Title: From the Lab to the Field: A Review of Tax Experiments

Citation: Mascagni, G. (2017) From the Lab to the Field: A Review of Tax Experiments, Journal of Economic Surveys (2018) Vol. 32, No. 2, pp. 273–301

Official URL: https://doi.org/10.1111/joes.12201

More details/abstract: Tax experiments have been gaining momentum in recent years, although this literature dates back several decades.With new developments in methods and data availability, tax experiments have gradually moved away from lab settings and towards the field. This movement from the lab to the field has happened against the background of the 'credibility revolution' in applied economics, which has seen more rigorous methods applied to policy relevant questions, and of the availability to researchers of administrative data from tax returns. These developments have allowed significant advances in the experimental literature on tax compliance. This paper reviews this literature, giving particular attention to field experiments using administrative data, but putting them in the broader context of the compliance literature. A particular effort is made to take a global perspective, in a literature that is only recently seeing the emergence of evidence from Africa, Latin America and Asia.

JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12201

FROM THE LAB TO THE FIELD: A REVIEW OF TAX EXPERIMENTS

Giulia Mascagni*

Institute of Development Studies

Abstract. Tax experiments have been gaining momentum in recent years, although this literature dates back several decades. With new developments in methods and data availability, tax experiments have gradually moved away from lab settings and towards the field. This movement from the lab to the field has happened against the background of the 'credibility revolution' in applied economics, which has seen more rigorous methods applied to policy relevant questions, and of the availability to researchers of administrative data from tax returns. These developments have allowed significant advances in the experimental literature on tax compliance. This paper reviews this literature, giving particular attention to field experiments using administrative data, but putting them in the broader context of the compliance literature. A particular effort is made to take a global perspective, in a literature that is only recently seeing the emergence of evidence from Africa, Latin America and Asia.

Keywords. Administrative data; Randomisation; Tax compliance; Tax experiments

1. Introduction

Tax non-compliance is one of the greatest obstacles that revenue administrations face. However, increasing compliance is not only important to collect more government revenue. The way that objective is achieved has implications also on the practice of tax administration. Collecting taxes through aggressive enforcement and coercion is more costly than encouraging taxpayers to collaborate with tax collectors. Furthermore, tax compliance is deeply intertwined with the political environment that shapes the social contract between citizens and the state. Low corruption and government accountability have been defined as pre-conditions for countries' successful tax performance (Bird *et al*., 2008).

Clearly, in any country, taxpaying is the result of both quasi-voluntary compliance and enforcement. It would be simplistic to think that taxpayers comply with tax laws just because they think it is a good thing to do. An element of coercion is present in most cases. For example, withholding procedures make it very difficult for employees to evade taxes on their wages. Therefore, most of them do not have the choice of whether to evade or not (Kleven *et al*., 2011). At the same time, it is also clear that the observed level of tax compliance is too high to be only the result of coercion (Alm *et al*., 1992b; Andreoni *et al*., 1998). Most countries simply do not have the capacity to control and audit a large number of taxpayers. Therefore, the right question to ask may be 'Why do people pay tax?' rather than 'Why do they evade?' (Alm *et al*., 1992b). The answer is likely to be a mix of deterrence, related to the probability of audit and sanctions, and a range of factors often referred to as tax morale. Although there is no uniform definition of tax morale, a recent review (Luttmer and Singhal, 2014) unpacked

*Corresponding author contact email: g.mascagni@ids.ac.uk Tel: +44(0)1273 915765

five main mechanisms through which it can operate: (1) intrinsic motivation, (2) reciprocity, (3) peer effects and social influences, (4) culture and (5) information imperfections and deviations from expected utility.

A recent surge in empirical studies has contributed to shed more light on the role of these factors in determining tax compliance and on what tax administrators can do to improve it. This recent surge has seen the emergence of large-scale field experiments, that are arguably more rigorous and policy relevant than previous studies. This paper aims to provide a review of this relatively new field of research, which has seen an exponential expansion in the past 15 years, since Torgler's review of 2002, and continues to grow rapidly. In the 2 years since the publication of a similar review in 2014 (Hallsworth, 2014), several new papers have been published, looking at a broader range of questions and countries than ever before. This paper aims at taking stock of this literature, as described in more detail in Section 1.2.

The literature on tax compliance is vast and diverse, spanning from quantitative to qualitative analyses and from surveys to experiments. Despite several differences in methods and approaches, the kind of questions asked, the underlying hypotheses and the theoretical foundations are very similar. In most cases, the theoretical framework can be traced back to the model of taxpayer behaviour of Allingham and Sandmo (1972). These authors model tax evasion based on the theoretical framework used in the economics of crime (such as Becker, 1968) and the economics of uncertainty (such as Mossin, 1968). They hold that a taxpayer's decision on whether to evade, and to what extent, is influenced by the expected economic returns to evasion. These returns are determined by the benefits of evasion, affected by the tax rate and taxable income; and by its costs, influenced by the probability of being detected and the size of sanctions if found guilty. However, researchers soon realised that if taxpayers acted purely according to this model, we would observe much higher levels of tax evasion. This observation has spurred researchers to include other factors in the model, such as social and moral considerations, which are likely to affect the decision to evade (Erard and Feinstein, 1994; Myles and Naylor, 1996; Traxler, 2010). A detailed discussion of recent theoretical models of tax compliance is available in the review article by Hashimzade *et al*. (2013).

Based on these theoretical frameworks, the empirical literature has tested the validity of such determinants of compliance in practice. The obvious problem in doing so is that it is difficult to get honest answers about dishonest behaviour. For this reason, survey-based studies have focused on attitudes to tax compliance rather than actual behaviour. For example, the World Value Survey asks: "Please tell me for each of the following actions whether you think it can always be justified, never be justified, or something in between. Cheating on taxes if you have a chance."[1] In other cases, surveys have adopted indirect approaches to capture evasion, as discussed for example in Gerxhani (2007) in the context of transition countries. There are already a number of good reviews of the survey-based literature on tax compliance, such as Slemrod (2007) and, more specifically on African countries, Fjeldstad *et al*. (2012). This literature offers useful insights into the question of why people pay tax. However, it suffers from a major drawback: the attitudes that people reveal in surveys may not accurately reflect their actual behaviour when faced with the real consequences of tax evasion and avoidance (Ajzen and Fishbein, 1977).

Tax experiments (TEs) address this drawback by looking at real taxpayer behaviour rather than perceptions – although Section 1.1 makes important distinctions on the level of realism of different types of TE. This paper reviews the evidence emerging from this literature, paying particular attention to the growing number of large-scale field experiments. It proceeds as follows. Section 1.1 provides an overview of different types of TEs and their evolution in time. Against this background, Section 1.2 states the specific aim and relevance of this review. Section 2 represents the core of this paper and summarises the main findings in the literature, classified by determinant of compliance. Finally, Section 3 critically takes stock of this literature and looks forward to its future.

1.1 *Speaking to Theorists, Searching for Facts and Whispering in the Ears of Princes*

TEs use experimental methods to identify causal effects of the determinants of compliance. The authors in the early TE literature (Friedland *et al*., 1978; Spicer and Thomas, 1982) argue that such causal links could be identified in a lab setting, where it is possible to control the environment and independent variables better than in real-life or survey-based research. They also argue that the artificiality of the lab allowed them to circumvent the problem of getting honest answers about dishonest behaviour by working within hypothetical settings. The more recent TEs instead turned to administrative data from anonymised tax returns, and the causal links are established thanks to randomisation. These are two separate types of TE, as discussed in more detail in this section. While the latter (field TE) is more relevant to understand real taxpaying behaviour and inform policy, the former (lab TE) helps to understand the origins and theoretical foundations of this literature.

In his 2002 review of TEs, Benno Torgler classified the first type of TEs in two categories labelled ‘speaking to theorists’ and ‘searching for facts’, based on the definition provided by Roth (1995) (Torgler, 2002). In both cases, the primary issue of interest is tax compliance, with a common focus on income taxation – although similar arguments can be made for other taxes. The first category includes lab experiments aimed primarily to test theories, and whose results feed back into the theoretical debate and literature. In particular, several early studies tested the empirical validity of the economic model of taxpayer behaviour of Allingham and Sandmo (1972). These *speaking to theorists* studies therefore focused primarily on the role of deterrence, sanctions and expected economic returns to explain compliance (Friedland *et al*., 1978; Spicer and Thomas, 1982; Spicer and Hero, 1985; Beck *et al*., 1991). In this case, experimenters speak to theorists. The typical lab experiment would involve a group of students who are faced with the decision on how much income to report, based on a set of given variables including, for example, the tax rate, the probability of audit and sanctions. The artificial lab setting then allows the researcher to change these variables and to observe how the participants’ decisions vary in response.

The second category, *searching for facts*, includes experiments that analyse variables that are not explicitly included in the traditional economic models, but that are motivated instead by the results of other experiments and by observations of what actually happens in the real world. One of these observations is that the level of compliance in many countries is much higher than predicted by purely economic models. This has prompted researchers to explore the role of public goods, and moral and social factors in tax compliance, in addition to the traditional economic considerations. In this case, experimenters are talking to each other (Torgler, 2002). Although experiments looking at these issues are searching for the facts, they speak to theorists too, since the new observed patterns can feed back into the theory to develop existing models further.

Speaking to theorists and *searching for facts* studies have some common characteristics. First, they typically use students as experimental subjects. While students respond to the incentives set out in the experiment, they are not real taxpayers and they may not have any experience at all with taxpaying in real life. Second, these studies usually rely on small samples, often involving only a small number of students in repeated rounds. This presents problems related to statistical power, and the resulting estimates may not be stable or reliable. Finally, they use largely artificial settings, despite efforts to move to more realistic environments. In a lab experiment, the individuals involved obviously know that they are part of a study and they may behave differently than they would under the pressures of real life. Behaviour in the artificial environment of the lab is likely to differ from actual decisions when faced with a real social context and real consequences of wrongdoing. These challenges have largely prevented lab TE from having much policy impact. Therefore, lab TE have been well suited to understand the theoretical mechanisms that drive compliance, but the relevance of their findings in real taxpaying situations has to be tested in the field.

In an attempt to achieve greater realism in TEs, some researchers have moved the lab to the field (lab-in-the-field TE). Although this helped in getting additional insights, for example into cross-cultural differences (Cummings *et al*., 2009), they still suffer from very similar methodological shortcomings.

Many of the drawbacks of lab TE were overcome with the emergence of large-scale field TE. Following the categorisation proposed by Roth (1995), these TE can be labelled as ‘whispering in the ears of princes’, as experimenters aim to speak to policymakers. To do so, the lab setting is substituted with real-life situations: the participants in the study are real taxpayers, the data come from real returns and the experiment occurs in the normal taxpaying environment where the main counterpart is the tax administration. The typical field TE is based on a comparison of randomly selected treatment and control groups. The treatment often consists of the provision of information that is meant to change key determinants of tax compliance, like perceptions on the probability of audit, fairness and tax morale (with all the aspects it encompasses). This information is typically delivered through a letter that highlights, for example, the fact that the tax administration routinely carries out audits and that those who are caught evading face severe fines; or the importance of paying taxes to finance important public services such as schools and hospitals. Because the participants in the treatment and control groups are chosen based on randomisation, they are very similar with respect to all other aspects such as age, income and location. The only difference between groups is therefore whether or not they received the informational treatment in the form of a letter. By observing the differences in taxpayer behaviour between groups, and before and after receiving the letters, researchers can establish the causal effect of different determinants on compliance.

By doing this, they can explore questions similar to those asked by lab experiments, for example if taxpayers are more affected by the probability of being caught or by an understanding of the social value of taxpaying. The results on these common questions are broadly consistent between field and lab TE. However, by being embedded in real taxpaying situations, field TE can also provide evidence on new factors. In addition to deterrence and tax morale, new questions have focused around more practical issues, such as the incentives of tax collectors and compliance costs for taxpayers. Field TE can explore compliance in the context of previous experience, beliefs and perceptions – all of which influence subjective views on tax that can influence compliance behaviour substantially. Moreover, they can sometimes uncover unexpected reactions that measures to increase compliance can generate in practice, for example backfiring effects found in Slemrod *et al*. (2001) or potential crowding out of intrinsic motivation (Del Carpio, 2014). More generally, field TE can offer a more nuanced understanding of real taxpaying experiences, including differences among taxpayers related to sectors, opportunity to evade and income level. While lab TE are useful as a ‘proof of concept’, providing the psychological and theoretical foundations to this literature, field TE are more suitable to test concrete policy options. However, lab experimenters have made greater efforts to approximate real-life conditions by using tax-specific language, simulations of exchanges of resources and participants with experience of taxpaying rather than students. Still, the lab remains a largely artificial setting and usually allows much smaller samples than field TE.

