

The use of field experiments to increase tax compliance

Michael Hallsworth*

Abstract Governments have become increasingly interested in the ‘explosion’ of research into taxpayer behaviour. This article briefly reviews two main theories of tax compliance (‘deterrence’ and ‘non-deterrence’), before discussing the recent rapid rise of natural field experiments (NFEs) in this area. These NFEs represent a ‘win-win’ for policy-makers and academics, since they can test theories while also evaluating the impact of a specific intervention in the real world. The evidence from NFEs shows that deterrence approaches improve compliance, while the case for non-deterrence approaches is less conclusive at the moment. NFEs could be improved by paying more attention to findings from behavioural science that apparently incidental factors, such as timing, framing, and complexity, substantially affect whether compliance occurs or not. New directions for future studies include: network effects, tax payments (in addition to declarations), the behaviour of firms, the provision of public goods, and the relationship between attitudes and behaviour.

Key words: tax compliance, tax evasion, natural field experiments, behavioural economics

JEL classification: H260, C93

I. Introduction

Improving tax compliance is a major policy goal for developed economies. An eroded tax base constrains a government’s choice of economic strategies, forcing it to consider higher and more distortionary taxes, increased borrowing, or reduced provision of public goods. The public perception that others are not paying their share may increase disrespect for the law and diminish trust between individuals. The costs of administering the tax system will rise, even as policy-makers find that their ability to measure—and hence steward—the economy has been significantly weakened. In short, effective tax collection is fundamental to a well-functioning state.

On this basis, developed economies may have reason to be concerned. The latest US Internal Revenue Service estimates show that in 2006 \$450 billion was not paid when it should have been, representing 16.9 per cent of total revenue owed ([Internal Revenue Service, 2012](#)). For some types of earnings, particularly those not subject to third-party reporting, over half of income is not reported to the government ([Slemrod,](#)

* Imperial College London, e-mail: m.hallsworth11@imperial.ac.uk

I would like to thank Ivo Vlaev for his comments, advice, and support. I would also like to thank Robert Metcalfe, Cameron Hepburn, Ara Darzi, the other authors in this issue, and two anonymous reviewers.
doi:10.1093/oxrep/gru034

© The Author 2014. Published by Oxford University Press.

For permissions please e-mail: journals.permissions@oup.com

2007). The European Commission estimates that 20–25 per cent of GDP in Italy and Greece is in the ‘shadow economy’ and thus not visible to the tax authorities (European Commission, 2013). The pressure on national finances since 2008 has provided greater impetus to address these issues, as have concerns that the globalization of economic activity is creating new opportunities for tax evasion (Alm, 2012).

In parallel, recent decades have seen an ‘explosion’ of theoretical and empirical research into taxpayer behaviour (Slemrod and Weber, 2012). It is perhaps unsurprising, therefore, that policy-makers have become increasingly interested in the findings from this research, and how they might be applied in practice (OECD, 2010; Shaw *et al.*, 2010; Cabinet Office, 2012). I contribute to this dialogue by giving an overview of how the field has developed, and then suggesting what questions should be explored next. One of my main arguments is that these questions should be addressed by policy-makers and academics collaborating on natural field experiments. This is an exciting practice that is expanding rapidly: eight such natural field experiments on tax compliance were published in the year to May 2014, which means that the field has roughly doubled in size (see Table 1).

Tax collection may be a crucial task, but the terms used to discuss it are often unclear. This article focuses on ‘tax compliance’ problems: ‘the unintentional failure of taxpayers to pay their taxes correctly’ (Webley *et al.*, 1991). Compliance includes three main obligations, not all of them applicable to all actors: (i) filing tax returns on time; (ii) making accurate reports on these returns; (iii) paying any tax owed on time (US Treasury, 2009). Three points are worth making about this definition. First, it means that my focus is broader than ‘tax evasion’, and includes the sizeable chunk of revenue (around 10 per cent of the US tax gap) that is declared but not paid on time, as well as those actors who fail to file on time despite owing no money (Slemrod 2007). Second, legal but potentially controversial ‘tax avoidance’ measures are excluded. Third, I am concerned with whether a payment or report is made, not the spirit or willingness with which it is made (cf. James and Alley, 2004).

The structure of the article is as follows. Section II sets out how theoretical approaches to explaining tax compliance have evolved, and how they remain the subject of intense academic debate. Section III shows how empirical analyses of non-compliance increasingly turned to lab and field experiments. Section IV gives an overview of natural field experiments to date. Section V identifies the questions that they should address next. Section VI concludes.

II. Theoretical approaches

Contemporary theoretical analyses of tax behaviour are generally accepted to have begun with the studies of Allingham and Sandmo (1972) and Srinivasan (1973). Over time, a central debate has emerged between two schools of thought, which I refer to as the ‘deterrence’ and ‘non-deterrence’ approaches (McGraw and Scholz, 1991; Smith, 1992; Alm and Martinez-Vazquez, 2003). This division reflects wider debates in the field of regulation and compliance (Ayles and Braithwaite, 1992).

The deterrence approach sees taxpayers as rational utility maximizers who are concerned solely with advancing their private economic interests. In Allingham and

Table 1: Summary of natural field experiments in tax compliance

Author(s) Year of publication Tax year Country	Subjects Setting Treatments	Dependent variable(s)	Main findings
Schwartz and Orleans (1967) 1962 USA	173 individual taxpayers. 7–11 questions asked of participants by experimenters in the field. Severity of legal sanctions and likelihood they would be applied; appeal to conscience and civic duty; 'placebo' unrelated questions.	Changes in levels of reported income, total deductions, and income tax after credits.	Only the appeal to conscience and civic duty was more effective than the placebo questions. Sanctions produced a larger increase in total deductions (i.e. lower compliance) than the placebo questions.
Coleman (1996) 1994 and 1995 USA <i>Note: includes same experiments as Slemrod et al. (2001).</i>	47,000 individual taxpayers. Mailed letters or revised tax form. Increased threat of audit; offer of enhanced service and help; public services; descriptive social norms; revised tax form.	Changes in reported income, deductions, and taxes paid. Differences in reported income between treatment groups.	Audit treatment significantly increased reported income of low- and middle-income taxpayers; effect on higher-income taxpayers was 'mixed'. No significant effect of enhanced services treatment. Social norms treatment was marginally effective at increasing declared income, and was significantly different from public services letter. Public services letter did not significantly differ from control. Revised form resulted in higher deductions.
Slemrod et al. (2001) 1995 USA <i>Note: Article based on study also contained in Coleman (1996).</i>	1,724 individual taxpayers. Mailed letters. Increased threat of audit.	Changes in reported income, deductions, and taxes paid.	Audit treatment significantly increased reported income of low- and middle-income taxpayers (with larger effects for those with greater opportunities to evade), but significantly reduced reported income of high-income taxpayers.
Blumenthal et al. (2001) 1995 USA	60,061 individual taxpayers. Mailed letters. Public services; descriptive norms.	Change in federal taxable income reported and Minnesota tax liability paid.	No statistically significant effects.
Wenzel and Taylor (2004) 2000 Australia	9,000 rental property owners. Mailed letter or letter and booklet. Offer of help; threat of penalties and audits; requirement to return schedule; information booklet.	Annual rental property deductions and claimable tax deductions.	For first-time participants, the harder tone led to significantly lower overall deduction claims than a softer letter. Inclusion of information booklet had no effect on deductions. Requirement to return schedule led to marginally lower rental deductions.

