Simplification treatment, Standard letter treatment	Rate of tax payments, Amount of taxes paid
Hernandez, Karver, Negre, and Perng (2019)	Kosovo	11,603 individual taxpayers; 23,622 firms in June 2018; 22,828 firms in August 2018	Standard letter reminder, Standard e-mail reminder, Public goods e-mail reminder, Standard SMS reminder, Social norm SMS reminder	Rate of tax declaration, Amount of taxes paid
Hiscox et al. (2018)	Australia	1,279 existing businesses (Experiment 1); 328 new businesses (Experiment 2)	Standard reminder	Tax compliance rate within the DGST scheme, Tax payments
John and Blume (2018)	UK	11,880 individual taxpayers (Experiment 1); 56,568 individual taxpayers (Experiment 2)	Simplification treatment, Social norm treatment	Full tax payment or partial tax payment

Kettle, Hernandez, Sanders, Hauser, and Ruda (2017)	Guatemala	627,242 individual taxpayers	Honesty declaration treatment, Public goods treatment, Public goods and active choice treatment, Active choice treatment	Amount of online tax declaration
Kettle, Hernandez, Ruda, and Sanders (2016)	Guatemala	43,387 individual taxpayers	Standard reminder, National pride treatment, Social norm treatment	Tax declaration rate, Log amount conditional on payment, Amount unconditional on payment
Koessler, Torgler, Feld, and Frey (2019)	Switzerland	2,201 individual taxpayers	Financial and non-financial rewards	Rate of tax payments
Larkin, Sanders, Andresen, and Algate (2019)	UK	9,130 individual taxpayers	Social norm treatment	Rate of tax payments, Amount of taxes paid
Mascagni, Nell, and Monkam (2018)	Rwanda	10,800 business and self-employed taxpayers	Public goods and Standard reminders sent via letter, E-Mail, or SMS	Rate of reduced income tax liabilities
Meiselman (2018)	USA	9,523 individual taxpayers	Compliance cost treatment, Civic pride treatment, Simple letter	Rate of remitted tax payments
Ortega and Sanguinetti (2013)	Venezuela	6,000 firms	General public goods message, Public health services message, Moral duty message, Simple letter	Difference-in-difference analysis of tax balance between accruements and payments
Perez-Truglia and Troiano (2018)	USA	34,334 individual taxpayers	Shaming treatments (higher and lower visibility to neighbouring households), Peer information treatment	Probability of leaving the tax delinquent list, Debt compared to amount of initial debt before treatment
Pomeranz (2015)	Chile	102,000 small and medium-sized firms receiving a letter	Tax morale letter, Simple letter	Monthly VAT payments
Sanders, Reckers, and Iyer (2008)	USA	277 use tax paying firms; 1,278 B&O tax paying firms	Signing affidavit that no tax was due	Reported use and B&O tax base
Schwartz and Orleans (1967)	USA	173 individual taxpayers	Appeal to conscience questions, Placebo questions	Changes in levels of reported income, tax deductions, income tax after credits
Shimeles, Gurara, and Woldeyes (2017)	Ethiopia	3,730 businesses	Appeal to tax morale letter	Profit tax payment
Torgler (2004)	Australia	580 individual taxpayers	Appeal to civic duty	Timely filing of tax return, Timely payment of tax liability
Wenzel (2005)	Australia	1,500 individual taxpayers	Survey only, Survey and social norms feedback letter	Deductions claimed for work-related expenses, deductions claimed for other expenses

Note. Column 2 shows the country, where the field experiment was conducted. Column 3 shows the sample of taxpayers as denoted in the underlying field experiment. Columns 4 and 5 show the interventions and dependent variables as denoted in the underlying field experiment.

3.3.1.3 Intensive Margin. The intensive margin of tax compliance coded in this review concerned the amount of taxes or income reported or paid, that is, it measured whether receiving the intervention led to a higher or lower amount of taxes paid or income reported within a given period of time. The amount was usually estimated either in absolute values in the national currency of the country where the study took place or on a logarithmic scale. As this systematic review focused entirely on capturing whether the intervention had a significant effect on the respective outcome, it was not differentiated between absolute and logarithmic values. This outcome was coded 127 times in 25 studies and accounted for almost half of all categorized outcomes.

3.3.1.4 Tax Deductions / Claims. In addition to the above-mentioned extensive and intensive margin outcomes, five studies examined whether the intervention affected deductions from gross revenues as well as credit claims available through different tax systems. In terms of categorization, it was decided to code the intervention as a *Positive Effect* if the intervention had a significant positive effect on tax deductions or credit claims. However, it should be noted that a significant positive effect on tax deductions or credit claims after receiving the intervention could be interpreted as a negative effect for the tax authority, as tax money is flowing back to taxpayers. The outcome was observed 22 times in 5 studies.

3.3.2 Effectivity

The main category *Effectivity* was considered the dependent variable in this systematic review and contained the sub-categories *Negative Effect*, *Zero Effect*, and *Positive Effect*. An intervention was categorized as a *Negative Effect* if the intervention had a significant backfiring or negative effect on the respective outcome in the study ($p < .05$), that is, if the intervention led to a reduction in tax compliance or tax deductions as measured by the outcome. Accordingly, an intervention was categorized as a *Positive Effect* if it has a significant positive effect on the respective outcome ($p < .05$), that is, if the intervention led to an increase in tax compliance or tax deductions. Finally, interventions were categorized as *Zero Effect* if no significant effect of the intervention was found in the underlying study. Likewise, significant effects found at $p > .05$ to $p < .10$ were also coded as *Zero Effect*, primarily to account for a reasonable significance level within a psychological context, as well as to provide a higher measure of confidence in the analysis.

3.3.3 Nudging Interventions

The interventions used in the sample of studies were categorized either into letters including deterrence or letters including non-deterrence nudges. As mentioned above, the current thesis focused on the subsample of interventions using non-deterrence nudging elements (e.g., social norms or rewards). The non-deterrence sub-categories in the coding table relevant for this analysis will be described in detail below.

3.3.3.1 Information. An intervention was coded as an *Information* nudge if the intervention provided some form of information about the tax system intended to encourage tax compliance, for example, informing about the availability of a customer service hotline (Coleman, 1996) or offering some form of advice (Hasseldine et al., 2007). Overall, 14 information nudges were observed in 6 studies.

3.3.3.2 Moral Appeal. Interventions, whose wording appealed to taxpayers' honesty or morality, were categorized as *Moral Appeal* (e.g., "For democracy to work, all citizens need to pay their fair share of taxes for community services"; Chirico et al., 2019). Moral appeal interventions were observed 36 times in 12 studies.

3.3.3.3 Public Goods. Interventions categorized as *Public Goods* informed taxpayers about how their tax money was spent on, for example, communal services (e.g., "Your income tax dollars are spent on services that we Minnesotans depend on"; Blumenthal et al., 2001). Some interventions also included detailed information about how tax money was distributed into different sectors (e.g., Chirico et al., 2016). Overall, 50 observations were coded in 16 studies, showing that many studies included some form of information about how tax money is spent in their interventions.

3.3.3.4 Reward. An intervention was coded as *Reward* if the letter included either a monetary or material incentive, or reward (e.g., "If you pay on time, you will be automatically entered in a lottery to win a year free of property tax payments"; Dunning et al., 2015). It is important to note that, although the outcome was observed relatively frequently in the sample (i.e., 44 times), interventions including an incentive or reward were only used in 5 studies. The implications of this will be discussed in detail in Section 3.4.

3.3.3.5 Social Norm. Interventions were assigned to the *Social Norm* category if the wording in the intervention used some form of social norm, for example, "nine out of ten people pay their tax on time" (Hallsworth et al., 2017) or "over 95% of Lambeth residents pay their council tax" (John & Blume, 2018). Social norm interventions were most frequently observed in the sample of studies and were coded 61 times in 19 studies.

3.3.3.6 Simplification. Contrary to previously described interventions, those categorized as *Simplification* did not contain a specific wording. Rather, the nudging elements for a simplification intervention were, for example, a simpler presentation of information or highlighting important information through visual stimuli (e.g., De Neve et al., 2019). Simplification interventions were coded 20 times in 8 studies.

3.3.3.7 Interaction. The sub-category *Interaction* was introduced to account for interventions that included more than one nudging element in the same intervention. For example, letters in the study by Shimeles et al. (2017) incorporated appeals to patriotic duties, as well as listing projects financed by taxpayers' money, compliance-based rewards, and incentives. Interaction interventions with more than one nudging element were observed 4 times in 3 studies.

3.3.3.8 Simple Letter / Simple Reminder. The sub-category *Simple Letter* contained interventions not including a nudging element as defined above, that is, letters sent out to taxpayers to inform them about their tax obligations. Additionally, reminders that informed taxpayers about outstanding tax liabilities without using nudging elements were coded separately in the sub-category *Simple Reminder* (e.g., "We want to remind you that the second payment of property taxes is due in July"; De Neve et al., 2019). In most studies, simple letter and reminder interventions were either compared to a control group that did not receive an intervention or were used as a control group for letters containing a nudging element. Nudging interventions that were not clearly defined or could not be categorized into one of the above-mentioned sub-categories were categorized as *Other* (i.e., 25 interventions in 5 studies).

3.3.4 Population

The main category *Population* contains several sub-categories that allowed for comparisons between different populations of taxpayers in terms of compliance and occupational context. Coded within the sub-category *Company* were interventions that were conducted on smaller or larger companies or businesses with multiple employees. The sub-category *Self-Employed*, in contrast, included sole proprietors (i.e., unincorporated businesses with a single owner) and micro-enterprises with a small number of employees – a rough cut-off used was a number of employees of less than five. When categorizing the studies, however, a clear allocation to one of the two categories was sometimes difficult to agree on. This was also due to the fact that the samples of some studies included a certain number of self-employed as well as corporate taxpayers (e.g., Biddle et al., 2018; Mascagni et al., 2018). Therefore, it was decided to assign samples with a predominant share of one population (> 70% of, for example,

self-employed taxpayers) to the respective sub-category. Additionally, if a clear allocation was still not possible, it was considered how the underlying study defined its sample, as for example, a 'small' company might be perceived differently regarding its size in varying national contexts. Finally, the third sub-category *Individuals* comprised the population of taxpayers who paid their taxes outside of an occupational context (e.g., personal income taxpayers or property taxpayers). In order to compare populations in terms of their compliance history, a non-compliance sub-category was added for every population sub-category (e.g., *Individual / Non-Compliant*). A sample was categorized as *Non-Compliant* if it was explicitly classified as 'at risk', if it had already missed payments in the past or if it declared income but did not pay taxes according to the underlying study. If there was no explicit mention that the sample was previously non-compliant, it was not known whether some taxpayers had a history of non-compliance. Therefore, it was referred to as a 'general population' of taxpayers that was supposed to include previously compliant and non-compliant taxpayers (see *Mixed* sample in the data analysis; see Section 3.4.2).

3.3.5 Region

Additional to coding the samples into population categories, a main category was created to differentiate between various regional contexts in which the studies were conducted. The sub-categories *Africa*, *Australia / New Zealand*, *Middle East,* and *North America* included studies that were carried out in countries in the respective continental regions. The sub-category *South America* included studies conducted in South as well as Central American countries. As most studies of this sample were implemented in European countries, Europe was further divided into three different sub-categories to account for potential regional differences within as, for example, trust in governments or individuals' support for democracy (see OECD, 2013). Field studies conducted in Scandinavian or Baltic countries were categorized as *Northern Europe*, whereas studies from Slavic countries and countries on the Balkan Peninsula were categorized as *Southern Europe*. Finally, the sub-category *Western Europe* included Central and Western European countries that did not fall into either of the aforementioned regional sub-categories.

3.3.6 Tax

Though no specific assumptions were made regarding different effects of interventions on different forms of taxes in the context of this thesis, it was decided to implement a main category to account for the different tax contexts for potential future work with the dataset.

The sub-categories accounted for the main tax contexts that could be clearly distinguished. Most studies were conducted in the context of income tax (19 out of 40), VAT (6 out of 40), and property tax (5 out of 40). Furthermore, separate sub-categories were created for Washington State Business & Occupation Tax as well as Use Tax, Norwegian Foreign Income Tax, German Church Tax, and Austrian TV license fee. Tax contexts that were not clearly defined, or could not be differentiated, were categorized as *Other*. A detailed description of each tax context can be found in Table 3 in Appendix B.

3.3.7 Comparison

As mentioned above, an extra category was added that specified which type of control group the intervention was compared to, that is, whether the control group was exposed to an intervention containing no nudging element (i.e., *Simple Letter* or *Simple Reminder*) or not exposed to an intervention at all (i.e., *No Letter*). Furthermore, various studies compared different interventions containing nudging elements with each other. These comparisons were subsequently summarized in the sub-category *Other*. Comparisons between interventions that contained nudging elements were not of interest for the current review and will not be discussed in detail in the empirical analysis.

3.4 Empirical Analysis

In the following section, the findings of the systematic review will be presented. The section will contain various frequency distributions of the underlying dataset, displaying which non-deterrence nudging interventions have most frequently shown negative, zero, or positive effects in various sub-categories, as for example, different populations of taxpayers or different regions. The analysis aimed at providing indicators about why some nudging interventions work in some contexts but not in others.

3.4.1 Overall Frequencies

As a first step, the frequency distribution of negative, zero, and positive effects within each sub-category was displayed in order to gain some insight into which sub-categories were coded relatively frequently compared to others (see Table 2). For example, interventions using some form of social norm were observed most frequently, being coded 61 times in 19 field experiments. Additionally, Table 2 provided some initial indicators on which sub-categories should be examined more closely in terms of their frequency distribution.

Table 2

Number of Studies and Frequencies of Effects per Sub-Category

	Nr. of Studies	Negative Effect (N=25)	No Effect (N=179)	Positive Effect (N=101)	Total (N=305)
Outcome					
Extensive Margin	21	11 (12.0%)	47 (51.1%)	34 (37.0%)	92 (100.0%)
Intensive Margin	25	7 (5.5%)	90 (70.9%)	30 (23.6%)	127 (100.0%)
Other	4	0 (0.0%)	3 (30.0%)	7 (70.0%)	10 (100.0%)
Tax Deductions/Claims	5	5 (22.7%)	13 (59.1%)	4 (18.2%)	22 (100.0%)
Timely Payment	10	2 (3.7%)	26 (48.1%)	26 (48.1%)	54 (100.0%)
Nudge					
Information	6	1 (7.1%)	12 (85.7%)	1 (7.1%)	14 (100.0%)
Interaction	1	0 (0.0%)	0 (0.0%)	1 (100.0%)	1 (100.0%)
Moral Appeal	13	1 (2.7%)	23 (62.2%)	13 (35.1%)	37 (100.0%)
Other	5	4 (16.0%)	17 (68.0%)	4 (16.0%)	25 (100.0%)
Public Goods	16	6 (11.8%)	31 (60.8%)	14 (27.5%)	51 (100.0%)
Reward	5	5 (11.4%)	29 (65.9%)	10 (22.7%)	44 (100.0%)
Simple Letter	7	1 (7.1%)	8 (57.1%)	5 (35.7%)	14 (100.0%)
Simple Reminder	11	2 (5.3%)	22 (57.9%)	14 (36.8%)	38 (100.0%)
Simplification	8	0 (0.0%)	7 (35.0%)	13 (65.0%)	20 (100.0%)
Social Norm	19	5 (8.2%)	30 (49.2%)	26 (42.6%)	61 (100.0%)
Population					
Company	9	2 (4.9%)	27 (65.9%)	12 (29.3%)	41 (100.0%)
Company / Non-Compliant	2	0 (0.0%)	9 (64.3%)	5 (35.7%)	14 (100.0%)
Individual	13	3 (4.8%)	52 (82.5%)	8 (12.7%)	63 (100.0%)
Individual / Non-Compliant	16	19 (15.3%)	61 (49.2%)	44 (35.5%)	124 (100.0%)
Self-Employed	7	1 (3.2%)	20 (64.5%)	10 (32.3%)	31 (100.0%)
Self-Employed / Non-Compliant	3	0 (0.0%)	10 (31.2%)	22 (68.8%)	32 (100.0%)
Region					
Africa	2	0 (0.0%)	4 (44.4%)	5 (55.6%)	9 (100.0%)
Australia / New Zealand	4	0 (0.0%)	16 (88.9%)	2 (11.1%)	18 (100.0%)
Middle East	1	0 (0.0%)	1 (50.0%)	1 (50.0%)	2 (100.0%)
North America	10	5 (7.6%)	44 (66.7%)	17 (25.8%)	66 (100.0%)
Northern Europe	2	0 (0.0%)	8 (61.5%)	5 (38.5%)	13 (100.0%)
South America	9	2 (2.6%)	50 (65.8%)	24 (31.6%)	76 (100.0%)
South / Eastern Europe	2	8 (22.9%)	12 (34.3%)	15 (42.9%)	35 (100.0%)
Western Europe	10	10 (11.6%)	44 (51.2%)	32 (37.2%)	86 (100.0%)
Tax					
B&O Tax	1	0 (0.0%)	1 (100.0%)	0 (0.0%)	1 (100.0%)
Church Levy	2	5 (15.6%)	23 (71.9%)	4 (12.5%)	32 (100.0%)
Council Tax	2	1 (16.7%)	1 (16.7%)	4 (66.7%)	6 (100.0%)
Foreign Income Tax	1	0 (0.0%)	5 (55.6%)	4 (44.4%)	9 (100.0%)

Income Tax	19	15 (10.3%)	70 (48.3%)	60 (41.4%)	145 (100.0%)
Other	3	2 (6.7%)	17 (56.7%)	11 (36.7%)	30 (100.0%)
Property Tax	5	0 (0.0%)	33 (73.3%)	12 (26.7%)	45 (100.0%)
TV License Fee	1	2 (40.0%)	2 (40.0%)	1 (20.0%)	5 (100.0%)
Use Tax	1	0 (0.0%)	0 (0.0%)	1 (100.0%)	1 (100.0%)
VAT	8	0 (0.0%)	27 (87.1%)	4 (12.9%)	31 (100.0%)
Comparison					
No Letter	27	2 (1.3%)	98 (64.1%)	53 (34.6%)	153 (100.0%)
Other	5	15 (31.9%)	30 (63.8%)	2 (4.3%)	47 (100.0%)
Simple Letter	15	6 (9.0%)	43 (64.2%)	18 (26.9%)	67 (100.0%)
Simple Reminder	5	2 (5.3%)	8 (21.1%)	28 (73.7%)	38 (100.0%)
Sample Size					
<100	3	1 (11.1%)	3 (33.3%)	5 (55.6%)	9 (100.0%)
100-1000	12	2 (3.6%)	38 (69.1%)	15 (27.3%)	55 (100.0%)
1001-5000	18	5 (4.3%)	76 (65.5%)	35 (30.2%)	116 (100.0%)
5001-10000	10	7 (13.0%)	24 (44.4%)	23 (42.6%)	54 (100.0%)
10001-50000	9	10 (19.6%)	21 (41.2%)	20 (39.2%)	51 (100.0%)
>50000	3	0 (0.0%)	17 (85.0%)	3 (15.0%)	20 (100.0%)

Note. Column 2 shows the frequency of studies containing the sub-category.

In addition to overall frequencies of effects, column two displayed the number of field experiments per sub-category, which was already discussed for nudging interventions in the previous section. This provided an important overview when interpreting the distributions, as for example, some interventions, populations, or regions were coded relatively often, but within relatively few studies (e.g., *Rewards* or *Self-Employed / Non-Compliant*). This prevented overweighting the effectiveness of such interventions due to a lack of evidence from different contexts and should be kept in mind when reading the tables below.

When examining the distribution of the outcome sub-categories, it could be observed that more positive effects were found for extensive margin outcomes than for intensive margin outcomes. The sub-category *Timely Payment* in Table 2 also demonstrated that positive effects occurred more frequently within a period of fewer than two months after receiving the intervention.

In terms of nudging interventions, Table 2 showed that 65% of interventions using simplification elements displayed positive effects, while 85% of nudges containing an information element were ineffective across all other main categories (i.e., outcomes, populations, regions, taxes etc.). Furthermore, around 11% of interventions containing rewards or information about the provision of public goods and services triggered negative or backfiring effects.

In terms of the effectiveness of interventions within the population sub-categories, most of the categories had relatively similarly distributed frequencies, exceptions being the *Individual* sub-category as well as the *Self-Employed / Non-Compliant* sub-category. 83% of nudging interventions were ineffective in a population of individual taxpayers with no explicit history of non-compliance, whereas 69% were effective in a population of self-employed individuals or small business owners with a history of non-compliance. However, frequencies within the population of *Self-Employed / Non-Compliant* should be interpreted with caution as this population was coded less frequently (i.e., in three studies).

When examining the region sub-categories, it could be observed that most studies were conducted in North America (i.e., mainly in the US), South America, or Europe, all of which had relatively similar frequency distributions in terms of the effectiveness of nudging interventions. Few studies were conducted in African and Middle Eastern countries, and, unfortunately, no data from international studies was available for South or East Asian countries.

The *Comparison* sub-categories showed that most interventions containing nudging elements were compared to a control group that received no intervention. They were also shown to exhibit more frequent positive effects in this sub-category compared to a control group that received a letter containing no nudging element. This might not be surprising, as receiving a letter from the authority might have already influenced taxpayers' behaviour. As such, the included nudging element in the treatment group might not have affected changes in the behaviour much more strongly when the control group already received a letter themselves. Interestingly, however, the nudging interventions seem to be frequently effective when compared to a control group that received a reminder with no nudging element. The differences could potentially be attributed to the underlying study design, as only a small number of studies included a control group that received a reminder.

Table 2 also displayed frequency distributions for different sub-categories of taxes, as well as different sample sizes. These categories were added to provide additional information about the sample of field experiments for potential future work with the dataset. However, in the context of this thesis, no assumptions were made regarding the effectiveness of interventions, neither in different tax contexts nor due to different sample sizes. As such, frequency distributions in neither main category will be discussed in detail. Furthermore, the *Other* sub-categories contained various outcomes, interventions and taxes that did not occur frequently enough to form a separate sub-category. Thus, they were excluded from subsequent analyses as well.

3.4.2 Compliance vs. Mixed Sample

The main point of interest of this systematic review was to give a comprehensive overview of the effectiveness of various non-deterrence interventions within different populations of taxpayers across field studies. Therefore, frequency distributions of negative, zero, and positive effects were examined in a population of taxpayers with no stated history of previous non-compliance, which will be denoted as *Mixed* sample, and compared with a population of taxpayers that had a history of tax non-compliance in the past, denoted as the *Non-Compliant* sample. For this purpose, the population sub-categories *Individuals*, *Self-Employed*, and *Company* were recoded as either *Mixed* or *Non-Compliant* based on whether they were previously tagged as non-compliant. Additionally, various nudging intervention sub-categories were combined according to the underlying non-deterrence mechanism that they targeted (see Section 1.4) to present a condensed result. The sub-categories *Information*, *Simple Letter*, and *Simple Reminder* have in common that the interventions contain some form of information about the taxpaying process and were, therefore, combined. Furthermore, *Moral Appeals* and *Social Norms* were assumed to trigger psychological costs in the taxpayers, whereas *Rewards* and *Public Goods* aimed at a reciprocal mechanism. The interventions were grouped together respectively. Interventions coded as *Interaction* were excluded from the analysis, due to their heterogeneous nature making them difficult to interpret.

Figure 2 depicted the distribution showing how frequently the nudging interventions, combined in terms of the underlying mechanism they target, were effective, both within the mixed sample of taxpayers, as well as within the non-compliant sample of taxpayers. The plot demonstrated that non-deterrence interventions were more frequently effective on tax compliance in the non-compliant sample of taxpayers compared to the mixed sample across all outcomes. The figure also showed that psychological cost interventions like moral appeals or social norms were frequently ineffective in the mixed sample of taxpayers (19% of the interventions in a mixed sample were effective). However, compared to the mixed sample, non-deterrence interventions focusing on the mechanism of psychological costs showed effects on tax compliance more frequently in the non-compliant sample (48%). Interestingly, 82% of all simplification interventions in the non-compliant sample of taxpayers were effective. However, it should be kept in mind that the sample size for simplification interventions was relatively small. Finally, it was also noteworthy that the interventions rarely led to negative or backfiring effects in the mixed sample (3%), whereas negative or backfiring effects occurred more often in the non-compliant sample (11%). Especially interventions targeting the reciprocal mechanism seemed to elicit negative responses after receipt of the intervention.

Figure 2. Balloon plot showing the frequency distribution of combined non-deterrence interventions in the mixed and non-compliant sample. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

As mentioned above, the effectiveness of the interventions in Figure 2 was depicted across all outcomes. This is important to mention, as the same intervention might have been effective for more than one outcome, for example, both at the extensive and intensive margin. Thus, interventions that were coded for more than one outcome weigh-in more strongly than others, which could be checked in the coding table (see Appendix E). This should be kept in mind when interpreting subsequent distributions where the effects on different outcomes were not separately investigated.