The types of TE described in this section are summarised in Table 1, which necessarily makes some over-simplifications – such as dividing the determinants of compliance into three general categories of deterrence, tax morale and practical issues.

Table 1. Types of Tax Experiments

	Lab	Field	Deterrence	Tax morale	Practical issues
Speaking to theorists	√		√		
Searching for facts	√		√	√	
Whispering in the ears of princes		√	√	√	√

Field experiments, by investigating behavioural factors of taxpaying, can potentially provide guidance on how to raise tax revenues in a cost-effective way by changing perceptions and the information available to taxpayers. This is particularly important because credibly increasing enforcement is very costly and not always desirable. With this information, policymakers can weigh their policy options and target their compliance strategies more specifically to the factors that generate the larger responses at the most reasonable cost. In some cases, researchers have studied existing initiatives of the tax administration, or variations in the law or other relevant variables, to evaluate their effect on compliance. These studies are natural experiments or quasi-experiments, which are also included in this review. Although they do not use randomisation by assigning taxpayers to control and treatment groups, they are useful to evaluate policy and are often able to establish causal effects.

1.2 *Aim and Relevance of This Review*

Against the background of Section 1.1, this paper aims to update Torgler's review of 2002 by adding TEs that are *whispering in the ears of princes*. At the time Torgler (2002) was published there was not an extensive literature of this sort – with the exception of Slemrod *et al*. (2001) – which justified its exclusion from his review. However, in the last 15 years field experiments have been thriving, forming a growing strand of literature that can no longer be ignored. This review aims to fill this gap, and is particularly relevant for three reasons.

The first one has to do with methods and to the 'credibility revolution' that applied economics has been experiencing in recent years (Angrist and Pischke, 2010). Field experiments combine rigorous methods with high policy relevance better than other types of tax compliance studies. For this reason, Slemrod and Weber (2012) argue that they are the most credible method to evaluate policy interventions to address tax compliance and, as such, they should be used as a benchmark. Despite being at the basis of this 'credibility revolution', randomisation has attracted criticism related to external validity and the possibility to scale up interventions (Deaton, 2009; Basu, 2013). Although these issues remain a source of concern, field experiments have an important advantage over the typical randomised control trial. Because they can use population-wide data, they can evaluate options for real policy interventions 'at scale'. Therefore, if proven effective, these interventions are easier to implement at the national level.

Secondly, field TEs use anonymised administrative data from tax returns. The use of these data by researchers is a relatively new development in many countries, which partly explains why this literature has expanded only recently. This type of data has at least two important advantages. First, it allows the observation of actual taxpaying behaviour in real life. However, since tax returns only capture what is known by the tax administration, they do not allow the direct observation of tax evasion and avoidance.[2] Experimental methods allow researchers to address this problem, but only partially. Those taxpayers who are completely informal – not registered for tax purposes – are still missed by field TE. Some recent studies started to fill this gap by looking at non-filing (Gangl *et al*., 2015) and non-registration (De Giorgi *et al*., 2015), but evidence from field TE is still limited in this area. Secondly, the effective use of administrative data requires a high degree of commitment from the local tax administration. Therefore, these studies have a built-in connection with policymakers, which is beneficial both to design relevant research projects and to make the results useful for supporting policy.

Finally, this review makes a particular effort to include studies that focus on middle- and low-income countries. Most of the TE literature, as well as the compliance literature more generally, has been largely concerned with Europe and the USA. However, more and more studies are being published with evidence from Latin America and Asia, while Africa remains largely under-represented. Tax compliance can be seen as a global issue, which all revenue administrations face to different extents, and whose possible determinants are likely to be similar across countries. However, there are at least three reasons that so far prevented field experiments from spreading more widely outside Europe and the USA. The first one is related to data availability. In the past two decades, most low- and middle-income countries have undergone great transformation in their tax administration, by establishing semi-autonomous revenue authorities and, more generally, by modernising the way tax collection is organised. This has entailed the digitalisation of tax records, which became available to researchers only recently. Still, data quality in these countries is a greater concern than in more established administrations. Secondly, the high degree of government commitment required to conduct field TE is not always available. Most tax administrations in low-income countries operate under great capacity constraints, so it can be particularly challenging for them to effectively collaborate on tax research. Thirdly, because revenue administrations in these countries are relatively new, they are developing only recently the practice of working with researchers – this includes practical issues such as the appropriate anonymisation of data, but also, more generally, trust and appreciation of the value of policy-relevant research.

The explicit effort in including evidence from low- and middle-income countries is one of the distinctive features of this paper, compared with other recent reviews (Hallsworth *et al.*, 2017; Slemrod, 2016). In particular, Hallsworth *et al.* (2017) surveys 23 natural field experiments on tax compliance, of which five are from middle-income countries. My review includes 14 field TE experiments from low- and middle-income countries, for a total of about 30 field experiments reviewed in detail. In addition, this paper also includes several lab experiments and quasi-experimental studies to support and complement the findings from the field. Most of recent literature was published after the review of Hallsworth *et al.* (2017), confirming the fast expansion of this field. A second distinctive feature is that this review tries to link field TE to the broader compliance literature, both experimental (lab, field) and survey based. This allows me to take stock of what we know on compliance, validating and complementing the results from field TE with key findings from the broader literature.

2. Review of tax experiments

This section reviews the TE literature and is organised around eight issues that matter for tax compliance. They are often interrelated, sometimes overlapping and almost always co-exist in various combinations. The main focus is on field TE, but evidence from lab TE is also included. Although most lab TE belong to the early literature, they have recently made advances in integrating methods from psychology to get additional insights into taxpayer behaviour. Many of the hypotheses tested in field TE come from the lab literature, and it therefore seems relevant to draw a clear connection between these distinct but highly interrelated strands of literature. Furthermore, the main results reviewed here are sometimes cross-validated against non-experimental studies that use data from tax returns. Since TEs are still a relatively narrow strand of literature, linking them to the broader context helps to support the empirical validity of their results.

2.1 *Traditional Economic Factors of Deterrence*

The traditional economic factors set out in Allingham and Sandmo (1972) include variables that influence the costs and benefits of evasion, such as: the tax rate, the probability of audit, sanctions and enforcement. Most of the early *speaking to theorists* lab TE test the effects of these variables on the probability of evading (the extensive margin), and the amount of taxes evaded through income under-reporting (the intensive margin). The pioneering simulation study by Friedland *et al.* (1978) analyses changes in the tax-paying behaviour of 15 Israeli psychology students when these economic factors vary. When the tax rate rose above a certain level, the participants responded by sharply decreasing their reported income. Moreover, large sanctions appeared to be more effective than frequent audits in improving compliance. Finally, decisions on the extensive and intensive margins are driven by different factors. While the tax rate seems to have the largest impact on the former, socioeconomic factors are more important to determine the latter. However, conclusions from this experiment need to be taken with caution, because of the small sample size and related power problems. An additional drawback is that in the lab environment all the key variables are fully known to participants, while in real life probability of audit is not known precisely and even the exact amount of sanctions can be uncertain.

Confirming the importance of uncertainty, Spicer and Thomas (1982) found that the predicted deterrence effect of the probability of audit is only empirically valid when taxpayers have full information about it. However, even when information is available, an increase in the probability of audit only affects the decision to evade and not its extent. Beck *et al.* (1991) reinforce these results by showing that, although evasion is lower when audit probabilities and sanctions are higher, uncertainty weakens these effects. Moreover, when actual variables are not known due to uncertainty, perceptions play an important role. Spicer and Hero (1985) test two hypotheses with a group of 36 psychology students from the University of Colorado: first, taxpayers evade more when they perceive that the level of evasion in general is higher; secondly, those who have been audited in the past perceive a higher probability of audit and therefore tend to evade less. The authors found no support for the former, but did find evidence for the latter. Although these results still stem from highly artificial settings (including other studies such as Mittone, 1997), with small samples of students, they help to understand the possible reasons why compliance in real life is higher than in the theoretical model of Allingham and Sandmo (1972). In the presence of uncertainty, taxpayers tend to evade less even if audit probabilities and sanctions are low. This may be partly due to upwardly biased perceptions of the risk of being caught.

Turning to field experiments, Slemrod *et al.* (2001) carry out a pioneering study of this kind based on a sample of 22,368 tax returns from taxpayers in Minnesota, USA. The authors, in close collaboration with the Department of Revenue, collected income tax data for two subsequent years: 1993 and 1994. The treatment consisted of a letter stating that the taxpayer had been selected to be part of a study and that their state and federal tax returns would be closely examined. The letter also stated that taxpayers would be contacted regarding any discrepancy and that if any irregularities were found, their previous tax returns would be scrutinised as well. The letters were sent by the Commissioner of Revenue in January 1995, at the beginning of the filing season for 1994, to a treatment group of about 1500 taxpayers. The letter was meant to increase the perceived probability of audit and therefore to affect the accuracy of income tax reporting. A difference-in-difference methodology was used to analyse the changes in reported taxable income between 1993 and 1994 in the treatment and control groups. Changes in reported income were then used as an indication of changes in compliance, as data on the audit results (the close examination) were not available to the authors.

Before discussing the results, it is worth noting that Slemrod *et al.* (2001) introduced two methodological elements that are common to the rest of this literature. First, the letters are sent directly from the revenue administration. This helps to make the experiment real and therefore to capture real behaviour. Secondly, the treatment does not change actual tax parameters, such as the rate or the number of audits. Instead it only varies taxpayers' perceptions and the information available to them. On the other hand, Slemrod *et al.* (2001) differ from other studies in the literature on at least two methodological choices. First, in this case the treatment is relatively strong, as it hints with some degree of certainty to an upcoming scrutiny of the taxpayer's return. Other experiments have adopted softer approaches that only increase the salience of audits without hinting to an upcoming audit for any specific taxpayer (Castro and Scartascini, 2013). Secondly, Slemrod *et al.* (2001) decide to let taxpayers know that they were part of a study. Other researchers, such as Kleven *et al.* (2011), instead prefer to conduct the study as if the treatment was a normal letter from the tax administration. By doing this, they believe that the message may be taken more seriously. Clearly, taxpayers' reactions to the treatment are a mix of both things: they may know it is part of a study, but if the treatment is strong enough, as in the case of Slemrod *et al.* (2001), they are still likely to respond.

Indeed Slemrod *et al.* (2001) find that the letter had a significant impact on reported income, in some cases increasing compliance by 12%. However, this effect varies across categories of taxpayers, who were stratified along two dimensions: income and opportunity to evade. The latter is defined according to the type of income being taxed: high opportunity occurs in presence of self-reported income, or income from rents and royalties. In a situation of high opportunity to evade, tax payments increased by 12% for low- and middle-income taxpayers. Among the high-opportunity taxpayers, those with self-reported income represented a major component of the treatment effect. For low-opportunity taxpayers in the low- and middle-income groups, the letter had a much smaller effect, although still positive. As far as high-income taxpayers are concerned, the results instead are somewhat perverse: reported income decreased as a result of the 'threat of audit' letter, indicating lower compliance as a result of increased audit probability. One possible explanation is that high-income taxpayers sought professional advice in response to the letter, which may have helped them to find legal ways to reduce their taxable income.

The authors underline a couple of caveats of the experiment design, which again are relevant for other field TE as well. First, some taxpayers may already have thought, before receiving the letter, that the authority closely examines most returns, including theirs. In this case, the taxpayers would not respond to the letter, because this treatment would not change their perceptions on the probability of detection. Similarly, they may not believe that such close inspection is actually capable of uncovering any wrongdoing, and therefore they may still not modify their behaviour in response to the message. As highlighted in Section 1 and further discussed in Section 2.5, the credibility of enforcement is a key issue. In the long term, a high perceived probability of audit can only be sustained by a relatively high number of actual audits and detections. Secondly, the indication in the letter that previous returns could also be inspected may have induced taxpayers to maintain their reporting pattern to avoid uncovering a history of non-compliance.

Kleven *et al*. (2011) largely confirmed the importance of deterrence by using administrative data from the Danish tax collection agency, involving 42,800 randomly selected taxpayers in 2007 and 2008. In 2007, a group of these taxpayers was randomly selected for unannounced tax audits. Irregularities were corrected and penalised according to the law. In 2008, both audited and non-audited taxpayers were selected for pre-announced audits. Two treatments were tested against a control group: one letter informed them that their tax return would certainly be audited; and the other that half the people in the group would be audited. By doing this, the treatment exogenously changed audit probabilities. In this case, participants were not aware of being part of a study. The threats of audit had a large effect on compliance, which was entirely driven by self-reported income. An important advantage of Kleven *et al*. (2011) in comparison to other field TE is that they had access to audit data and could therefore directly measure variations in compliance. Importantly, audited taxpayers were selected randomly and not based on risk, which would be endogenous to their evasion behaviour. Using these data, the authors show that compliance rates in Denmark are high despite relatively high tax rates. Consistent with the importance of having the opportunity to evade, third-party reporting environments display close to full compliance, while substantial evasion occurs when income is self-reported. In addition, the authors used a quasi-experimental setting to assess the effect of changes in the tax rate on compliance. In this case, the research design exploits salient kinks in the income tax schedule, and compares taxpayers on each side. Marginal tax rates do not seem to have a large effect on compliance.