Table 1: *Continued*

Author(s) Year of publication Tax year Country	Subjects Setting Treatments	Dependent variable(s)	Main findings
Torgler (2004) 2001 Switzerland	580 individual taxpayers. Mailed letter. Public services and civic duty.	On-time filing of tax return. On-time payment of tax liability.	No significant effects.
Wenzel (2005a) 2000 Australia	1,500 individual taxpayers. Mailed survey or mailed survey followed by feedback letter. Injunctive social norms.	Deductions claimed for work-related expenses; deductions claimed for other expenses.	No significant effects for work-related expenses. For other expenses, combined treatment effects are marginally significant. Deduction claims were significantly lower for the survey and letter group than for no-contact group and survey-only group (i.e. survey and letter led to increased compliance).
Wenzel (2006) 2001 Australia	2,052 individuals who had not filed Annual Statement on time. Mailed letter. Interpersonal fairness; informational fairness.	Filing of Annual Statement.	Interpersonal fairness letter increased compliance, significant at 0.05 level; informational letter increased compliance, with marginal significance; combined effect of both letters was significant.
Hasseldine et al. (2007) 2001 UK	7,307 sole proprietors with a turnover between £14,000 and £15,000 for 2 consecutive years. Mailed letter. Offer of help; descriptive norms and public services; increased risk of audit; increased risk of audit and details of penalties; notification of pre-selection for audit.	Absolute level of reported turnover; change in net profit over 2-year period.	All treatments significantly increased the proportion of sole proprietors filing above £15,000. Sanction treatments led to significant increases in reported turnover and profit. Sanction letters significantly more effective than norm and public services letters. Results vary according to whether an accountant is used.
Iyer et al. (2010) 2003 USA (Washington State)	1,000 construction industry firms. State taxes. Mailed letters. Penalties; increased risk of detection.	Reported income levels for two specific taxes ('Business and Occupation' and 'Use' taxes).	Main effects not significant at 0.05 level for either form of tax. However, pooled effect of the two treatments is significant at 0.05 level for Use Tax.
Servicio de Administración Tributaria (2010) 2006 Mexico Reported in OECD (2010)	31,754 individual taxpayers, stratified by risk category. Mailed letter. Audit probability.	Level of income declared. Filing of tax returns.	Declarations and return filings were significantly higher in the treatment group. However, outcomes were not significantly different for groups designated as high risk.

Table 1: *Continued*

Author(s) Year of publication Tax year Country	Subjects Setting Treatments	Dependent variable(s)	Main findings
Kleven <i>et al.</i> (2011) 2007 Denmark	42,784 taxpayers (17,764 self-employed). Randomly allocated audits; mailed letters giving future audit probability (100 per cent, 50 per cent, no information.)	Reported income tax levels.	Audits increased reported income in the subsequent year by 1 per cent of income (effect driven entirely by self-reported income). Threat of audit increases reported income, with 100 per cent probability having roughly twice the effect of 50 per cent probability.
Ariel (2012) 2006 Israel	4,395 corporations. Mailed letters. Deterrence; moral persuasion focusing on societal costs of non-compliance.	Gross sales values reported; VAT payments; VAT deductions.	Deterrence had no statistically significant effects on any of the dependent variables. Moral persuasion had no effect on sales values or payments, but significantly increased VAT deductions (i.e. backfired).
Torgler (2012) 2001 Switzerland	578 individual taxpayers. Mailed letters. Public services and civic duty.	Reported income levels for income tax. Levels of tax deductions.	Treatment had a small positive effect on reported income levels, but not statistically significant. Treatment appeared to increase deductions (i.e. backfire), but again not statistically significant.
Pomeranz (2013) 2008 Chile	408,600 Chilean firms, of whom 102,000 received a letter. Mailed letters. Implied increased audit probability.	VAT payments. 'Spillover' effects on suppliers of companies threatened with audit.	Threat of audit letters increased VAT payments. Suppliers of treated firms increased payments, but clients of these firms did not. Both results show that the VAT 'paper trail' increases compliance.
Fellner <i>et al.</i> (2013) 2005 Austria	50,498 individuals identified as potential television licence fee evaders. Mailed letters. Standard letter; enforcement threat; moral appeal; descriptive social norm; norm and threat; moral appeal and threat.	Proportion of recipients registering for licence or updating contract details.	Reminder letter increases registrations significantly compared to control (no letter condition). Threat treatment significantly increases registrations, but morality and norm treatments did not. In high evasion areas, evidence for positive effect of norm information and negative effect of morality message.

Table 1: Continued

Author(s) Year of publication Tax year Country	Subjects Setting Treatments	Dependent variable(s)	Main findings
Harju <i>et al.</i> (2013) 2012 Finland	4,800 small businesses, of whom 1,800 were hairdressers experiencing an exogenous rise in VAT rates. Mailed letters. High (33 per cent) and low (5 per cent) probability of audit; no letter for control group.	Turnover reported to tax authority in relation to VAT.	High-probability letter significantly increased the turnover declared to the tax authority, but only for firms experiencing an increase in VAT rate. Low-probability letter did not significantly affect evasion.
Castro and Scartascini (2013) 2011 Argentina	23,195 individual property owners in a single province. Mailed letters. Enforcement; descriptive social norms; public services.	Rate of payment of property tax.	Enforcement significantly increases payments, but descriptive social norms and public goods do not. Heterogeneous effects by geographic area, property size, past taxpayer behaviour.
Ortega and Sanguinetti (2013) 2011 Venezuela	6,000 firms in an urban municipality who were liable for the local business tax. Mailed letters. Enforcement; moral duty; public services related to firm operation; public services enabling social inclusion; basic information about address of tax office; no letter for control group.	Difference-in-difference analysis of tax balance of firm before and after receipt of letters.	Letters concerning enforcement and public services related to firm operation significantly reduced the tax balance. Effects of the moral duty and basic information letters were significant only at the 0.1 level. No significant effects from the letter referring to social inclusion public services. No significant differences between coefficients for all letter treatments, and therefore the receipt of a letter <i>per se</i> may account for the observed effects.
Del Carpio (2013) 2012 Peru	22,318 individual property owners in two municipalities. Mailed letters. Descriptive social norms; probability of enforcement; norms and enforcement; standard reminder.	Rate of payment of property tax.	All letters increased payment rates relative to no letter (by around 5 percentage points). Norms treatment appeared to produce higher payment rates than enforcement, enforcement and norm, and standard reminder—but this difference was not significant at the 0.05 level.

Table 1: Continued

Author(s) Year of publication Tax year Country	Subjects Setting Treatments	Dependent variable(s)	Main findings
Hallsworth <i>et al.</i> (2014) 2011–12 UK	220,993 income tax filers. Mailed letters. Descriptive social norms; injunctive social norms; public services; moral duty; interest charges; payment information; standard reminder.	Occurrence and timing of payments to tax authority.	Reminder letters <i>per se</i> increase tax payments. All message treatments significantly increased occurrence and speed of payments compared to standard reminder. Descriptive norms more effective than injunctive norms.
Dwenger <i>et al.</i> (2014) 2012 Germany	39,782 individuals liable for the church tax in a metropolitan region of Bavaria. Mailed letters. Letter simplification; statement of zero probability of audit; (varied) probability of audit; social and monetary rewards; social norms; moral appeal.	Rate of payment of church tax, compared to amount due.	20.9 per cent of individuals comply despite zero probability of audit. Letter simplification significantly increases compliance. Probability of audit significantly increases compliance overall, with effects driven by baseline non-compliers. No significant differences between audit probabilities. Social and monetary rewards reduce compliance among baseline non- compliers, but increase it among compliers. Zero probability of audit, social norms, and moral appeal do not significantly affect compliance.
Gangl <i>et al.</i> (2014) 2012 Austria	1,721 newly established firms in 'high-risk' sectors. Supervision by tax authority: face-to-face visit giving advice and informing firm that constant auditing would take place. Control group not contacted.	Proportion of firms paying tax liability on time; amount owed by late-paying firms.	Supervision reduces compliance rate. Some evidence that, for those who are late, supervision reduces the amount they owe (i.e. increases compliance).