Extensive Margin and Intensive Margin Outcomes. In a next step, it was investigated whether the interventions had a different impact on the two main types of outcomes within the mixed and non-compliant samples. Therefore, a closer look was taken at the subsamples where tax compliance was measured at the extensive as well as the intensive margin, which were displayed similarly in a balloon plot (see Figures 3 and 4). For this representation, it was not of relevance whether the effect was observed within two months or less of receiving the

intervention. As both sub-categories were defined as outcomes measured at the extensive margin (see Section 3.3), the outcome sub-category *Timely Payment* was recoded into *Extensive Margin*. Outcomes measuring the impact of the intervention on tax deductions or claims, as well as other compliance indicators, were not of further interest for the current thesis due to the limited amount of studies with such outcomes available and will be not presented in detail.

Figure 3 depicted the distribution that showed how frequently the nudging interventions were effective within both samples in terms of tax compliance measured at the extensive margin. Information nudges (65%), moral, and social nudges (52%), as well as simplification nudges (100%) all seemed to be frequently effective on the non-compliant sample of taxpayers, whereas only 24% of nudges targeting the reciprocal mechanism were effective. In the mixed sample, on the other hand, only 23% of all treatments showed positive effects, thus pointing towards a higher frequency of positive effects of letter treatments in the non-compliant sample for extensive margin outcomes.

Figure 3. Balloon plot showing the frequency distribution of combined non-deterrence interventions in the mixed and non-compliant sample measured at the extensive margin. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Figure 4 depicted the frequency distribution for the effectiveness of nudging interventions on tax compliance measured at the intensive margin in the two samples. The plot showed that the interventions were overall less likely to have a positive impact on increasing the amount of income reported or taxes paid (i.e., only 24% of all interventions showed positive effects), compared to 41 % for increasing the likelihood of reporting, filing, or paying taxes (i.e., extensive margin outcomes). In terms of interpretation, it should be taken into account that intensive margin outcomes subsumed a larger variety of outcomes focusing on the amount of income reported or taxes paid both in absolute values and on a logarithmic scale.

Figure 4. Balloon plot showing the frequency distribution of combined non-deterrence interventions in the mixed and non-compliant sample measured at the intensive margin. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

No Letter and Simple Letter Comparisons. Additional to taking a closer look at the different outcomes within the samples, it was also investigated whether the nudging interventions exhibited different patterns when compared to a control group that received no intervention or a control group that received an intervention without a nudging element.

Figure 5 depicted the frequency distribution for the effectiveness of nudging interventions when compared to a no-intervention control group in the mixed and non-compliant sam-

ples. The distribution showed that psychological cost interventions were most frequently ineffective in the mixed sample (92%) when compared to a no-letter control group. However, reciprocal interventions seemed to be frequently ineffective in the non-compliant sample (80%). It was also noteworthy that the interventions rarely had negative or backfiring effects when compared to a control group that received no intervention, with zero backfiring effects in the non-compliant sample. Receiving a message from the authority might therefore at least not have negatively affected tax compliance, compared to not receiving a message at all.
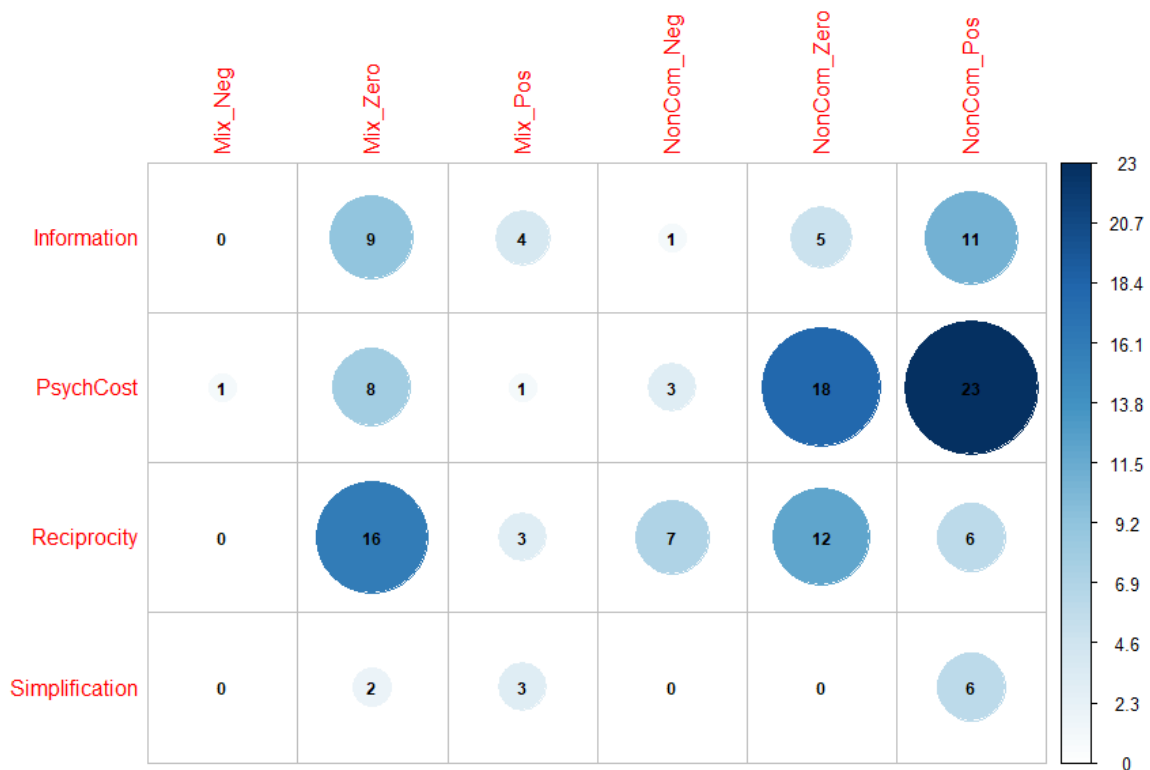

Figure 5. Balloon plot showing the frequency distribution of combined non-deterrence interventions in the mixed and non-compliant sample compared to a no-letter control group. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Figure 6 showed the distribution for the effectiveness of nudging interventions when compared to a control group that received a message with no nudging element. As *Simple Letter* and *Simple Reminder* control groups received a message with no nudging element, they were grouped together for this distribution. Interestingly, both psychological cost and reciprocal interventions showed a higher number of positive effects in the non-compliant sample than in the mixed sample, when compared to a control group that has also been exposed to a

message. Comparing to a control group that received a message did not seem to diminish the frequency of effects of the nudging treatments. The interventions still showed similarly frequent positive effects as when compared to a no-letter control group.
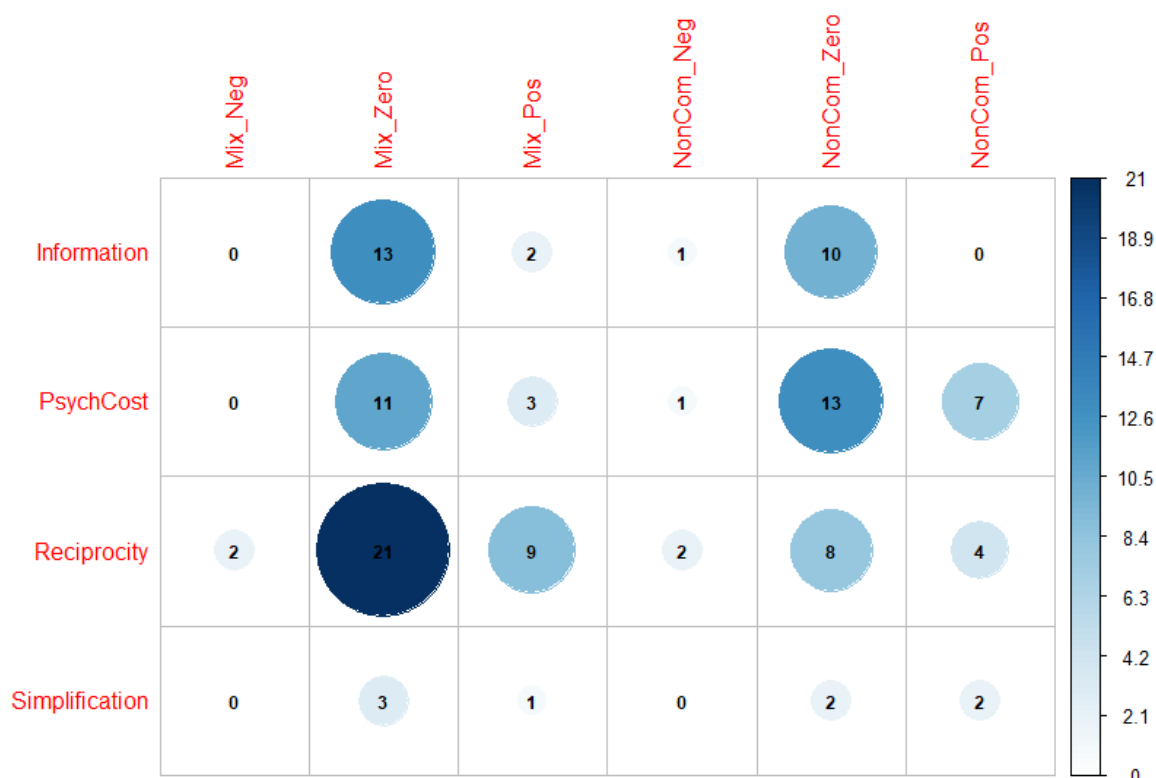

Figure 6. Balloon plot showing the frequency distribution of combined non-deterrence interventions in the mixed and non-compliant sample compared to a simple letter control group. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Types of Interventions in Detail. To explore the different types of nudging interventions more in detail, it was also investigated how frequently each of the separate types of interventions exhibited negative, zero, or positive effects in the mixed and non-compliant samples.

Figure 7 showed the frequency distribution for each individual type of nudging treatment. In this distribution, it could be observed that 74% of the positive effects of interventions inducing psychological costs in the non-compliant sample came from social norm interventions. Still, Table 7A showed that, when compared to a control group that received a message, 75% of moral appeal interventions in the non-compliant sample displayed positive effects. Furthermore, Figure 7 showed that information nudges were frequently ineffective, with just

one intervention displaying a positive effect. In terms of interventions targeting the reciprocal mechanism, the plot showed that 28% of the public goods interventions displayed positive effects compared to 23% of the reward interventions. Moreover, public goods interventions were more frequently effective in the non-compliant sample. Both types of interventions, however, appeared to be frequently ineffective in the mixed sample.

Figure 7. Balloon plot showing the frequency distribution of individual non-deterrence interventions in the mixed and non-compliant sample. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.

The effects on tax compliance measured at the extensive and intensive margins were also compared for each type of intervention (see Table 10A and 11A). An interesting finding was that 48% of the standard letters in the non-compliant sample (i.e., simple letters and reminders grouped together) had a positive effect on tax compliance when measured at the ex-

tensive margin. This further highlighted that receiving any kind of message from the authority might already be sufficient to affect taxpayers' behaviour. Additional frequency tables showing differences between the control groups for each type of intervention can be found in Appendix B (see Table 12A and 13A).

3.4.3 Individual vs. Self-Employed vs. Companies

Besides comparing the samples in terms of their compliance history, it was also of interest for this review to investigate differences between individuals, self-employed, and companies in terms of how nudging interventions affected their tax compliance. Therefore, the mixed and non-compliant samples of each population sub-category (i.e., *Individual*, *Self-Employed*, and *Company*) were combined and differences on an occupational level were investigated. Corresponding to the frequency distributions above, the different types of interventions were first consolidated to groups based on the underlying mechanism they target and thereafter investigated individually for each type of treatment.

Figure 8 showed how frequently the nudging interventions were effective in each population sub-category. It could be observed that information treatments were frequently ineffective in each sub-category. Psychological cost interventions exhibited the most positive effects within a population of self-employed individuals and micro-sized businesses. Simplification interventions, on the other hand, showed the most positive effects in a population of individual taxpayers. It should be noted that simplification interventions were so far almost exclusively tested in a sample of individual taxpayers. Reciprocal interventions, like public goods or rewards, showed frequent zero but also backfiring effects within an individual population, with only 16% of the treatments displaying positive effects in this population. In general, most backfiring effects occurred within the individual population of taxpayers. Finally, no group of intervention appeared to be particularly effective within the company sample.

When further investigating the population sub-categories for differences in extensive and intensive margin outcomes, it could be observed that both psychological cost and simplification interventions were more frequently effective when measured at the extensive, but not at the intensive margin (see Table 15A and 16A).

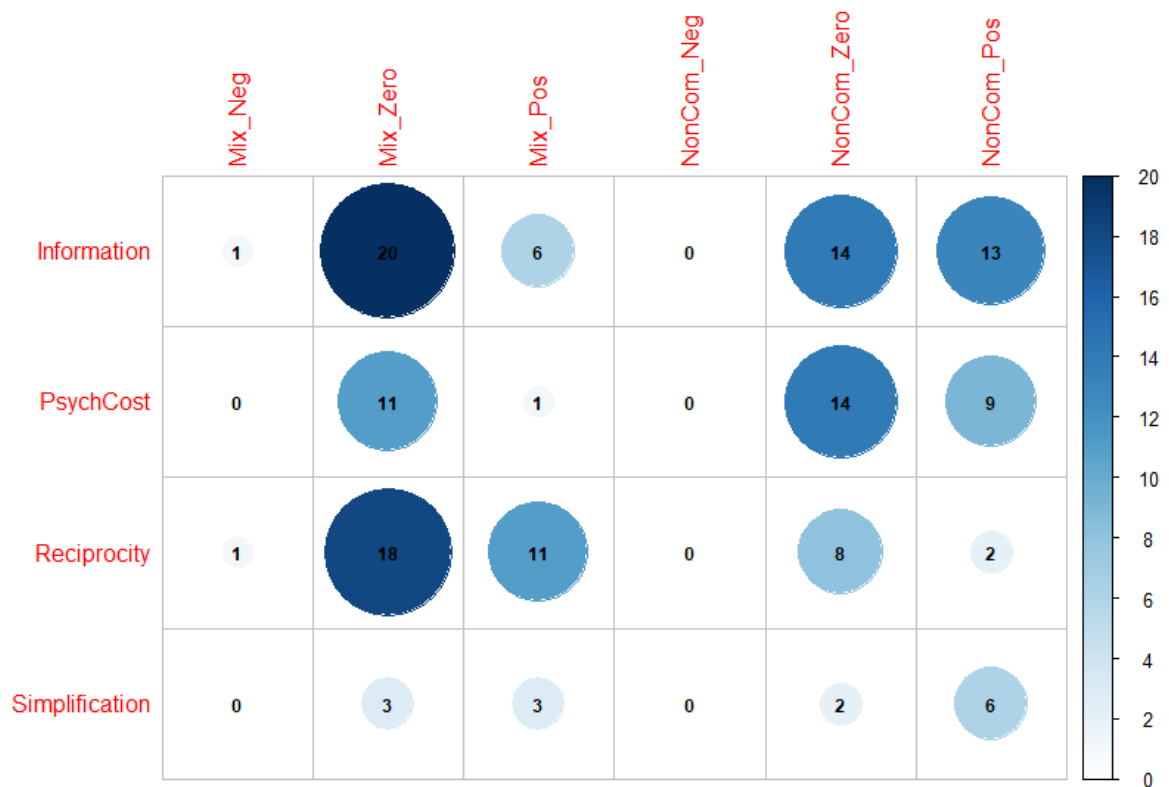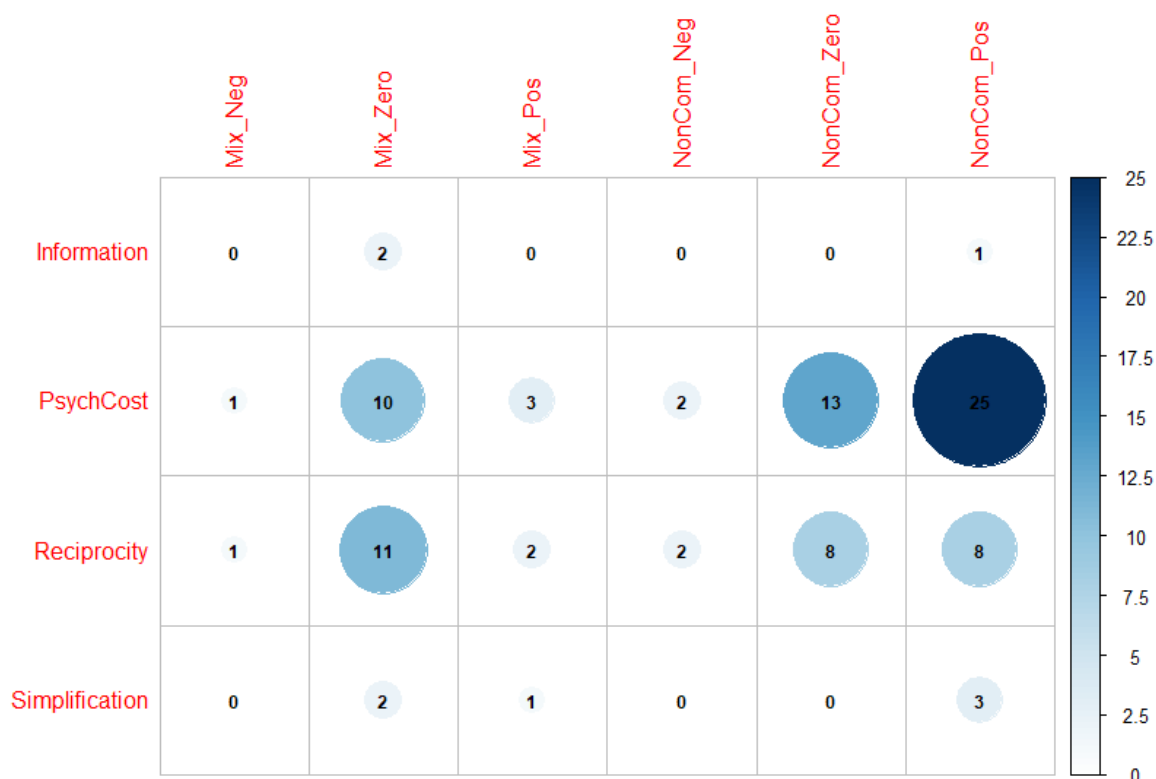*Figure 8.* Balloon plot showing the frequency distribution of combined non-deterrence interventions in the samples of company, individual, and self-employed taxpayers. C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

The results further highlighted that non-deterrence interventions more frequently affected the probability of paying, filing, or reporting taxes than the actual amount of income reported or taxes paid, both in terms of taxpayers' compliance history and their occupational status. Again, additional frequency tables comparing differences between control groups can be found in Appendix B (see Table 17A and 18A).

Types of Interventions in Detail. Similar to the procedure above, the distribution for each type of intervention was also investigated within the occupation sub-categories. Figure 9 further confirmed that information nudges seemed to be frequently ineffective, both in terms of taxpayers' compliance history and their occupational status. The plot also showed that 76% of the psychological cost interventions, that were effective in the self-employed and micro-enterprise sample, were made up of social norm interventions. However, only 29% of social norm treatments in the individual population of taxpayers were effective, with 14% of the treatments exhibiting backfiring effects. Nevertheless, it could be observed that social norms accounted for more frequent positive effects on tax compliance outcomes than moral appeals within the subpopulations (i.e., in terms of compliance history and occupation). The distribution in Figure 9 showed no type of intervention that exhibited frequent positive effects in the company subsample. Evidence within this subsample was generally limited due to few interventions in field experiments performed in a corporate setting. Additional frequency tables,

comparing differences in terms of outcome measurements as well as between control groups, can be found in Appendix B (see Tables 20A to 23A).

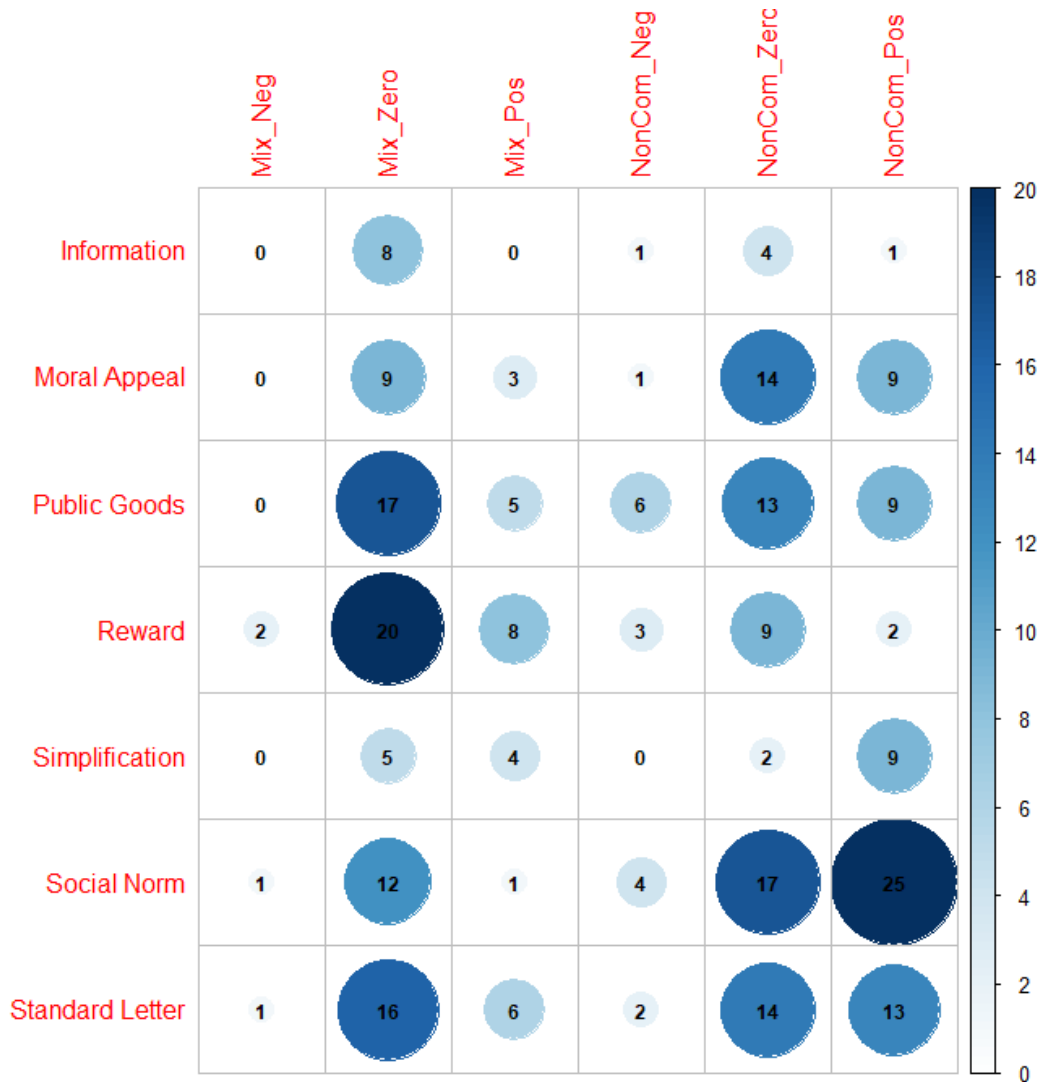*Figure 9.* Balloon plot showing the frequency distribution of individual non-deterrence interventions in the samples of company, individual, and self-employed taxpayers. C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

3.4.4 Compliance and Occupation

In order to shed more light on how the interventions affected various subpopulations, the effectiveness for each occupation sub-category (i.e., *Individual*, *Self-Employed*, and *Company*) was investigated while taking into account previous compliance history. This provided a clearer picture about what groups of interventions as well as individual interventions worked in more specific subpopulations (e.g., individuals with a history of non-compliance). At the same time, however, in this display, the frequencies in each field decreased with many fields having zero cases. The low frequencies per field should be kept in mind when interpreting the data.

Figure 10 depicted how frequently the nudging interventions were effective in each possible population sub-category, taking into account their compliance history. First, the plot

showed that the interventions seemed to have the most frequent positive impact on tax compliance in the population of self-employed with a history of non-compliance. However, as could be observed in Table 2, evidence for this subpopulation only came from three studies, thus limiting the generalizability of this finding. Nevertheless, it showed that this population might be more susceptible to non-deterrence nudging interventions than other populations of taxpayers. The plot also showed that 69% of the simplification interventions, that displayed positive effects in the population of individual taxpayers, were effective when they had a history of non-compliance. Furthermore, it could be observed that 81% of all backfiring effects were found in the population of individual taxpayers with a history of non-compliance, most frequently for reciprocity interventions. Interestingly, only 12% of non-deterrence interventions had a positive effect on tax compliance in the sample of individual taxpayers with no history of non-compliance. Additional analyses were included in Appendix B (see Tables 24A to 33A). Due to the relatively low frequencies when further subdividing the samples as well as little additional information gained from their interpretation, those distributions will not be discussed in detail.

3.4.5 Regional Differences

Since the importance of considering regional variation was introduced in the context of this review, the frequency distribution of nudging interventions within the regional sub-categories was also examined. As only two studies were conducted in South-Eastern as well as Northern European countries, they were grouped together with studies from Western European countries into the sub-category *Europe* for the subsequent analysis. Similar to the frequency distributions above, the interventions were first grouped together and then investigated individually to examine what mechanisms and interventions showed frequent effects in various regional contexts. Due to the limited amount of evidence, especially from regions like Africa or the Middle East, there were various fields with low or zero frequencies. Again, this should be taken into account when interpreting the data.