Castro and Scartascini (2013) evaluated the effectiveness of similar deterrence letters in Latin America. They analysed property taxes in a municipality in Argentina to test the effect of different factors on taxpayer compliance, including deterrence but also equity and fairness. This property tax is based on the linear front size of the property, making the tax base observable by the authorities. Therefore in this case, as in Kleven *et al*. (2011) though with a different method, the authors can directly measure compliance. Importantly, this tax is linked to the provision of public goods by the same level of government that levies the tax. In total, 23,000 individual taxpayers were randomly divided into four groups, three of which received a treatment letter influencing their perceived probability of detection or their beliefs. All letters provided information: the first one about enforcement and possible fines; the second one about other taxpayers' behaviour, thus affecting beliefs on equity and the third about how government resources are used, thus underlining fiscal exchange. Compared with Slemrod *et al*. (2001) and Kleven *et al*. (2011), the deterrence message here is much softer, along the following lines: 'did you know that if you do not pay your taxes on time you will have to pay sanctions and you may be prosecuted?'. To make the message clearer and more salient, the authors included a monetary example of underpayment with the relative amount of fines potentially due. This message increased compliance by almost 5 percentage points, while the other treatments did not have any effects compared with the control group.

Other field TE typically include a deterrence treatment along with other factors, generally finding deterrence to be effective (Cummings *et al*., 2009; Ariel, 2012; Fellner *et al*., 2013; Bott *et al*., 2014; Dwenger *et al*., 2016).[3] Moreover, a meta-analysis of lab TE argues that '[this] literature unambiguously shows that raising either the fine rate or the probability of audit will increase tax compliance' (Blackwell, 2007). Despite empirical support for the effectiveness of deterrence factors, three considerations are in order. First, only few studies, such as Kleven *et al*. (2011), can evaluate the impact of actual audits and access the related data. In most countries, particularly low- and middle-income ones, audit capacity is limited and it therefore focuses on riskier taxpayers, rather than a random selection. This makes it harder to conduct rigorous evaluations. Secondly, most studies only nudge perceptions of the probability of being audited rather than real probabilities. This is partly because audit plans are difficult to change and partly it is related to the credibility of threats of audit (see Sections 1 and 2.5). Still, studies that can make a concrete threat of an imminent audit (as in Kleven *et al*., 2011; or Slemrod *et al*., 2001) find a larger treatment effect than those using more hypothetical statements (Castro and Scartascini, 2013). Thirdly, as shown in Slemrod *et al*. (2001), deterrence messages can backfire and thus decrease compliance. This possibility has been documented in a field TE looking at a friendly version of deterrence, where new taxpayers are closely supervised by officials from the tax authority (Gangl *et al*., 2014). Whether deterrence has a positive or backfiring effect also depends on the broader nature of interaction between taxpayers and revenue authorities, which can range from an antagonistic climate characterised by 'cops and robbers' attitudes to a more service-oriented synergistic relation (Kirchler *et al*., 2008).

2.1.1 *Delivery Methods*

Most studies in this literature, including all of those reviewed so far, test the effects of potential determinants of compliance by using physical letters. While this helps in testing the effectiveness of different message contents, it leaves open questions on what is the most effective *way* to deliver these messages. In principle, revenue administrations can contact taxpayers with a variety of methods, including SMS, e-mails, phone calls or personal visits. Physical letters, while being more official and formal than many other delivery methods, are therefore not the only option. In low- and middle-income countries in particular, it can be challenging to deliver letters, as they are simply not a common way to communicate or because addresses in the taxpayer registry are not updated.

Two recent TEs tested this issue in the field. The first one was carried out in Slovenia and it involved a relatively small sample of 142 small accounting firms (Doerrenberg and Schmitz, 2015). The same message, a mix of deterrence and moral suasion, was sent to 110 firms in two ways: with a physical letter sent by post and through a visit of a tax official. In the latter case, officials were instructed to read the letter to the firm's representative, but to avoid answering other questions or providing additional information. The control group received no message. Both treatment groups report higher taxable income after the intervention than the control group, more so in the case of visits than letters. However, none of these differences is statistically significant, probably due to the small sample size.

The second field experiment on delivery methods overcomes this limitation and includes a large number of taxpayers in Colombia (Ortega and Scartascini, 2016a). This study tests the effectiveness of three delivery methods: letters, e-mails and personal visits. The authors hypothesise that personal visits in particular may be most effective, both based on the findings of the related 'get-out-the-vote' literature and because ‘actions may speak louder than words’ (Ortega and Scartascini, 2016a). According to their theoretical framework, taxpayers may update their perceived probability of detection depending on the delivery method chosen by the authority, so that more selective and more costly ones may have a higher deterrence effect. The experiment involved over 20,000 eligible taxpayers with unpaid tax liabilities between 2011 and 2013. They were divided in four groups that received either the same message, with the three different methods, or no message. Letters and e-mails contained exactly the same text, which was developed in line with the key lessons learned from the broader behavioural economics literature (see BIT, 2012). The message started by giving information about taxpayers’ account balance and other details of their unpaid tax. An explicit element of deterrence was included by underlining the consequences of not paying tax, such as penalties and potential prosecution. Finally, an appeal to national pride and social duty (‘Colombia, a commitment we can’t evade’) concluded the message. Personal visits by tax inspectors delivered the same information, following a precise protocol. Moreover, a related paper (Ortega and Scartascini, 2016b) complements these treatments by looking at the effect of phone calls in a similar experimental setting. Generally, the results show that all messages are effective in encouraging the payment of outstanding taxes. However, personal visits have a much higher effect, particularly when the high level of non-compliance with the treatment is taken into account. The estimated effect on the probability of making a payment, for those who were actually treated, was 8, 17 and 88, respectively for letters, e-mails and visits. Despite their high effectiveness, scaling up personal visits to a large number of taxpayers is not feasible in many countries. However, the study shows that a portfolio of delivery methods, that includes personal visits, can generate large revenue gains. In addition, this experiment opens new avenues for research and encourages future studies to randomise both messages and delivery methods.

This was done in a recent study, where the same messages were delivered through three delivery methods (Mascagni et al., 2017). This study, reviewed in more detail in section 2.4, shows that non traditional delivery methods like emails and SMS can be highly effective compared to letters in low-income countries.

2.2 *Social Norms, Equity and Moral Appeals*

As noted in Section 1, tax morale is a broad concept that encompasses various factors. Based on the classification of Luttmer and Singhal (2014), this section focuses particularly on intrinsic motivation, culture and peer effects and social influences.

Some of the studies reviewed in the previous section also test the importance of social and moral factors, in addition to traditional economic ones. Castro and Scartascini (2013) do not find any significant effect from the messages on equity and public services. In addition, a study by Blumenthal *et al.* (2001) uses a very similar design and method as Slemrod *et al.* (2001) and was also conducted in Minnesota. This time, however, the authors investigated the effect of moral appeals, using two treatments. The first one is a ‘support valuable services’ message, communicating how taxes are spent and encouraging taxpayers to comply voluntarily to support them. The second one is a ‘join the compliant majority’ message, indicating that the majority of citizens comply with tax laws and that one should be compliant in order to be part of that majority. Contrary to Castro and Scartascini (2013), who included their messages directly in the tax bill, Blumenthal *et al.* (2001) sent dedicated letters. As in the rest of this literature, the letters would start and end with the same text, a reminder of deadlines and indications on where to get additional information, while the middle paragraph reported one of the two messages above. The authors found little or no evidence that the letters resulted in any changes in compliance behaviour, with high-opportunity taxpayers (defined as in Slemrod *et al.*, 2001) being even less likely to respond to normative appeals.

The lack of empirical support for moral and social factors may be due to various reasons. Moral suasion may not have a large effect if the elasticity of compliance to perceptions is low. Moreover, evaders may have generally lower trust in government, therefore making the moral appeals less effective for them.

However, variations in compliance come from precisely this group, as they have some margin to increase reported income. From a methods perspective, Castro and Scartascini (2013) suggest that their 'equity' letter may have actually changed beliefs in unexpected ways. Taxpayers may have discovered that evasion is more widespread than they thought, therefore being encouraged to evade more. This is probably related to the specific message, illustrating the importance of the context and design of field TEs. While for the USA, Blumenthal *et al*. (2001) could state that 'people who file tax returns report correctly and pay voluntarily 93 per cent of the income taxes they owe'; the message in Castro and Scartascini (2013) reads 'did you know that only 30 per cent of taxpayers do not pay the [property tax]?'. Since compliance levels are much lower for the Argentinian property tax than for the U.S. income tax, similar information messages may have very different effects: a feeling of unfairness towards the 30% that do not pay in the former case and a feeling of social shame for being part of a small minority (7%) in the latter.

Moral appeals are also explored by Torgler (2004b) in a field experiment conducted in collaboration with the local tax administration in Trimbach, Switzerland. Of 580 randomly selected taxpayers, in 2002 a subgroup of them received a letter signed by the commune's fiscal commissioner. The letter was sent in a separate envelope and printed on a pink sheet to increase taxpayers' awareness of it. Participants were not informed they were part of a study, to elicit true behaviour. The relatively complex text of the letter alluded to factors such as: active citizen participation in the life of the commune, citizens' sense of responsibility and the importance of tax compliance in 'preserving this democratic and liberal structure' and to 'keep Trimbach attractive for its inhabitants'. The effect of the letter was assessed against two compliance indicators: timely filing and timely paying. From a methodological perspective, these compliance indicators have the advantage of being observable – unlike the level of under-reported income used in other studies. The results largely confirm the previous findings that moral suasion has virtually no effect on compliance. The treatment group did not significantly change the timeliness of their returns or payments in response to the moral appeal. A related article by the same author, looking at the same treatment but using different dependent variables (over-deductions, under-declaration of income and under-declaration of wealth), also fails to find any effect of moral suasion (Torgler, 2013).[4] The relatively complex treatment message may also be a reason for this lack of results.

Evidence on TV licences from Austria further supports the previous findings (Fellner *et al*., 2013). This experiment involved a sample of 50,498 individuals that, contrary to the other studies reviewed so far, is composed exclusively of potential evaders. This design feature has implications for the interpretation of results since, as discussed above, evaders may be less motivated by moral appeals. All taxpayers, except a control group of about 5% of the total sample, received a letter from the Fee Info Service, the managing agency for the TV licence. One group received a control letter containing a standard, neutral text. This standard text was then complemented with three treatment messages: a threat, social information emphasising the high level of compliance and a moral appeal stressing fairness. Moreover, for some participants, the threat treatment was combined with either social information or a moral appeal, making a total of six different letters. The messages clearly indicated that participants were selected because they were suspected of evasion, making treatments stronger. Compliance here is measured in terms of new registrations, as a binary variable indicating whether an evader started to comply or not. This feature makes this study different than the typical field TE, as tax administrations usually do not know who the non-registered evaders are. Generally, the letters were effective in increasing new registrations, compared to the 'no letter' control group. However, the bulk of the effect came from the neutral letter, indicating that the pure fact of receiving any letter may be more effective than the specific contents. Nonetheless, among the different treatments, the threat message was the only one to have any effect in addition to the neutral mail. A survey designed to investigate the mechanisms of increased compliance, shows that it is mainly due to a signal of surveillance and to increased sanction risk. Although the social information did not increase compliance overall, there is substantial heterogeneity across regions. Where compliance is relatively high, the message had a weakly negative effect, while where it is relatively low, it had a weakly positive one.

Further evidence on VAT payments by corporations in Israel supports the apparent ineffectiveness of moral suasion (Ariel, 2012). Of 4395 Israeli corporations, two groups of about 730 taxpayers each received either a deterrence message or a moral appeal. The latter stressed the social implications of non-compliance ('we all suffer from noncompliance') and informed companies how public money is spent. The study used gross sales values reported to the authority, tax payments and tax deductions as measures of compliance. None of the messages significantly increased compliance. However, corporations receiving the moral appeal claimed higher deductions, implying a decrease in government revenues. This backfiring effect recalls the perverse finding of Slemrod *et al.* (2001) where high-income taxpayers decreased compliance as a result of the deterrence letter, and it may be motivated by similar reasons. Corporations, much like high-income individuals, have large resources to dedicate to aggressive tax planning.