Sandmo's (1972) model, which draws on Becker (1968), a taxpayer's decision whether to evade tax is based on their expected utility after considering the probability of audit, the size of fine, tax rates, and income. In other words, compliance is seen as a gamble with the risk of punishment. Accordingly, the way to curb non-compliance is through vigilant monitoring, the threat of sanctions and penalties, and restrictions on the opportunities to cheat. Effective enforcement procedures are the routes to success for the authorities; without these procedures (and given the opportunity), taxpayers will not comply.

Deterrence has been the dominant approach in economic studies and revenue departments (Feld and Larsen, 2012). However, a common objection is that this model significantly over-predicts non-compliance, given the very low probability of audit. This has led to the remark that, if the model is valid, the greater puzzle is why people *pay* taxes, not why they evade them (Alm, 2012). Proponents of the deterrence approach generally make three main responses. One is that the decision is based on the *perceived* risk of detection, not the true audit rate, and there is much evidence that we are prone to overweight the small probability of this happening—one study found that participants estimated the rate as 2,000 per cent higher than it was (Alm *et al.*, 1992; Bobek *et al.*, 2013). The second is that the audit rate does not represent the true probability of detection, since tax authorities can and do use third-party information to check the veracity of taxpayer reports. Empirical evidence supports this claim: evasion is lower when third-party reporting is present (Slemrod, 2007; Kleven *et al.*, 2011). Finally, taxpayers may be aware that enforcement is not random, and that their own actions or characteristics may increase the likelihood they will suffer action; these endogeneities can provide sufficient incentives to comply even when the enforcement parameter levels are low (Phillips, 2014).

The non-deterrence approach claims that the taxpaying decision does not rest solely on the financial decisions, but rather is influenced by factors such as social norms, perceptions of fairness, tax morale, and the provision of public goods (Erard and Feinstein, 1994; Kirchler, 2007; Torgler, 2007). Essentially, two variants of this argument exist. The first holds that these factors can be integrated into the utility decision of the individual, thus representing a certain ‘cost’ to them (Gordon, 1989). Indeed, Allingham and Sandmo themselves suggested that this would be an appropriate development. The second variant is a more radical departure, since it rejects the premise that taxpayers seek to maximize their utility in a calculating manner. Rather, it starts from the position that taxpayers have ‘a primary disposition to comply with tax laws’ (Ariel, 2012). This viewpoint posits a fundamentally cooperative relationship between tax authority and taxpayer, with the latter asking ‘what should I do?’ rather than ‘what can I get away with?’ (McGraw and Scholz, 1991). The deterrence model is critiqued as overly individualistic, and neglectful of the fact that taxpayers are grounded in a set of interactions and identities stretching beyond themselves (Taylor, 2003; Alm *et al.*, 2012).

In terms of administrative policy, the non-deterrence approach argues that taxpayers should be treated fairly and with respect, given clear and helpful information, and provided with a competent service to make compliance easy. In order to reduce evasion, administrations should persuade taxpayers by emphasizing that tax compliance is ethical, practised by the great majority of people, and creates valued public goods (Kirchler, 2007). Traditional deterrence procedures, on the other hand, will create an adversarial relationship in which taxpayers react against the attempt to control their behaviour, which then undermines (or ‘crowds out’) voluntary compliance (Brehm and Brehm, 1981; Hessing *et al.*, 1992; Sheffrin and Triest, 1992).

It is, however, rare to encounter the position that deterrence approaches should be dispensed with altogether.¹ The more common stance in the non-deterrence camp is that ‘a duality of deterrent fears and civic obligations as motivation to comply is required’,

¹ It is important to note that the deterrence approach does suggest that using approaches like social norms and fairness may reduce compliance because they signal to taxpayers that the authority is ‘toothless’ (Ariel, 2012).

with the two elements reinforcing one another (Smith, 1992; Kirchler, 2007). The main question to be answered, which represents a major source of current theoretical debate, is how this duality should be realized. Kirchler *et al.* (2008) have proposed the ‘slippery slope’ framework as a means of understanding how the two factors interact: tax behaviour is influenced by ‘trust in authorities’ and ‘power of authorities’, both of which exist on a sliding scale. The theory of tax compliance remains in flux, however, with an integrated and uncontested model of behaviour some way off yet.

III. Evolving methodologies

Another narrative concerns the means by which academics have advanced the theories above through empirical analyses. One thing on which commentators agree is that such analyses are plagued by difficulties. Recent reviews have talked of the ‘extraordinary challenges’ and ‘severe measurement problems’ involved, which mean that ‘we are still trying to answer many *basic questions* on measuring, explaining, and controlling evasion’ (Slemrod and Weber, 2012; Alm, 2012). The main issue is obtaining reliable measures of non-compliant behaviour, not least since this is something that people wish to conceal.

The first empirical analyses applied econometric techniques to existing data sources. Perhaps the most direct means of doing so is to audit a sample of tax returns, then extrapolate the results. This approach has its drawbacks, however. Audits are expensive to conduct; officials may not actually discover undeclared income; auditors may not be able to distinguish between fraud and unintentional errors; judgements are not always consistent from one auditor to another; and audits do not address the ‘ghosts’ who simply fail to file a tax return (Elffers *et al.*, 1992; Kleven *et al.*, 2011). Tax amnesties have also been used as data sources, but they suffer from selection bias because participation is voluntary. For the academic researcher, there is also the issue that data may be confidential and thus likely to be guarded carefully by the tax authority.

These challenges have led researchers to innovate. They have found new ways of observing tax behaviour, such as whether eBay retailers collect sales tax (Alm and Melnik, 2010). They have developed new ways of identifying ‘traces’ of the shadow economy in the visible economy, such as the proportion of bank notes issued that are of large denomination, or suspicious patterns in the distribution of declared income (Gutmann, 1977; Nigrini, 1996). They have even used apparently unrelated datasets, such as the degree of luminescence observed from space, in order to assess true levels of economic activity (Henderson *et al.*, 2009). Despite their advantages, the explanatory power of these studies is often limited.

Surveys are another approach that has attracted much attention. They may be used to estimate levels of evasion behaviour through self-reports, or explore beliefs and attitudes towards the subject, such as perceived probability of detection, acceptability of evasion, and views about the prevalence of non-compliance (Vogel, 1974). Again, surveys have disadvantages. People may simply not be able to remember their past behaviour, or they may interpret ‘compliance’ in varying ways (Hessing *et al.*, 1989; Webley *et al.*, 1991). Behavioural intentions, particularly those derived from hypothetical situations, may not translate into actual tax behaviour (Elffers *et al.*, 1987). Most obviously,

people may not give honest responses: they may wish to appear consistent with previous answers; to appear to comply with social norms; or simply to avoid incriminating themselves (Wenzel, 2005b).

The limitations of the empirical approaches above have led researchers to see laboratory experiments as particularly useful methods for analysing tax compliance. These experiments usually proceed along similar lines. At the beginning of a round, subjects are either given an income or have to earn it through a task. They then have to decide how much income to declare, and thus how much tax to pay, with the knowledge that undeclared income has a certain probability of being discovered and attracting a fine. Several rounds take place. At the end of the experiment, the participant is paid the amount they succeeded in retaining (Alm and Torgler, 2011). From the experimenter's point of view, various changes can be introduced (for example, to the size of the fine) and the ensuing effects on behaviour measured.

Experiments of this kind present several advantages. Most obviously, they address the measurement issue: evasion becomes visible. They also allow for the particular variable in question to be isolated and manipulated, and its causal effect measured, all without interference. This is particularly valuable for tax evasion studies, since there are major challenges to exogenously varying the penalty rate or the public goods that are provided through taxation, for example. Experiments are also inexpensive: the decision environment can be configured and reconfigured easily (Webley *et al.*, 1991). As a result, 'virtually all aspects of compliance have been examined in some way in experimental work' (Alm, 2012).