Figure 11 showed the frequency distribution of the interventions for the regional sub-categories across all outcomes. Even though there did not seem to be a region where non-deterrence interventions worked particularly well in terms of relative frequencies, small regional differences could still be identified. For example, compared to 40% of non-deterrence interventions showing positive effects in European countries, only 27% of the treatments showed positive effects in studies from North America (i.e., studies from the US). Especially psychological cost as well as simplification treatments showed frequent positive effects in
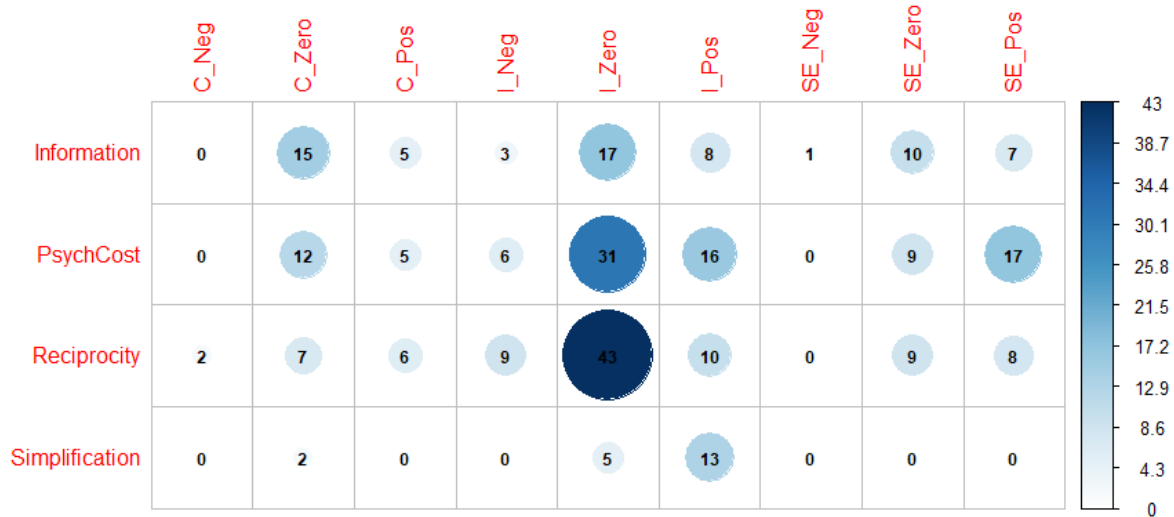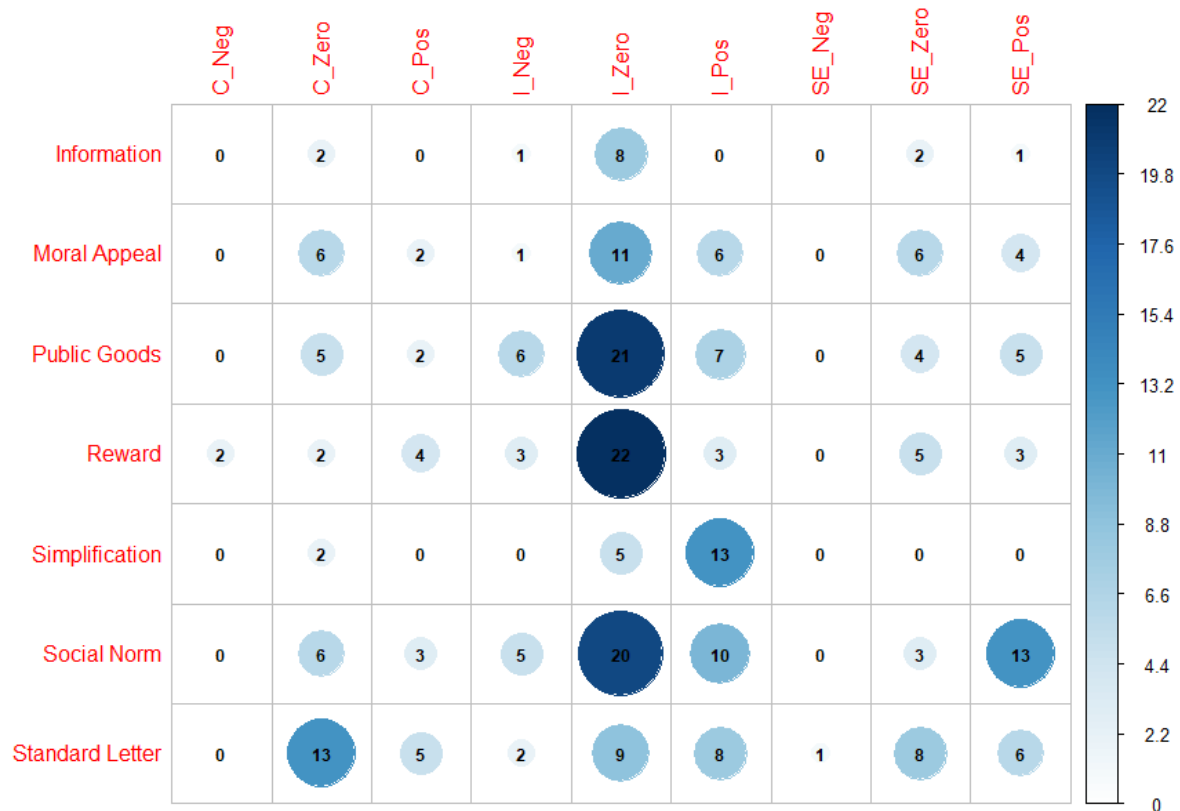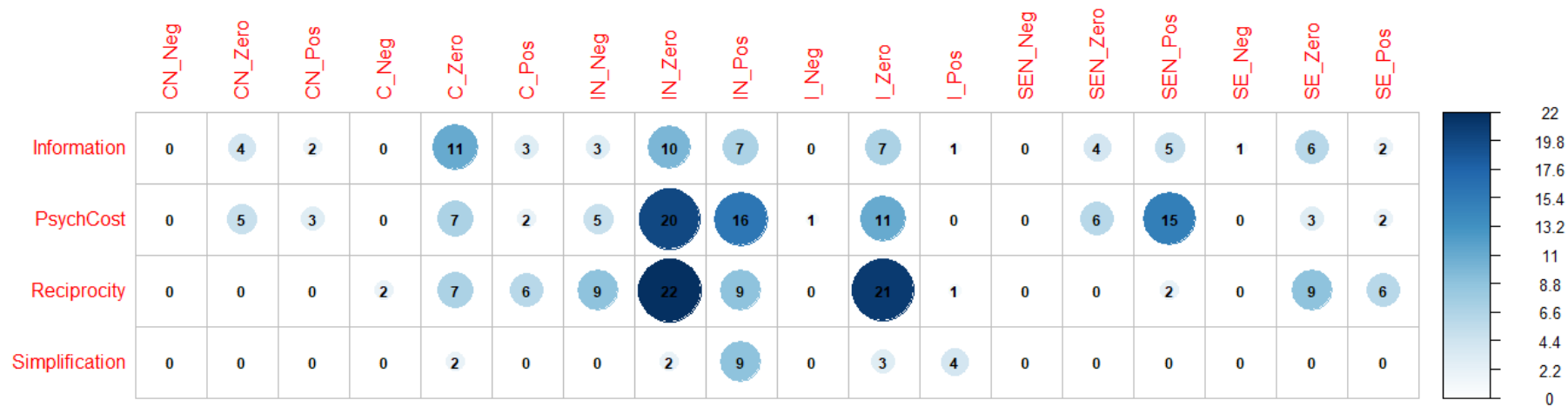
Figure 10. Balloon plot showing the frequency distribution of combined non-deterrence interventions in the samples of company, individual, and self-employed taxpayers in combination with their compliance history. CN = Company / Non-Compliant. C = Company. IN = Individual / Non-Compliant. I = Individual. SEN = Self-Employed / Non-Compliant. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

studies from European countries. However, 76% of all backfiring effects also stemmed from studies conducted in Europe.

When examining the types of interventions in more detail, it could be observed that public goods interventions were frequently ineffective both in North and South American countries, Reward treatments, on the other hand, displayed zero positive effects in European countries. However, they showed more promising effects in South American countries, especially when measured at the intensive margin (see Table 36A). It was also remarkable that moral appeal treatments, while showing no noteworthy effects in any population sub-category, seemed to display some positive effects in both Europe and the US. Though this was an observation based on low frequencies, it was nevertheless a notable finding in terms of previous literature on how trust in governments can affect tax morale (see Section 1.3). Finally, it was worth mentioning that social norm treatments hardly had a positive impact on tax compliance in the US across various studies. Additional analyses showing frequency distributions across different outcomes as well as different control groups were included in Appendix B (see Tables 34A to 43A).

3.4.6 Correspondence Analysis

Though the main focus of the Results section rested on interpreting the frequencies of effects, there were different ways of analysing and interpreting frequency distributions. Correspondence analysis was one alternative method that should be discussed in this review, which graphically displayed associations between the treatments and population, as well as region sub-categories. However, correspondence analysis is not as straightforward to understand as might be assumed when looking at the graph (see, for example, Figure 12). The interpretation of the correspondence analysis will therefore be discussed in more detail based on the results of this review. Figure 12 showed the correspondence analysis for the frequency table displayed in the balloon plot in Figure 2, whereas the blue points represented the rows (i.e., the interventions), and the red triangles represented the columns (i.e., the population and region sub-categories). The correspondence analysis only displayed relative associations between the rows and columns.

First, it is important to note that the further a row point or column triangle is situated from the origin (i.e., where the x- and y-axis cross), the more discriminating it is. For example, simplification interventions showed the most variation of frequencies in the columns compared to other treatments and were therefore more discriminating and further from the origin. Row points (or column triangles) can be compared with each other based on the
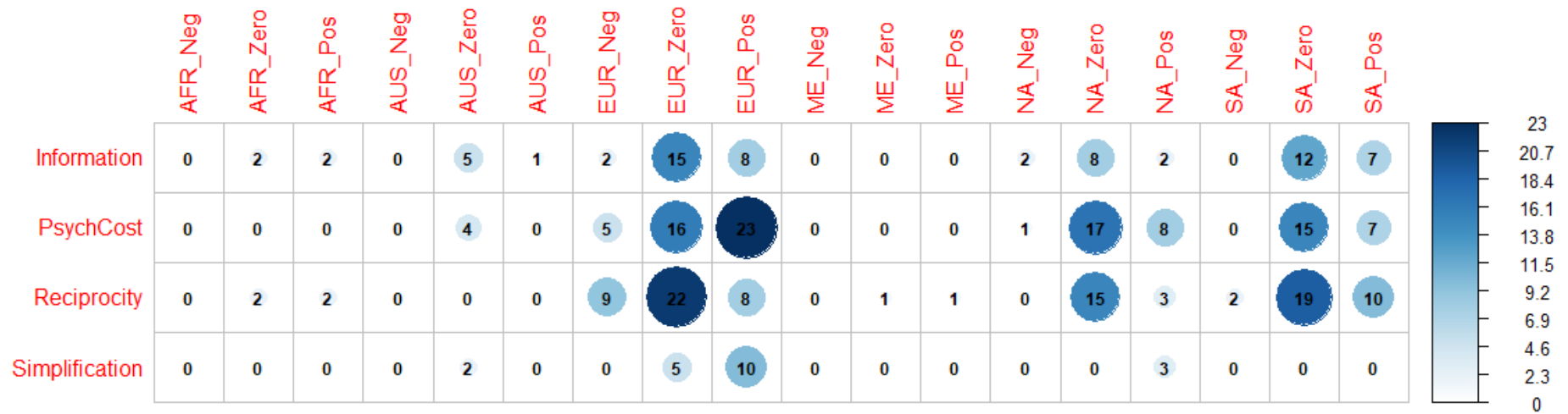
Figure 11. Balloon plot showing the frequency distribution of combined non-deterrence interventions in the regional sub-categories. AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

distance to the origin. For example, negative effects had similar frequencies in both samples (i.e., mixed and non-compliant samples) across the interventions and were therefore closer to each other. Additionally, they also displayed a higher variation across the treatments further affected by their lower overall frequencies, which led to them being placed further from the origin (see Figure 2 as reference).

Figure 12. Correspondence analysis for frequencies of combined non-deterrence interventions in Table 4A. Blue points represent the rows (i.e., the interventions). Red triangles represent the columns (i.e., the population sub-categories). Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

However, the interpretation when comparing row points with column triangles is a bit different. In this case, a larger distance to the origin indicates a stronger positive or negative association with a triangle and vice versa. Additionally, it is important to look at the angle formed between a line drawn from the origin to a row point and a column triangle. Small angles indicate an association, as for example, between reciprocal treatments and zero effects in

the mixed sample. Angles around 90 degrees indicate no association, and angles around 180 degrees point towards a negative association between row points and column triangles. Figure 13 displayed the lines drawn to each row and column element. Based on this information, it could be observed that simplification treatments and positive effects in the non-compliant sample were both discriminating. Furthermore, based on their large distance to the origin and an angle below 45 degrees, their association was relatively strong. At the same time, it could be seen that there was a strong negative association between simplification treatments and negative effects in both samples, corresponding to the distribution in Figure 2.
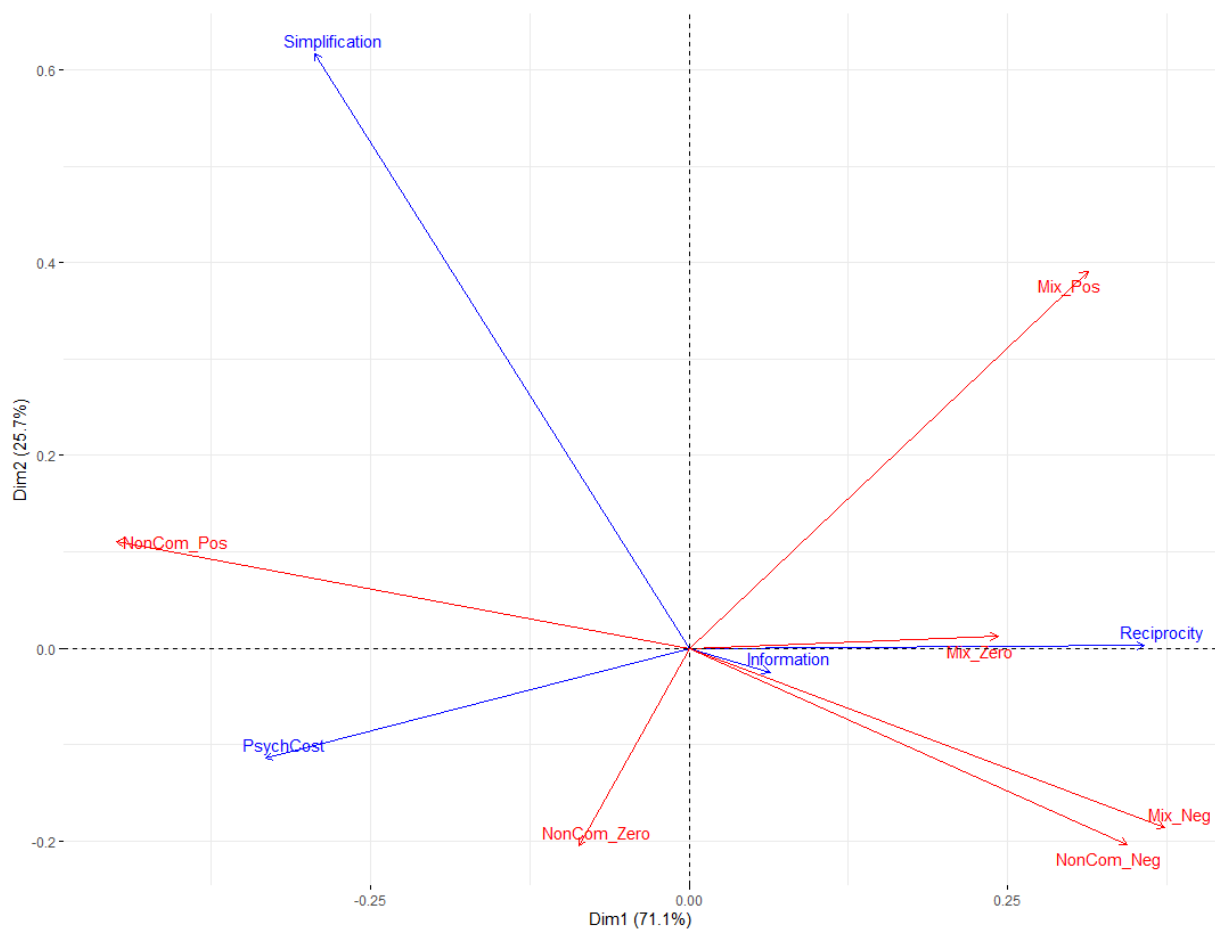
Figure 13. Asymmetric correspondence analysis for frequencies of combined non-deterrence interventions in Table 4A. Blue lines represent the rows (i.e., the interventions). Red lines represent the columns (i.e., the population sub-categories). Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Finally, the percentages in the axes labels of the chart show the amount of variance explained in the horizontal and vertical dimension. In Figure 12, the horizontal dimension explained around 71% of the variance in the data, while the vertical dimension explained around 26%. The percentage being higher on the horizontal than on the vertical dimension indicated that the points varied more strongly on the horizontal dimension. Thus, the horizontal dimension explained more relative variance. Taken together, both dimensions explained roughly 97% of the variance, showing that the analysis represented almost all of the information in the data, and few insights in terms of relative associations were missed. In contrast, if the plot explains a relatively small amount of variance, it means that data has been left out. For example, one intervention might be more differentiated on a dimension that is not relevant for most of the other interventions, which leads to the dimension not being taken into account.

Additional correspondence analyses for frequency distributions were included in the Appendix C (see Figures 14A to 28A). They can be read based on the previous example and underlying frequency tables found in Appendix B. As balloon plots and correspondence analyses can be interpreted similarly, additional correspondence analyses will not be discussed in this review. Further, limitations of correspondence analyses for the current data set will be discussed in Section 4.4.1.

Discussion

The current systematic review evaluated the effectiveness of non-deterrence nudging interventions on tax compliance based on their frequency of negative, zero, and positive effects from various studies conducted in the field. The effects were investigated for different subpopulations of taxpayers, taking into account their compliance history and occupational context. Additionally, the effects were analysed for various regions in which the field experiments were conducted. The review aimed to present the influences of these factors on the effectiveness of non-deterrence nudging interventions on tax compliance. Previous research to date has not yet systematically investigated how the effectiveness of non-deterrence interventions is influenced by individual and contextual factors. This review is the first that attempted to summarize various facets of field experiments (e.g., types of interventions, populations, and regions) in a standardized framework.

The results could show that non-deterrence nudging interventions were more frequently effective in increasing tax compliance in a previously non-compliant population of taxpayers compared to a general population. When looking at the interventions in detail, it was found

that most interventions including a simplification element were effective, however those effects mainly stemmed from a population of individual taxpayers. Social norm interventions also showed frequent positive effects, mainly in the population of previously non-compliant taxpayers. Reciprocity and information treatments did not show frequent positive effects in any sub-population or regional context. When investigating regional contexts in general, it was found that non-deterrence nudging interventions were more frequently effective in European than in North American field studies (i.e., mainly studies from the US). Additionally, there was only limited evidence from regions outside of Europe, North, and South America.

In the following sections, the effectiveness of the treatments in the population and region sub-categories will be discussed in detail, followed by a review of more general findings of this thesis. Finally, general and methodological limitations of this thesis, implications for future work, and policy implications will be discussed.

4.1 Effectiveness of Interventions in the Population Subsamples

For this review, the frequency of negative, zero, and positive effects of non-deterrence interventions was compared between a sample of taxpayers with an explicit history of non-compliance and a sample of taxpayers that had no explicit history of non-compliance (i.e., a general population of taxpayers). Furthermore, populations of taxpayers were compared in terms of their occupation, that is, whether they were paying their taxes individually, as self-employed, or in a corporate context. The interventions targeted various mechanisms, either aiming at creating psychological costs for the taxpayer, showing that their tax payments were reciprocated by the government or reducing their cognitive effort to pay taxes.

When analysing the frequency of effects, it was shown that simplification interventions were most frequently effective in an individual population of taxpayers with a history of non-compliance. Simplification treatments did not directly aim at the individuals' motivation to pay their taxes. Rather, they aimed at reducing cognitive barriers that might hinder their decision to pay. A history of non-compliance might therefore not only be existent due to a lack of motivation or a negative attitude towards paying taxes, but rather due to the taxpayers feeling overwhelmed with the amount of information that needed to be processed (Luttmer & Singhal, 2014). This could be one possible explanation as to why most simplification treatments showed relatively frequent positive effects within a population of previously non-compliant individuals. Furthermore, this would be in line with the argument that most taxpayers generally want to pay their taxes (Alm, 2019) and have a general disposition to comply with tax laws (Ariel, 2012). Still, it should be considered that, even though the results

stemmed from eight different studies, the overall number of simplification interventions implemented in field experiments was comparably small. Only one study thus far tested a simplification treatment on a corporate population of taxpayers, thereby finding no effects on tax compliance (see Biddle et al., 2018). As such, it would be interesting to further investigate how these interventions affect tax compliance, especially for self-employed and corporate taxpayers.

Interventions that aimed at inducing psychological costs for the taxpayer (i.e., social norms and moral appeals) were frequently ineffective in the general population of taxpayers (i.e., the mixed sample). However, they showed more frequent positive effects in a population of previously non-compliant taxpayers, particularly in the subpopulation of self-employed and micro-sized businesses. It should be noted that most of the positive effects in the non-compliant sample were produced by social norm treatments. Social norm interventions were argued to be effective if they update taxpayers' prior beliefs about tax non-compliance in the population (Del Carpio, 2014). Self-employed individuals may be especially prone to exhibiting non-compliance when believing that their peers are non-compliant, as they have a higher opportunity to cheat due to paying their own taxes (Alm et al., 2020a). As such, if taxpayers believed that the overall compliance rate in a given population was low, their psychological costs for displaying non-compliant behaviour might have been relatively low as well. Therefore, prior beliefs could be another explanation about why some taxpayers exhibited a history of non-compliance in the past, irrespective of their general attitudes towards paying taxes. Updating prior beliefs with actual compliance rates (e.g., Del Carpio, 2014) could then explain why social norm interventions showed more positive effects in a population with a history of non-compliance.

Moral appeals, which were frequently ineffective in the population subsamples, however, might not have been specific enough to update the beliefs of taxpayers. Authorities should consider taxpayers' former beliefs about overall compliance rates in the population so that these interventions can be tailored to the right population and to avoid backfiring effects in the process (Luttmer & Singhal, 2014; see Section 1.4). Though the findings indicate that specific subpopulations might be more susceptible to interventions inducing psychological costs, it should still be kept in mind that overall, most of these interventions did not result in changes in tax compliance. Besides the possibility of taxpayers already having a clear sense of overall compliance rates (Luttmer & Singhal, 2014), it could have very well been the case that this information did not fundamentally alter the taxpayers' perceptions or preferences. Furthermore, psychological cost treatments might have only had a short-term effect on tax

compliance (see, for example, Hallsworth et al., 2017) and an awareness of compliance rates might have not been salient across a longer period of time (Bott et al., 2019).

Looking at the frequency distributions for interventions that aimed at influencing tax compliance decisions through a reciprocal mechanism, it could be observed that, overall, the interventions were frequently ineffective in most subpopulations, both in terms of compliance history and occupation. Neither public goods nor reward treatments exhibited frequent positive effects in any subgroup of taxpayers, with public goods interventions being somewhat more effective in terms of the frequency of positive effects than reward treatments. Instead, those interventions triggered the most frequent backfiring effects compared to other nudging treatments, particularly in the subpopulation of individual taxpayers with a history of non-compliance. As mentioned in Section 1.4, the effectiveness of reciprocity treatments seems to be dependent on how taxpayers benefit in return for being tax compliant (Bordignon, 1993). According to the current findings, the interventions could therefore be considered a double-edged sword. Looking at public goods treatments, it is possible that the intervention backfired if the provision of public goods was made more salient, but taxpayers did not perceive that these goods were actually provided (OECD, 2013). It is also possible that individual taxpayers were simply not content with the sort of public goods that the intervention advertised (Wahl et al., 2010). For reciprocity interventions to positively affect tax compliance in general, it might therefore be helpful to first understand the needs of the target population that is addressed.

Information treatments were similarly often ineffective on tax compliance across population subsamples, with information nudges showing almost zero effectiveness overall. Due to the fact that the interventions conveyed various forms of information to the taxpayers (e.g., informing about customer service or offering advice), they were quite heterogeneous in nature, making it relatively difficult to specify reasons as to why information nudges did not seem to work in most studies. Positive effects were found in previous field experiments, however, only in selected subgroups within these studies (e.g., high- and low-income taxpayers; Coleman, 1996). Interestingly, even standard letters and reminders with no additional information or nudging element showed more frequent positive effects than information treatments, though it should be considered that standard letters were only compared to a control group that received no letter. Still, it might be possible that providing too much information in the information treatments overwhelmed the taxpayers, leading them to procrastinate their tax obligations (Luttmer & Singhal, 2014). It seems that tailoring information treatments to a target population is even more challenging than it is for other types of nudges, as there is a wide

variety of information regarding the taxpaying process that can be conveyed to taxpayers. Overall, the results indicated that receiving any kind of information from the authority (e.g., in the form of a standard letter) could already have a positive effect on tax compliance, especially for taxpayers with a history of non-compliance.