While these results may seem discouraging, lab experiments and recent field studies show that moral and social factors can have an important role to play in increasing compliance.

Building on the methods of the early 'searching for facts' experiments (Bosco and Mittone, 1997; Mittone, 1997), recent lab TEs linked the traditional economic factors even more closely with methods and theories from psychology. For example, Coricelli *et al.* (2010) analyse psychological costs and benefits of evasion using data from self-reports and skin conductance responses. The experiment, involving eight subjects in each repeated round, showed that taxpayers' decisions to evade are dictated by the moral and emotional implications of cheating, and not by monetary calculations alone. In addition, they show that public disclosure of deceptive behaviour, achieved by displaying the picture of cheaters, deters evasion. Looking at repeated rounds, they can show that audits decrease the extent of cheating in the next period, but larger fines increase the proportion of evaded income. Similarly, Dulleck *et al.* (2016) use heart rate variability to show that psychic stress, arising from the contemplation of the possibility to evade, can increase compliance. Still in a lab setting, Coricelli *et al.* (2014) investigate two types of shaming: one that results in stigmatisation and exclusion from the community, and another that leads to reintegration of the evader. In both cases, the shaming ritual involved displaying in public a picture of the evader, but in the reintegration treatment the picture immediately disappeared. Stigmatisation and exclusion from the community significantly increased evasion compared to reintegration. Christian and Alm (2014) further show that the promotion of empathy and sympathy can help to increase compliance.

Clearly, these studies suffer from low statistical power and artificiality of the lab setting. However, they offer interesting insights into the emotional dimension of tax evasion that may guide empirical investigations in the field. For example, some countries adopt policies of naming and shaming that are the real-world equivalent of displaying images of cheaters in the lab. The lab evidence shows that although they can be a powerful tool, the intense negative emotions associated with stigmatisation should also be taken into account.

As far as large-scale field experiments are concerned, a recent study conducted in the United Kingdom (Hallsworth *et al.*, 2017) unpacks moral factors into social norms and concerns over public goods. The basic message for the former was that the majority of people pay tax, while the latter underlined the importance of taxes to finance public goods for all. Respectively, three and two variations of these short messages were used as alternative treatments, in addition to a neutral control letter. For example, the public goods message was phrased in a positive ('Paying tax means we all gain from vital public services') and negative way ('Not paying taxes means ...'); while the three versions of the 'minority norm' message presented the same message in various degrees of specificity to the individual. The most specific and strongest one stated that 'nine out of ten people in the UK pay their tax on time. You are currently in the very small minority of people who have not paid us yet.' Similarly to Fellner *et al.* (2013), the letter is directed to those who failed to pay on time (a sample of 100,000 individual taxpayers). Thus, compliance can be fully measured, since the tax administration can observe whether and to what extent taxpayers have paid due taxes. All letters resulted in an increase in the probability to clear outstanding taxes, with the greatest effect from the strongest version of the 'minority norm' treatment. Although this treatment is designed around social norms, its effect may also be related to deterrence since it makes explicit that the revenue administration knows that the recipient is not compliant. Following these results, the authors carried out a second experiment where they compared the effect of descriptive norms (i.e. how other taxpayers behave) and injunctive norms (what other people think is the right thing to do). The former is found to have a significantly larger effect than the latter. These experiments show that large revenue gains can be made by increasing moral costs to non-compliance, with negligible additional costs to the revenue administration.

The importance of moral appeals is confirmed in a recent experiment carried out in Norway among taxpayers who earned income in a foreign country (Bott *et al*., 2014). This type of income is not pre-populated in the tax return, so compliance fully relies on self-declaration. However, the Norwegian Tax Administration recently gained access to third-party reports on this type of income. These reports allowed the identification of a sample of taxpayers whose self-reported amount was substantially different than the third-party reported one. In addition to a neutral letter, three treatments were used: (1) an equity message reminding taxpayers that most people pay their taxes fully and on time; (2) a message underlining the importance of taxes to finance public goods and services and (3) a deterrence letter. The main result is a large and significant effect of the moral appeal, particularly the equity treatment. The amount reported by those who received the moral treatment was almost double that of the base letter. Increasing the probability of detection also had a large effect on reported income, but in a different way. While moral appeals resulted in increasing the amount of income reported – the intensive margin, the detection message increased the number of people declaring foreign income – the extensive margin. A complementary survey explored how the letters were perceived by recipients and whether there were cross-contaminations whereby, for example, the moral treatment also affected the perceived probability of detection. The survey shows that moral appeals actually affected moral motivation rather than working through the probability of detection.

Del Carpio (2014) uses a mix of surveys and experimental methods to look at perceptions of compliance and enforcement, where the latter is very similar to the social or minority norm treatments used in the rest of the literature. The baseline survey reveals that in Peru the true levels of both compliance and enforcement are generally underestimated. Exploiting this feature, the experiment focuses on providing the true information on compliance and enforcement, both in separate treatments and jointly in the same letter, as well as a neutral message to remind of deadlines. The treatments' effect is evaluated on property tax payments in Peruvian municipalities. As in the case of Castro and Scartascini (2013), compliance is fully observable thanks to cadastral registries that allow the distinction between taxpayers who comply and those who do not. The strongest effect is generated by the information letter on true levels of compliance, which increases payments by 20% with respect to the no-letter control group. Comparing this effect with that of the neutral message reveals that about half of it is due to the pure reminder effect of the letter. Interestingly, the joint treatment, where both true compliance and enforcement levels are revealed, has a lower effect than the social norms treatment alone. This can be interpreted as a crowding out of intrinsic motivations to comply. Moreover a post-experiment survey shows that all treatments raise beliefs about compliance and enforcement (except for the pure reminder that only affects the former).

Other studies, which are reviewed in other sections, looked at moral suasion among other factors and found both positive effects (lab-in-the-field study by Torgler, 2003) and no effect (field TE by Dwenger *et al*., 2016). Generally, the literature is not fully conclusive on the effectiveness moral and social factors. Within countries, they are likely to play an important role in explaining the observed *level* of compliance.[5] However, *variations* in compliance, which is what field experiments capture, rely on evaders starting to comply. The fact that evaders may be less affected by moral and social appeals could therefore partly explain the lack of significant results in many studies. Across countries, the same appeal is likely to have different effects in different social, institutional and cultural settings. However, differences in design prevent direct cross-country comparisons, so we cannot identify how country-specific conditions influence results.

2.3 *Positive Incentives*

Positive incentives are closely related to moral and social factors, as they are meant to strike a chord related to feelings of social acceptance and recognition. However, while moral messages appeal to intrinsic motivations, positive rewards have an element of extrinsic motivation, as they can still be driven partly by self-interest. Contrary to deterrence, however, these incentives provide positive rewards to taxpaying, rather than negative sanctions for evaders.

In an early lab study, Alm *et al*. (1992a) tested four different options of positive rewards as incentives for compliance: (1) a lottery that could be accessed by those who were tax compliant, where the chance of winning was one in 25; (2) a fixed reward for compliant subjects; (3) an audit reduction and (4) a public good. The lottery had the largest effect on compliance, but it had the same expected value of the fixed reward. Audit reductions and public goods improved compliance but less so than other treatments. The authors therefore concluded that positive rewards need to be immediate and salient to have a significant effect. In an attempt to improve the realism of lab evidence, Torgler (2003) conducted a lab-in-the-field experiment in Costa Rica using 37 real taxpayers and real money, but neutral rather than tax-specific language. The treatments focus on fiscal exchange, moral suasion and positive rewards. Positive rewards are found to be most effective in increasing compliance, followed by the other two treatments. The number of subjects per treatment is very small (between 8 and 13) and hence, the statistical analysis may have suffered from power problems. While these two studies start shedding light on the role of positive rewards, they suffer from the typical limitations of the lab setting, including artificiality and small sample sizes.

A field experiment from Germany (Dwenger *et al*., 2016) points to the importance of distinguishing between intrinsic and extrinsic motivations of taxpaying. Intrinsic motivations are related to social norms and feelings of guilt and shame, while extrinsic motivations have more to do with enforcement and detection. The authors study a legally binding local church tax that encourages overpayments, so individuals can be evaders, compliers or donors. These groups can be clearly distinguished because the relevant tax base used by the Church, reported taxable income, is fully observable from tax declarations. While the Church in principle could cross-check this data from the tax administration, it has never done so. This feature allows the identification of taxpayers who comply entirely due to intrinsic motivations (20%) and those who evade in part or in full in absence of enforcement (80%). Both groups received three sets of treatments, in addition to a neutral letter, which aim to: (1) increase the salience of the tax and inform taxpayers of the zero audit probability, to correct possible misperceptions about audits; (2) underline deterrence and communicate positive audit probabilities and (3) appeal to moral values and provide positive rewards. The first letter just confirms that there are no misperceptions about the absence of deterrence, confirming that the group of compliers is intrinsically motivated. Increasing the probability of deterrence increases compliance substantially only for those taxpayers that are extrinsically motivated. Among the third set of treatments, a first subset included social norms and moral appeals similar to other studies reviewed in Section 2.2. These treatments do not produce any significant effect, in line with most of this literature. The second subset included public recognition and positive rewards. The former involved the publication of compliant contributors' names in the local newspaper, while the latter allowed compliant contributors to take part in a lottery and possibly gain a monetary reward. An additional treatment combined both the recognition and monetary reward in the same letter. These treatments had a positive effect on compliance for the intrinsically motivated, with larger effects when social recognition was part of the treatment. However rewards further reduced compliance for the extrinsically motivated – those who evade in absence of enforcement. A possible explanation is that such rewards reinforced a view that tax payments are voluntary rather than mandatory, and that enforcement is weak. These results show, importantly, that taxpayers' responses to various treatments depend on what motivates them to pay taxes to start with.

Koessler *et al*. (2016) provide further evidence on differences between monetary and non-monetary rewards, based on a recent field experiment in Switzerland. The main focus of this study is the role of promises in tax compliance, in the form of a voluntary commitment from taxpayers to pay their taxes on time. In return for this promise, they are included in a lottery that offers different types of rewards: one is purely monetary (cash) and the other is non-monetary (a wellness weekend). The authors find that these promises are more effective when they are associated with a non-monetary reward. This evidence reinforces the results in Dwenger *et al.* (2016) and confirm the importance of going beyond monetary rewards.

Several tax administrations around the globe have experimented the use of lotteries to increase compliance. Dunning *et al.* (2016) study the case of a lottery in the municipality of Montevideo (Uruguay) using a combination of natural, field and survey experiments. The authors start by evaluating the effect of winning the lottery on subsequent compliance behaviour. The setting is a natural experiment, exploiting the fact that the lottery randomly assigned a year-long tax holiday to eligible taxpayers – i.e. those who do not have any outstanding municipal taxes and paid on time in the previous year. The control group was generated by mimicking the lottery procedure, but selecting those eligible taxpayers who did not win. The final sample is composed of over 6000 good taxpayers, evenly divided into winners (treated) and non-winners (control). The results show that winning the lottery has a negative effect on compliance that lasts for 3 years and that reduces the proportion of good taxpayers by about 3 percentage points. Therefore, the overall effect is a revenue loss – not only because of the tax holidays, but also because the lottery lowers compliance after a win. Using a range of supporting evidence, the authors investigate a number of alternative explanations for this effect. They argue that the key mechanism is the interruption of the habit of paying taxes. Furthermore, the authors test the effect of knowledge of the lottery, as their survey shows that only 8% of taxpayers correctly identify it as a municipal reward scheme. To this aim, they conduct a field experiment involving a random sample of taxpayers, both good and bad. Three sets of letter treatments are tested: (1) a simple reminder of deadlines; (2) information about the existence of the lottery, in three variations underlining individual rewards, social rewards and the probability of winning and (3) deterrence messages highlighting the individual and social punishment for non-compliance. The messages focused on property tax, which is the most important tax levied by the municipality. The key result is that messages seem to have a positive effect on intended (i.e. checked web tax account) and actual compliance (paid on time). However, there is no discernible effect of the information about either the lottery or punishment, as compared to the control message. Finally, the survey experiment shows some evidence that the lottery could have a beneficial effect on attitudes towards taxation, despite the disappointing results on revenue generation. Overall, this study underlines the importance to carefully evaluate initiatives to increase compliance, as they can have unexpected effects. While recommending some caution in the use of lotteries, the authors also suggest practical ways to make them work better in practice, such as using in kind benefits or discounts rather than holidays that interrupt the good habit of taxpaying.