A common criticism of tax compliance lab experiments is that their findings do not apply to real-world tax behaviour (Elffers *et al.*, 1992). This lack of 'external validity' is a general criticism of lab experiments, reflected in other papers in this issue, so I focus on the aspects specific to tax compliance. The first is that the tasks in such experiments, and the environment in which they are performed, are quite different from real world tax compliance. Many of these experiments reduce the decision to a gamble in a game where the explicit goal is to maximize income. Perhaps understandably, this makes risk-taking a more entertaining option than compliance (Webley *et al.* 1991). Moreover, tax compliance presents a case where the stakes in the game are notably smaller than those in real life (Harrison and List, 2004). These task-related factors may create a bias towards non-compliance (Kirchler *et al.* 2010).

The experimental environment may also lack verisimilitude. Tax non-compliance is an activity that is likely to take place when actors are not being observed, or at least observation is not salient. In contrast, experiments are social situations involving heightened oversight (Levitt and List, 2007). Many such experiments have deliberately omitted any references to tax compliance, in order to focus on the core decision features being manipulated (Spicer and Thomas 1982). This strategy of abstraction makes it less likely that participants will apply the heuristics that would normally be triggered in the field, while 'there is *no control for the context that subjects might themselves impose on the abstract experimental task*' (Harrison and List, 2004).

The second main criticism is that the participants in these experiments—usually students—are not representative of the taxpaying population. As has been rehearsed elsewhere, students are likely to be younger, better-educated, and less experienced at paying tax than the population at large (Levitt and List, 2007). This lack of representativeness is concerning, given the evidence that (for example) older people are more likely to be

compliant with tax laws (Kirchler, 2007; Hallsworth *et al.* 2014). However, the few studies that specifically examine whether student and non-student responses differ in tax compliance experiments give mixed results. Some indicate that there are no significant differences, others that there are substantial differences; a key factor may be how representative the ‘non-student’ sample is of the taxpaying population (Alm *et al.*, 2013; Choo *et al.*, 2014).

We can see, therefore, that tax compliance researchers have been increasingly drawn to experimental approaches because they address some of the limitations of empirical analyses: they greatly improve measurement power, while also creating a counterfactual that allows the impact of potential policies to be measured. However, the external validity of these experiments has been questioned. As a result, there has been increasing interest in taking the experimental approach into the field.

IV. Natural field experiments

(i) Context

Harrison and List (2004) present the following typology for field experiments:

- an artefactual field experiment is the same as a conventional lab experiment but with a nonstandard [i.e. non-student] subject pool;
- a framed field experiment is the same as an artefactual field experiment but with field context in either the commodity, task, or information set that the subjects can use;
- a natural field experiment is the same as a framed field experiment but where the environment is one where the subjects naturally undertake these tasks and where the subjects do not know that they are in an experiment.

This section focuses mainly on natural field experiments (NFEs), since they hold the greatest promise for advancing tax compliance research—and, of the three approaches, they have generated the most activity recently. Of Harrison and List’s three variants, NFEs alone make the potentially decisive shift to the *environment* in which real decisions take place, thus allowing the normal cues and heuristics to operate. They are also able to measure the relative effects of different real-world policies *as implemented in practice*, which greatly increases their policy relevance (Slemrod *et al.*, 2001). ‘The result’, as one such study puts it, ‘is external validity at the highest level’ (Wenzel and Taylor, 2004). However, despite the idea of a full-scale NFE on tax compliance being suggested some 50 years ago (Schwartz, 1960), a review of the academic literature suggests that only two such experiments took place in the twentieth century (see Table 1).

There are many reasons for the dearth of studies. From an academic’s perspective, collaborating on a NFE has higher transaction costs than a lab experiment: time must be invested to identify the right partners, build relationships, understand the policy context, and discover the best opportunity to run the study (Feld *et al.*, 2006). Officials may also see few incentives to collaborate in academic studies.² With some Scandinavian

² My assertions about the motivations of officials are derived from my experience working as a tax official in the UK government.

exceptions, tax data in developed countries are protected by privacy laws, so careful work would be needed from the official side to clean and anonymize data. Officials may also be concerned about making their compliance options and strategies public, since this information could be used by uncooperative taxpayers to plan their non-compliance so that sanctions are avoided. A better option may seem to be to conduct a NFE but not release the results publicly. While these ‘private’ studies may have taken place, they may not be widespread because of governments’ reluctance (not specific to tax compliance) to evaluate policies through rigorous experimentation (Cook, 2003; Hallsworth *et al.*, 2011).

(ii) Overview of studies

Until very recently, it was common to assert that there was a serious lack of empirical evidence about how theory-based prescriptions for increasing compliance translate into the real world (e.g. Del Carpio, 2013). Recent developments mean that this view needs to be revised. The past few years have seen a sudden flourishing of NFEs in tax compliance: the number of available studies doubled between 2012 and 2014. Table 1 gives a summary of the NFEs in tax compliance published up until May 2014. To be included, a study needed to: be written in English; administer the treatment in the field, not the lab, even if measurement took place in the field (McGraw and Scholz, 1991); measure actual behaviour, rather than intentions or self-reports (Torgler, 2003; Hasseldine and Hite, 2003); deal with *compliance*, rather than other types of tax behaviour (Bhargava and Manoli, 2011); explicitly state that it employed randomization (Doyle *et al.*, 2009). For brevity, only main findings are reported.

(iii) Overview of findings

The following section gives a very brief overview of how findings from these NFEs have advanced our understanding of tax compliance. The focus on NFEs means that the section does not attempt to summarize the entire field of tax compliance research, which is large and growing rapidly (Andreoni *et al.*, 1998; Alm *et al.*, 2010). Instead, the section highlights the contribution of NFEs in three areas: deterrence approaches; non-deterrence approaches; and how intervention effects vary across populations.

In terms of deterrence, most interventions which focused on increasing the perceived probability of enforcement action, or the perceived severity of this action, did increase subsequent compliance. There is also limited evidence that taxpayers may be sensitive to different levels of audit probability. However, the size of these deterrence effects can be relatively small: Kleven *et al.* (2011) show that being audited increased reported income in the following year, but only by 1 per cent. Indeed, Kleven *et al.*’s study also clearly demonstrates the advantages of NFEs in tax compliance: the random allocation of audit eliminated any selection bias, thus establishing a reliable effect size; it also allowed a comparison between different policy options, revealing that expanding third-party reporting may be a more cost-effective option than further audits.

Another component of the Allingham–Sandmo model is the tax rate. In the original model, the tax rate had an ambiguous effect on non-compliance; in line with Yitzhaki (1974), if the size of the penalty is proportionate to the tax evaded (as it is in most

countries), then a higher tax rate will produce higher compliance. Empirical studies have suggested that the opposite is true, but an exogenous variation in the tax rate is needed to eliminate confounding factors and give an accurate estimated effect (Fisman and Wei, 2004; Marion and Muehlegger, 2008). Harju *et al.* (2013)'s experiment provides such variation, and finds that a higher VAT rate led to increased evasion among firms. Finally, Dwenger *et al.* (2014) show that a tax compliance rate of 20.9 per cent exists in a zero enforcement situation (the church tax in Germany), and thus conclude that the Allingham–Sandmo model's predictions were correct for nearly 80 per cent of their sample.