4.1.1 Discussing Populations in General

Previous field studies argued that non-deterrence interventions seem to be more effective in a general population of taxpayers (e.g., Castro & Scartascini, 2015; Del Carpio, 2014), as they were assumed to have a more positive orientation towards the authority than a previously non-compliant population of taxpayers. However, the findings of the current systematic review on tax compliance in field studies showed that more non-deterrence interventions positively affected tax compliance in a previously non-compliant sample. Similarly, self-employed individuals, especially those with a history of non-compliance, showed frequent changes in their compliance behaviour after receiving an intervention. Besides the above-mentioned intervention-related explanations, one possible reason could be that self-employed and previously non-compliant taxpayers in general were more inclined to feel exposed when getting an official letter from the tax authority. As it was assumed that those sub-groups either have a higher disposition towards non-compliance (Braithwaite, 2003) or a higher opportunity to cheat taxes (Kleven et al., 2011), it can be reasoned that there might have also been a higher risk for those subgroups if exposed. As such, this could have led them to adapt their behaviour and increase tax compliance more frequently in response to letters from the authority, regardless of the information contained in them. Additionally, it could be argued that previously non-compliant taxpayers had a higher potential to improve their compliance, as taxpayers who were compliant in the past might not have been able to be 'more compliant' in response to interventions. Contrary to previous assumptions, it can therefore be reasoned that taxpayers with a history of non-compliance might be more susceptible to nudging interventions than the general population.

Unfortunately, little is known about the population of companies with a history of non-compliance, as only two field studies so far examined effects of non-deterrence interventions on tax compliance in this subsample. Smaller to larger sized companies in the general population (i.e., the mixed sample) mostly did not change their compliance behaviour in response to the interventions. Companies usually employ professional tax practitioners or accountant departments to prepare their tax returns for them (Klepper et al., 1991). Though it was argued that drivers for company compliance might be similar to that of individuals (Alm, 2019), pro-

fessional tax practitioners usually do not act in their own interest and, furthermore, do not bear the risk of tax non-compliance themselves. Additionally, it can be assumed that they are more likely to have a clear sense of compliance and audit rates, as well as having knowledge about the taxpaying process due to their profession. Taken together, this might have diminished some of the positive effects for corporate taxpayers that were found, for example, for social norm and simplification interventions in other subsamples. Still, as mentioned above, evidence from corporate taxpayers was limited, which is surprising, as most of the tax revenue collected comes from larger companies in most of the countries (Hallsworth, 2018). Further investigating corporate samples might, therefore, have the highest potential benefits for tax authorities.

Finally, similar to the general population of companies, the general population of individual taxpayers hardly changed their compliance behaviour in response to non-deterrence interventions. If most individual taxpayers indeed have a positive orientation towards paying taxes and do pay their taxes most of the time (Alm, 2019), non-deterrence interventions may be perceived as redundant information asking them to comply when they eventually already did. Therefore, the positive effects of some treatments might rather be explained by individual preferences, or they could also be dependent on the actual wording of the intervention.

4.2 Discussing Regional Variation

Additional to comparing frequencies of negative, zero, and positive effects of non-deterrence treatments in various population subsamples, frequencies of effects for various geographic regions with similar structures in terms of their tax systems, based on previous OECD reports (OECD, 2013, 2019), were also compared.

It was found that nudging treatments were, overall, more often effective in European countries than in North America (i.e., the US). Especially interventions that included simplified information or induced psychological costs exhibited frequent positive effects within Europe. It can be argued that the effectiveness of simplification treatments might be rather region unspecific and probably more dependent upon individual differences, as limited cognitive resources and deviations from rational behaviour were found in studies in many different countries (Rau et al., 2020). However, there is not enough evidence from other regions in the context of tax compliance, as most simplification treatments thus far were implemented in European field experiments (e.g., De Neve et al., 2019; Dwenger et al., 2016). Interestingly, interventions inducing psychological costs were found to be more frequently effective in Europe than in the US. Specifically, social norm treatments displayed frequent zero effects in

studies conducted in the US. This might be related to differences in terms of taxpayer hetero-geneity (i.e., race, religion etc.) between the two regions. As mentioned in Section 1.6, tax non-compliance was found to be related to socioeconomic diversity and group heterogeneity, which led to different social norms held within these groups (Alm et al., 2016). As the US arguably has a more diverse population of taxpaying citizens due to the country's history than European countries (Alesina et al., 2005), this might explain some differences in the effec-tiveness of social norm interventions.

Furthermore, it is interesting that moral appeals, despite having shown frequent zero ef-fects in most population sub-categories, did show more frequent positive effects within Eu-rope and the US when comparing the regions. However, it is important to consider that this observation was based on relatively low frequencies. Still, it was shown that higher trust in authorities increased voluntary tax compliance in various regions (Batrancea et al., 2019), usually being higher in wealthier countries (OECD, 2013). Therefore, if the taxpayers in these regions had a higher trust in the authority, they might have felt a higher moral obligation to pay their taxes. Consequently, appealing to tax morale could have had a higher chance of in-creasing tax compliance in European countries or in the US. In past research, it was also found that the political alignment with the current ruling party, which is affiliated with the tax authority, was an influencing factor for the decision to be voluntarily tax compliant (Cullen et al., 2018). This finding could be particularly relevant for explaining the effectiveness of moral appeal nudges, as the moral obligation to pay taxes is likely to be lower if taxpayers are not aligned with values of the ruling party and interventions might, therefore, be less effective.

Reciprocity interventions did not show frequent positive effects in any specific region sub-category. Almost zero public goods interventions were effective in field studies in the US and South America, whereas zero reward interventions were effective in Europe. In terms of public goods interventions, this might be partly related to the actual provision of public goods. As mentioned in Section 1.4, taxpayers may be dissatisfied with the public goods and services that the government provides (OECD, 2013). This could be especially relevant for the US, as the provision of many public goods has been on a decline in the past decades (Chantrill, 2021). Even though satisfaction with, for example, health care services increased in recent years, the majority of American taxpayers still does not seem to be satisfied with health care in the US (Jones & Brenan, 2020). This, in turn, might have negatively affected taxpayers' compliance decisions when made salient through the treatments. It should also be kept in mind that the perception and actual provision of public goods can vary even within geograph-ic regions (OECD, 2019). Given that the satisfaction with public goods and services is highly

country-specific, it could be more useful to investigate the effectiveness of such interventions by comparing countries with similar contexts in terms of providing public goods instead of regional semblance.

Finally, taking a closer look at the reward treatments from European studies, the evidence should be interpreted with caution, as only two field studies implemented reward treatments in this region. Koessler et al. (2019) argued that the type of reward might be an explanation as to why the treatments were mostly ineffective in their study. They suspected that financial rewards might trigger the perception that the reward is a financial exchange instead of an acknowledgement for being compliant. However, findings from other regions like South America, where some financial reward interventions showed more frequent positive effects (e.g., Alm et al., 2019), lead to rather inconclusive evidence for those treatments. Therefore, it would be beneficial to further investigate which type of rewards have the most potential of influencing taxpayers' compliance decisions.

4.3 General Remarks

Although the systematic review provides indications of circumstances where non-deterrence nudging interventions can have a positive impact on tax compliance, it is important to consider that the vast majority of interventions were not effective in increasing tax compliance. The data also showed that the interventions were more frequently effective when measured at the extensive margin, that is, they were more often effective in increasing the probability to pay, file, or report taxes. In contrast, the treatments less frequently increased the amount of taxes paid or income reported within a given period of time (i.e., the intensive margin), which corresponds with evidence from previous meta-analyses (see Antinyan & Asatryan, 2020). Furthermore, the current findings showed that some nudging interventions were less frequently effective than sending standard letters or reminders that include no nudging element (see Figure 7). Still, nudging interventions rarely led to backfiring effects when compared with a control group that did not receive any message. Despite frequent zero effects, the review nevertheless illustrates the importance of considering individual, social, and environmental aspects when interpreting the effectiveness of non-deterrence interventions. As discussed above, these types of interventions are very context-sensitive, however, given the right implementation in the right circumstances, they can be very beneficial for increasing tax compliance.

4.4 Limitations

The current systematic review faced various limitations and constraints worthy of mention, both of methodological nature and related to the underlying sample of field studies. The limitations make up an important part of this review and should provide indication for improvements that can be made for future study designs. In the following segments, these limitations will be discussed in detail.

4.4.1 Methodological Limitations

An important limitation in regard to the explanatory power of the results was the decision not to summarize effect sizes of the interventions implemented in the field studies within this review. Instead, it was solely focused on categorizing whether the treatments significantly affected tax compliance for various outcomes ($\alpha = .05$). Though it was agreed to set the significance at a more restrictive level than some of the previous field studies to guarantee a higher measure of confidence, the data did not allow for reporting magnitudes of the treatment effects in the various subsamples that were discussed above. As some field studies in the review included large sample sizes, some of the statistically significant effects found in the studies might have had relatively small effect sizes, which could not be accounted for based on the current design. Evidence from previous meta-analyses on this subject showed that the magnitudes of treatment effects for non-deterrence interventions indeed seemed to be relatively small (Antinyan & Asatryan, 2020; Hallsworth, 2014).

Additionally, this review only allowed for interpreting frequency distributions based on absolute frequencies. Due to the low and zero frequencies in some cells of the frequency distributions, either due to rarely occurring negative effects or few cases in certain populations, the condition for using a chi-square statistic (i.e., no cells with fewer than 5 cases) to test relationships between interventions, population, and region sub-categories, was not met.[2] It would have been possible to conduct tests for some of the distributions when creating a dummy variable that takes the value 1 when there are positive effects and 0 when there are zero or negative effects. However, for this thesis, it was decided to focus on interpreting the distribution of effects instead of testing the statistical significance of associations.

Another limitation, already mentioned in Section 3.4.2 is that some field studies investigated the effects of their treatments on multiple compliance outcomes (e.g., at the extensive and intensive margin). Subsequently, the effects of some interventions were represented mul-

---

[2] see https://libguides.library.kent.edu/spss/chisquare for more information

tiple times, whereas the effects of other interventions were investigated for just one compli-
ance outcome. This could lead to overestimating the frequency of positive or negative effects
of certain treatments while, at the same time, underestimating the frequency of effects of oth-
ers. One way to avoid this could have been to give less weight to interventions that were
measured on multiple outcomes during categorization. However, the decision was to go along
without giving weights to interventions, as even though some interventions might have been
weighed more strongly in this case, the advantage was that all effects on relevant outcomes
were considered equally. Additionally, in the results, the problem was partly accounted for by
displaying the frequency of effects separately at the extensive and intensive margin to reduce
overweighting some of the treatment effects.

There are also some limitations in terms of categorizing and interpreting effects in the
population subsamples. For example, not every field experiment clearly denoted whether their
population of taxpayers had a history of non-compliance. This prevented the interpretation of
effects as clear differences between a previously non-compliant and a compliant population of
taxpayers, as it was not known for certain whether the 'compliant' population (i.e., the general
population or mixed sample) actually had no history of non-compliance in the past. Similarly,
differentiating between self-employed and corporate taxpayers was not always entirely clear
in the categorization process. Field studies varied not only in their definition of self-employed
taxpayers (e.g., based on income or number of employees), but also in terms of how much
information they provided about the sample in the first place. As mentioned in Section 3.3.4,
when differentiating between self-employed and companies, it was decided to focus on the
number of employees. However, this criterion might not have been the best suited for discrim-
ination and other researchers could come to different conclusions. It might have been more
sensible to differentiate between self-employed and companies in terms of whether they em-
ploy professional tax preparers. Yet, this information was rarely available in the present field
experiments.

Finally, it should be mentioned that the application of correspondence analyses on the
current data set was relatively limited. Correspondence analyses are mostly used for larger
datasets with enough frequencies in each cell of the frequency table, where they can give in-
teresting insights into interconnections between variables (Greenacre, 2016). However, some
of the frequency tables of the current data set were relatively small (e.g., Table 4) and there-
fore easier to interpret with a balloon plot instead of a correspondence analysis. Second, larger
frequency tables in the data set (e.g., Table 39A) exhibited zero frequencies in some cells,
mostly due to the relatively few negative effects of the interventions, which could distort the

visual representation of the table in the correspondence analysis. Finally, correspondence analyses in R exclude all rows or columns with zero frequencies in the visual representation (e.g., Column 4 in Figure 5), leading to an incomplete depiction of the underlying frequency tables. These considerations directed the focus to mainly interpreting balloon plots, which showed a relatively clear picture of how the nudging treatments were related to various population and region sub-categories. The additional inclusion of the correspondence analyses should provide further insights into the data.

At this point, it is important to note that not all interesting points from the various frequency distributions might have been discussed in detail. There are various ways in which investigating the data would have been insightful, not all of which have been included in this review. For this review, the focus was on aspects that were of apparent interest to the research question. However, additional visualizations of the data or findings of interest to other readers may have been overlooked. Therefore, further indications and conclusions could be drawn from the data set.

4.4.2 General Limitations

Additional to methodological constraints, there are some limitations attributable to the heterogeneity of the sample of field studies that affect the validity of the results. Cross-study comparisons like this come with a lot of methodological concerns regarding the standardization of, for example, dependent and independent variables across studies. One major advantage of this review is that the interventions were categorized based on how they were phrased (i.e., based on the information most salient in the treatment) and not based on how they were defined or intended by the experimenters (e.g., a 'citizenship' treatment, worded as a loss of public goods; Hasseldine et al., 2007). Although this made it easier to standardize interventions and at the same time allowed for attributing the effects directly to the choice of words, the process of categorization is even more dependent on subjective ratings. Though it should be mentioned that the current interrater reliability for the types of interventions was sufficiently high, it was nevertheless the lowest among the calculated reliabilities (see Section 3.2). Therefore, different raters might have come to different conclusions when categorizing interventions.

Furthermore, the underlying field studies were conducted in many different contexts, with countries employing different tax systems and taxpayers exhibiting diverse tax morale not only across countries (Alm et al., 1995), but even when thinking about different types of taxes within a country (Luttmer & Singhal, 2014). Including the fact that there is still a lack of

accessible evidence from various regions like Africa, the Middle East and, most strikingly, Asia, the problem of generalizing results on what interventions work in which contexts remains even within this framework.

Finally, it is important to note that the overall sample size for this review was relatively small, especially for certain interventions like simplification treatments. This was partly attributable to the fact that field experiments investigating behavioural interventions in the context of tax compliance are still scarce. As such, the explanatory power of the results is limited. However, there has been an increase in research on this topic within the last decade and field experiments in collaboration with authorities are becoming more and more attractive for both parties due to their relatively low implementation costs (Hallsworth, 2018).

4.5 Implications for Future Research

The current review attempted to analyse what non-deterrence nudging interventions work "for whom, when, and why" (Hallsworth, 2018, p. 446). However, to draw more general conclusions, it is necessary to conduct more behavioural research in the field of tax compliance. There is still limited evidence from corporate settings, in which professionals are employed to correctly report and pay taxes. Furthermore, there are only few studies so far that were conducted in developing countries, where tax morale and attitudes towards the authority can vary due to, for example, lower trust in governments (OECD, 2013). In addition to that, it would also be interesting to include research on tax compliance from East Asian countries, where attempts at increasing voluntary tax compliance already exist, but lots of research has not been reported publicly yet (Asian Development Bank, 2020).

Another venue for future research, especially for field experiments, would be to pay attention to giving more detailed information on the sample of taxpayers and the context in which the experiment was conducted. This would allow for a clearer allocation of treatment effects to certain populations and circumstances in individual studies and, consequently, would make it easier to generalize effects when comparing these studies on a macro-level. As mentioned in the limitations of this review, many difficulties during the categorization process related to the heterogeneous nature of the sample and similar problems were reported in other reviews (e.g., Antinyan & Asatryan, 2020).

Additionally, it might be beneficial to gather as much information as possible about the target population and the contextual framework before conducting a large-scale field experiment. This means, for example, acquiring more data on past behaviour of the target population, which could help to determine their general attitudes towards taxes as well as identify

reasons for what motivated the taxpayers to comply. This might also require closer cooperation between the tax authorities and researchers, enabling relevant information to be accessed more easily. Subsequently, non-deterrence interventions could be tailored more closely to the sample of interest. The current review indicated that a history of non-compliance, for instance, might not necessarily be only due to a negative attitude towards paying taxes, but also due to limited resources to properly fulfil tax obligations.

Finally, a few brief suggestions should be mentioned. First, few of the field studies in this review have investigated long-term effects of non-deterrence nudges, and those who did mostly found no lasting effects of the interventions in the next tax year (e.g., Chirico et al., 2019; Eerola et al., 2019). In other fields of research, it was often argued that nudges present only short-term incentives that dissipate fast when the intervention is no longer salient (e.g., in terms of smoking cessation; Cahill & Perera, 2008). However, further evidence is needed to conclude whether there are certain circumstances in which long-term effects of nudging interventions on tax compliance are more likely to occur.

Furthermore, although interventions were categorized based on their exact wording, it was not investigated in detail if interventions with frequent positive effects had a similar wording or used certain phrases that were particularly effective for increasing tax compliance. It might be interesting to examine if the phrasing of interventions itself can be linked to treatment effects, irrespective of the type of nudging element included in the intervention. Another potential approach would be to explore if a different wording of the same intervention has different effects on tax compliance.

Lastly, it would be beneficial to ensure that the materials used are translated and added to supplementary materials or made publicly available otherwise. This would be especially relevant for the interventions sent to taxpayers. In some of the present field studies, either no exact wording was provided, or the intervention included was not fully translated. Having access to accurate translations of the original wording used would be important for examining the effectiveness of interventions and enabling comparisons between various field studies in this context.

4.6 Practical Implication and Conclusion

This review adds to previous literature on tax compliance by trying to consolidate findings on the effectiveness of non-deterrence nudging interventions from various field experiments. Behavioural interventions were often criticised for their inconclusive evidence and strong context-dependency, thus not always being considered useful by authorities and policy-

makers (Hummel & Maedche, 2019). Many tax authorities still think of traditional approaches like audits and sanctions to be the most effective way in increasing tax compliance. However, not only are these approaches rather expensive (Hallsworth, 2014), they also potentially undermine trust in authorities (Batrancea et al., 2019; Leonidou et al., 2008) and intrinsic motivation to pay taxes (Frey, 1997). Therefore, as was already illustrated in the theory of responsive regulation (Braithwaite, 2003), authorities and policy-makers require a broader set of methods to respond appropriately to different types of taxpayers. The current work demonstrates that behavioural interventions can be a complementing and cost-effective alternative for policy-makers to increase tax compliance, given that they are implemented with care.

Acknowledgments

First, I would like to express my sincere gratitude to Žiga Puklavec, who continuously supported this work, gave insightful feedback, and bore with me through many discussions in the past year. This thesis would not be what it is now without his support and helpful remarks.

Second, I would like to thank Dr. Erich Kirchler for his knowledge, support, and continuous availability whenever questions arose. His share of expertise often gave me clarity during the implementation and execution of this thesis.

Special thanks to friends and colleagues, who supported me, listened to problems, provided solutions, and gave helpful feedback in the past year. I know you sometimes had to put up with me and my doubts and worries. I appreciate it dearly that you did. In alphabetical order: Paul Ballot, Mariana Balog, Samir Bellido, Kathrin Brückl, Sophia Burger, Qian Feng, Thomas Gebetsberger, Julia Höchtl, Barbara Kaiser, Hjørdis Kasamas, Regina Kolb, Katharina Koss, Sophie Melzer, Barbara Reifetshammer, Alissa Riffert, Nadine Schuhleitner, Christine Stegmann, Mathias Sternbauer, Remya Thottakara, and Amanda Wöhler.

Finally, I would like to thank my parents, who made all this possible through their continuous support and guidance whenever I needed it. You are the best!

Thank you.

References

Alesina, A. F., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. T. (2005). Fractionalization. *SSRN Electronic Journal,* (1959). https://doi.org/10.2139/ssrn.319762

Algan, Y., & Cahuc, P. (2010). Inherited trust and growth. *American Economic Review*, *100*(5), 2060–2092. https://doi.org/10.1257/aer.100.5.2060

Allingham, M. G., & Sandmo, A. (1972). Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics*, *1*, 323–338. https://doi.org/10.4324/9781315185194

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. (2019). What Motivates Tax Compliance? *Journal of Economic Surveys*, *33*(2), 353–388. https://doi.org/10.1111/joes.12272

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), 1–23. https://doi.org/10.3390/g10040043

Alm, J., Clark, J., & Leibel, K. (2016). Enforcement, socioeconomic diversity, and tax filing compliance in the United States. *Southern Economic Journal*, *82*(3), 725–747. https://doi.org/10.1002/soej.12106

Alm, J., Gerbrands, P., & Kirchler, E. (2020a). Using "responsive regulation" to reduce tax base erosion. *Regulation and Governance*, (2020). https://doi.org/10.1111/rego.12359

Alm, J., Jackson, B., & McKee, M. (1992). Estimating the Determinants of Taxpayer Compliance with Experimental Data. *National Tax Journal, 45*(1), 107-114. https://doi.org/10.1086/NTJ41788949.

Alm, J., Malézieux, A., & McKee, M. (2020b). 40 Years of Tax Evasion Games: a Meta-Analysis. *Experimental Economics 24*, 699-750. https://doi.org/10.1007/s10683-020-09679-3

Alm, J., Sanchez, I., & De Juan, A. (1995). Economic and Noneconomic Factors in Tax Compliance. *Kyklos*, *48*(1), 1–18. https://doi.org/10.1111/j.1467-6435.1995.tb02312.x

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. (2020). Nudging for Tax Compliance: A Meta-Analysis. *SSRN Electronic Journal*, (August). https://doi.org/10.2139/ssrn.3500744

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

Asian Development Bank. (2020). *A Comparative Analysis of Tax Administration in Asia and the Pacific: 2020 Edition*. Asian Development Bank. https://www.adb.org/publications/comparative-analysis-tax-administration-asia-pacific-2020

Ayres, I., & Braithwaite, J. (1992). *Responsive Regulation: Transcending the Deregulation Debate*. Oxford University Press.

Batrancea, L., Nichita, A., Olsen, J., Kogler, C., Kirchler, E., Hoelzl, E., … Zukauskas, S. (2019). Trust and power as determinants of tax compliance across 44 nations. *Journal of Economic Psychology*, *74*(July), 102191. https://doi.org/10.1016/j.joep.2019.102191

Bazart, C., & Pickhardt, M. (2011). Fighting income tax evasion with positive rewards. *Public Finance Review*, *39*(1), 124–149. https://doi.org/10.1177/1091142110381639

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., … Galing, S. (2017). Should Governments Invest More in Nudging? *Psychological Science*, *28*(8), 1041–1055. https://doi.org/10.1177/0956797617702501

Bérgolo, M., Ceni, R., Cruces, G., Giaccobasso, M., & Perez-Truglia, R. (2017). *Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment* (Working Paper No. 12335). National Bureau of Economic Research. https://doi.org/10.3386/w23631

Bhargava, S., & Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review*, *105*(11), 3489–3529. https://doi.org/10.1257/aer.20121493

Biddle, N., Fels, K. M., & Sinning, M. (2018). Behavioral insights on business taxation: Evidence from two natural field experiments. *Journal of Behavioral and Experimental Finance*, *18*, 30–49. https://doi.org/10.1016/j.jbef.2018.01.004

Blackwell, C. (2007). *A Meta-Analysis of Tax Compliance Experiments* (Working Paper No. 07-24). International Studies Programme. https://icepp.gsu.edu/files/2015/03/ispwp0724.pdf

Blumenthal, M., Christian, C., & Slemrod, J. (2001). Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal*, *54*(1), 125–136. https://doi.org/10.17310/ntj.2001.1.06

Boning, W. C., Guyton, J., Hodge, R., & Slemrod, J. (2020). Heard it through the grapevine: The direct and network effects of a tax enforcement field experiment on firms. *Journal of Public Economics*, *190*, 104261. https://doi.org/10.1016/j.jpubeco.2020.104261

Bordignon, M. (1993). A fairness approach to income tax evasion. *Journal of Public Economics*, *52*(3), 345–362. https://doi.org/10.1016/0047-2727(93)90039-V

Bott, K. M., Cappelen, A. W., Sørensen, E. O., & Tungodden, B. (2019). You've got mail: A randomized field experiment on tax evasion. *Management Science*, *66*(7), 2801–2819. https://doi.org/10.1287/mnsc.2019.3390

Boyer, P. C., Dwenger, N., & Rincke, J. (2016). Do norms on contribution behavior affect intrinsic motivation? Field-experimental evidence from Germany. *Journal of Public Economics*, *144*, 140–153. https://doi.org/10.1016/j.jpubeco.2016.10.003

Braithwaite, V. (2003). Dancing with tax authorities: Motivational postures and non-compliant actions. In V. Braithwaite (Ed.), *Taxing Democracy: Understanding Tax Avoidance and Evasion* (pp. 15–40). Routledge.