Taken together, the field studies in this section and Section 2.2 show that the response to social and moral factors is more complex and nuanced than previously documented in the lab. This response is likely to be influenced by existing beliefs on compliance and enforcement (Del Carpio, 2014), by taxpayers' initial motivations to comply (Dwenger *et al.*, 2016) and even by habit (Dunning *et al.*, 2016). Given this heterogeneity in taxpayers' responses, lab TE's small samples seem particularly unsuited to analyse the complexity of real life behaviour.

2.4 *Fiscal Exchange and Public Goods*

Fiscal exchange is another one of the mechanisms of tax morale identified by Luttmer and Singhal (2014). It refers to the hypothesis that taxpayers contribute to the public purse because they appreciate the goods, services and guarantees that the government gives them in return. This aspect is sometimes analysed by field TE in the context of moral appeals, when the message underlines the importance of taxes to finance public goods (Blumenthal *et al*., 2001; Castro and Scartascini, 2013). Some of the experiments discussed in section 2.2 include treatments related to fiscal exchange. They generally find little support for its effectiveness in increasing compliance, revealing either no effect (Blumenthal *et al*., 2001; Ariel, 2012), or smaller effects than other moral appeals (Torgler, 2003; Bott *et al*., 2014; Hallsworth *et al*., 2017). However, others find some evidence that public service delivery may be an important component. For example, Castro and Scartascini (2013) show that their treatments have heterogeneous effects based, among others, on the level of provision of public goods. While their evidence cannot be considered as causal, their findings suggest that a higher level of public goods provision is associated with higher compliance. Similarly, preliminary results from a recent field experiment in the municipality of Santa Fé (Argentina) also show heterogeneous effects depending on the level of public goods provision (Carrillo *et al*., 2016). In this case, the authors use a unique setting where a public good, namely, 400 individual sidewalks, is randomly assigned to those who did not have any outstanding property tax payments. Construction of the sidewalk increases the probability that taxpayers will keep complying, compared to other good taxpayers, and this effect lasts for over 3 years. Moreover, the effect was larger in those neighbourhoods where the provision of the public good was lower. Clearly, this study is closely related to the rewards literature reviewed in Section 2.3. However, as the reward here is a public good, its effect may occur through the reciprocity channel. An alternative explanation would be related to peer effects, since the reward in this case is highly visible.

Further evidence on fiscal exchange comes from a recent field experiment from Rwanda (Mascagni *et al*., 2017). In this context, as in low- and middle-income countries more generally, non-compliance with the treatment is a greater concern because taxpayer registries are often incomplete or not kept up to date.[7] Building on the study by Ortega and Scartascini (2016a), the set of treatments in this study vary both the content of the message (reminder, deterrence, public services) and the mode of delivery (letter, e-mail, SMS). The contents of letters and e-mails are identical, so they can be directly compared. They include a reminder of deadlines, which is combined with a message designed to underline deterrence or fiscal exchange. In line with the best practice in behavioural research (see BIT, 2012), the messages were personalised with taxpayers' names and accompanied by images: a justice hammer for deterrence and a graph showing how tax revenue is spent for fiscal exchange. All messages were signed by the Commissioner General and sent by the Rwanda Revenue Authority, through official channels, about a month before the declaration deadline at the end of March 2016. The SMS treatments are designed along similar lines, but are necessarily shorter and do not contain images. The results confirm that messaging taxpayers is an effective strategy to increase compliance in the African context as well. Taken together, the treatments increased compliance by about 20 percentage points. The combination of the public service message and SMS delivery mode was particularly effective, having a significantly higher impact than the control message. Deterrence instead seems to be relatively less effective in Rwanda, while reminders of deadlines significantly increased tax revenue. As far as delivery methods are concerned, e-mails and SMS proved to be highly effective as compared to letters, although SMS are harder to compare directly with other delivery methods. These results suggest that in a low-income context, with limited resources to support revenue administration, relatively cheap delivery methods like SMS and e-mails can be highly effective.

In addition to these field TE, evidence from the lab also provides insights into fiscal exchange. Building on previous lab studies showing that taxpayers evade less under conditions of uncertainty (see Section 2.1), Alm *et al*. (1992a) show that this relation is reversed when public goods are introduced in the experiment. When participants know that tax revenues are used to benefit everyone, to simulate a public good, then increased uncertainty leads to less compliance. This finding highlights the importance of the institutional environment in which the compliance decision is made. Other lab studies argue that rational utility-maximising taxpaying behaviour must take into account both tax payments and the transfers that result from the redistribution of the tax yield. Two lab studies (Becker *et al*., 1987; Mittone, 1997) confirm this hypothesis by showing that when taxpayers benefit from public spending, they are less likely to evade taxes. In addition, Becker *et al*. (1987) find that the perceived equity in government transfers helps to fight evasion. Still in the lab setting, Alm and Jackson (1993) analyse fiscal exchange, the use of tax revenue and the decision processes for deciding such uses. Individuals tend to comply more when they know that the tax yield is spent on programmes they approve of and when they take an active part in the decision on spending. Compliance is higher when the outcome of the vote reflects widespread support rather than a close win. On the contrary, compliance is reduced when unpopular expenditure programmes are imposed top–down on participants. Similarly, the lab study by Lamberton *et al*. (2014) shows that allowing taxpayers to indicate preferences on government spending increases tax compliance. These findings are confirmed in a lab in the field study conducted in Costa Rica and Switzerland (Torgler, 2004a). The author shows that voting has a positive effect on compliance, which depends on the benefits obtained in return of taxes.

Cummings *et al*. (2009) use a lab-in-the-field experiment to analyse the role of governance quality on compliance. Recognising the limitations of the lab, they also use survey data from Afrobarometer to corroborate the results. To test the hypothesis that perceptions on the quality of governance increase compliance, two parallel experiments are carried out in Botswana and South Africa – two countries with similar tax systems but different political histories and institutions. This hypothesis is clearly related to tax morale more broadly, whereby the motivation to pay tax derives from a social and moral obligation to contribute to society. However, the link between taxpaying and public goods, through service delivery, depends on the quality of governance. The authors show that a lower level of governance is associated with lower compliance. Moreover, while increasing enforcement always increases compliance, this effect is larger in presence of better governance. This finding is echoed in Torgler *et al*. (2010) who combine findings obtained using survey data, macroeconomic indicators and experimental design. Not only do they argue that governance quality matters for compliance; but also they identify the quality of political institutions as one of the most important determinants. This broad indicator includes aspects such as accountability, rule of law and control of corruption.

Other studies have focused on the relation between taxpaying and public accountability using experimental or quasi-experimental methods, such as Paler (2013), Martin (2014) and Gadenne (2017). However, they have been more concerned with testing the hypothesis that reliance on taxes has beneficial effects on public accountability, than with the importance of the latter as a motive for tax compliance. The first two studies are lab-in-the-field experiments, respectively, in Indonesia and Uganda, while Gadenne (2017) uses municipality-level data from Brazil in a non-experimental setting. While they are beyond the scope of this review, it is worth mentioning that they provide evidence of a positive relation of tax revenues with public accountability and the quality of public spending. In addition, survey-based studies also support the fiscal exchange hypothesis (Fjeldstad and Semboja, 2001; Ali *et al*., 2015a). Therefore, the broader literature seems to confirm the potential importance that this aspect can play in increasing compliance, even if the evidence from field TE is still limited. As such, the case of fiscal exchange is a good example of the complementarity of different types of studies, among which are field and lab TE. While we know from the latter (both lab and lab-in-the-field) that fiscal exchange is a potential mechanism, it is unclear whether it can be leveraged by revenue administrations to increase compliance in practice. Still, the effectiveness (or lack thereof) of specific treatments in field TE does not overturn the conceptual arguments and the evidence from the broader literature. Lab TE provide a useful ‘proof of concept’ for the relation between compliance and public goods and services. Further studies, including but also beyond field TE, need to explore how this relation may play out in the complexity of real taxpaying situations.

Clearly, the fiscal exchange motive involves important considerations about the quality of the institutional environment, as well as the level of public services and goods offered by the government. The evidence so far is still too limited and scattered to draw clear cross-country comparisons. Still, low-income countries generally have lower levels of public goods and services (both in quantity and quality), which can make it particularly challenging to leverage fiscal exchange to increase tax compliance. However, recent field TE show that fiscal exchange could be a viable strategy in this type of context, and perhaps particularly where the level of public good provision is lower (Carrillo *et al*., 2016; Mascagni *et al*., 2017). Moreover, different types of public goods may generate different taxpayers' responses. For example, countries where violent crime is more widespread may respond more to appeals on the importance of taxes to finance policing than to support other public services. A recent study based on Afrobarometer data compares four frican countries and explores fiscal exchange, among others (Ali *et al*., 2015a). It shows that the link between compliance and public service delivery is based on specific services that are different in each country.

2.5 *Information and Third-Party Reports*

One way in which revenue authorities can enforce tax laws is by making full use of information, and particularly of third-party reports. This aspect is closely linked with the issue of deterrence and probability of detection, since the effective use of appropriate information increases the probability of detection. Some of the studies reviewed in previous sections specifically show that the treatments are only or mostly effective on those who self-report income (Slemrod *et al*., 2001; Bott *et al*., 2014; Kleven *et al*., 2011). Kleven *et al*. (2011) argue that compliers are unable, rather than unwilling, to cheat. In particular, taxpayers who are only subject to third-party reporting (e.g. those only receiving wage income) may not be able to cheat even if they wanted to. Therefore, for this group the probability of audit is a poor proxy for the probability of detection, as the authorities could easily uncover any irregularity simply by matching information reports to tax returns. Consistently, Kleven *et al*. (2011) find the audit letters to be largely ineffective among this group. Using audit data, they also show that while evasion in the third-party reporting environment is close to zero, it is instead substantial in the case of self-reporting. Interestingly, this finding holds even considering individuals who earn both types of income, as they fully comply for the portion of income that is third party reported while they tend to evade on self-reported income. This finding suggests that it is information reporting, rather than an emotional or social aversion to cheat that drives compliance.

Pomeranz (2015) conducted a field experiment to look further into information as a key determinant of compliance in Chile, in close collaboration with the Chilean tax authority. The study focused on VAT, which is often praised for its built-in incentive structure, triggered by the verifiable paper trail it generates. If the self-enforcement properties of the VAT are effective, taxpayers' response to increased probability of audit would be very low due to the preventive deterrence effect of the paper trail. To evaluate the validity of this hypothesis, transactions that are covered by the paper trail (those between firms) are compared to those that are not (those with final consumers). A first experiment involved sending letters to over 100,000 randomly selected firms, indicating increased audit probabilities. Consistent with the hypothesis, the letters generated a strong increase in tax payments that is largely driven by sales to final consumers, rather than by transactions between firms. Moreover, small firms appear to respond more, perhaps because they tend to have higher shares of retail sales. A second 'spillover' experiment looked particularly into the mechanism of the paper trail. In this case, half the sample received a pre-announcement of an upcoming audit. The full sample was then audited and the information was made available to the authors. In addition, data were collected for the trading partners of the audited firms, before the audit. In presence of the paper trail, one should observe a spillover effect of the audit on the suppliers of the audited firms, which will increase compliance as a result. Finding empirical support for this mechanism, this study argues that the effectiveness of the paper trail gets multiplied thanks to spillover effects. The paper trail allows the tax authority to easily identify discrepancies in reporting between the supplier and the audited firm by cross-checking records, therefore encouraging suppliers to also comply more.

Carrillo *et al*. (2016) show a fundamental limit of the role of third-party reporting, particularly in countries where constraints on information and enforcement are high. The study is based on a natural experiment in Ecuador and focuses on corporate income taxes. While third-party information has been available in Ecuador for some time, it is only recently that the tax authority has started using it to increase compliance. In 2011 and 2012, the tax administration notified about 8000 firms of discrepancies between their self-reported revenue and third-party information. Because the notification concerned tax returns for previous years, the response is not due to changes in actual behaviour but only to changes in reporting. Therefore, even without a control group, the reported values before the notification provide a counterfactual for the true, post-notification reports, therefore allowing the identification of causal effects. Third-party reports show higher taxable income than the self-reported amount in 24% of the cases. Interestingly, under-reporting also seems to occur on firms' expenditure, although third-party information on costs is much more limited than for revenues. This experiment presents two important results. First, a substantial proportion of firms fail to file an amendment even after receiving the notification. Under weak credible enforcement, firms still under-report revenues because they do not believe the revenue authority has the ability to punish them even when it has detected under-reporting. Increasing credibility would require the tax administration to follow up on the notifications about discrepancies with actual audits. Failing that, taxpayers do not seem to fully respond to the notifications. However, audits are costly and many countries simply do not have the capacity to follow up every case where evasion is suspected. Therefore, when traditional enforcement mechanisms are weak, third-party information may not be effective.