Perhaps surprisingly, the number of NFEs that feature non-deterrence factors (15) is similar to those including deterrence factors (17). Of these 15, seven showed a significant increase in compliance (Schwartz and Orleans, 1967; Wenzel, 2005a, 2006; Hasseldine *et al.*, 2007; Del Carpio, 2013; Hallsworth *et al.*, 2014; Dwenger *et al.*, 2014); two showed partial effectiveness (Coleman, 1996; Ortega and Sanguinetti, 2013); six showed no significant effects (Blumenthal *et al.*, 2001; Wenzel and Taylor, 2004; Torgler, 2004, 2012; Fellner *et al.*, 2013; Castro and Scartascini, 2013); and one showed a significant increase in non-compliance (Ariel, 2012). There are also a few studies that show a backfire from deterrence interventions, in line with the non-deterrence approach (Schwartz and Orleans, 1967; Slemrod *et al.*, 2001; Gangl *et al.*, 2014).

These results suggest that, when NFEs alone are considered, the evidence that the non-deterrence approach can be used to increase compliance is mixed. However, there are important qualifications to be made. The non-deterrence approach involves a greater variety of concepts (norms, fairness, public goods, etc.), compared to the more uniform deterrence options (audits, fines, etc.), which means that a greater variety of results is perhaps unsurprising. Since many of these concepts are more abstract, they may be more dependent on the exact way they are communicated, a factor that has generally been underappreciated in NFEs to date (see below). Moreover, they may be more dependent on their context (whether this means the type of tax concerned, the type of actor, or the cultural and geographic situation) than deterrence factors. Much work has analysed how tax morale varies by country, for example (Torgler, 2007). Finally, from a policy-maker's perspective, it is worth noting that non-deterrence approaches are generally less expensive than deterrence options; they may even have zero marginal cost (Hallsworth *et al.*, 2014).

NFEs are also helping to identify how the impact of deterrence and non-deterrence approaches varies according to the recipient's past levels of compliance. There is emerging evidence that those with a history of non-compliance are more likely to respond to deterrence approaches, while non-deterrence interventions are more effective for those who have complied in the past (Castro and Scartascini, 2013; Del Carpio, 2013; Dwenger *et al.*, 2014). As Fellner *et al.* (2013) point out, this means that tax compliance studies which only focus on a previously non-compliant population risk underestimating the effect of non-deterrence approaches, since they may produce a larger effect in the general taxpaying population.

Social norms are something of a special case here, since their effect depends greatly on recipients' prior beliefs. Fellner *et al.* (2013) find (weak) evidence that social norm information reduces evasion in high-evasion areas, and weak evidence that it increases evasion in low-evasion areas. Similarly, Castro and Scartascini (2013) find that a social norms message backfired among those who had complied in the past. In both cases, it

is likely that the crucial factor is whether the social norm information updates beliefs about non-compliance in a positive or negative direction. This effect has been found in other policy areas as well (Ringold, 2002; Allcott, 2011). The policy implication here is that governments need to assess current beliefs before deciding whether to use a social norms approach. It is noticeable that the positive effect of social norms found by Del Carpio (2013) occurred in an environment where taxpayers underestimated the true level of compliance.

At this point, it is worth returning to the question of how to ensure ‘a duality of deterrent fears and civic obligations as motivation to comply’. The wider policy implication from these NFEs may be to use non-deterrence approaches to maintain tax morale in the general population (the ‘intrinsically motivated’), while deploying deterrence approaches to address likely evasion in the previously non-compliant (the ‘extrinsically motivated’). Sophisticated data analytics could be used to assess the risk of future non-compliance (or the likelihood it has occurred) and tailor interventions accordingly. Without this kind of segmentation, deterrence approaches may signal to the compliant that others are evading, while non-deterrence options may imply weakness to the non-compliant. The practical challenge is that ‘compliant’ and ‘non-compliant’ are not stable and clearly delineated categories. Taxpayers are likely to adopt various ‘motivational postures’ from situation to situation, and evasion may be a spontaneous, context-driven act (Braithwaite, 2003). Moreover, it is very difficult to ensure that messages aimed at one putative group do not reach the other. Nevertheless, it could be that segmentation of this kind forms the basis of a compliance strategy that integrates deterrence and non-deterrence approaches.

(iv) Assessing existing studies

Natural field experiments offer many advantages. However, it is important to reflect on the shortcomings of existing studies and how they could be addressed. First, the most obvious restriction is that NFEs only examine factors that can be operationalized through discrete interventions. This means that certain factors are usually excluded: for example, fiscal decentralization appears to affect tax compliance, but it is difficult to vary exogenously and thus does not feature in the table above (Güth *et al.*, 2005). It also means that NFEs can only illuminate the behaviour that occurs in the period after a policy intervention (as opposed to a period of stasis). However, there is an argument that existing academic studies could do more to analyse the long-term effects of their interventions, and to show whether their initial results can be replicated. Many of them have a short-term focus that is only concerned with the immediate impact of the change being studied (‘did it work?’). In contrast, administrators are incentivized to consider longer time horizons, seeing themselves as the stewards of a system in which repeated games take place. They are likely to be concerned with questions such as: ‘Will it work next time?’, ‘Will it lead to a backfire in the next tax cycle?’, ‘What will happen if we stop taking this action, now we have started?’ On these questions, most studies are silent.

Second, studies should feature more explicit and sophisticated analyses of the costs and benefits of acting. In most developed economies, governments commit (at least ostensibly) to conducting *ex ante* cost–benefit analyses to ensure they are deploying

resources efficiently (HM Treasury, 2003). However, these kind of analyses are surprisingly uncommon in published tax compliance NFEs: relatively few actually monetize the benefits they describe (with accelerated revenue producing smaller benefits than ‘new money’, for example), or properly account for the costs of acting. These considerations are likely to be highly salient to administrators, and many studies do not meet this demand sufficiently. I recognize that these kind of analyses may have been conducted, but not published. In that case, we are dealing with a different problem: the creation of effective channels for the communication of tax compliance findings.

Finally, NFEs could profit by drawing more on the findings from behavioural science. There has been a tendency to conflate ‘psychology’ with non-deterrence approaches, and to adopt an implicit model of behaviour that involves changing attitudes to tax compliance in order to change behaviour. For example, Trivedi *et al.* (2003) conclude that ‘to increase compliance, policy-makers need to take positive steps to increase the moral reasoning and value orientation of low moral reasoning and low value-oriented subjects’. A behavioural science perspective would, in contrast, focus on the *mechanisms* by which an intervention (deterrence or non-deterrence) results in the target behaviour, with particular regard to the power of context and environmental cues (Dolan *et al.*, 2012). Thus, ‘psychology’ may more usefully be seen as a means of engendering behaviour, rather than as a worldview that proposes certain motivations for tax behaviour.

This use of behavioural science suggests a greater focus on *how* tax authorities intervene and deal with taxpayers more generally. As Hallsworth *et al.* (2014) show, significantly different results can be obtained by the way in which a policy (e.g. sending reminder letters) is implemented, even though the application of the policy itself is not randomized (everyone receives a letter).³ Compliance may therefore be affected by a range of apparently incidental factors such as timing, framing, complexity, tone, visual presentation, and so on. Clearly, this perspective opens up a wide range of opportunities to understand how the specific interactions between tax authority and taxpayer influence compliance. It also opens up a source of concern about existing findings. Academics are drawing generalizations about the effectiveness of certain concepts from specific presentations of those concepts in NFEs. If, as a behavioural science perspective suggests, the details of this presentation may have a significant effect, then three points need to be borne in mind.