Braithwaite, V. (2009). *Defiance in Taxation and Governance*. Edward Elgar Publishing Limited. https://doi.org/10.4337/9781848449077

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

Burtless, G. (1995). The Case for Randomized Field Trials in Economic and Policy Research. *Journal of Economic Perspectives*, *9*(2), 63–84. https://doi.org/10.1257/jep.9.2.63

Cahill, K., & Perera, R. (2008). Competitions and incentives for smoking cessation. *Cochrane Database of Systematic Reviews*, (3). https://doi.org/10.1002/14651858.CD004307.pub3

Camerer, C. F. (2012). The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1977749

Carlsson, F., Gravert, C., Johansson-Stenman, O., & Kurz, V. (2018). Nudging as an environmental policy instrument. *Review of Environmental Economics and Policy*, *2473*(756), 31.

Carrillo, P., Pomeranz, D., & Singhal, M. (2017). Dodging the Taxman. *American Economic Journal: Applied Economics*, *9*(2), 144–164.

Cartwright, N., & Hardie, J. (2013). *Evidence-based policy: a practical guide to doing it better* (Vol. 50). Oxford University Press. https://doi.org/10.5860/choice.50-5831

Casal, S., Kogler, C., Mittone, L., & Kirchler, E. (2016). Tax compliance depends on voice of taxpayers. *Journal of Economic Psychology*, *56*(June), 141–150. https://doi.org/10.1016/j.joep.2016.06.005

Castro, L., & Scartascini, C. (2015). Tax compliance and enforcement in the pampas evidence

from a field experiment. *Journal of Economic Behavior and Organization*, *116*, 65–82. https://doi.org/10.1016/j.jebo.2015.04.002

Chantrill, C. (2021). *Total US Government Spending*. US Government Spending. http://www.usgovernmentspending.com/

Chirico, M., Inman, R. P., 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

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

Coleman, S. (1996). The Minnesota income tax compliance experiment: State tax results. *MPRA Paper*, *4*(April), 528–541. Retrieved from http://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

Cullen, J., Turner, N., & Washington, E. (2018). *Political Alignment, Attitudes Toward Government, and Tax Evasion* (NBER Working Paper No. 24323). National Bureau of Economic Research. https://doi.org/10.3386/w24323

Cummings, R. G., Martinez-Vazquez, J., McKee, M., & Torgler, B. (2009). Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment. *Journal of Economic Behavior and Organization*, *70*(3), 447–457. https://doi.org/10.1016/j.jebo.2008.02.010

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. (2014). *Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru* [Unpublished Manuscript], Princeton University.

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, *33*(1), 264–277. https://doi.org/10.1016/j.joep.2011.10.009

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

Erard, B., & Feinstein, J. S. (1994). The Role of Moral Sentiments and Audit Perceptions in Tax Compliance. *Public Finance/Finances Publiques*, *49*(August), 70–89.

Erard, B. (1997). Self-selection with measurement errors a microeconometric analysis of the decision to seek tax assistance and its implications for tax compliance. *Journal of Econometrics*, *81*(2), 319–356. https://doi.org/10.1016/S0304-4076(97)86570-0

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315. https://doi.org/10.1016/j.geb.2005.03.001

Feld, L., Frey, B., & Torgler, B. (2006). *Rewarding honest taxpayers? Evidence on the impact of rewards from field experiments* (CREMA Working Paper No. 2006-06). CREMA, Center of Research. http://ww.w.crema-research.ch/papers/2006-16.pdf

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. L., Levin, B., & Paik, M. C. (2004). The Measurement of Interrater Agreement. In J. L. Fleiss, B. Levin, & M. C. Paik (Eds.), *Statistical Methods for Rates and Proportions* (pp. 598–626). Wiley Online Library. https://doi.org/10.1002/0471445428.ch18

Frey, B. S. (1997). A constitution for knaves crowds out civic virtues. *Economic Journal*, *107*(443), 1043–1053. https://doi.org/10.1111/j.1468-0297.1997.tb00006.x

Gillitzer, C., & Sinning, M. (2020). Nudging businesses to pay their taxes: Does timing matter? *Journal of Economic Behavior and Organization*, *169*, 284–300. https://doi.org/10.1016/j.jebo.2019.11.020

Gordon, J. P. F. (1989). Individual Morality and Reputation Costs as Deterrents to Tax Evasion. *European Economic Review, 33*, 797-805.

Gould, S. J. (1997). *Full House: The Spread of Excellence from Plato to Darwin.* (Vol. 314). Harmony Rooks. https://doi.org/10.1136/bmj.314.7082.761a

Greenacre, M. (2016). *Correspondence Analysis in Practice.* Taylor & Francis Inc. https://doi.org/10.1201/9781315369983

Grimshaw, J. M., Eccles, M. P., Lavis, J. N., Hill, S. J., & Squires, J. E. (2013). Knowledge translation of research findings. *Implementation Science*, *409*(8856), 1–17.

Guiso, L., Sapienza, P., & Zingales, L. (2006). Does Culture Affect Economic Outcomes? *The Journal of Economic Perspectives*, *20*(2), 23–48.

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

Hansen, P. G. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation*, *7*(1), 155–174. https://doi.org/10.1017/S1867299X00005468

Harju, J., Kosonen, T., & Ropponen, O. (2018). *Do honest hairdressers get a haircut? On tax rate and tax evasion*. Government Institute for Economic Research (VATT). https://www.ntanet.org/wp-content/uploads/proceedings/2014/143-kosonen-harju-ropponen-honest-hairdressers-get-haircut-tax.pdf

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., Korczyc, 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*. Australian Taxation Office.

http://creativecommons.org/licensesby/4.0/deed.enhttp://www.itsanhonour.gov.au/coat-arms

Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics* , *80*(March), 47–58. https://doi.org/10.1016/j.socec.2019.03.005

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Morano, J., … Puranen, B. (2014). *World Values Survey: Round Five - Country-Pooled Datafile Version*. World Values Survey. https://www.worldvaluessurvey.org/WVSDocumentationWV5.jsp.

Iyer, G. S., Reckers, P. M. J., & Sanders, D. L. (2010). 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

John, P., & Blume, T. (2018). How best to nudge taxpayers? The impact of message simplification and descriptive social norms on payment rates in a central London local authority. *Journal of Behavioral Public Administration*, *1*(1). https://doi.org/10.30636/jbpa.11.10

Jones, J. M., & Brenan M. (2020). *Americans' Satisfaction With Health Costs at New High*. Gallup. https://news.gallup.com/poll/327686/americans-satisfaction-health-costs-new-high.aspx

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.4324/9780203070710.pt2

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*(4), 341–350. https://doi.org/10.1017/CBO9780511803475.002

Kettle, S., Hernandez, M., Ruda, S., & Sanders, M. (2016). *Behavioral Interventions in Tax Compliance - Evidence from Guatemala*. World Bank Policy Research. https://ssrn.com/abstract=2811337

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. (2007). *The economic psychology of tax behaviour*. Cambridge University Press. https://doi.org/10.1017/CBO9780511628238

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., Muehlbacher, S., Hoelzl, E., & Webley, P. (2009). Effort and aspirations in tax evasion: Experimental evidence. *Applied Psychology*, *58*(3), 488–507. https://doi.org/10.1111/j.1464-0597.2009.00403.x

Kirchler, E., Muehlbacher, S., Kastlunger, B., & Wahl, I. (2010). Why pay taxes?: A review of tax compliance decisions. In J. Alm, J. Martinez-Vazquez, & B. Torgler (Eds.), *Developing Alternative Frameworks for Explaining Tax Compliance* (pp. 15–31). Routledge. https://doi.org/10.4324/9780203851616

Klepper, S., Mazur, M., & Nagin, D. (1991). Expert Intermediaries and Legal Compliance: The Case of Tax Preparers. *The Journal of Law and Economics*, *34*(1), 205–229. https://doi.org/10.1086/467224

Klepper, S., & Nagin, D. (1989). Tax Compliance and Perceptions of the Risks of Detection and Criminal Prosecution. *Law & Society Review*, *23*(2), 209. https://doi.org/10.2307/3053715

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

Kleven, H. J. (2014). How can scandinavians tax so much? *Journal of Economic Perspectives*, *28*(4), 77–98. https://doi.org/10.1257/jep.28.4.77

Koessler, A. K., Torgler, B., Feld, L. P., & Frey, B. S. (2019). Commitment to pay taxes: Results from field and laboratory experiments. *European Economic Review*, *115*, 78–98. https://doi.org/10.1016/j.euroecorev.2019.02.006

Lapinski, M. K., & Rimal, R. N. (2005). An Explication of Social Norms. *Communication Theory*, *15*(2), 127–147. https://doi.org/10.1111/j.1468-2885.2005.tb00329.x

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

Leonidou, L. C., Talias, M. A., & Leonidou, C. N. (2008). Exercised power as a driver of trust and commitment in cross-border industrial buyer-seller relationships. *Industrial Marketing Management*, *37*(1), 92–103. https://doi.org/10.1016/j.indmarman.2007.08.006

Luttmer, E. F. P., & Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives*,

28(4), 149–168. https://doi.org/10.1257/jep.28.4.149

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

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7). https://doi.org/10.1371/journal.pmed.1000097

Muehlbacher, S., & Kirchler, E. (2016). About the External Validity of Laboratory Experiments in Tax Compliance Research. *Die Betriebswirtschaft, 76*(1), 7– 19.

OECD (2013). *What drives tax morale?* OECD Publishing. https://www.oecd.org/tax/tax-global/what-drives-tax-morale.pdf

OECD (2019). *What Motivates Tax Morale in Individuals and Businesses?* OECD Publishing. https://doi.org/10.1787/f3d8ea10-en

Ortega, D., & Sanguinetti, P. (2013). Deterrence and reciprocity effects on tax compliance: experimental evidence from Venezuela. *CAF Working Paper*. https://doi.org/10.1016/j.bbapap.2013.06.007

Ortiz-Ospina, E., & Roser, M. (2016). *Trust*. Our World in Data. https://ourworldindata.org/trust

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

Plumley, A. H. (2002). The impact of the IRS on voluntary tax compliance: Preliminary empirical results. In *Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association* (pp. 355–363). National Tax Association.

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

Rau, P.-L. P., Guo, Z., Qie, N., Lei, X., & Zhang, A. (2020). The Cognitive Bias in Cross-Cultural Design. In E. Vanderheiden, C.-H. Mayer (Eds.), *Mistakes, Errors, and Failures across Cultures* (pp. 455–475). Springer. https://doi.org/10.1007/978-3-030-35574-6

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

Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, *69*, 172–177. https://doi.org/10.1016/j.jesp.2016.09.009

Scholz, J. T., & Pinney, N. (1995). Duty, Fear, and Tax Compliance: The Heuristic Basis of Citizenship Behavior. *American Journal of Political Science, 39*(2), 490-512. https://doi.org/10.2307/2111622

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: Papers & Proceedings*, *107*(5), 420–424.

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, *57*(4), 904–954. https://doi.org/10.1257/jel.20181437

Slemrod, J., & Velayudhan, T. (2017). Do Firms Remit at Least 85 Percent of Tax Everywhere? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2970229

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

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press. https://doi.org/10.2307/j.ctt1t89fcp.13

Torgler, B. (2004). Moral suasion: An alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland. *Economics of Governance*, *5*(3), 235–253. https://doi.org/10.1007/s10101-004-0077-7

Traxler, C. (2010). Social Norms and Conditional Cooperative Taxpayers. *European Journal of Political Economy*, *26*(1), 89–103.

Tyran, J.-R., & Feld, L. P. (2001). *Why People Obey the Law: Experimental Evidence from the Provision of Public Goods* (CESifo Working Paper Series No. 651). CESifo. https://doi.org/10.2307/j.ctv1j66769

Vainre, M., Aaben, L., Paulus, A., Koppel, H., Tammsaar, H., Telve, K., … Uusberg, A. (2020). Nudging towards tax compliance: A fieldwork-informed randomised controlled trial. *Journal of Behavioral Public Administration*, *3*(1), 1–10.

https://doi.org/10.30636/jbpa.31.84

Wahl, I., Kastlunger, B., & Kirchler, E. (2010). Trust in authorities and power to enforce tax compliance: An empirical analysis of the "slippery slope framework." *Law and Policy*, *32*(4), 383–406. https://doi.org/10.1111/j.1467-9930.2010.00327.x

Webley, P., Robben, H., Elffers, H., & Hessing, D. (1991). *Tax evasion. An experimental approach*. Cambrige University Press.

Wenzel, M. (2005). Misperceptions of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology*, *26*(6), 862–883. https://doi.org/10.1016/j.joep.2005.02.002

Appendix

Appendix A

Table 3

Description of Sub-Categories

Category	Nr. of Studies	Nr. of Observations	Description
Outcome			
Extensive Margin	21	92	Probability to pay, file, or report taxes in any given time horizon without a deadline
Intensive Margin	25	127	Amount of reported taxes or reported income
Tax Deductions / Claims	5	22	Expenses deducted from gross revenues subject to tax or credit claims due to specific tax programmes
Timely Payment	10	54	Probability to pay, file, or report taxes within a deadline of up to 2 months
Other	4	10	Other compliance indicators (e.g., a compliance indicator obtained from different taxes; Del Carpio, 2014)
Nudge			
Information	6	14	Intervention contains some form of information regarding the tax system (e.g., the availability of customer service; Coleman, 1996)
Moral Appeal	12	36	Interventions contain some form of appeal to the taxpayer's morale or honesty (e.g., "For democracy to work, all citizens need to pay their fair share of taxes for community services"; Chirico et al., 2019)
Public Goods	16	50	Intervention contains information on how the tax money is used (e.g., "Your income tax dollars are spent on services that we Minnesotans depend on"; Blumenthal et al., 2001)
Reward	5	44	Intervention contains a reward or incentive (e.g., "If you pay on time, you will be automatically entered in a lottery to win a year free of property tax payments"; Dunning et al., 2015)
Social Norm	19	61	Intervention contains a social norm (e.g., "Nine out of ten people pay their tax on time"; Hallsworth et al., 2017)
Simplification	8	20	Intervention contains a simplified presentation of information or the introduction of visual stimuli (e.g., by using colours, reducing the amount of information etc.; see, for example, De Neve et al., 2019)
Interaction	3	51	Different nudging elements used in one intervention (see, for example, Vainre et al., 2020)
Simple Letter	7	14	Intervention does not contain a nudging element
Simple Reminder	11	38	Intervention contains a reminder regarding the obligation to pay taxes without containing a nudging element (e.g., "We want to remind you that the second payment

			of property taxes is due in July", De Neve et al., 2019)
Other	6	25	Intervention contains some other form of nudging element (e.g., framing, active choice etc.)
Population			
Company	9	41	Smaller to larger companies with several employees
Company / Non-Compliant	2	14	Smaller to larger companies with several employees that have been classified as 'at risk' or have missed payments in the past
Individual	13	63	Individual taxpayers unaffiliated with some form of business
Individual / Non-Compliant	16	124	Individual taxpayers unaffiliated with some form of business that have been classified as 'at risk' or have missed payments in the past
Self-Employed	7	31	Sole proprietors or 'micro-enterprises' with few employees
Self-Employed / Non-Compliant	3	32	Sole proprietors or 'micro-enterprises' with few employees that have been classified as 'at risk' or have missed payments in the past
Region			
Africa	2	9	Studies conducted in African countries
Australia / New Zealand	4	18	Studies conducted in either Australia or New Zealand
Middle East	1	2	Studies conducted in Middle Eastern countries
North America	10	66	Studies conducted in the US or Canada
South America	9	76	Studies conducted in Middle or South American countries
Northern Europe	2	13	Studies conducted in Scandinavian or Baltic countries
South / Eastern Europe	2	35	Studies conducted in Slavic or Balkan countries
Western Europe	10	86	Studies conducted in Middle or Western European countries
Tax			
Business & Occupation Tax (B&O Tax)	1	1	Excise tax on gross revenues in the Washington state (see Iyer et al., 2010)
Church Levy / Tax	2	32	Legal tax obligation for all members of the Catholic and Protestant churches in Germany or Austria (see Dwenger et al., 2016)
Council Tax	2	6	Taxes administered at the subnational level (e.g., in the UK; see Larkin et al., 2019)
Foreign Income Tax	1	9	Self-report of foreign income in Norway (Bott et al., 2019)
Income Tax	19	145	Tax imposed on individuals or legal entities in respect of income or profits earned (including corporate and personal income tax)

Property Tax	5	45	Tax on the value of a property (see Chirico et al., 2016)
TV License Fee	1	5	Payment required for the reception of television broadcasts in various countries (see Fellner et al., 2013)
Use Tax	1	1	Form of a sales tax levied in the US (see Iyer et al., 2010)
VAT	6	31	Tax levied on the price of a product or service at each stage of production, distribution, or sale to the end consumer (see Ariel, 2012)
Other	3	30	Other forms of taxes (e.g., health or pension payments; Alm et al., 2019)
Effectivity			
Negative Effect	8	24	The intervention has a significant negative or backfiring effect on the outcome of interest ($p < .05$)
Zero Effect	36	179	The intervention has no significant effect on the outcome of interest ($p < .05$)
Positive Effect	29	102	The intervention has a significant positive effect on the outcome of interest ($p < .05$)
Compared against			
No Letter	27	153	Control group that did not receive any intervention
Simple Letter	15	67	Control group that did receive an intervention without nudging elements
Simple Reminder	5	38	Control group that did receive a reminder without nudging elements
Other	5	47	Control group that did receive an intervention containing nudging elements

Note. Column 2 shows the frequency of studies containing the sub-category. Column 3 shows the frequency of observation of each sub-category.

Appendix B

Table 4A

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

Nudge	Mix_Neg	Mix_Zero	Mix_Pos	NonCom_Neg	NonCom_Zero	NonCom_Pos
Information	1	24	6	3	18	14
PsychCost	1	21	4	5	31	34
Reciprocity	2	37	13	9	22	11
Simplification	0	5	4	0	2	9

Note. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 5A

Frequency Distribution of Combined Non-Deterrence Interventions in the Mixed and Non-Compliant Sample measured at the Extensive Margin

Nudge	Mix_Neg	Mix_Zero	Mix_Pos	NonCom_Neg	NonCom_Zero	NonCom_Pos
Information	0	9	4	1	5	11
PsychCost	1	8	1	3	17	23
Reciprocity	0	16	3	7	12	6
Simplification	0	2	3	0	0	6

Note. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 6A

Frequency Distribution of Combined Non-Deterrence Interventions in the Mixed and Non-Compliant Sample measured at the Intensive Margin

Nudge	Mix_Neg	Mix_Zero	Mix_Pos	NonCom_Neg	NonCom_Zero	NonCom_Pos
Information	0	13	2	1	10	0
PsychCost	0	11	3	1	13	7
Reciprocity	2	21	9	2	8	4
Simplification	0	3	1	0	2	2

Note. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 7A

Frequency Distribution of Combined Non-Deterrence Interventions in the Mixed and Non-Compliant Sample compared to a No Letter Control Group

Nudge	Mix_Neg	Mix_Zero	Mix_Pos	NonCom_Neg	NonCom_Zero	NonCom_Pos
Information	1	20	6	0	14	13
PsychCost	0	11	1	0	14	9
Reciprocity	1	18	11	0	8	2
Simplification	0	3	3	0	2	6

Note. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 8A

Frequency Distribution of Combined Non-Deterrence Interventions in the Mixed and Non-Compliant Sample compared to a Simple Letter Control Group

Nudge	Mix_Neg	Mix_Zero	Mix_Pos	NonCom_Neg	NonCom_Zero	NonCom_Pos
Information	0	2	0	0	0	1
PsychCost	1	10	3	2	13	25
Reciprocity	1	11	2	2	8	8
Simplification	0	2	1	0	0	3

Note. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 9A

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

Nudge	Mix_Neg	Mix_Zero	Mix_Pos	NonCom_Neg	NonCom_Zero	NonCom_Pos
Information	0	8	0	1	4	1
Moral Appeal	0	9	3	1	14	9
Reward	0	17	5	6	13	9
Public Goods	2	20	8	3	9	2
Simplification	0	5	4	0	2	9
Social Norm	1	12	1	4	17	25
Standard Letter	1	16	6	2	14	13

Note. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Table 10A

Frequency Distribution of Individual Non-Deterrence Interventions in the Mixed and Non-Compliant Sample measured at the Extensive Margin

Nudge	Mix_Neg	Mix_Zero	Mix_Pos	NonCom_Neg	NonCom_Zero	NonCom_Pos
Information	0	3	0	0	1	1
Moral Appeal	0	2	0	1	7	5
Reward	0	7	2	4	8	5
Public Goods	0	9	1	3	4	1
Simplification	0	2	3	0	0	6
Social Norm	1	6	1	2	10	18
Standard Letter	0	6	4	1	4	10

Note. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Table 11A

Frequency Distribution of Individual Non-Deterrence Interventions in the Mixed and Non-Compliant Sample measured at the Intensive Margin

Nudge	Mix_Neg	Mix_Zero	Mix_Pos	NonCom_Neg	NonCom_Zero	NonCom_Pos
Information	0	5	0	0	1	0
Moral Appeal	0	5	3	0	7	2
Reward	0	10	2	2	4	4
Public Goods	2	11	7	0	4	0
Simplification	0	3	1	0	2	2
Social Norm	0	6	0	1	6	5
Standard Letter	0	8	2	1	9	0

Note. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Table 12A

Frequency Distribution of Individual Non-Deterrence Interventions in the Mixed and Non-Compliant Sample compared against a No Letter Control Group

Nudge	Mix_Neg	Mix_Zero	Mix_Pos	NonCom_Neg	NonCom_Zero	NonCom_Pos
Information	0	4	0	0	0	0
Moral Appeal	0	5	0	0	9	3
Reward	0	13	5	0	5	1
Public Goods	1	5	6	0	3	1
Simplification	0	3	3	0	2	6
Social Norm	0	6	1	0	5	6
Standard Letter	1	16	6	0	14	13

Note. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Table 13A

Frequency Distribution of Individual Non-Deterrence Interventions in the Mixed and Non-Compliant Sample compared against a Simple Letter Control Group

Nudge	Mix_Neg	Mix_Zero	Mix_Pos	NonCom_Neg	NonCom_Zero	NonCom_Pos
Information	0	2	0	0	0	1
Moral Appeal	0	4	3	1	1	6
Reward	0	4	0	2	8	8
Public Goods	1	7	2	0	0	0
Simplification	0	2	1	0	0	3
Social Norm	1	6	0	1	12	19

Note. Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Table 14A

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

Nudge	C_Neg	C_Zero	C_Pos	I_Neg	I_Zero	I_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	15	5	3	17	8	1	10	7
PsychCost	0	12	5	6	31	16	0	9	17
Reciprocity	2	7	6	9	43	10	0	9	8
Simplification	0	2	0	0	5	13	0	0	0

Note. C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 15A

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

Nudge	C_Neg	C_Zero	C_Pos	I_Neg	I_Zero	I_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	5	4	1	6	7	0	3	4
PsychCost	0	4	2	4	18	9	0	3	13
Reciprocity	0	3	1	7	24	5	0	1	3
Simplification	0	1	0	0	1	9	0	0	0

Note. C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 16A

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

Nudge	C_Neg	C_Zero	C_Pos	I_Neg	I_Zero	I_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	8	0	1	8	0	0	7	2
PsychCost	0	8	1	1	11	6	0	5	3
Reciprocity	2	4	4	2	17	4	0	8	5
Simplification	0	1	0	0	4	3	0	0	0

Note. C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 17A

Frequency Distribution of Combined Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers compared against a No Letter Control Group

Nudge	C_Neg	C_Zero	C_Pos	I_Neg	I_Zero	I_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	13	5	0	11	8	1	10	6
PsychCost	0	7	4	0	12	3	0	6	3
Reciprocity	1	6	4	0	16	3	0	4	6
Simplification	0	0	0	0	5	9	0	0	0

Note. C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 18A

Frequency Distribution of Combined Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers compared against a Simple Letter Control Group

Nudge	C_Neg	C_Zero	C_Pos	I_Neg	I_Zero	I_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	2	0	0	0	0	0	0	1
PsychCost	0	5	1	3	15	13	0	3	14
Reciprocity	1	1	2	2	13	6	0	5	2
Simplification	0	2	0	0	0	4	0	0	0

Note. C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 19A

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

Nudge	C_Neg	C_Zero	C_Pos	I_Neg	I_Zero	I_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	2	0	1	8	0	0	2	1
Moral Appeal	0	6	2	1	11	6	0	6	4
Reward	0	5	2	6	21	7	0	4	5
Public Goods	2	2	4	3	22	3	0	5	3
Simplification	0	2	0	0	5	13	0	0	0
Social Norm	0	6	3	5	20	10	0	3	13
Standard Letter	0	13	5	2	9	8	1	8	6

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

Table 20A

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

Nudge	C_Neg	C_Zero	C_Pos	I_Neg	I_Zero	I_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	1	0	0	2	0	0	1	1
Moral Appeal	0	2	0	1	6	4	0	1	1
Reward	0	3	1	4	11	3	0	1	3
Public Goods	0	0	0	3	13	2	0	0	0
Simplification	0	1	0	0	1	9	0	0	0
Social Norm	0	2	2	3	12	5	0	2	12
Standard Letter	0	4	4	1	4	7	0	2	3