Secondly, among firms that do file an amendment, reported income increases by large amounts. When firms are notified the exact amount in the third-party reports they tend to amend for larger amounts (93% of the discrepancy) rather than when they do not know the specific amount (36% on average). However, for every dollar of revenue adjustment, costs tend to increase by 96 cents. The line item of increased costs most frequently adjusted is 'other administrative costs', which is particularly difficult to verify because of its miscellaneous nature. By amending both revenue and costs, post-amendment profits remain similar and therefore increases in tax payments are limited. In other words, firms seem to adjust less monitored margins (i.e. costs) as a response to increased monitoring on income. This result illustrates how, in presence of incomplete information on all relevant dimensions of compliance, third-party reporting may not have a large impact on revenue mobilisation. Importantly, the same would happen when the information is available but taxpayers know the tax administration cannot use it, typically because of administrative constraints.

While administrative constraints are certainly a greater concern in low- and middle-income countries, Slemrod *et al*. (2015) show that similar behaviour as documented by Carrillo *et al*. (2016) occurs in the USA too. Since 2011, firms that process electronic payments (e.g. credit cards) are requested to report this information to the Internal Revenue Service (IRS). These reports are then used by IRS to cross-verify the income declared by small businesses. Although a clear effect of these information reports is difficult to identify, they seem to result in higher declared income particularly for those firms that did not file in the previous year. Firms bunch at the exact amount of the information reports, either because they fear that reporting less would trigger an audit or because that amount accurately reflects true income. However, many businesses report equal amounts for income and costs, resulting in zero profits. This study reports some evidence that the new information reports made firms more likely to do so. Therefore, as in the case of Ecuador, the potential benefits from increased declared revenues were partly offset by increased costs. Also in this case the adjustment occurred mostly through items that are difficult to verify, such as 'other expenses'. This study shows that using third-party reporting to improve compliance of businesses may be challenging in all contexts. Contrary to wage income, business income involves complex reporting requirements that offer more opportunities to minimise tax payments through less verifiable margins.

Despite these limitations, there is increasing evidence that a more effective use of information and IT can be beneficial for revenue administrations in low-income countries. For example, two recent non-experimental studies looked at the effectiveness of electronic billing machines (EBMs) in Africa using data from VAT tax returns (Eissa and Zeitlin, 2014; Ali *et al*., 2015b). They found that both in Rwanda and in Ethiopia, firms substantially increased VAT payments after EBM adoption, although the size of the increase varies across firms' type, size and sectors. However, it is still unclear whether these increased payments have been offset by other adjustments.

2.6 *Tax Collectors and Their Incentives*

In addition to problems of capacity and enforcement, tax collection is likely to be constrained by the presence of widespread corruption in tax administrations, particularly in low- and middle-income countries. In countries where the use of IT systems and third-party reporting is still limited, contact with tax collectors can be an important factor in explaining compliance. Tax administrations may offer incentives to collectors to minimise the leakage of revenue into corruption. The incentive may be, for example, to allow collectors to cash a share of tax revenues, thus increasing their overall salary. In countries where wages are relatively low, this monetary incentive can be very salient. However, an adverse consequence of such incentives may be to encourage collectors to use coercion and aggressive methods to raise tax. There is evidence that this may have the counterproductive effect of generating tax resistance and confrontation between taxpayers and the authorities, thus eventually decreasing compliance (Fjeldstad and Semboja, 2001). Although 'tax farming' may be effective in increasing revenue in the short run, it also bears some potential political costs that may be detrimental in the longer run.

A recent field TE investigates the effect of incentives for tax collectors both on revenues and on taxpayers' satisfaction with the practices of the tax department. Khan *et al*. (2016) focus on property taxes in Punjab, Pakistan, and use both tax returns data and a satisfaction survey. The treatments consist of three incentive schemes for collectors. The first one is a pure revenue incentive, where they receive a percentage of any increase beyond a benchmark level of revenue. This incentive is relatively large, as it can almost double their base salary. The second incentive adds to the first one non-revenue outcomes, by adjusting the reward according to measures of taxpayer satisfaction and the accuracy of tax assessments. The third scheme aims to take into account the complexities of tax collection, which may not be easily captured by objective measures. To do so, it is based on subjective assessments of managers who distribute a fixed amount of bonus to a pool of good collectors.

The incentives are expected to interact with existing bribery in two possible ways: they might result either (1) in higher tax payments with a lower level of bribes, or (2) in a constant level of tax payments but higher bribes, as collectors now want to be compensated for a higher foregone gain (the incentive). The results show that on average the treatments are effective in increasing revenue, while taxpayer satisfaction does not change significantly. The most effective treatment is the pure revenue incentive, while the second one has a smaller effect and the third no effect. The additional revenue seem to come primarily from an increase in the tax base, through new assessments and reassessments, and from the recovery of arrears. Most of the increases come from high-value properties, suggesting that collectors may be maximising the amount of revenue from a few large taxpayers. This would also explain why the perception of corruption and satisfaction does not decrease significantly on average. Overall, this study suggests that incentives can provide an effective way to increase revenue. Although they seem to do so without much political cost, it is still unclear whether such costs may occur in a longer time frame than the one of the study.

2.7 *Complexity and Compliance Costs*

Tax systems are typically complex and it is often difficult for taxpayers to fully understand them in detail. This can result both in misperceptions about the system and in compliance costs, due to the time and resources dedicated to tax filing. Therefore, complexity and compliance costs may be an important element in explaining misreporting and non-compliance, in addition to plain tax evasion. Compliance costs are particularly relevant for small taxpayers and businesses, as well as first-time filers. There is evidence that small businesses face a cost of up to 15% of their turnover, while this percentage is lower for larger businesses (Coolidge, 2012). Moreover, taxpayers in low- and middle-income countries are particularly affected, due to lower literacy and less availability of trained accountants. As a result, they may file only partially, not file at all, or not do so in time, therefore being partially or fully non-compliant. Modern revenue authorities are making greater efforts to simplify their procedures and to act as tax educators. However, complexity and compliance costs remain real, practical constraints in many countries.

A recent natural experiment used administrative data to show that taxpayers are often confused by complex tax systems (Feldman *et al.*, 2016). This study, conducted in the USA, analyses labour responses to the loss of Child Tax Credit when a household's child turns 17. The loss of Child Tax Credit results in a lump sum increase in tax liabilities but not in a change of the marginal tax rate. However, the authors show that confused taxpayers respond to this change by interpreting it as an increase in their marginal tax rate, thus documenting a substantial degree of misperception of the tax system.

In addition to misperceptions, taxpayers may actually lose out because of the complexity of the system and the compliance costs associated with it. Benzarti (2015) uses administrative data from the IRS to show that taxpayers forego large amounts of money because of the compliance cost associated with reporting deductions. Taxpayers in the USA have two choices: they can itemise all deductions or claim a standard amount. Itemising results in a monetary benefit when the sum of all itemised deductions is larger than the standard one. If tax filing is costly, rational taxpayers should only itemise when the benefits are higher than the costs of doing so. This is mostly a time cost, due to two elements: record keeping and actual filing. The study reveals that the average taxpayer foregoes $617 that could have been saved by itemising deductions. However, there is large heterogeneity among taxpayers. Richer households lose more than poorer ones through this mechanism, perhaps because they value their time more (as a result of wage differences). These compliance costs are higher than previously documented, either because of actual compliance costs or a psychological bias against tax filing.

Similarly, Gillitzer and Skov (2013) report a large amount of under-reporting of charitable contributions, which could be deducted, with a resulting loss for taxpayers. The natural experiment in this case comes from the introduction of third-party reports for donations by the Danish revenue administration. Since 2008, these reports were used to pre-populate tax returns. This measure resulted in doubling the number of tax deductions claimed, and a 15% increase in the value of these claims. This shows that taxpayers were foregoing tax benefits, as there is no evidence of an actual change in charitable giving around the time of the reform.

2.8 *Learning and Long-Term Perspective*

Some of the letter treatments reviewed so far, and particularly those related to deterrence (Section 2.1), have large effects on compliance. However, is that effect sustainable or is it a simple nudge? Do taxpayers learn from these treatments? In a way, these questions are related to the issue of credibility of enforcement discussed by Carrillo *et al*. (forthcoming). Taxpayers may react to a threat of audit, but they may go back to their previous behaviour if no action follows that threat. Field TEs typically evaluate only the immediate effects of treatments, since obviously it takes time to observe longer-run effects. However, for policy purposes, it is important to know whether the same measure can be repeated every year with similar results; whether taxpayers will learn from such interventions and therefore do not need regular communications; and perhaps if such interventions should be used only for new taxpayers. All these questions bring into the discussion a longer-term perspective and considerations on sustainability.

A few of the papers reviewed in previous sections discuss this issue and provide some initial evidence. Carrillo *et al*. (2016) also show that the effect of allocating the randomised public good last for at least 3 years after the intervention. Moreover, Dunning *et al*. (2016) show that the negative effect of the lottery lasts for 3 years after winning. Based on this finding, along with a range of other supporting evidence, the authors argue that habit plays an important role in explaining compliance. If habit is a key mechanism, then can the type of treatment messages of the typical field TE change it or do they simply nudge taxpayers into short term behavioural changes?

A recent paper explored this issue in detail by looking at take-up of the Earned Income Tax Credit (EITC) in the USA (Manoli and Turner, 2014). Naturally, the paper focuses on a country with a longer history of TEs, therefore building on previous work. In order to increase take-up of the EITC, the IRS sends out notices to taxpayers who seem eligible but who have not claimed the tax credit. The authors are particularly interested in understanding whether these notices have purely short-term nudge effects, or if they also produce longer-term learning effects. First, a descriptive analysis of tax returns data show that information and knowledge of the benefit schedule is indeed linked to low take-up. The study is then based on three experimental components. The first one exploits a computer glitch that resulted in the failure to notify a number of taxpayers who should have been notified. By comparing the group who received the notice and those who did not (but should have), the authors show large nudge effect of the notices on take-up, with a substantial gain of about 4% of labor income as a result of the credit. However, such effect gradually disappears, decreasing drastically in the year following the notice and disappearing completely 3 years later. The second component largely confirms these results by using data from EITC-eligible taxpayers in California in 2009. It looks at different informational treatments, rather than just the difference between receiving a notice or not. The sample is composed of taxpayers who had already received a notice previously but did not respond. The simplest and most salient notice has the largest nudge effect, as can be expected. However, in the years after the treatment the differences between treatment notices disappear, as none of them induces meaningful learning effects. The third component looks at taxpayers who did not claim their tax credit and who were selected for a random audit. Following the audit, they were found to be eligible for the EITC. This can be seen as a more heavy-handed treatment than notices, and may be expected to have a larger learning effect for taxpayers. Although learning effects appear to be larger than those of the notices, they still seem to fade away 2 years after the audit. While the smaller sample size in this component calls for some caution, the results are in line with the previous ones. Generally, they suggest that regular notices may be needed to increase take-up, as it seems very difficult to generate learning effects about the tax code.

Importantly, learning does not happen only through the revenue administration or through official measures more generally. A recent natural experiment (Alstadsaeter *et al.*, 2014) shows that family networks are an important channel for transmitting information about the tax system. The authors look at a well-defined, legal tax shelter in Norway. Using tax returns data, they exploit a discontinuity in the availability of the tax shelter that allows them to compare similar individuals who have or do not have access to it. By doing this, they show that similar individuals, with similar social networks, have different taxpaying behaviour according to whether or not they are part of a network including an eligible individual. These results suggest that taxpayers affect each other along social networks, which are therefore an important element in explaining tax avoidance and evasion.

The evidence on learning through networks also raises important questions about spillover effects in all the field TE literature. If the information contained in messages does not only reach taxpayers in the treatment group but also other taxpayers in the control group, estimates of treatment effects may be biased. Many field experiments do not explicitly take this issue into account. However, a recent study analysed spillovers in detail by evaluating the effect of the treatments of a previous field experiment (Fellner *et al.*, 2013) on untreated taxpayers (Drago *et al.*, 2015). Contrary to the sidewalks analysed by Carrillo *et al.* (2016), who also find large spillover effects, letter treatments or the behavioural reactions they may trigger are not observable by other taxpayers. Therefore, any effect on the untreated should be due to communication among taxpayers. A related survey confirms that communication patterns exist particularly among neighbours, which adds a novel dimension to the family networks analysed by Alstadsaeter *et al.* (2014). This study shows large spillover effects, as taxpayers who did not receive the letters, but were part of the same network of treated ones, were more likely to switch from evading to complying. These spillover effects are of similar magnitude as the direct treatment effects. Furthermore, this study shows how information about treatments spreads within networks, which has implications for policymakers who may want to maximise the spillover effects from their communication strategies. This study certainly opens new avenues for future research, including new questions on networks as well as the possibility to make better use of geographical information to identify spillovers.