First, the specific treatment may not adequately represent the general concept that is being tested (i.e. its ‘construct validity’ may be low). For example, it should be noted that the social norm message created by Blumenthal *et al.* (2001) begins by stating that ‘many Minnesotans believe other people routinely cheat on their taxes’, which introduces a negative perceived norm that risks undermining the positive descriptive norm that follows. Second, treatments often introduce other variations apart from the specific element they claim to test. For example, in Del Carpio (2013) the information about compliance levels is communicated through a bar graph, while a pie chart is used for information about enforcement levels. It may be that one of these options is significantly easier to understand than the other, which may produce significantly different behavioural effects. The final concern is that, as Hasseldine (2000) notes, the effectiveness of a treatment may not be sustained if a policy-maker changes an apparently insignificant element of its presentation. These points need to be acknowledged

³ This has been defined as an ‘indirect use’ of randomized experiments (Khawaja and Mian, 2011).

and addressed more frequently by those running tax compliance NFEs. One solution, as Jackson (1992) suggests, is to use multiple wordings that attempt to represent a single factor that is being tested.

V. Future directions

Some 15 years ago, Andreoni *et al.* (1998) noted that empirical research on tax compliance had ‘blossomed’ over the preceding decade. However, they also argued that such research was still immature, and ‘many of the most important behavioral hypotheses and policy questions [are] yet to be adequately investigated’. I argue that this judgement of the field in general now applies to NFEs specifically: although their use is growing rapidly, there are many questions still to be addressed in the field. Below I briefly indicate five issues that academics and officials should be using NFEs to examine in the future (excluding the methodological concerns outlined above). Note that these are meant to represent the priorities for tax-compliance NFEs specifically, rather than for the study of tax compliance as a whole.

Individuals versus firms

Most of the theoretical and attention has been directed towards individuals rather than organizations (Alm and McClellan, 2012). Although the number of NFEs in this area has increased recently, there is still a significant imbalance if one considers the significance of corporate tax compliance in most developed economies. In the UK, for example, ‘self assessment’ (which involves individuals filing a tax return) constituted only 4.7 per cent of receipts in 2012–13, with the great majority of the remainder paid through companies (HMRC, 2014). The issue is particularly pressing because there is some evidence that firms have lower tax ‘morale’ than individuals (Torgler, 2007). More work is required to establish whether and how the findings discussed above apply differently to firms versus individuals.

The relationship between attitudes and behaviour

A few previous studies have examined the relationship between attitudes measured through surveys or in the lab and actual tax behaviour (Elffers *et al.*, 1987, Elffers *et al.*, 1992). They have generally concluded that this relationship is weak or even non-existent. However, changes in attitudes are rarely measured in NFEs, so there is little evidence of whether they contribute to observed changes in behaviour. I believe that Del Carpio (2013) is the only NFE to have approached this issue, which produced the interesting result that enforcement messages (as well as social norm messages) increased perceived levels of compliance. Clearly, more work should be done in this area.

Tax payments

Until recently, the vast majority of NFEs had focused on the honest declaration of taxes. Declaration is only one component of compliance, however, since taxes still have to be paid; moreover, for some types of tax (such as property taxes) payment is the only action required. Focusing on payment eliminates many of the measurement problems discussed above. Over the past year, a set of studies have emerged that focus on payment

and thus have been able to access richer and more accurate datasets (Del Carpio, 2013; Castro and Scartascini, 2013; Hallsworth *et al.*, 2014; Dwenger *et al.*, 2014; Gangl *et al.*, 2014).

Public goods provision

The impact of varying public goods provision has been studied extensively in lab experiments. Most NFEs have resorted to measuring the impact of reminding taxpayers of the public goods they receive, given the obvious difficulties with translating this strategy to the field. These difficulties are not insurmountable, however: Gonzalez-Navarro and Quintana-Domeque (2013) show how a step-wedge design can be used to randomize street maintenance so its effect on property tax compliance can be measured. Future studies should consider how this approach could be applied to other public goods.

Network effects

Recent years have seen renewed interest in the extent to which behaviours spread through social networks (Fowler and Christakis, 2010). This is a valid question to ask of tax compliance interventions, not least because developed economies are investing in analytic tools that employ network analyses for this purpose (Alm, 2012; HMRC, 2013). Drago *et al.* (2014) provide the first NFE evidence that the effects of compliance interventions can be transmitted through geographic proximity; additional work here could open up new policy opportunities.

More generally, there is an opportunity to understand if and how these research findings are feeding back into administrative practice, and whether they are improving tax compliance as a result. The OECD has made a useful start in this regard (OECD, 2010) and some governments have stated exactly how they are incorporating NFEs into their practices (Australian Public Service Commission, 2013; HMRC, 2013). However, it would be useful to understand the mechanisms by which this happens, and the success factors involved, since there are various ways that research can translate into policy (Weiss, 1979).

VI. Conclusion

If designed and executed well, then tax compliance NFEs represent a win–win situation for academics and policy-makers. As Alm (2012) points out, ‘measuring, explaining, and controlling evasion are fundamentally and inextricably bound together’: a good NFE will allow policy-makers to evaluate the effectiveness and cost-effectiveness of specific interventions that also advance the theory of tax compliance. Recent years have seen momentum building behind the use of such NFEs. We remain, however, in a situation where we are drawing conclusions from opportunistic and isolated studies that deal with specific types of tax, taxpayer groups, and geographical areas.

The next stage is to ensure sustained engagement with a tax authority so that a coherent body of interrelated findings can be produced. From the academic side, this will require a major commitment to negotiating and relationship building, and tolerance of the arbitrary occurrences that can derail real-world experiments. From the policy side, this will require open-mindedness and tolerance of the possibility that conclusive

evidence will emerge that the intervention backfired. In reality, it is likely that academics will need to make the first move. A good way of doing this is to ensure any proposed study explicitly analyses the costs and benefits of intervening, since this is a major concern for administrators. Given the importance of tax compliance as a policy issue, and the results that have already been achieved from NFEs, there should be a compelling case for collaboration.

References

- Allcott, H. (2011), 'Social Norms and Energy Conservation', *Journal of Public Economics*, **95**, 1082–95.
- Allingham, M. G., and Sandmo, A. (1972), 'Income Tax Evasion: A Theoretical Analysis', *Journal of Public Economics*, **1**, 323–8.
- Alm, J. (2012), 'Measuring, Explaining, and Controlling Tax Evasion: Lessons from Theory, Experiments, and Field Studies', *International Tax and Public Finance*, **19**, 54–77.
- McClellan, C. (2012), 'Tax Morale and Tax Compliance from the Firm's Perspective', *Kyklos*, **65**, 1–17.
- Martinez-Vazquez, J. (2003), 'Institutions, Paradigms, and Tax Evasion in Developing and Transition Countries', in J. Martinez-Vazquez (ed.), *Public Finance in Developing and Transitional Countries—Essays in Honor of Richard Bird*, Cheltenham, Edward Elgar.
- Melnik, M. I. (2010), 'Do Ebay Sellers Comply with State Sales Taxes?', *National Tax Journal*, **63**, 215–36.
- Torgler, B. (2011), 'Do Ethics Matter? Tax Compliance and Morality', *Journal of Business Ethics*, **101**, 635–51.
- Bloomquist, K. M., and McKee, M. (2013), 'On the External Validity of Laboratory Policy Experiments', Appalachian State University Working Paper.
- Jackson, B. R., and McKee, M. (1992), 'Estimating the Determinants of Taxpayer Compliance with Experimental Data', *National Tax Journal*, **45**, 107–14.
- Kirchler, E., and Muehlbacher, S. (2012), 'Combining Psychology and Economics in the Analysis of Compliance: From Enforcement to Cooperation', *Economic Analysis and Policy*, **42**, 133–51.
- Martinez-Vazquez, J., and Torgler, B. (eds) (2010), *Developing Alternative Frameworks for Explaining Tax Compliance*, Abingdon, Routledge.
- Andreoni, J., Erard, B., and Feinstein, J. (1998), 'Tax Compliance', *Journal of Economic Literature*, **36**, 818–60.
- Ariel, B. (2012), 'Deterrence and Moral Persuasion Effects on Corporate Tax Compliance: Findings from a Randomized Controlled Trial', *Criminology*, **50**, 27–69.
- Australian Public Service Commission (2013), *State of the Service 2012–13*, Canberra, Australian Public Service Commission.
- Ayres, I., and Braithwaite, J. (1992), *Responsive Regulation: Transcending the Deregulation Debate*, Oxford, Oxford University Press.
- Becker, G. (1968), 'Crime and Punishment: An Economic Approach', *The Journal of Political Economy*, **76**, 169–217.
- Bhargava, S., and Manoli, D. (2011), 'Why Are Benefits Left on the Table? Assessing the Role of Information, Complexity, and Stigma on Take-up with an IRS Field Experiment', Chicago, IL, University of Chicago Working Paper.
- 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**, 125–36.
- Bobek, D. D., Hageman, A. M., and Kelliher, C. F. (2013), 'Analyzing the Role of Social Norms in Tax Compliance Behavior', *Journal of Business Ethics*, **115**, 451–68.
- Braithwaite, V. (2003), 'Dancing with Tax Authorities: Motivational Postures and Non-compliant Actions', in V. Braithwaite (ed.), *Taxing Democracy: Understanding Tax Avoidance and Evasion*, 177–202, Aldershot, Ashgate.