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

Table 21A

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

Nudge	C_Neg	C_Zero	C_Pos	I_Neg	I_Zero	I_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	1	0	0	4	0	0	1	0
Moral Appeal	0	4	1	0	4	2	0	4	2
Reward	0	2	0	2	9	4	0	3	2
Public Goods	2	2	4	0	8	0	0	5	3
Simplification	0	1	0	0	4	3	0	0	0
Social Norm	0	4	0	1	7	4	0	1	1
Standard Letter	0	7	0	1	4	0	0	6	2

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

Table 22A

Frequency Distribution of Individual Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers compared against a No Letter Control Group

Nudge	C_Neg	C_Zero	C_Pos	I_Neg	I_Zero	I_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	0	0	0	2	0	0	2	0
Moral Appeal	0	3	1	0	6	1	0	5	1
Reward	0	5	2	0	10	1	0	3	3
Public Goods	1	1	2	0	6	2	0	1	3
Simplification	0	0	0	0	5	9	0	0	0
Social Norm	0	4	3	0	6	2	0	1	2
Standard Letter	0	13	5	0	9	8	1	8	6

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

Table 23A

Frequency Distribution of Individual Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers compared against a Simple Letter Control Group

Nudge	C_Neg	C_Zero	C_Pos	I_Neg	I_Zero	I_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	2	0	0	0	0	0	0	1
Moral Appeal	0	3	1	1	1	5	0	1	3
Reward	0	0	0	2	11	6	0	1	2
Public Goods	1	1	2	0	2	0	0	4	0
Simplification	0	2	0	0	0	4	0	0	0
Social Norm	0	2	0	2	14	8	0	2	11

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

Table 24A

Frequency Distribution of Combined Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers, including Compliance History

Nudge	CN_Neg	CN_Zero	CN_Pos	C_Neg	C_Zero	C_Pos	IN_Neg	IN_Zero	IN_Pos	I_Neg	I_Zero	I_Pos	SEN_Neg	SEN_Zero	SEN_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	4	2	0	11	3	3	10	7	0	7	1	0	4	5	1	6	2
PsychCost	0	5	3	0	7	2	5	20	16	1	11	0	0	6	15	0	3	2
Reciprocity	0	0	0	2	7	6	9	22	9	0	21	1	0	0	2	0	9	6
Simplification	0	0	0	0	2	0	0	2	9	0	3	4	0	0	0	0	0	0

Note. CN = Company / Non-Compliant. C = Company. IN = Individual / Non-Compliant. I = Individual. SEN = Self-Employed / Non-Compliant. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 25A

Frequency Distribution of Combined Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers, including Compliance History measured at the Extensive Margin

Nudge	CN_Neg	CN_Zero	CN_Pos	C_Neg	C_Zero	C_Pos	IN_Neg	IN_Zero	IN_Pos	I_Neg	I_Zero	I_Pos	SEN_Neg	SEN_Zero	SEN_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	1	1	0	4	3	1	3	6	0	3	1	0	1	4	0	2	0
PsychCost	0	1	1	0	3	1	3	13	9	1	5	0	0	3	13	0	0	0
Reciprocity	0	0	0	0	3	1	7	12	4	0	12	1	0	0	2	0	1	1
Simplification	0	0	0	0	1	0	0	0	6	0	1	3	0	0	0	0	0	0

Note. CN = Company / Non-Compliant. C = Company. IN = Individual / Non-Compliant. I = Individual. SEN = Self-Employed / Non-Compliant. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 26A

Frequency Distribution of Combined Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers, including Compliance History measured at the Intensive Margin

Nudge	CN_Neg	CN_Zero	CN_Pos	C_Neg	C_Zero	C_Pos	IN_Neg	IN_Zero	IN_Pos	I_Neg	I_Zero	I_Pos	SEN_Neg	SEN_Zero	SEN_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	3	0	0	5	0	1	4	0	0	4	0	0	3	0	0	4	2
PsychCost	0	4	0	0	4	1	1	6	6	0	5	0	0	3	1	0	2	2
Reciprocity	0	0	0	2	4	4	2	8	4	0	9	0	0	0	0	0	8	5
Simplification	0	0	0	0	1	0	0	2	2	0	2	1	0	0	0	0	0	0

Note. CN = Company / Non-Compliant. C = Company. IN = Individual / Non-Compliant. I = Individual. SEN = Self-Employed / Non-Compliant. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 27A

Frequency Distribution of Combined Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers, including Compliance History compared against a No Letter Control Group

Nudge	CN_Neg	CN_Zero	CN_Pos	C_Neg	C_Zero	C_Pos	IN_Neg	IN_Zero	IN_Pos	I_Neg	I_Zero	I_Pos	SEN_Neg	SEN_Zero	SEN_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	4	2	0	9	3	0	6	7	0	5	1	0	4	4	1	6	2
PsychCost	0	5	3	0	2	1	0	5	3	0	7	0	0	4	3	0	2	0
Reciprocity	0	0	0	1	6	4	0	8	2	0	8	1	0	0	0	0	4	6
Simplification	0	0	0	0	0	0	0	2	6	0	3	3	0	0	0	0	0	0

Note. CN = Company / Non-Compliant. C = Company. IN = Individual / Non-Compliant. I = Individual. SEN = Self-Employed / Non-Compliant. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 28A

Frequency Distribution of Combined Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers, including Compliance History compared against a Simple Letter Control Group

Nudge	CN_Neg	CN_Zero	CN_Pos	C_Neg	C_Zero	C_Pos	IN_Neg	IN_Zero	IN_Pos	I_Neg	I_Zero	I_Pos	SEN_Neg	SEN_Zero	SEN_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0
PsychCost	0	5	1	2	11	13	1	4	0	0	2	12	0	1	2	0	5	1
Reciprocity	1	1	2	2	8	6	0	5	0	0	0	2	0	5	0	1	1	2
Simplification	0	2	0	0	0	3	0	0	1	0	0	0	0	0	0	0	2	0

Note. CN = Company / Non-Compliant. C = Company. IN = Individual / Non-Compliant. I = Individual. SEN = Self-Employed / Non-Compliant. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 29A

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

Nudge	CN_Neg	CN_Zero	CN_Pos	C_Neg	C_Zero	C_Pos	IN_Neg	IN_Zero	IN_Pos	I_Neg	I_Zero	I_Pos	SEN_Neg	SEN_Zero	SEN_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	0	0	0	2	0	1	4	0	0	4	0	0	0	1	0	2	0
Moral Appeal	0	3	1	0	3	1	1	8	6	0	3	0	0	3	2	0	3	2
Public Goods	0	0	0	0	5	2	6	13	7	0	8	0	0	0	2	0	4	3
Reward	0	0	0	2	2	4	3	9	2	0	13	1	0	0	0	0	5	3
Simplification	0	0	0	0	2	0	0	2	9	0	3	4	0	0	0	0	0	0
Social Norm	0	2	2	0	4	1	4	12	10	1	8	0	0	3	13	0	0	0
Standard Letter	0	4	2	0	9	3	2	6	7	0	3	1	0	4	4	1	4	2

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

Table 30A

Frequency Distribution of Individual Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers, including Compliance History measured at the Extensive Margin

Nudge	CN_Neg	CN_Zero	CN_Pos	C_Neg	C_Zero	C_Pos	IN_Neg	IN_Zero	IN_Pos	I_Neg	I_Zero	I_Pos	SEN_Neg	SEN_Zero	SEN_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0
Moral Appeal	0	1	0	0	1	0	1	5	4	0	1	0	0	1	1	0	0	0
Public Goods	0	0	0	0	3	1	4	8	3	0	3	0	0	0	2	0	1	1
Reward	0	0	0	0	0	0	3	4	1	0	9	1	0	0	0	0	0	0
Simplification	0	0	0	0	1	0	0	0	6	0	1	3	0	0	0	0	0	0
Social Norm	0	0	1	0	2	1	2	8	5	1	4	0	0	2	12	0	0	0
Standard Letter	0	1	1	0	3	3	1	2	6	0	2	1	0	1	3	0	1	0

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

Table 31A

Frequency Distribution of Individual Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers, including Compliance History measured at the Intensive Margin

Nudge	CN_Neg	CN_Zero	CN_Pos	C_Neg	C_Zero	C_Pos	IN_Neg	IN_Zero	IN_Pos	I_Neg	I_Zero	I_Pos	SEN_Neg	SEN_Zero	SEN_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	0	0	0	1	0	0	1	0	0	3	0	0	0	0	0	1	0
Moral Appeal	0	2	0	0	2	1	0	3	2	0	1	0	0	2	0	0	2	2
Public Goods	0	0	0	0	2	0	2	4	4	0	5	0	0	0	0	0	3	2
Reward	0	0	0	2	2	4	0	4	0	0	4	0	0	0	0	0	5	3
Simplification	0	0	0	0	1	0	0	2	2	0	2	1	0	0	0	0	0	0
Social Norm	0	2	0	0	2	0	1	3	4	0	4	0	0	1	1	0	0	0
Standard Letter	0	3	0	0	4	0	1	3	0	0	1	0	0	3	0	0	3	2

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

Table 32A

Frequency Distribution of Individual Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers, including Compliance History compared against a No Letter Control Group

Nudge	CN_Neg	CN_Zero	CN_Pos	C_Neg	C_Zero	C_Pos	IN_Neg	IN_Zero	IN_Pos	I_Neg	I_Zero	I_Pos	SEN_Neg	SEN_Zero	SEN_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	4	2	0	9	3	0	6	7	0	3	1	0	4	4	1	4	2
Moral Appeal	0	0	0	1	1	2	0	3	1	0	3	1	0	0	0	0	1	3
Public Goods	0	0	0	0	5	2	0	5	1	0	5	0	0	0	0	0	3	3
Reward	0	2	2	0	2	1	0	2	2	0	4	0	0	1	2	0	0	0
Simplification	0	3	1	0	0	0	0	3	1	0	3	0	0	3	1	0	2	0
Social Norm	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0
Standard Letter	0	0	0	0	0	0	0	2	6	0	3	3	0	0	0	0	0	0

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

Table 33A

Frequency Distribution of Individual Non-Deterrence Interventions in the Samples of Company, Individual, and Self-Employed Taxpayers, including Compliance History compared against a Simple Letter Control Group

Nudge	CN_Neg	CN_Zero	CN_Pos	C_Neg	C_Zero	C_Pos	IN_Neg	IN_Zero	IN_Pos	I_Neg	I_Zero	I_Pos	SEN_Neg	SEN_Zero	SEN_Pos	SE_Neg	SE_Zero	SE_Pos
Information	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0
Moral Appeal	0	0	0	0	3	1	1	1	5	0	0	0	0	0	1	0	1	2
Public Goods	0	0	0	0	0	0	2	8	6	0	3	0	0	0	2	0	1	0
Reward	0	0	0	1	1	2	0	0	0	0	2	0	0	0	0	0	4	0
Simplification	0	0	0	0	2	0	0	0	3	0	0	1	0	0	0	0	0	0
Social Norm	0	0	0	0	2	0	1	10	8	1	4	0	0	2	11	0	0	0

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

Table 34A

Frequency Distribution of Combined Non-Deterrence Interventions in the Regional Sub-Categories

Nudge	AFR_Neg	AFR_Zero	AFR_Pos	AUS_Neg	AUS_Zero	AUS_Pos	EUR_Neg	EUR_Zero	EUR_Pos	NA_Neg	NA_Zero	NA_Pos	SA_Neg	SA_Zero	SA_Pos
Information	0	2	2	0	5	1	2	15	8	0	0	0	2	8	2
PsychCost	0	0	0	0	4	0	5	16	23	0	0	0	1	17	8
Reciprocity	0	2	2	0	0	0	9	22	8	0	1	1	0	15	3
Simplification	0	0	0	0	2	0	0	5	10	0	0	0	0	0	3

Note. AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects.

Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 35A

Frequency Distribution of Combined Non-Deterrence Interventions in the Regional Sub-Categories measured at the Extensive Margin

Nudge	AFR_Neg	AFR_Zero	AFR_Pos	AUS_Neg	AUS_Zero	AUS_Pos	EUR_Neg	EUR_Zero	EUR_Pos	NA_Neg	NA_Zero	NA_Pos	SA_Neg	SA_Zero	SA_Pos
Information	0	2	2	0	1	1	1	5	8	0	1	2	0	5	2
PsychCost	0	0	0	0	2	0	4	11	18	0	7	4	0	5	2
Reciprocity	0	2	2	0	0	0	7	12	4	0	6	1	0	8	2
Simplification	0	0	0	0	1	0	0	1	7	0	0	2	0	0	0

Note. AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects.

Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 36A

Frequency Distribution of Combined Non-Deterrence Interventions in the Regional Sub-Categories measured at the Intensive Margin

Nudge	AFR_ Neg	AFR_ Zero	AFR_ Pos	AUS_ Neg	AUS_ Zero	AUS_ Pos	EUR_ Neg	EUR_ Zero	EUR_ Pos	ME_ Neg	ME_ Zero	ME_ Pos	NA_ Neg	NA_ Zero	NA_ Pos	SA_ Neg	SA_ Zero	SA_ Pos
Information	0	0	0	0	2	0	1	9	0	0	0	0	0	5	0	0	7	2
PsychCost	0	0	0	0	2	0	1	3	5	0	0	0	0	9	4	0	10	1
Reciprocity	0	0	0	0	0	0	2	9	4	0	1	0	0	8	1	2	11	8
Simplification	0	0	0	0	1	0	0	4	3	0	0	0	0	0	0	0	0	0

Note. AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 37A

Frequency Distribution of Combined Non-Deterrence Interventions in the Regional Sub-Categories compared against a No Letter Control Group

Nudge	AFR_ Neg	AFR_ Zero	AFR_ Pos	AUS_ Neg	AUS_ Zero	AUS_ Pos	EUR_ Neg	EUR_ Zero	EUR_ Pos	ME_ Neg	ME_ Zero	ME_ Pos	NA_ Neg	NA_ Zero	NA_ Pos	SA_ Neg	SA_ Zero	SA_ Pos
Information	0	2	2	0	3	1	0	11	7	0	0	0	1	6	2	0	12	7
PsychCost	0	0	0	0	0	0	0	3	1	0	0	0	0	9	2	0	13	7
Reciprocity	0	2	2	0	0	0	0	3	1	0	1	1	0	9	1	1	11	8
Simplification	0	0	0	0	0	0	0	5	8	0	0	0	0	0	1	0	0	0

Note. AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 38A

Frequency Distribution of Combined Non-Deterrence Interventions in the Regional Sub-Categories compared against a Simple Letter Control Group

Nudge	AUS_Neg	AUS_Zero	AUS_Pos	EUR_Neg	EUR_Zero	EUR_Pos	NA_Neg	NA_Zero	NA_Pos	SA_Neg	SA_Zero	SA_Pos
Information	0	2	0	0	0	1	0	0	0	0	0	0
PsychCost	0	4	0	3	9	22	0	8	6	0	2	0
Reciprocity	0	0	0	2	6	7	0	5	1	1	8	2
Simplification	0	2	0	0	0	2	0	0	2	0	0	0

Note. AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Table 39A

Frequency Distribution of Individual Non-Deterrence Interventions in the Regional Sub-Categories

Nudge	AFR_Neg	AFR_Zero	AFR_Pos	AUS_Neg	AUS_Zero	AUS_Pos	EUR_Neg	EUR_Zero	EUR_Pos	NA_Neg	NA_Zero	NA_Pos	SA_Neg	SA_Zero	SA_Pos
Information	0	0	0	0	2	0	0	6	1	0	0	0	1	4	0
Moral Appeal	0	0	0	0	2	0	1	7	5	0	0	0	0	5	5
Public Goods	0	2	2	0	0	0	6	7	8	0	1	1	0	14	2
Reward	0	0	0	0	0	0	3	15	0	0	0	0	0	1	1
Simplification	0	0	0	0	2	0	0	5	10	0	0	0	0	0	3
Social Norm	0	0	0	0	2	0	4	9	18	0	0	0	1	12	3
Standard Letter	0	2	2	0	3	1	2	9	7	0	0	0	1	4	2

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

Table 40A

Frequency Distribution of Individual Non-Deterrence Interventions in the Regional Sub-Categories measured at the Extensive Margin

Nudge	AFR_Neg	AFR_Zero	AFR_Pos	AUS_Neg	AUS_Zero	AUS_Pos	EUR_Neg	EUR_Zero	EUR_Pos	NA_Neg	NA_Zero	NA_Pos	SA_Neg	SA_Zero	SA_Pos
Information	0	0	0	0	1	0	0	3	1	0	0	0	0	0	0
Moral Appeal	0	0	0	0	1	0	1	4	3	0	2	2	0	2	0
Public Goods	0	2	2	0	0	0	4	5	4	0	6	1	0	2	0
Reward	0	0	0	0	0	0	3	7	0	0	0	0	0	6	2
Simplification	0	0	0	0	1	0	0	1	7	0	0	2	0	0	0
Social Norm	0	0	0	0	1	0	3	7	15	0	5	2	0	3	2
Standard Letter	0	2	2	0	0	1	1	2	7	0	1	2	0	5	2

Note. AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects.

Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Table 41A

Frequency Distribution of Individual Non-Deterrence Interventions in the Regional Sub-Categories measured at the Intensive Margin

Nudge	AFR_Neg	AFR_Zero	AFR_Pos	AUS_Neg	AUS_Zero	AUS_Pos	EUR_Neg	EUR_Zero	EUR_Pos	ME_Neg	ME_Zero	ME_Pos	NA_Neg	NA_Zero	NA_Pos	SA_Neg	SA_Zero	SA_Pos
Information	0	0	0	0	1	0	0	3	0	0	0	0	0	2	0	0	0	0
Moral Appeal	0	0	0	0	1	0	0	2	2	0	0	0	0	2	3	0	7	0
Public Goods	0	0	0	0	0	0	2	1	4	0	1	0	0	8	1	0	4	1
Reward	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	2	7	7
Simplification	0	0	0	0	1	0	0	4	3	0	0	0	0	0	0	0	0	0
Social Norm	0	0	0	0	1	0	1	1	3	0	0	0	0	7	1	0	3	1
Standard Letter	0	0	0	0	1	0	1	6	0	0	0	0	0	3	0	0	7	2

Note. AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects.

Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Table 42A

Frequency Distribution of Individual Non-Deterrence Interventions in the Regional Sub-Categories compared against a No Letter Control Group

Nudge	AFR_ Neg	AFR_ Zero	AFR_ Pos	AUS_ Neg	AUS_ Zero	AUS_ Pos	EUR_ Neg	EUR_ Zero	EUR_ Pos	ME_ Neg	ME_ Zero	ME_ Pos	NA_ Neg	NA_ Zero	NA_ Pos	SA_ Neg	SA_ Zero	SA_ Pos
Information	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0
Moral Appeal	0	0	0	0	0	0	0	2	0	0	0	0	0	3	1	0	9	2
Public Goods	0	2	2	0	0	0	0	3	1	0	1	1	0	9	1	0	3	1
Reward	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	8	7
Simplification	0	0	0	0	0	0	0	5	8	0	0	0	0	0	1	0	0	0
Social Norm	0	0	0	0	0	0	0	1	1	0	0	0	0	6	1	0	4	5
Standard Letter	0	2	2	0	3	1	0	9	7	0	0	0	1	4	2	0	12	7

Note. AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects.

Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Table 43A

Frequency Distribution of Individual Non-Deterrence Interventions in the Regional Sub-Categories compared against a Simple Letter Control Group

Nudge	AUS_Neg	AUS_Zero	AUS_Pos	EUR_Neg	EUR_Zero	EUR_Pos	NA_Neg	NA_Zero	NA_Pos	SA_Neg	SA_Zero	SA_Pos
Information	0	2	0	0	0	1	0	0	0	0	0	0
Moral Appeal	0	2	0	1	1	5	0	2	4	0	0	0
Public Goods	0	0	0	2	4	7	0	5	1	0	3	0
Reward	0	0	0	0	2	0	0	0	0	1	5	2
Simplification	0	2	0	0	0	2	0	0	2	0	0	0
Social Norm	0	2	0	2	8	17	0	6	2	0	2	0

Note. AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects.

Zero = Frequency of zero effects. Pos = Frequency of positive effects.
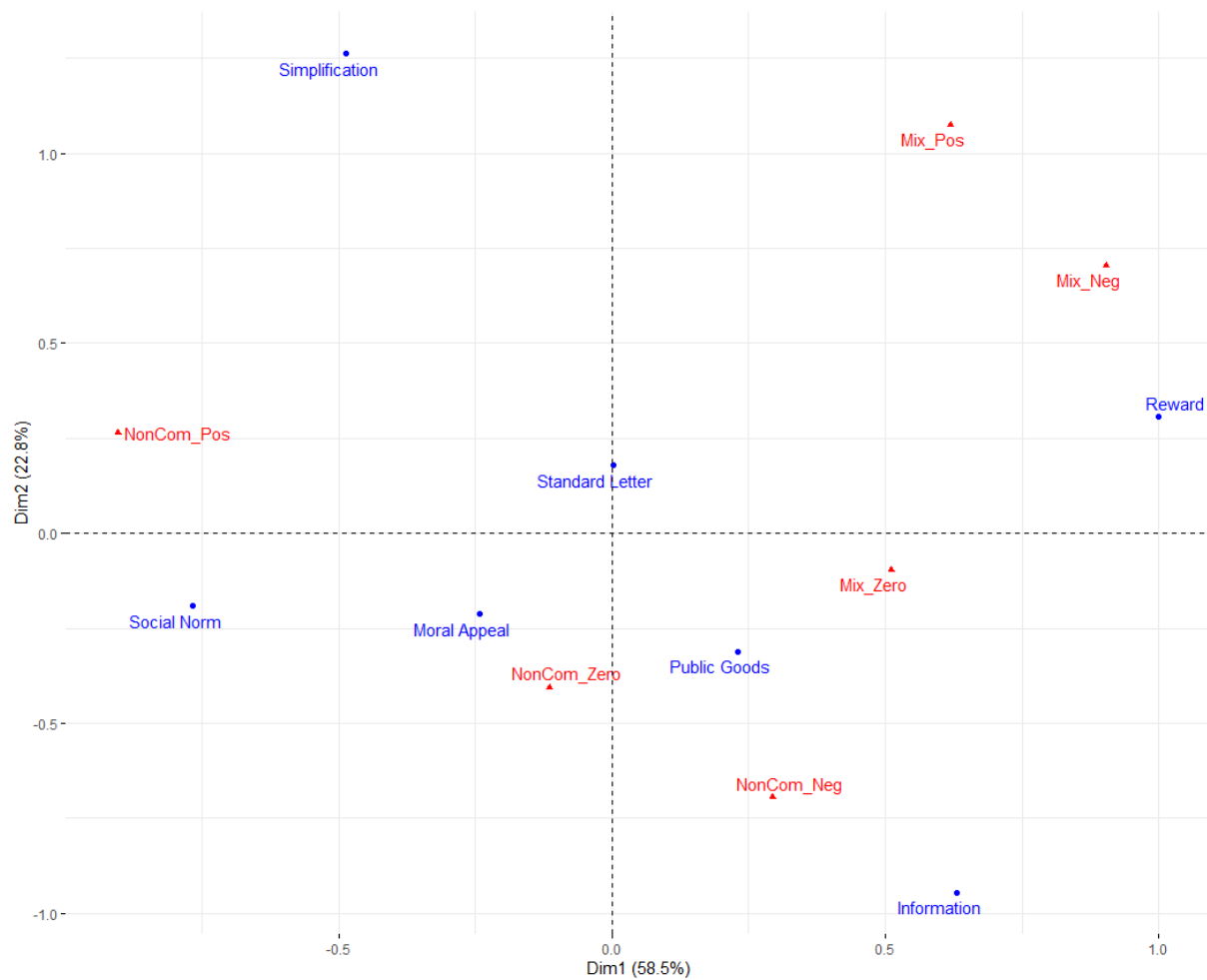
Appendix C