3. Conclusions

The TE literature has developed greatly in the last decade, moving gradually from the lab to the field. The latter have seen a fast expansion in recent years and have often become the gold standard in compliance research. Indeed, they present a few important advantages: they are fully embedded in real taxpaying situations, thus being more suitable as a basis for policy advice; they benefit from more statistical power, as samples are much larger than what the lab previously allowed; and they provide further insights into the heterogeneity of taxpayers' responses to various determinants of compliance. This paper reviewed this growing literature and allows drawing a few concluding reflections both on the progress done so far and on the future of field TE.

As far as research design is concerned, field TE still present important differences. Although some studies follow the best practice of behavioural research, as summarised for example in BIT (2012), these insights are not applied uniformly across the literature. For example, some studies include images in the messages (Castro and Scartascini, 2013; Mascagni *et al.*, 2017), while others do not (Torgler, 2004b). Moreover, similar messages are presented with various degrees of complexity in different studies. While Hallsworth *et al.* (2017) and Castro and Scartascini (2013) only make small variations to a standard message, Slemrod *et al.* (2001) or Torgler (2004b) changed entire paragraphs. Even the dependent variable, compliance, is often measured in different ways, in terms of outstanding payments (Hallsworth *et al.*, 2017), as new registrations (Fellner *et al.*, 2013), or as increases in reported income (Slemrod *et al.*, 2001). Other differences regard the presentation of treatment messages within the tax bill (Castro and Scartascini, 2013) or, as in most cases, as a separate letter; as well as the tax types considered, such as VAT (Pomeranz, 2015), personal income taxes (Bott *et al.*, 2014), corporate taxes (Ariel, 2012) and property taxes (Castro and Scartascini, 2013). Although compliance *issues* are similar across tax types, the compliance and enforcement *levels* differ, with possible implications on the results (see Section 2.2).

Importantly, differences in design can influence the results and their interpretation – making rigorous and context-appropriate research designs crucial to their credibility. However, it is usually difficult to disentangle specific features of design and their impact on outcomes, since a combination of various features are usually present in each experiment. For example, Torgler (2004b) uses separate pink letters to increase salience, but the text is relatively complex. Other studies have simpler text, but the message is included in the tax bill (Castro and Scartascini, 2013). Since experiment designs differ on more than one feature, it is difficult to evaluate the effect of a specific design feature on results. One of the papers reviewed here used various phrasings for the same broad message, showing that stronger messages have larger effects (see Section 2.2). This variation is encouraged in other studies, if the sample allows for multiple sub-treatments. In addition to the content of the message, the way in which it is delivered may matter for results. Some studies have included this aspect explicitly as a research question, showing that messages delivered by tax officials may have a larger effect (see Section 2.1.1). However, many studies do not make it clear who delivers the letters. Along similar lines, most recent field TE use a neutral 'control' letter to isolate the effect of receiving any letter from the specific content of the letter. This has proven particularly important, since indeed the control letter has been found to account for a large portion of the results in some cases (Del Carpio, 2014; Dunning *et al.*, 2016). If a control letter is not included, the authors may wrongly infer that any detected effect is due to the content of the letter, whereas it is instead due to the pure fact of receiving it. Moreover, complex letters may be responsible for non-significant or weakly significant results, as taxpayers may not pay enough attention or fail to understand the intended message. In addition, more nuanced results can be shown in studies that analyse subgroups separately. The absence of any effect at the aggregate level may hide heterogeneous responses across groups (Fellner *et al.*, 2013). Aside from questions on optimal intervention design, these differences make it difficult to directly compare results across the literature.

Despite these differences across individual studies, this literature points to a few common results on what matters for compliance. Deterrence factors seem to be the most effective method to increase compliance in the short run, particularly regarding those taxpayers with self-reported income (Slemrod *et al.*, 2001; Kleven *et al.*, 2011). However, its effect may not be sustainable over time (Manoli and Turner, 2014) and in presence of weak follow-up enforcement (Carrillo *et al.*, forthcoming). The evidence on moral appeals and social norms is less conclusive, as evidence is available both in favour (Hallsworth *et al.*, 2017) and against (Blumenthal *et al.*, 2001) its effectiveness. However, tax morale interacts in complex ways with the institutional environment and preferences on public services, among others (see Section 2.2). Factors other than deterrence are particularly important for two reasons. The first one, mentioned before, has to do with the sustainability of threats in weak enforcement environments. Increasing actual enforcement is very costly and may not be optimal for many tax administrations. Therefore, it makes sense to explore voluntary compliance strategies that are more cost effective. Secondly, an approach based only on deterrence could fuel a confrontational relationship between citizens and the state, which in turn may increase taxpayers' resistance and lower compliance. Last but not least, there are some important practical issues that act as obstacles to compliance, such as the complexity of tax systems and compliance costs. Moreover, several studies show that compliance behaviour is more nuanced than documented before and depends, for example, on existing beliefs on enforcement and compliance (Del Carpio, 2014; Carrillo *et al.*, forthcoming) and taxpayers' initial motivations to comply (Dwenger *et al.*, 2016).

Nonetheless, some issues are under- or unresearched. Hallsworth (2014) identified some new avenues for research, for example, the differential responses to treatments by firms or individuals, the relation between attitudes and behaviour, the effectiveness of public goods provision and the role of networks. Some of these questions have only been explored very recently, such as the effect of networks (Drago *et al.*, 2015), different methods of communication (Mascagni *et al.*, 2017; Ortega and Scartascini, 2016a), the effectiveness of public goods (Carrillo *et al.*, 2016) and positive rewards (Dwenger *et al.*, 2016; Dunning *et al.*, 2016). However, other topics remain largely unexplored. For example, tax advisors are likely to substantially affect compliance, but their influence is currently under-researched (Slemrod, 2016). Moreover, the longer-term effects of communication strategies to increase compliance are still uncertain. Although a few studies consider longer-term effects (Manoli and Turner, 2014; Dunning *et al.*, 2016), the majority of field experiments focus on the short term, often the same year of the intervention. Moreover, more evidence is needed on new registrations. This is particularly important in countries with large informal sectors, where compliance behaviour often takes the form of non-registration. Although some evidence is available from non-experimental studies (De Giorgi *et al.*, 2015; Gangl *et al.*, 2015), field experiments have not fully explored this issue. The literature could usefully expand by including these factors in the future, to provide a more complete evidence base on compliance.

Furthermore, more evidence is required from low- and middle-income countries. This review has made a particular effort to include studies from all regions of the world, in a literature that is largely dominated by Europe and the USA. The increasing availability of tax returns data from low- and middle-income countries is allowing researchers to draw a more global picture than ever before. Still, only one field tax experiment is currently available for Africa or for any low-income country, although others are underway. Generally, the types of issues that emerge as key determinants of compliance are very similar across countries. However, there are important context-specific factors that potentially affect research design, results and policy implications. First, moral appeals often include information on existing compliance levels. While they may be effective in countries with low evasion, they can backfire where evasion is widespread (see Section 2.2). Lower compliance levels in low-income countries make it harder to use 'join the compliant majority' messages. Secondly, the lower level of public goods provision can have mixed implications for treatments related to fiscal exchange. On the one hand, it may make it hard to use it as a leverage to encourage taxpayers to comply. On the other hand, it may make promises related to public goods more salient, precisely because they are more scarce. Two recent studies from low- and middle-income countries provide some support support for the fiscal exchange hypothesis (Carrillo *et al*., 2016; Mascagni *et al*., 2017). Thirdly, low-income countries have typically a lower level of administrative capacity and often sending physical letters can be particularly burdensome. For this reason, the recent TE from Africa used a variety of delivery methods, showing that low-cost methods, such as SMS and e-mails, can be highly effective (Mascagni *et al*., 2017). Using this type of technology may facilitate the expansion of field TE in Africa. In addition, while in rich countries threats of audits are likely to be more credible, they may not be taken seriously in contexts where administrative constraints are well known. Again, heterogeneity matters as large taxpayers are still less likely to be affected by such threats in any country. Despite these initial comparisons between low- and high-income countries, it is still difficult to identify systematic differences as evidence for the former is still relatively scarce. As highlighted earlier, cross-country comparisons are generally hard because any difference in results may be due to variations in context, design or both.

In conclusion, this literature certainly promises to provide further useful insights on compliance both for the research community and for policymakers. There are at least three directions for further research emerging from this review. First, there is still a lack of evidence from Asia, Latin America and, particularly, Africa. A few ongoing studies are starting to fill this gap, and more work needs to be done to build a truly global body of knowledge. Secondly, this literature would benefit from more standardised methods across studies. Although research design needs to take the context into account, researchers should make greater efforts to achieve comparability across studies. Finally, there are a number of practical issues that are known to matter for compliance but lack sufficient rigorous evidence. For example, we know that complexity and compliance costs are important constraints, particularly in low-income countries. However, there is still relatively little experimental evidence, mostly from natural experiments, on what tax administrations can do to tackle them. Moreover, the issue of informality remains largely untouched in this literature. Although administrative data naturally exclude the informal sector, it may be possible to look at new registrations, potentially by matching tax data with other sources of information. Finally, further studies should keep making efforts to disaggregate effects along dimensions of heterogeneity that can offer a more nuanced picture of compliance.

Acknowledgements

This paper was produced with the support of the International Centre for Tax and Development (ICTD), which is financed by DFID and NORAD. I am particularly grateful for the comments provided by Mick Moore and by anonymous reviewers. All remaining errors are my own.

Notes

1. From the official questionnaire of the World Value Survey Wave 6 2010–2014, available at www.worldvaluessurvey.org.
2. The exception to this is data from audits, that however are not available to researchers in many countries.
3. These studies are reviewed in other sections of this review, as their main results are related to the other factors explored.
4. Note that the author had access to both declarations and audit data, which are used as a basis for the calculation of the dependent variables.

5. For example, the review paper by Alm and Torgler (2011) underlines the role of 'ethics' in explaining compliance decisions and argues that it should be taken in account when designing strategies to increase compliance.
6. As such, this experiment was preceded by a pilot. Results and lessons learned from the pilot are available as a separate paper (Mascagni *et al*., 2016).
7. For example, Ortega and Scartascini (2016a) report that in Colombia only 38% of those assigned to the letter treatment received it, while the same figures for e-mails and personal visits are respectively, 88% and 25%.
8. This includes taxpayers who pay self-declared income taxes, namely, corporate and personal income taxpayers, but excluding those who only pay withholding taxes on their wages.

References

Ajzen, I. and Fishbein, M. (1977) Attitude-behavior relations: a theoretical analysis and review of empirical research. *Psychological Bulletin* 84(5): 888–918.

Ali, M., Fjeldstad, O.-H. and Sjursen, I.H. (2015a) To pay or not to pay? citizens' attitudes towards taxation in Kenya, Tanzania, Uganda and South Africa. *World Development* 64: 828–842.

Ali, M., Shifa, A., Shimeles, A. and Woldeyes, F. (2015b) Information technology and fiscal capacity in a developing country: evidence from Ethiopia. Working paper 31, International Centre for Taxation and Development.

Allingham, M.G. and Sandmo, A. (1972) Income tax evasion: a theoretical analysis. *Journal of Public Economics* 1(3–4): 323–338.

Alm, J. and Jackson, B. (1993) Fiscal exchange, collective decision institutions and tax compliance. *Journal of Economic Behavior and Organisation* 22(3): 285–303.

Alm, J. and Torgler, B. (2011) Do ethics matter? Tax compliance and morality. *Journal of Business Ethics* 101(4): 635–651.

Alm, J., Jackson, B. and McKee, M. (1992a) Deterrence and beyond: towards a kinder, gentler IRS. In J. Slemrod (ed), *Why People Pay Taxes*. Ann Arbor, USA, University of Michigan Press.

Alm, J., McClelland, G.H. and Schulze, W.D. (1992b) Why do people pay taxes? *Journal of Public Economics* 48(1): 21–38.

Alstadsaeter, A., Kopczuk, W. and Telle, K. (2014) Social networks and tax avoidance: evidence from a well-defined Norwegian tax shelter. Mimeo, University of Oslo, Columbia University, and Statistics Norway.

Andreoni, J., Erard, B. and Feinstein, J. (1998, June) Tax compliance. *Journal of Economic Literature* 36(2): 818–860.

Angrist, J.D. and Pischke, J.-S. (2010) The credibility revolution in empirical economics: how better research design is taking the con out of econometrics out of econometrics. *Journal of Economic Perspectives* 24(2): 3–30.