- Brehm, S. S., and Brehm, J. W. (1981), *Psychological Reactance: A Theory of Freedom and Control*, New York, Academic Press.
- Cabinet Office (2012), *Applying Behavioural Insights to Fraud, Error and Debt*, London, The Stationery Office.
- Castro, L., and Scartascini, C. (2013), 'Tax Compliance and Enforcement in the Pampas. Evidence from a Field Experiment', Inter-American Development Bank Working Paper Series, Washington, DC, Inter-American Development Bank.
- Choo, L., Fonseca, M., and Myles, G. (2014), 'Lab Experiment to Investigate Tax Compliance: Audit Strategies and Messaging', Her Majesty's Revenue and Customs Research Report, No. 308.
- Coleman, S. (1996), 'The Minnesota Income Tax Compliance Experiment: State Tax Results', MPRA Paper No. 4827, University of Munich.
- Cook, T. D. (2003), 'Why Have Educational Evaluators Chosen Not to Do Randomized Experiments?', *Annals of the American Academy of Political and Social Science*, **589**, 114–49.
- Del Carpio, L. (2013), 'Are the Neighbors Cheating? Evidence from a Social Norm Experiment on Property Taxes in Peru', Princeton, NJ, Princeton University Working Paper.
- Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., and Vlaev, I. (2012), 'Influencing Behaviour: The Mindspace Way', *Journal of Economic Psychology*, **33**, 264–77.
- Doyle, E., Gallery, K., and Coyle, M. (2009), 'Procedural Justice Principles and Tax Compliance in Ireland: A Preliminary Exploration in the Context of Reminder Letters', *Journal of Finance and Management in Public Services*, **49**.
- Drago, F., Mengel, F., and Traxler, C. (2014), 'Compliance Behavior in Networks: Evidence from a Field Experiment', presented at 'Taxation, Social Norms and Compliance' conference, Nuremberg, March.
- Dwenger, N., Kleven, H., Rasul, I., and Rincke, J. (2014), 'Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany', Working Paper, Max Planck Institute for Tax Law and Public Finance.
- Elffers, H., Robben, H. S., and Hessing, D. J. (1992), 'On Measuring Tax Evasion', *Journal of Economic Psychology*, **13**, 545–67.
- Weigel, R. H., and Hessing, D. J. (1987), 'The Consequences of Different Strategies for Measuring Tax Evasion Behavior', *Journal of Economic Psychology*, **8**, 311–37.
- Erard, B., and Feinstein, J. S. (1994), 'The Role of Moral Sentiments and Audit Perceptions in Tax Compliance', *Public Finance-Finances Publiques*, **49**, 70–89.
- European Commission (2013), *Europe 2020: Shadow Economy and Undeclared Work*, Brussels, European Commission.
- Feld, L. P., and Larsen, C. (2012), 'Self-perceptions, Government Policies and Tax Compliance in Germany', *International Tax and Public Finance*, **19**, 78–103.
- Frey, B. S., and Torgler, B. (2006), 'Rewarding Honest Taxpayers? Evidence on the Impact of Rewards from Field Experiments', Center for Research in Economics, Management and the Arts (CREMA) Working Paper 16.
- Fellner, G., Sausgruber, R., and Traxler, C. (2013), 'Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information', *Journal of the European Economic Association*, **11**, 634–60.
- Fisman, R., and Wei, S. (2004), 'Tax Rates and Tax Evasion', *Journal of Political Economy*, **112**, 471–96.
- Fowler, J. H., and Christakis, N. A. (2010), 'Cooperative Behavior Cascades in Human Social Networks', *Proceedings of the National Academy of Sciences*, **107**, 5334–8.
- Gangl, K., Torgler, B., Kirchler, E., and Hoffmann, E. (2014), 'Effects of Supervision on Tax Compliance', *Economics Letters*, **123**(3), 378–82.
- Gonzalez-Navarro, M., and Quintana-Domeque, C. (2013), 'Local Public Goods and Property Tax Compliance: Evidence from Residential Street Pavement', paper presented at 'Taxation, Social Norms and Compliance' conference, Nuremberg, March.
- Gordon, J. P. F. (1989), 'Individual Morality and Reputation Costs as Deterrents to Tax Evasion', *European Economic Review*, **33**, 797–805.
- Güth, W., Levati, V., and Sausgruber, R. (2005), 'Tax Morale and (De-) Centralization: An Experimental Study', *Public Choice*, **125**, 171–88.

- Gutmann, P. M. (1977), 'The Subterranean Economy', *Financial Analysts Journal*, 26–34.
- Hallsworth, M., Rutter, J., and Parker, S. (2011), *Policy Making in the Real World*, London, Institute for Government.
- List, J. A., Metcalfe, R. D., and Vlaev, I. (2014), 'The Behaviorist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance', National Bureau of Economic Research Working Paper 20007.
- Harju, J., Kosonen, T., and Ropponen, O. (2013), 'Do Honest Hairdressers Get a Haircut? On Tax Rate and Tax Evasion', Government Institute for Economic Research (VATT) Working Paper, available at http://www.tax.mpg.de/files/pdf2/Paper_Harju_Harju_Tuomas_Kosonen_and_Olli_Ropponen21.pdf
- Harrison, G. W., and List, J. A. (2004), 'Field Experiments', *Journal of Economic Literature*, 1009–55.
- Hasseldine, J. (2000), 'Using Persuasive Communications to Increase Tax Compliance: What Experimental Research Has (and Has Not) Told Us', *Australian Tax Forum*, 15, 227–42.
- Hite, P. A. (2003), 'Framing, Gender and Tax Compliance', *Journal of Economic Psychology*, 24, 517–33.
- — James, S., and Toumi, M. (2007), 'Persuasive Communications: Tax Compliance Enforcement Strategies for Sole Proprietors', *Contemporary Accounting Research*, 24, 171–94.
- Henderson, J. V., Storeygard, A., and Weil, D. N. (2009), 'Measuring Economic Growth from Outer Space', National Bureau of Economic Research Working Paper 15199.
- Hessing, D., Robben, H., and Elffers, H. (1989), 'The Relationship between Self-reported and Documented Behaviour in the Case of Fraud with Unemployment Benefits', 1989 Annual Meeting of the Law and Society Association, Madison, WI.
- Elffers, H., Robben, H., and Webley, P. (1992) 'Does Deterrence Deter? Measuring the Effect of Deterrence on Tax Compliance in Field Studies and Experimental Studies', in J. Slemrod (ed.), *Who Pays Their Taxes and Why? Tax Compliance and Enforcement*, Ann Arbor, MI, University of Michigan Press.
- HMRC (2013), *Levelling The Playing Field: Compliance Progress Report*, Her Majesty's Revenue and Customs, London, The Stationery Office.
- (2014), *HM Revenue and Customs Receipts*, Her Majesty's Revenue and Customs, London, The Stationery Office.
- HM Treasury (2003), *The Green Book: Appraisal and Evaluation in Central Government*, Her Majesty's Treasury, London, The Stationery Office.
- Internal Revenue Service (2012), *Tax Gap for Year 2006: Overview*, available at <http://www.irs.gov/uac/The-Tax-Gap>
- Iyer, G. S., Reckers, P. M., and Sanders, D. L. (2010), 'Increasing Tax Compliance in Washington State: A Field Experiment', *National Tax Journal*, 63(1), 7–32.
- Jackson, S. A. (1992), *Message Effects Research: Principles of Design and Analysis*, New York, Guilford Press.
- James, S., and Alley, C. (2004), 'Tax Compliance, Self Assessment, and Tax Administration', *Journal of Finance and Management in Public Services*, 2, 27–42.
- Khwaja, A. I., and Mian, A. (2011), 'Rent Seeking and Corruption in Financial Markets', *Annual Review of Economics*, 3, 579–600.
- Kirchler, E. (2007), *The Economic Psychology of Tax Behaviour*, Cambridge, Cambridge University Press.
- Hoelzl, E., and Wahl, I. (2008), 'Enforced Versus Voluntary Tax Compliance: The "Slippery Slope" Framework', *Journal of Economic Psychology*, 29, 210–25.
- Muehlbacher, S., Kastlunger, B., and Wahl, I. (2010), 'Why Pay Taxes? A Review of Tax Compliance Decisions', in J. Alm, J. Martinez-Vazquez, and B. Torgler (eds), *Developing Alternative Frameworks for Explaining Tax Compliance*, London, Routledge, 15–31.
- 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, 651–92.
- Levitt, S. D., and List, J. A. (2007), 'What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?', *Journal of Economic Perspectives*, 21(2), 153–74.