Figure 14A. Correspondence analysis for frequencies of individual non-deterrence interventions in Table 9A. Blue points represent the rows (i.e., the interventions). Red triangles represent the columns (i.e., the population sub-categories). Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.
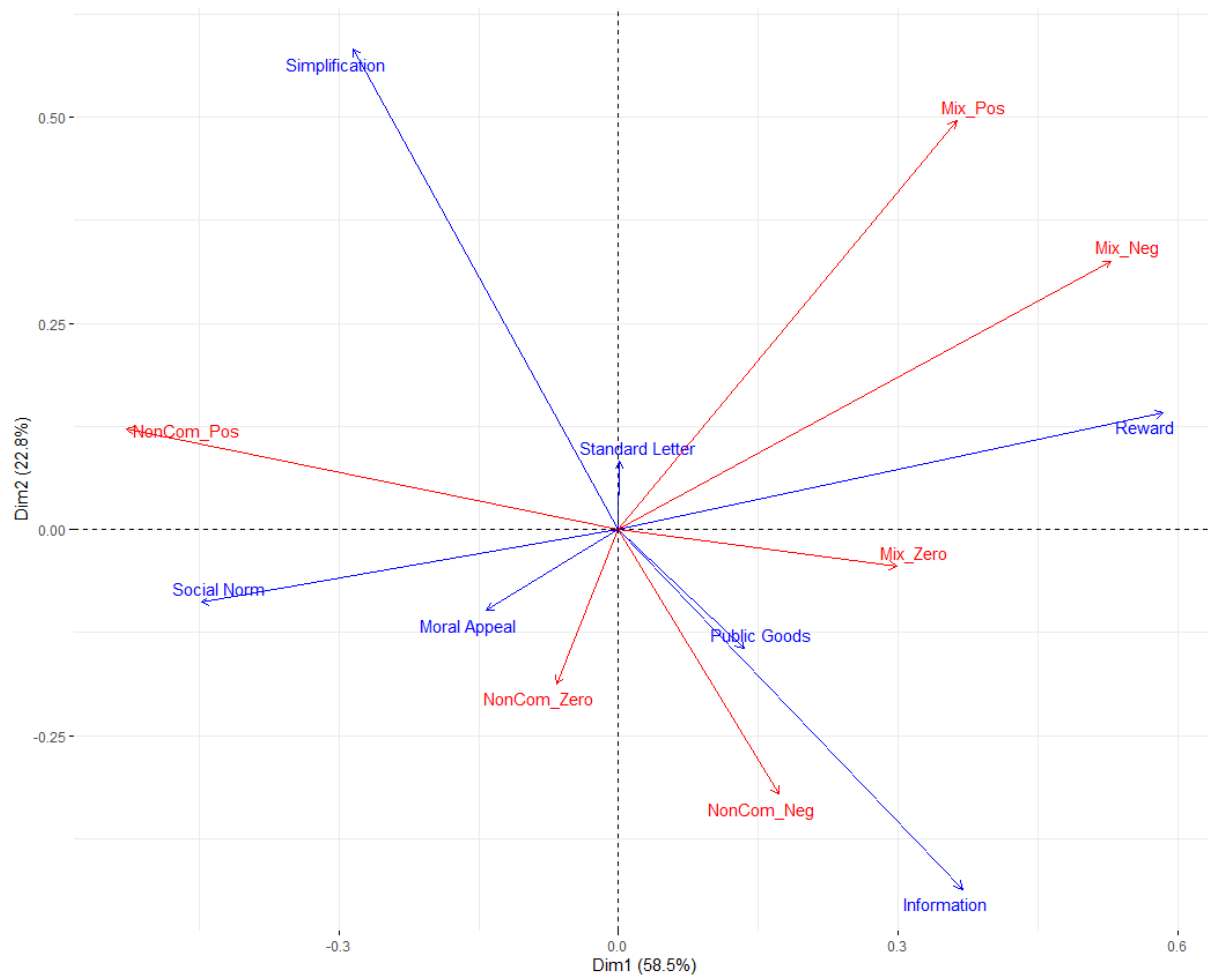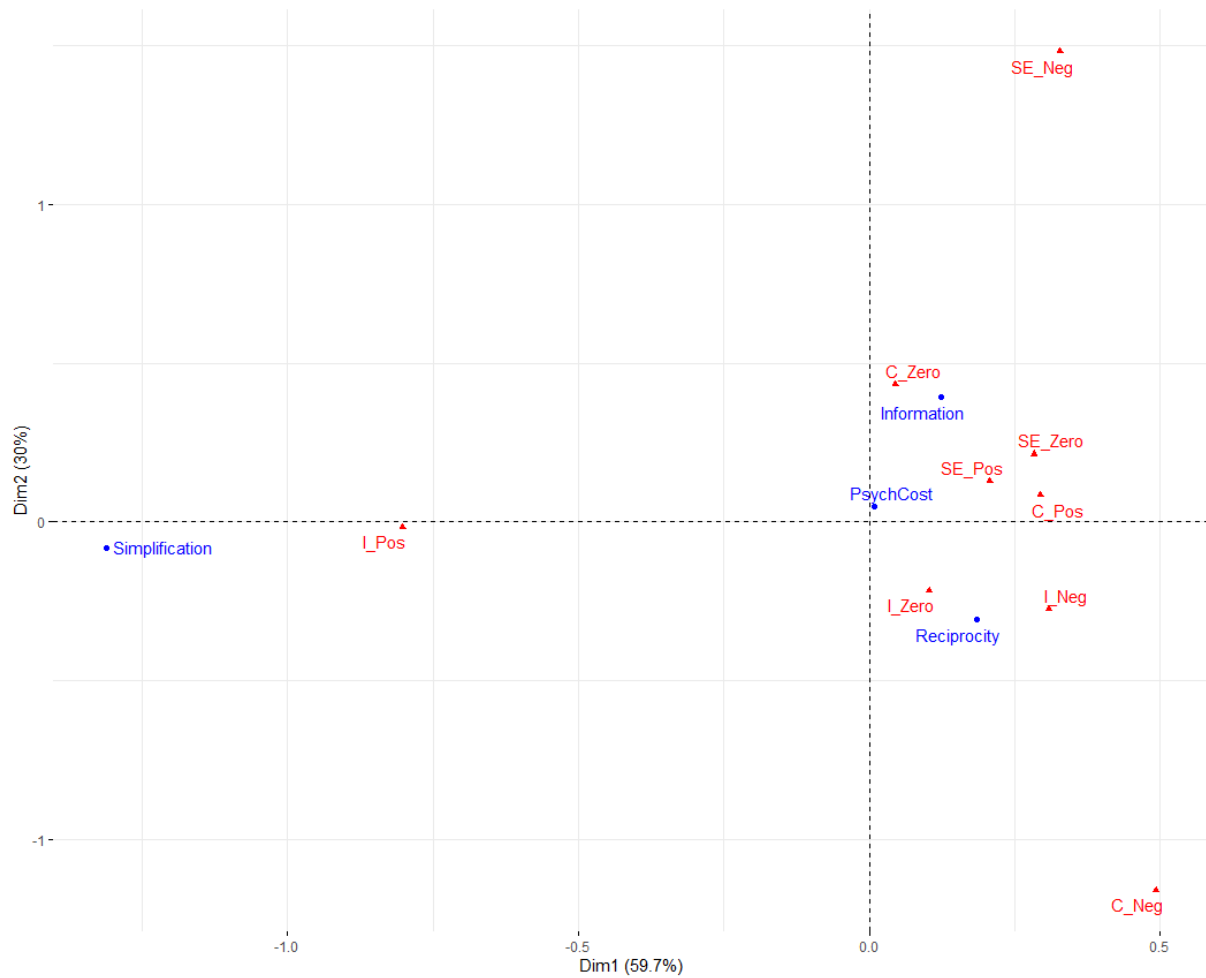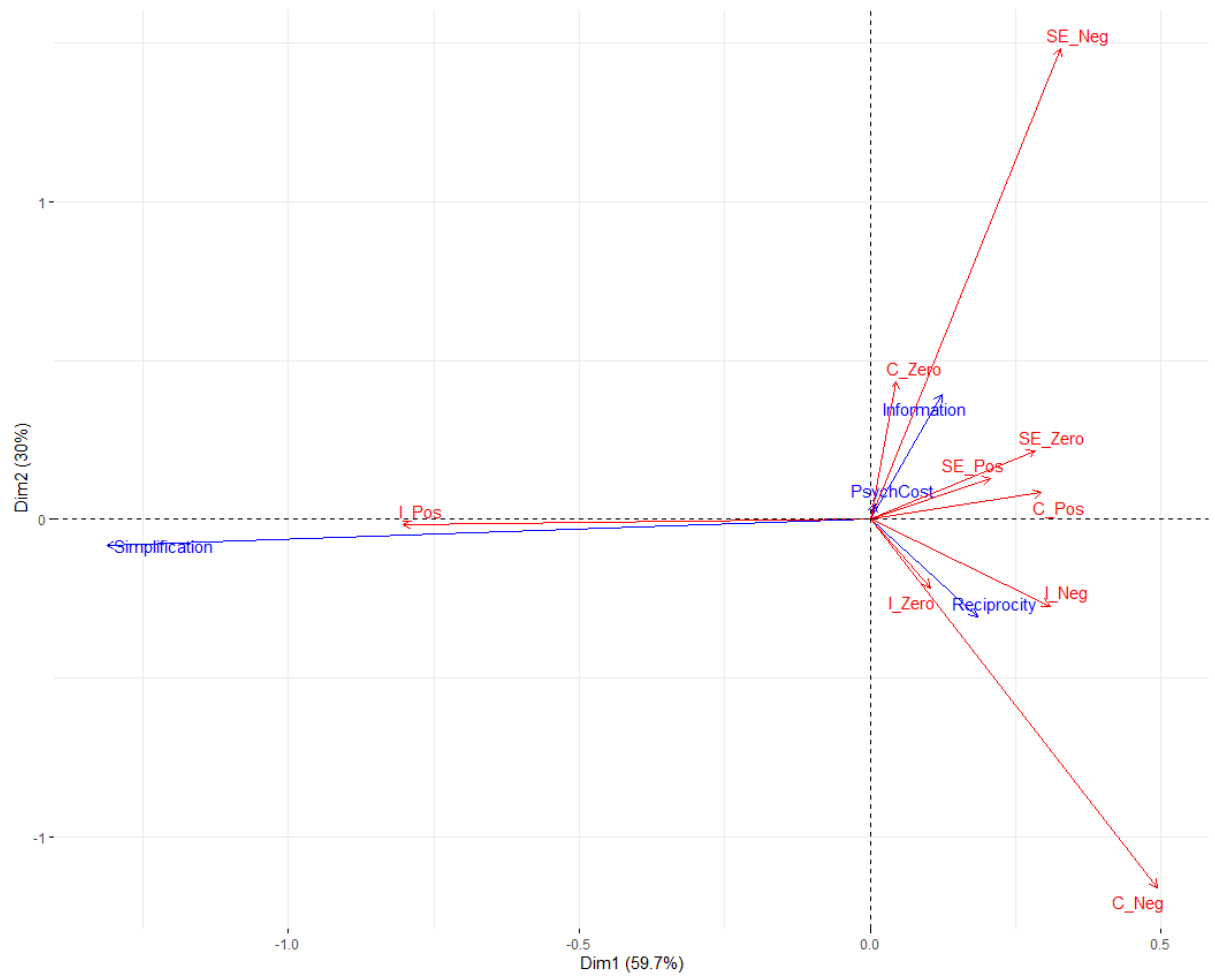
Figure 15A. Asymmetric correspondence analysis for frequencies of individual non-deterrence interventions in Table 9A. Blue lines represent the rows (i.e., the interventions). Red lines represent the columns (i.e., the population sub-categories). Mix = Mixed. NonCom = Non-Compliant. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Figure 16A. Correspondence analysis for frequencies of combined non-deterrence interventions in Table 14A. Blue points represent the rows (i.e., the interventions). Red triangles represent the columns (i.e., the population sub-categories). C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.
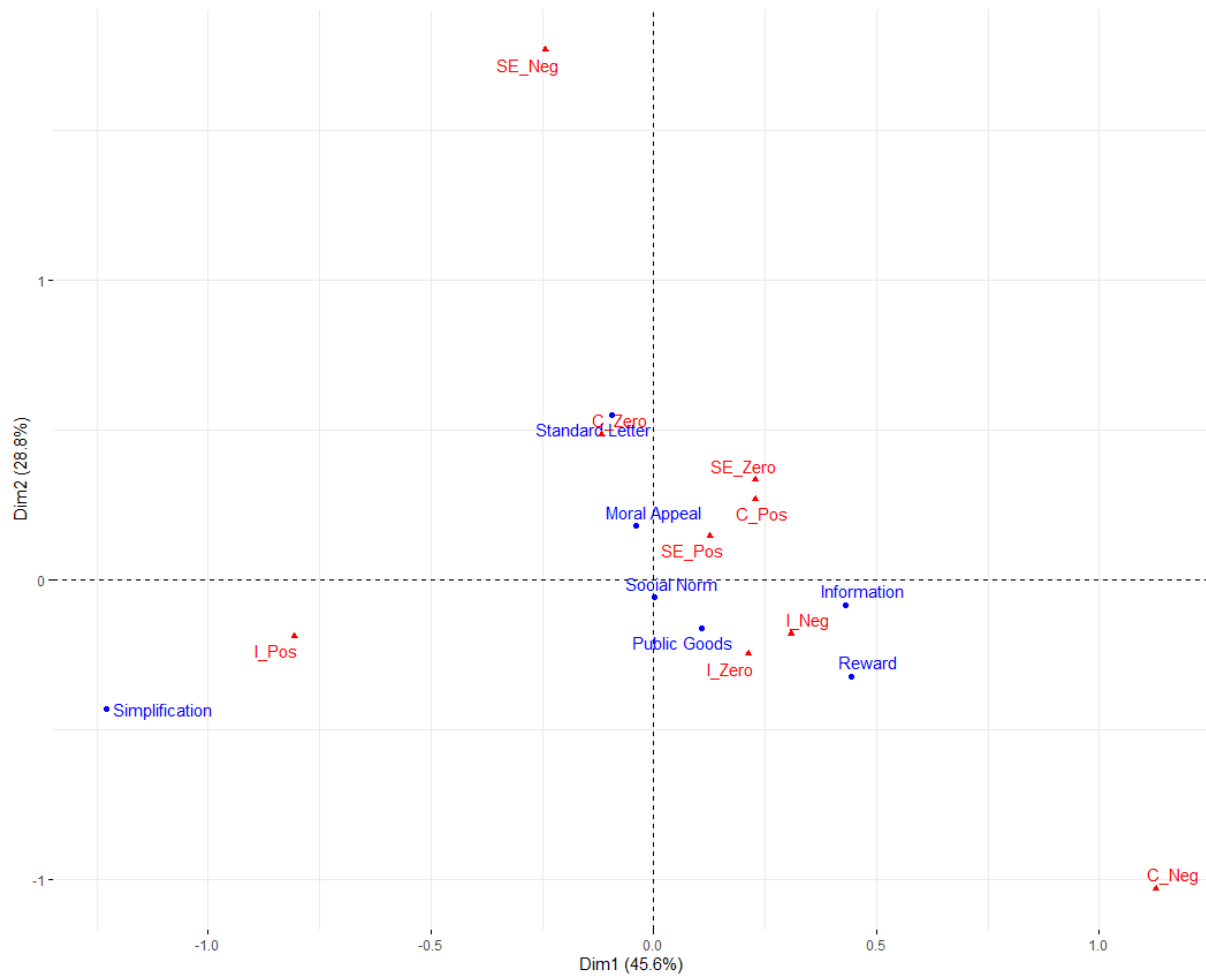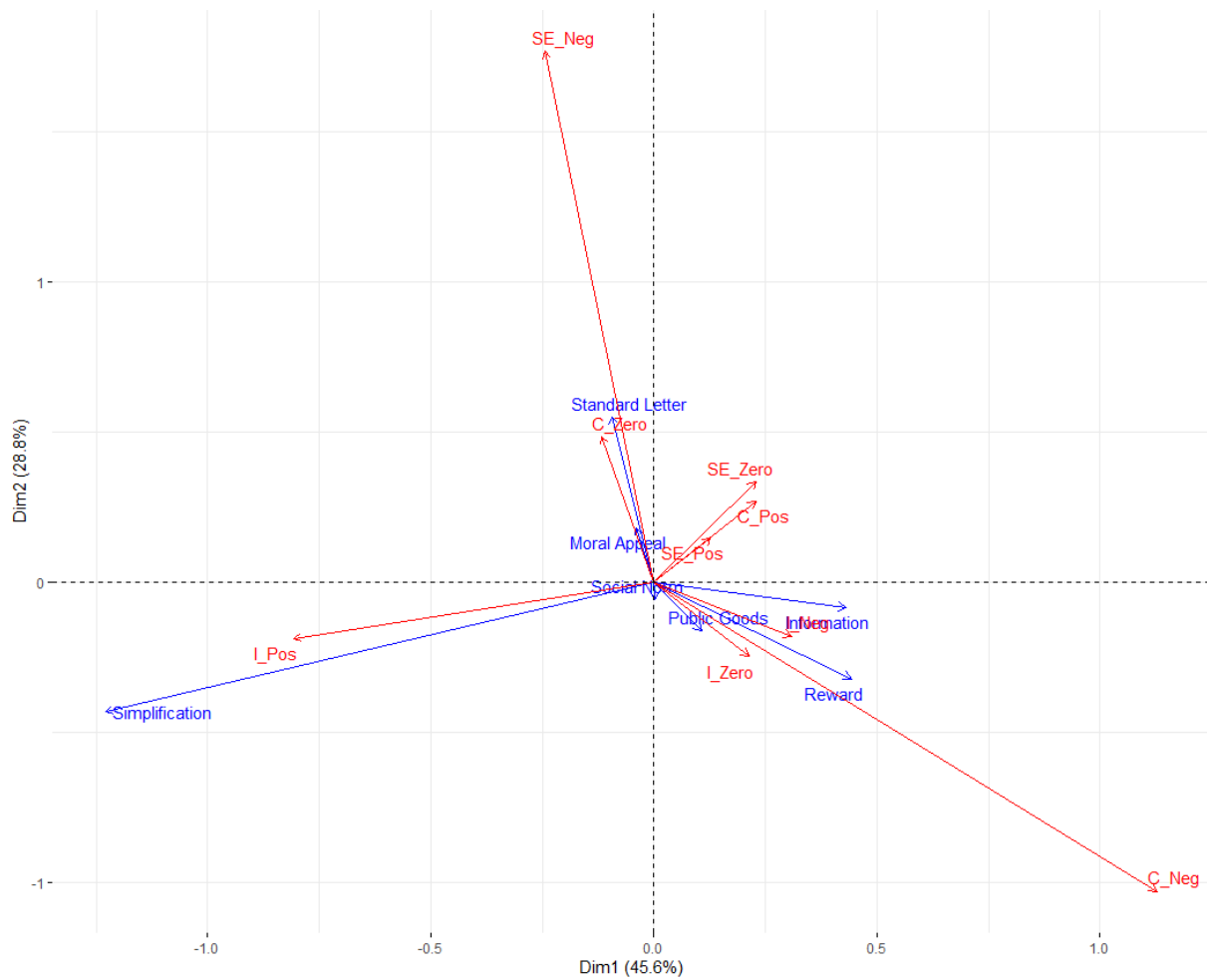
Figure 17A. Asymmetric correspondence analysis for frequencies of combined non-deterrence interventions in Table 14A. Blue lines represent the rows (i.e., the interventions). Red lines represent the columns (i.e., the population sub-categories). C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.