Ariel, B. (2012). Deterrence and moral persuasion effects on corporate tax compliance: findings from a randomized controlled trial. *Criminology* 50(1): 27–69.

Basu, K. (2013). The method of randomisation, economic policy, and reasoned intuition. Policy Research Working Paper 6722, World Bank.

Beck, P.J., Davis, J.S. and Jung, W.-O. (1991) Experimental evidence on taxpayer reporting under uncertainty. *Accounting Review* 66(3): 535–558.

Becker, G.S. (1968) Crime and punishment: an economic approach. *Journal of Political Economy* 76: 169–217.

Becker, W., Buchner, H.-J. and Sleeking, S. (1987) The impact of public transfer expenditures on tax evasion : an experimental approach. *Journal of Public Economics* 34(2): 243–252.

Benzarti, Y. (2015) How taxing is tax filing? Leaving money on the table because of hassle costs. Mimeo, University of California, Berkeley.

Bird, R.M., Martinez-Vazquez, J. and Torgler, B. (2008) Tax effort in developing countries and high income countries: the impact of corruption, voice and accountability. *Economic Analysis and Policy* 38(1): 55–71.

BIT. (2012) Applying behavioural insights to reduce fraud, error and debt. Technical report, Behavioural Insights Team, Cabinet Office, UK.

Blackwell, C. (2007) A meta-analysis of tax compliance experiments. Working paper 07-24, Andrew Young School of Policy Studies.

Blumenthal, M., Christian, C. and Slemrod, J. (2001) Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal* 54(1): 125–138.

Bosco, L. and Mittone, L. (1997) Tax evasion and moral constraints: some experimental evidence. *Kyklos* 50(3): 297–324.

Bott, K., Cappelen, A.W. and Sørensen, E.Ø. (2014) You've got mail: a randomised field experiment on tax evasion. Discussion Paper 26/2014, Department of Economics, Norwegian School of Economics.

Christian, R.C. and Alm, J. (2014) Empathy, sympathy, and tax compliance. *Journal of Economic Psychology* 40(C): 62–82.

Carrillo, P., Castro, E. and Scartascini, C. (2016) Do rewards work to maintain and increase tax compliance? Evidence from the randomization of public goods. World Bank ABCDE 2016 Conference, World Bank, Washington, DC.

Carrillo, P., Pomeranz, D. and Singhal, M. (forthcoming) Dodging the taxman: firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics*.

Castro, L. and Scartascini, C. (2013) Tax compliance and enforcement in the Pampas: evidence from a field experiment. *Journal of Economic Behavior and Organization* 116: 65–82.

Coolidge, J. (2012) Findings of tax compliance cost surveys in developing countries. *eJournal of Tax Research* 10(2): 250–287.

Coricelli, G., Joffily, M., Montmarquette, C. and Villeval, M. (2010) Cheating, emotions, and rationality: an experiment on tax evasion. *Experimental Economics* 13(2): 226–247.

Coricelli, G., Rusconi, E. and Villeval, M.C. (2014) Tax evasion and emotions: an empirical test of re-integrative shaming theory. *Journal of Economic Psychology* 40(C): 49–61.

Cummings, R.G., Martinez-Vazquez, J., McKee, M. and Torgler, B. (2009, June) Tax morale affects tax compliance: evidence from surveys and an artefactual field experiment. *Journal of Economic Behavior & Organization* 70(3): 447–457.

De Giorgi, G., Ploenzke, M. and Rahman, A. (2015) Small firms' formalization. The stick treatment. Policy Research Working Paper 7318, World Bank.

Deaton, A. (2009) Instruments of development: randomization in the tropics, and the search for the elusive keys to economic development. Working Paper 14690, NBER.

Del Carpio, L. (2014) Are the neighbors cheating? Evidence from a social norm experiment on property taxes in Peru. Mimeo, INSEAD.

Doerrenberg, P. and Schmitz, J. (2015) Tax compliance and information provision: a field experiment with small firms. IZA Discussion Papers 9013, Institute for the Study of Labor (IZA).

Drago, F., Mengel, F. and Traxler, C. (2015) Compliance behavior in networks: evidence from a field experiment. Discussion Paper 9443, IZA.

Dulleck, U., Fooken, J., Newton, C., Ristl, A., Schaffner, M. and Torgler, B. (2016) Tax compliance and psychic costs: behavioral experimental evidence using a physiological marker. *Journal of Public Economics* 134: 9–18.

Dunning, T., Monestier, F., Pi neiro, R., Rosenblatt, F. and Tu nón, G. (2016) Is paying taxes habit forming? Experimental evidence from Uruguay. Mimeo.

Dwenger, N., Kleven, H., Rasul, I. and 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.

Eissa, N. and Zeitlin, A. (2014) Using mobile technologies to increase VAT compliance in Rwanda. Mimeo, McCourt School of Public Policy, Georgetown University.

Erard, B. and Feinstein, J.S. (1994) Honesty and evasion in the tax compliance game. *RAND Journal of Economics* 25(1): 1–19.

Feldman Naomi E., Katuščák Peter, Kawano Laura (2016) Taxpayer Confusion: Evidence from the Child Tax Credit. *American Economic Review* 106(3): 807–835.

Fellner, G., Sausgruber, R. and Traxler, C. (2013) Testing enforcement strategies in the field: legal threat, moral appeal and social information. *Journal of the European Economic Association* 11(3): 634–660.
Fjeldstad, O.-H. and Semboja, J. (2001) Why people pay taxes: the case of the development levy in Tanzania. *World Development* 29: 2059–2074.
Fjeldstad, O.-H., Schulz-Herzenberg, C. and Sjursen, I.H. (2012) People's views of taxation in Africa: a review of research on determinants of tax compliance. Working Paper 8, International Centre for Taxation and Development.
Friedland, N., Maital, S. and Rutenberg, A. (1978) A simulation study of income tax evasion. *Journal of Public Economics* 10(1): 107–116.
Gadenne, L. (2017) Tax me, but spend wisely? Sources of public finance and government accountability. *American Economic Journal: Applied Economics* 9(1): 274–314.
Gangl, K., Torgler, B., Kirchler, E. and Hofmann, E. (2014) Effects of supervision on tax compliance: evidence from a field experiment in Austria. *Economics Letters* 123(3): 378–382.
Gangl, K., Kirchler, E., Lorenz, C. and Torgler, B. (2015) Wealthy tax non-filers in a developing nation: the roles of taxpayer knowledge, perceived corruption and service orientation in Pakistan. CREMA Working Paper Series 2015–08, Center for Research in Economics, Management and the Arts (CREMA).
Gerxhani, K. (2007) 'Did you pay your taxes?' How (not) to conduct tax evasion surveys in transition countries. *Social Indicators Research* 80(3): 555–581.
Gillitzer, C. and Skov, P. (2013) Evidence on unclaimed charitable contributions from the introduction of third-party information reporting in Denmark. EPRU Working Paper Series 2013–04, Economic Policy Research Unit (EPRU), University of Copenhagen. Department of Economics.
Hallsworth, M. (2014) The use of field experiments to increase tax compliance. *Oxford Review of Economic Policy* 30(4): 658–679.
Hallsworth, M., List, J.A., Metcalfe, R.D. and Vlaev, I. (2017) The behavioralist as tax collector: using natural field experiments to enhance tax compliance. *Journal of Public Economics* 148: 14–31.
Hashimzade, N., Myles, G.D. and Tran-Nam, B. (2013) Applications of behavioural economics to tax evasion. *Journal of Economic Surveys* 27(5): 941–977.
Khan, A.Q., Khwaja, A.I. and Olken, B.A. (2016) Tax farming redux: experimental evidence on performance pay for tax collectors. *Quarterly Journal of Economics* 131(1): 219–271.
Kirchler, E., Hoelzl, E. and Wahl, I., (2008) Enforced versus voluntary tax compliance: the 'slippery slope' framework. *Journal of Economic Psychology* 29: 210–225.
Kleven, H.J., Knudsen, M.B., Kreiner, C.T., Pedersen, S. and Saez, E. (2011) Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica* 79(3): 651–692.
Koessler, A.K., Torgler, B., Feld, L.P. and Frey, B.S. (2016) Commitment to pay taxes: a field experiment on the importance of promise. Tax and Transfer Policy Institute Working Paper 10/2016.
Lamberton, C., De Neve, J.-E. and Norton, M.I. (2014) Eliciting taxpayer preferences increases tax compliance. Working Paper 14-106, Harvard Business School.
Luttmer, E. and Singhal, M. (2014) Tax morale. *Journal of Economic Perspectives* 28(4): 149–168.
Manoli, D.S. and Turner, N. (2014) Nudges and learning: evidence from informational interventions for low-income taxpayers. Working Paper 20718, NBER.
Martin, L. (2014) Taxation, loss aversion, and accountability: theory and experimental evidence for taxation's effect on citizen behavior. Mimeo, Yale University.
Mascagni, G., Nell, C. and Monkam, N. (2017) One size does not fit all: a field experiment on drivers of compliance and delivery methods in Rwanda. ICTD Working Paper N. 58. Brighton, UK.
Mascagni, G., Nell, C., Monkam, N. and Mukama, D. (2016) The carrot and the stick: evidence on voluntary tax compliance from a pilot field experiment in Rwanda. ICTD Working Paper N. 57. Brighton, UK.
Mittone, L. (1997) Subjective versus objective probability: results from seven experiments on fiscal evasion. CEEL Working Papers 4, CEEL.
Mossin, J. (1968) Taxation and risk-taking: an expected utility approach. *Economica* 35(137): 74–82.
Myles, G.D. and Naylor, R.A. (1996, April) A model of tax evasion with group conformity and social customs. *European Journal of Political Economy* 12(1): 49–66.
Ortega, D. and Scartascini, C. (2016a) Don't blame the messenger: a field experiment on delivery methods for increasing tax compliance. Mimeo, Inter-American Development Bank.

Ortega, D. and Scartascini, C. (2016b) Who's calling? The effect of phone calls as a deterrence mechanism. Mimeo, Inter-American Development Bank.
Paler, L. (2013) Keeping the public purse: an experiment in windfalls, taxes, and the incentives to restrain government. *American Political Science Review* 104(7): 706–725.
Pomeranz, D. (2015) No taxation without information: deterrence and self-enforcement in the Value Added Tax. *American Economic Review* 105(8): 2539–2569.
Roth, A.E. (1995) Introduction to experimental economics. In J.H. Kagel and A.E. Roth (eds.), *The Handbook of Experimental Economics*. Princeton, USA, Princeton University Press.
Slemrod, J. (2007) Cheating ourselves: the economics of tax evasion. *Journal of Economic Perspectives* 21(1): 25–48.
Slemrod, J. (2016) Tax compliance and enforcement: new research and its policy implications. Ross School of Business Working Paper 1302, University of Michigan.
Slemrod, J. and Weber, C. (2012) Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance* 19(1): 25–53.
Slemrod, J., Blumenthal, M. and 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.
Slemrod, J., Collins, B., Hoopes, J., Reck, D. and Sebastiani, M. (2015) Does credit-card information reporting improve small-business tax compliance? NBER Working Paper 21412, NBER.
Spicer, M.W. and Hero, R.E. (1985) Tax evasion and heuristics : a research note. *Journal of Public Economics* 26(2): 263–267.
Spicer, M.W. and Thomas, J.E. (1982) Audit probabilities and the tax evasion decision: an experimental approach. *Journal of Economic Psychology* 2(3): 241–245.
Torgler, B. (2002) Speaking to theorists and searching for facts: tax morale and tax compliance in experiments. *Journal of Economic Surveys* 16(5): 657–683.
Torgler, B. (2003) Beyond punishment: a tax compliance experiment with taxpayers in Costa Rica. *Revista de Analisis Economico - Economic Analysis Review* 18(1): 27–56.
Torgler, B. (2004a) Cross-culture comparison of tax morale and tax compliance: evidence from Costa Rica and Switzerland. *International Journal of Comparative Sociology* 45(1–2): 17–43.
Torgler, B. (2004b) Moral suasion: an alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland. *Economics of Governance* 5(3): 235–253.
Torgler, B. (2013) A field experiment in moral suasion and tax compliance focusing on underdeclaration and overdeduction. *FinanzArchiv: Public Finance Analysis* 69(4): 393–411.
Torgler, B., Schaffner, M. and Macintyre, A. (2010) Tax compliance, tax morale and governance quality. In J. Alm, J. Martinez-Vazquez, and B. Torgler (eds.), *Developing Alternative Frameworks for Explaining Tax Compliance*. Routledge, UK.
Traxler, C. (2010) Social norms and conditional cooperative taxpayers. *European Journal of Political Economy* 26(1): 89–103.