- McGraw, K. M., and Scholz, J. T. (1991), 'Appeals to Civic Virtue Versus Attention to Self-Interest—Effects on Tax Compliance', *Law and Society Review*, **25**, 471–98.
- Marion, J., and Muehlegger, E. (2008), 'Measuring Illegal Activity and the Effects of Regulatory Innovation: Tax Evasion and the Dyeing of Untaxed Diesel', *Journal of Political Economy*, **116**, 633–66.
- Nigrini, M. (1996), 'A Taxpayer Compliance Application of Benford's Law', *Journal of the American Taxation Association*, **18**, 72–92.
- OECD (2010), 'Understanding and Influencing Taxpayers' Compliance Behaviour', Paris, Organization for Economic Cooperation and Development.
- Ortega, D., and Sanguinetti, P. (2013), 'Deterrence and Reciprocity Effects on Tax Compliance: Experimental Evidence from Venezuela', CAF Working Paper No. 2013/08.
- Phillips, M. (2014), 'Deterrence vs Gamesmanship: Taxpayer Response to Targeted Audits and Endogenous Penalties', *Journal of Economic Behavior and Organization*, **100**, 81–98.
- Pomeranz, D. (2013), 'No Taxation Without Information: Deterrence and Self-enforcement in the Value Added Tax', Working Paper No. 19199, National Bureau of Economic Research.
- Ringold, D. J. (2002), 'Boomerang Effects in Response to Public Health Interventions: Some Unintended Consequences in the Alcoholic Beverage Market', *Journal of Consumer Policy*, **25**, 27–63.
- Schwartz, R. D. (1960), 'Field Experimentation in Sociolegal Research', *Journal of Legal Education*, **13**, 401.
- Orleans, S. (1967), 'On Legal Sanctions', *The University of Chicago Law Review*, **34**, 274–300.
- Shaw, J., Slemrod, J., and Whiting, J. (2010), 'Administration and Compliance', in J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba (eds), *Dimensions of Tax Design: The Mirrlees Review*, Oxford, Oxford University Press.
- Sheffrin, S., and Triest, R. (1992), 'Can Brute Deterrence Backfire? Perceptions and Attitudes in Taxpayer Compliance', in J. Slemrod (ed.), *Why People Pay Taxes: Tax Compliance and Enforcement*, Ann Arbor, MI, University of Michigan Press.
- Slemrod, J. (2007), 'Cheating Ourselves: The Economics of Tax Evasion', *Journal of Economic Perspectives*, **21**, 25–48.
- 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**, 25–53.
- 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**, 455–83.
- Smith, K. W. (1992), 'Reciprocity and Fairness: Positive Incentives for Tax Compliance', in J. Slemrod (ed.), *Why People Pay Taxes: Tax Compliance and Enforcement*, Ann Arbor, MI, University of Michigan Press.
- Spicer, M. W., and Thomas, J. E. (1982), 'Audit Probabilities and the Tax Evasion Decision: An Experimental Approach', *Journal of Economic Psychology*, **2**(3), 241–5.
- Srinivasan, T. N. (1973), 'Tax Evasion: A Model', *Journal of Public Economics*, **2**, 339–46.
- Taylor, N. (2003), 'Understanding Taxpayer Attitudes through Understanding Taxpayer Identities', in V. Braithwaite (ed.), *Taxing Democracy: Understanding Tax Avoidance and Tax Evasion*, Aldershot, Ashgate.
- Torgler, B. (2003), 'Tax Morale in Transition Countries', *Post-Communist Economies*, **15**, 357–81.
- (2004), 'Moral Suasion: An Alternative Tax Policy Strategy? Evidence from a Controlled Field Experiment in Switzerland', *Economics of Governance*, **5**, 235–53.
- (2007), *Tax Compliance and Tax Morale: A Theoretical and Empirical Analysis*, Cheltenham, Edward Elgar.
- (2012), 'A Field Experiment on Moral Suasion and Tax Compliance Focusing on under-Declaration and over-Deduction', School of Economics and Finance, Queensland University of Technology, Discussion Paper and Working Paper Series.
- Trivedi, V. U., Shehata, M., and Lynn, B. (2003), 'Impact of Personal and Situational Factors on Taxpayer Compliance: An Experimental Analysis', *Journal of Business Ethics*, **47**, 175–97.

- US Treasury (2009), *Update on Reducing the Federal Tax Gap and Improving Voluntary Compliance*.
- Vogel, J. (1974), 'Taxation and Public Opinion in Sweden—Interpretation of Recent Survey Data', *National Tax Journal*, **27**, 499–513.
- Webley, P., Robben, H., Elffers, H., and Hessing, D. (1991), *Tax Evasion: An Experimental Approach*, Cambridge, Cambridge University Press.
- Weiss, C. H. (1979), 'The Many Meanings of Research Utilization', *Public Administration Review*, **39**, 426–31.
- Wenzel, M. (2005a), 'Misperceptions of Social Norms About Tax Compliance: From Theory to Intervention', *Journal of Economic Psychology*, **26**, 862–83.
- (2005b), 'Motivation or Rationalisation? Causal Relations between Ethics, Norms and Tax Compliance', *Journal of Economic Psychology*, **26**, 491–508.
- (2006), 'A Letter from the Tax Office: Compliance Effects of Informational and Interpersonal Justice', *Social Justice Research*, **19**, 345–64.
- Taylor, N. (2004), 'An Experimental Evaluation of Tax-reporting Schedules: A Case of Evidence-based Tax Administration', *Journal of Public Economics*, **88**, 2785–99.
- Yitzhaki, S. (1974), 'Income Tax Evasion: A Theoretical Analysis', *Journal of Public Economics*, **3**, 201–2.