Figure 18A. Correspondence analysis for frequencies of individual non-deterrence interventions in Table 19A. Blue points represent the rows (i.e., the interventions). Red triangles represent the columns (i.e., the population sub-categories). C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.
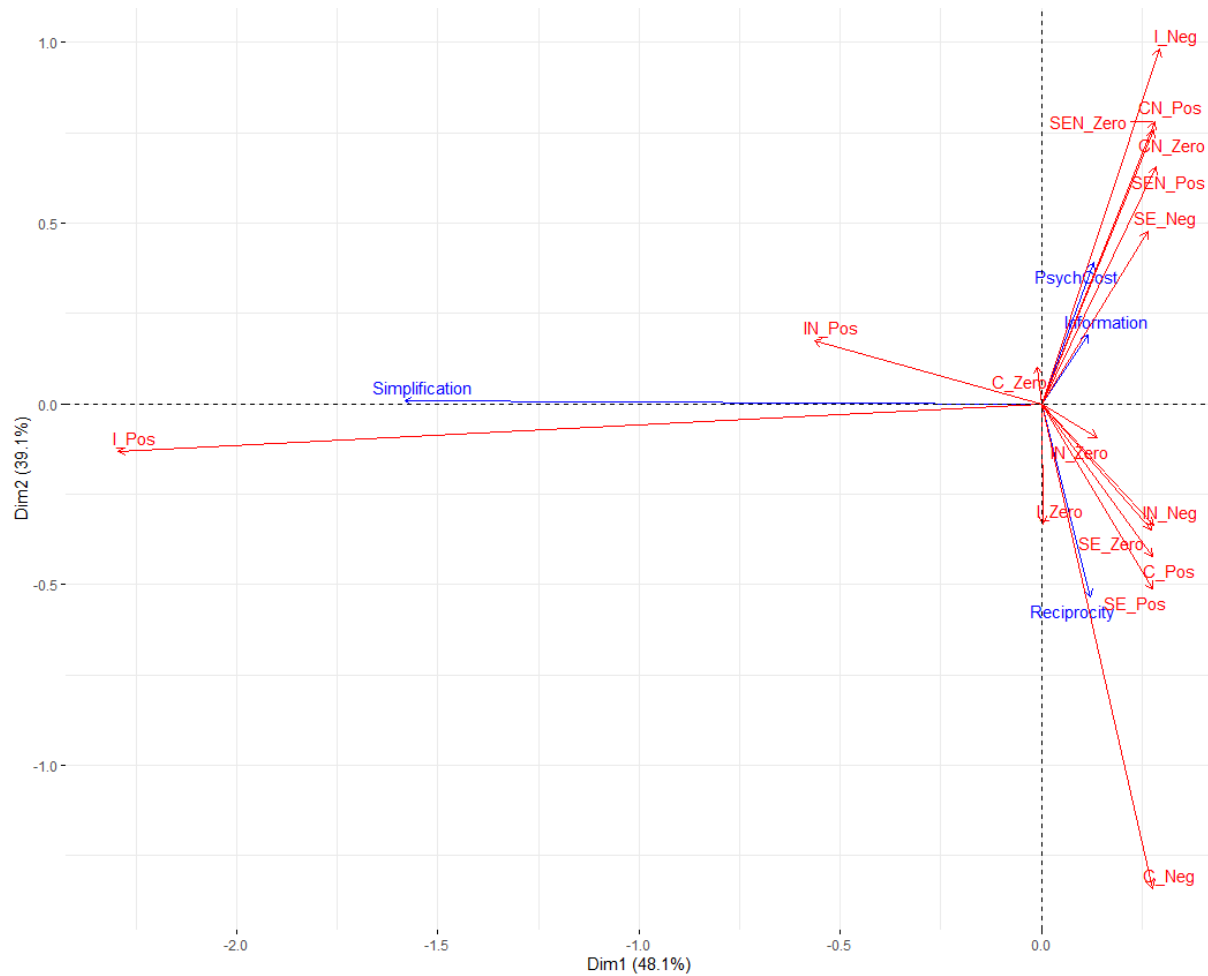
Figure 19A. Asymmetric correspondence analysis for frequencies of individual non-deterrence interventions in Table 19A. Blue lines represent the rows (i.e., the interventions). Red lines represent the columns (i.e., the population sub-categories). C = Company. I = Individual. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Figure 20A. Correspondence analysis for frequencies of combined non-deterrence interventions in Table 24A. Blue points represent the rows (i.e., the interventions). Red triangles represent the columns (i.e., the population sub-categories). CN = Company / Non-Compliant. C = Company. IN = Individual / Non-Compliant. I = Individual. SEN = Self-Employed / Non-Compliant. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.
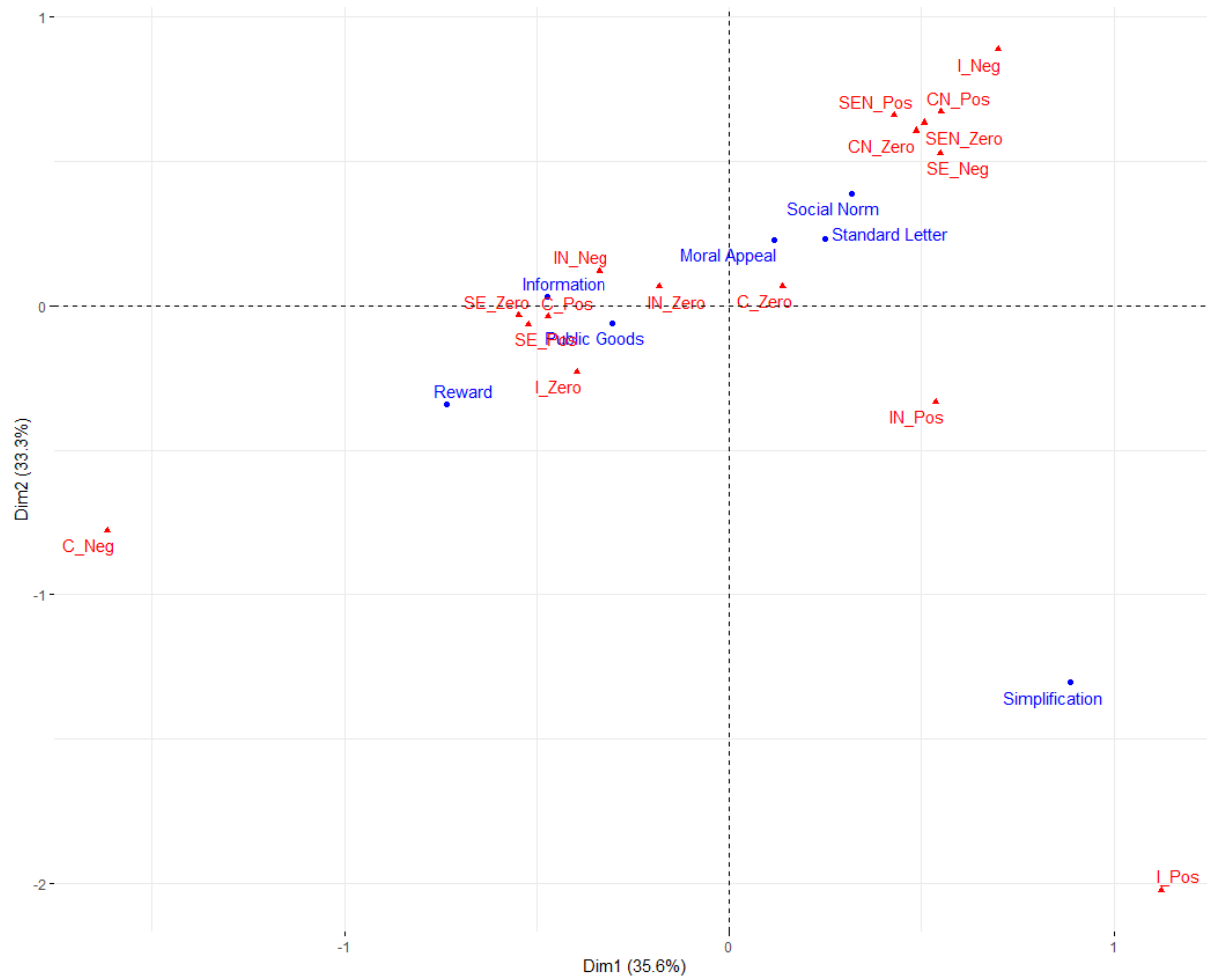
Figure 21A. Asymmetric correspondence analysis for frequencies of combined non-deterrence interventions in Table 24A. Blue lines represent the rows (i.e., the interventions). Red lines represent the columns (i.e., the population sub-categories). CN = Company / Non-Compliant. C = Company. IN = Individual / Non-Compliant. I = Individual. SEN = Self-Employed / Non-Compliant. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.
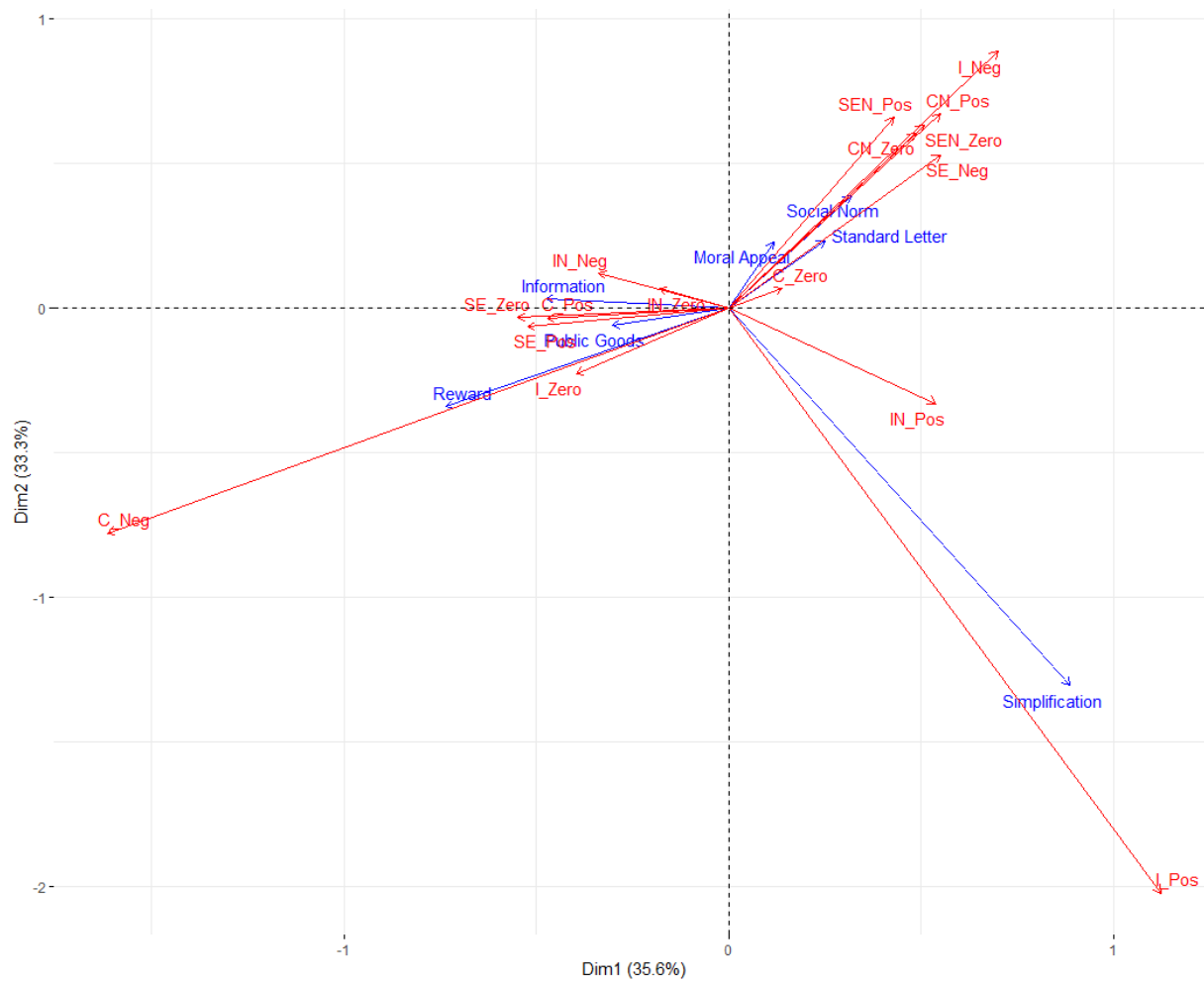
Figure 22A. Correspondence analysis for frequencies of individual non-deterrence interventions in Table 29A. Blue points represent the rows (i.e., the interventions). Red triangles represent the columns (i.e., the population sub-categories). CN = Company / Non-Compliant. C = Company. IN = Individual / Non-Compliant. I = Individual. SEN = Self-Employed / Non-Compliant. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Figure 23A. Asymmetric correspondence analysis for frequencies of individual non-deterrence interventions in Table 29A. Blue lines represent the rows (i.e., the interventions). Red lines represent the columns (i.e., the population sub-categories). CN = Company / Non-Compliant. C = Company. IN = Individual / Non-Compliant. I = Individual. SEN = Self-Employed / Non-Compliant. SE = Self-Employed. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.
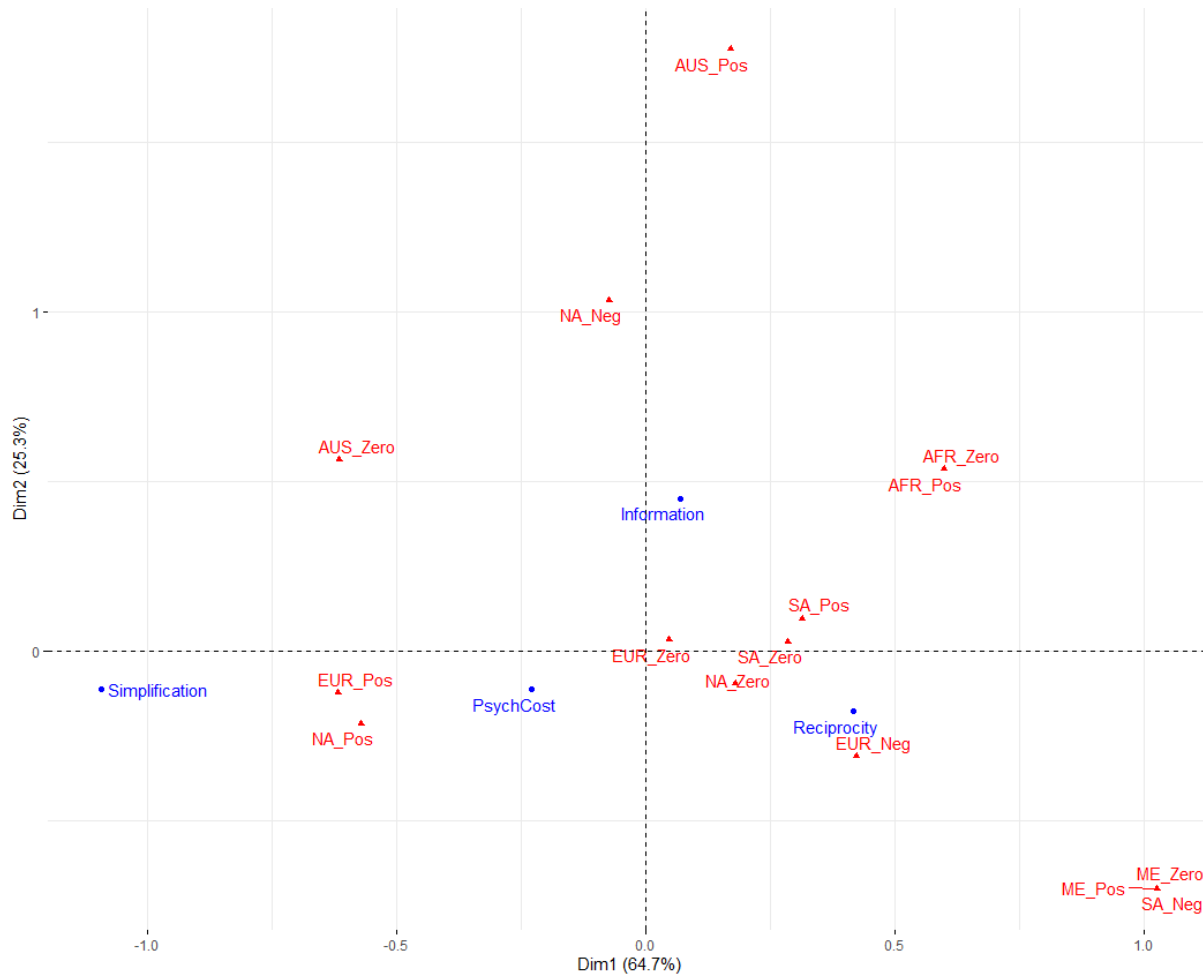
Figure 24A. Correspondence analysis for frequencies of combined non-deterrence interventions in Table 34A. Blue points represent the rows (i.e., the interventions). Red triangles represent the columns (i.e., the region sub-categories). AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.
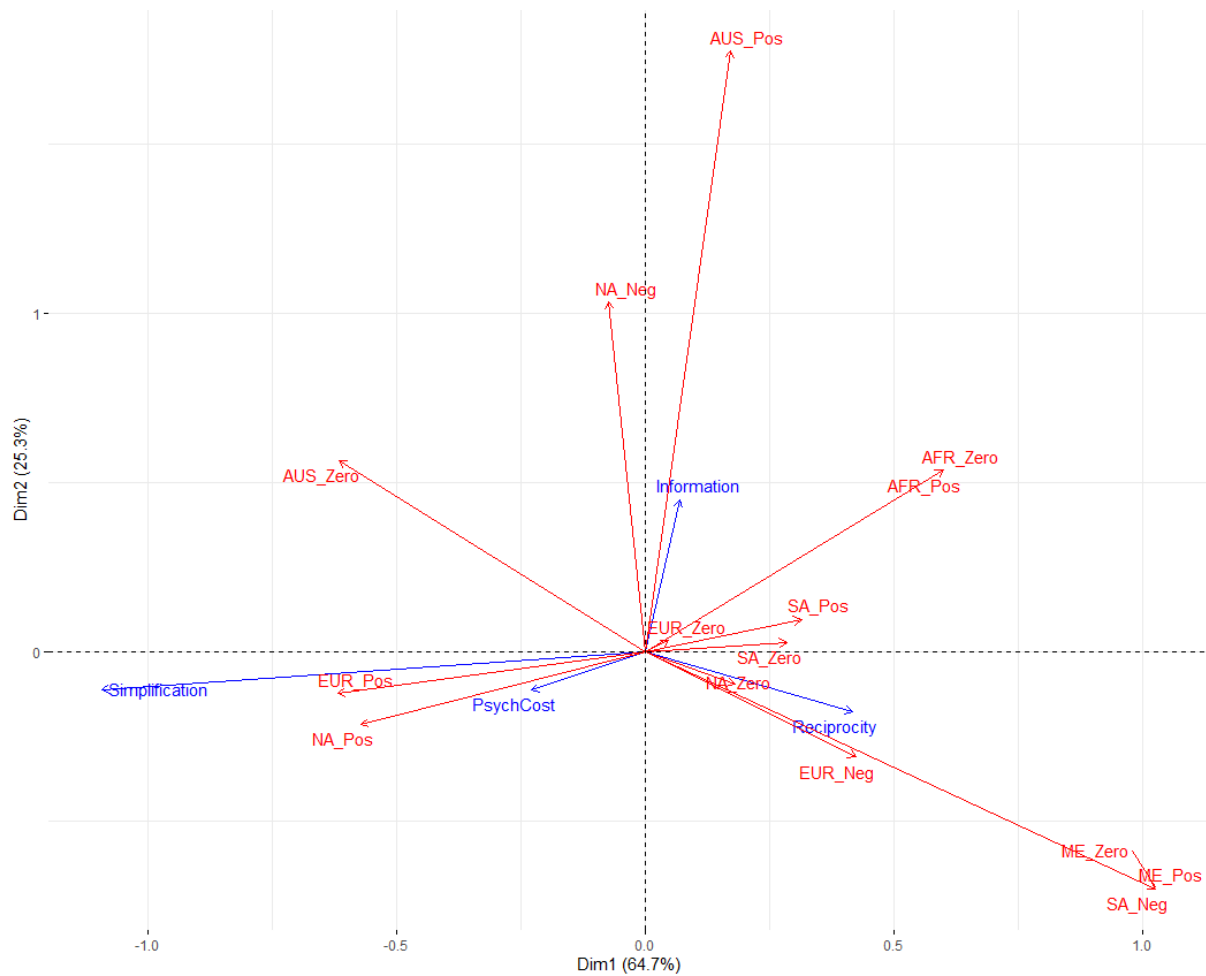
Figure 25A

CA_Region_Sum_Arrow

Figure 26A. Asymmetric correspondence analysis for frequencies of combined non-deterrence interventions in Table 34A. Blue lines represent the rows (i.e., the interventions). Red lines represent the columns (i.e., the region sub-categories). AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects. PsychCost = Interventions assumed to trigger psychological costs.
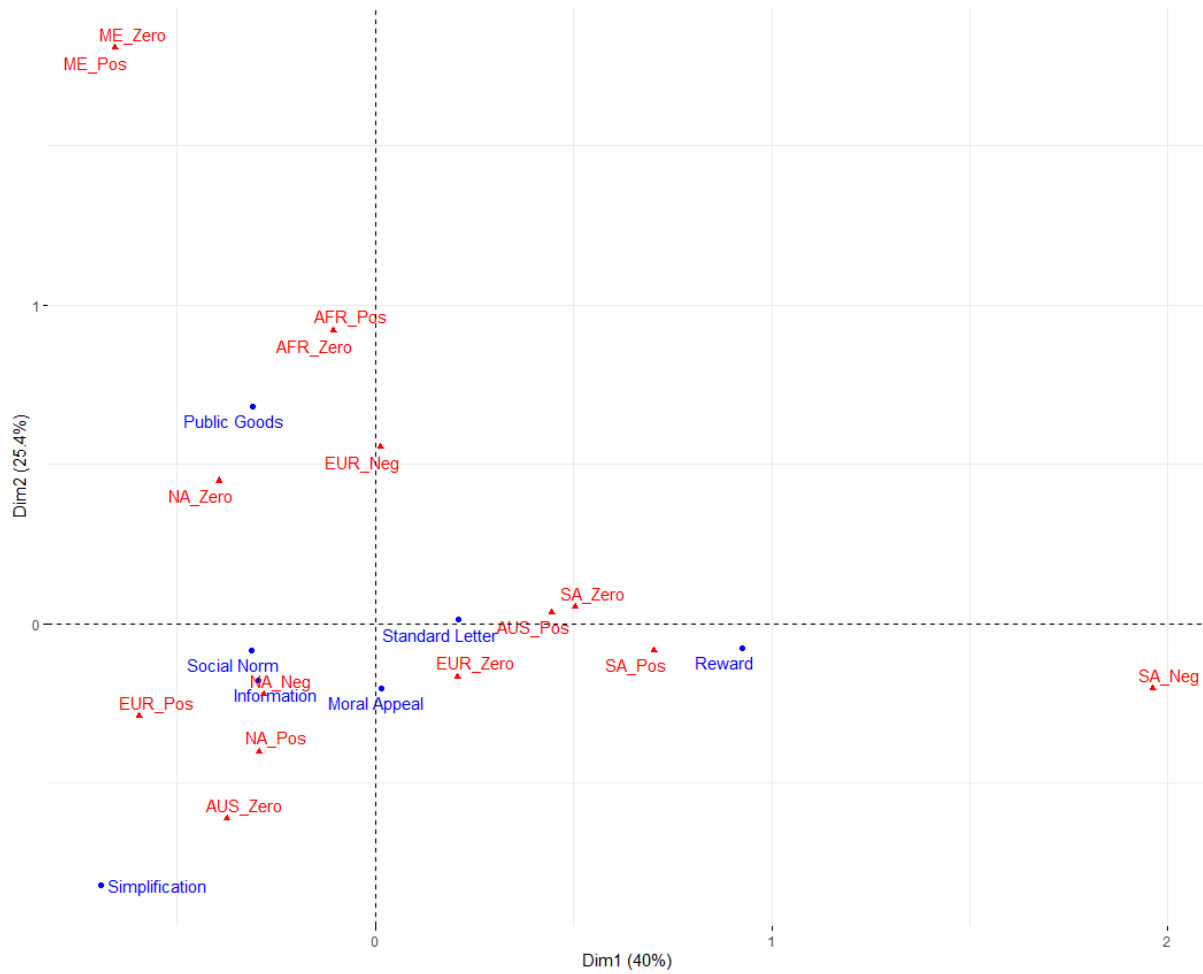
Figure 27A. Correspondence analysis for frequencies of individual non-deterrence interventions in Table 39A. Blue points represent the rows (i.e., the interventions). Red triangles represent the columns (i.e., the region sub-categories). AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.

Figure 28A. Asymmetric correspondence analysis for frequencies of individual non-deterrence interventions in Table 39A. Blue lines represent the rows (i.e., the interventions). Red lines represent the columns (i.e., the region sub-categories). AFR = Africa. AUS = Australia / New Zealand. EUR = Europe. ME = Middle East. NA = North America. SA = South America. Neg = Frequency of negative effects. Zero = Frequency of zero effects. Pos = Frequency of positive effects.
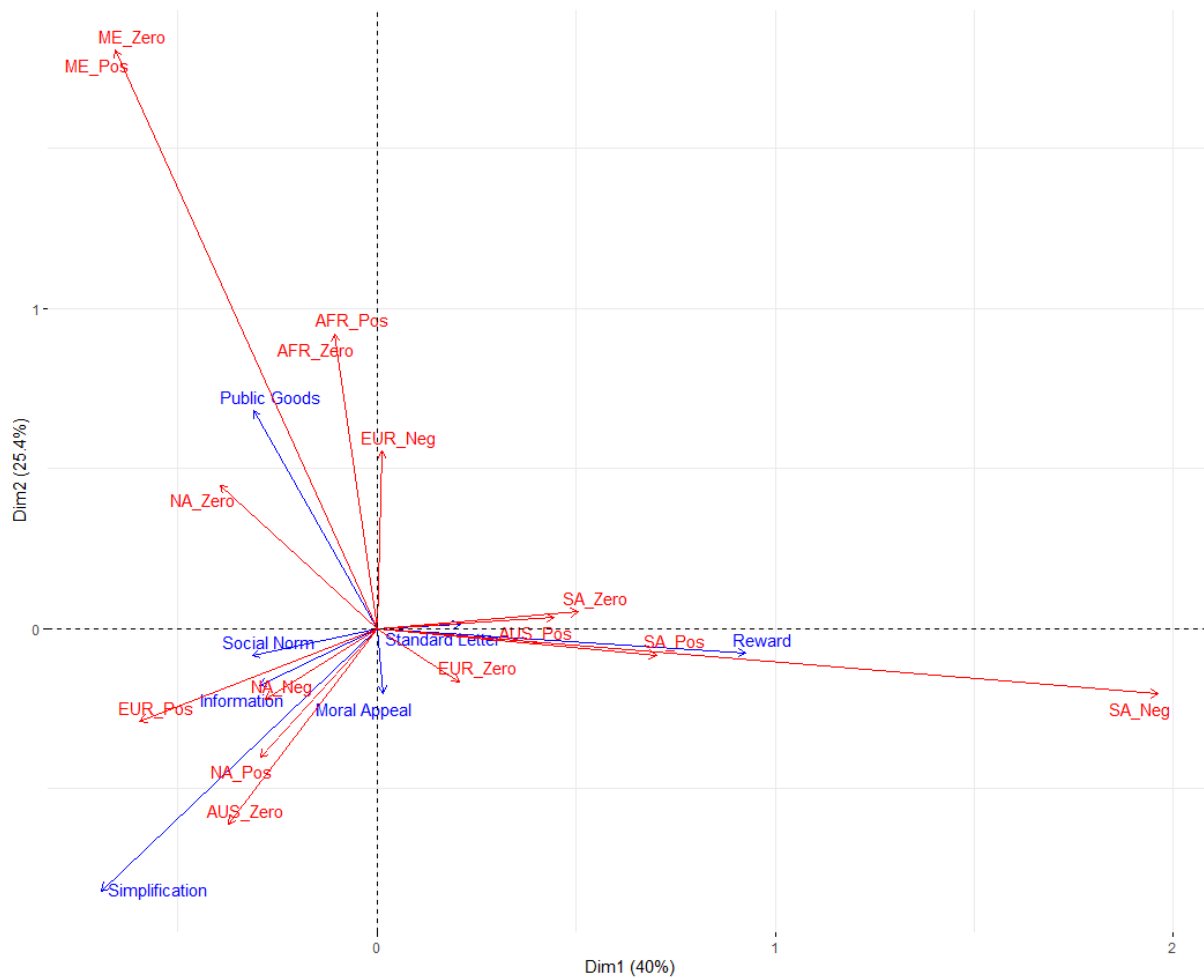
Appendix D

Figure 29A

Fleiss' Kappa for Two Independent Raters across all Three Categories (R Output)

```
 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 30A

Fleiss' Kappa for Two Independent Raters for the Main Category 'Type of Nudge Element' (R Output)

```
 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 31A

Fleiss' Kappa for Two Independent Raters for the Main Category 'Target Population' (R Output)

```
 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 32A

Fleiss' Kappa for Two Independent Raters for the Main Category 'Effectivity of the Intervention' (R Output)

```
 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
```

Appendix E

Link to Data Extraction Sheet

Link to Coding Table

Link to Fleiss' Kappa Rating Tables

Zusammenfassung

In zahlreichen Feldexperimenten in den letzten Jahren wurde die Wirksamkeit von Non-Deterrence Nudging-Interventionen auf Steuercompliance als Alternative zur Durchführung von Audits und der Androhung von Strafen untersucht. In der bisherigen Forschung wurde jedoch noch nicht systematisch untersucht, wie die Wirksamkeit dieser Interventionen durch individuelle und kontextuelle Faktoren beeinflusst wird. Daher wurden in der vorliegenden Arbeit Ergebnisse verschiedener Feldexperimente in ein standardisiertes Schema integriert. Die Effekte der Studien wurden hinsichtlich verschiedener Subpopulationen von Steuerzahlern untersucht, wobei deren beruflicher Kontext sowie das vergangene Steuerverhalten berücksichtigt wurden. Zusätzlich wurden die Effekte hinsichtlich verschiedener Regionen untersucht, in denen die Feldexperimente durchgeführt wurden. Die vorhandene Literatur wurde nach den PRISMA-Richtlinien ausgewertet. Die Ergebnisse zeigen, dass Non-Deterrence Nudging-Interventionen häufiger in einer Population von Steuerzahlern wirksam waren, die in der Vergangenheit bereits einmal Steuervorschriften missachtet haben – verglichen mit einer allgemeinen Gruppe von Steuerzahlern. Zusätzlich konnten die Ergebnisse zeigen, dass Nudging-Interventionen in europäischen Feldstudien häufiger wirksam waren als in nordamerikanischen Feldstudien. Wenn man sich die Effekte der einzelnen Interventionen ansieht, zeigte sich, dass die meisten Interventionen, die auf eine Vereinfachung des Inhalts abzielten, wirksam waren, wobei die Effekte hauptsächlich auf eine Population von individuellen Steuerzahlern zurückzuführen waren. Interventionen, die auf soziale Normen abzielten, zeigten ebenfalls häufig positive Effekte, vor allem in der Population der zuvor bereits einmal nicht gesetzestreuen Steuerzahler. Reziprozitäts- und Informationsinterventionen zeigten kaum positive Effekte, weder in einer bestimmten Subpopulation noch in einem bestimmten regionalen Kontext. Weitere Forschung ist besonders hinsichtlich der Steuercompliance von Unternehmen, sowie für Länder außerhalb von Europa und Amerika erforderlich.

Stichworte: Steuercompliance, Non-Deterrence, Nudging, Steuerverhalten, Regionale Unterschiede