Strengthening tax compliance by implementing behavioural insights for revenue administrations in Belgium, Austria, and Portugal

Deliverable 2 – Methodological handbook

Technical Support Instrument

Supporting reforms in 27 Member States

Funded by the European Union

Federal Public Service FINANCE

Federal Ministry Republic of Austria Finance

AT autoridade tributária e aduaneira

This document was produced with the financial assistance of the European Union. Its content is the sole responsibility of the author(s). The views expressed herein can in no way be taken to reflect the official opinion of the European Union.
The project is funded by the European Union via the Technical Support Instrument, managed by the European Commission Directorate-General for Structural Reform Support.
This report has been delivered in February 2024, under the EC Contract No. REFORM/2021/OP/0006-06. It has been delivered as part of the project "Strengthening tax compliance by implementing BI in Austria, Belgium and Portugal."

The main authors of this methodological handbook are Stefan Elbers, John Gustavsson, Piotr Lipinski and Raphaël De Landsheer

Directorate-General for Structural Reform Support
REFORM@ec.europa.eu
+32 2 299 11 11 (Commission switchboard)
European Commission
Rue de la Loi 170 / Wetstraat 170
1049 Brussels, Belgium

Table of Contents

Introduction

This handbook stems from project funded by the European Union via the Structural Reform Support Programme and implemented by Verian in cooperation with the European Commission's Structural Reform Support (DG Reform).

Background

Recent advancements in behavioural sciences have changed our comprehension of human behaviour. Behavioural Insights (BI) – rooted in psychology, economics, cognitive science, and other social sciences - has emerged as a powerful approach for policymakers to improve public policy[1]. BI assumes that individuals' determination and attention spans are both limited and that they do not incorporate all information available to them into their decision-making processes. Instead, they use "mental short-cuts" (known as heuristics) to make decisions, with the understanding that not all decisions align with underlying preferences. Furthermore, the reasons why individuals make certain choices are complicated and influenced not just by risk/rewards but by intrinsic values such as having a sense of duty, altruism, and the feeling of peer pressure. A typical BI approach makes use of inductive methods to gather insights from observational and experimental data and applies these findings to design and implement interventions that support long-term goals and societal benefits.

Understanding the complexities of human decision making has profound implications for tax compliance. Governments globally are adopting this approach to improve their services by simplifying information or changing the so called "choice architecture"[2] faced by individuals making a decision. The 'architecture' refers to the way that a choice is presented to decision makers such as the taxpayers (e.g., whether or not tax returns utilise default options). Drawing on evidence from applications of BI in the area of tax compliance worldwide, this handbook explores how understanding human behaviour within its social context can offer innovative, cost-effective solutions to challenges in tax filing, payment, and accuracy.

It is important to acknowledge that behavioural interventions in the context of tax compliance should not be considered as substitutes for other kinds of policies, but as tools complementary with the tax law and power instruments of the authorities. The "slippery slope" framework suggests that the relationship between taxpayers and tax authorities is dynamic, influenced by trust and power. Accordingly, effective tax compliance is achieved through a balance of coercive power and trust, emphasising that while behavioural interventions are beneficial, they must operate within a framework that includes both legal enforcement and trust-building measures[3]. The purpose of this handbook is not to offer a "silver bullet", as no such thing exists. There are simply far too many contextual factors influencing tax compliance that cannot be targeted with behavioural interventions, such as poverty[4]. No behavioural intervention can solve the issue of a taxpayer being outright unable to pay their taxes due to liquidity constraints. In fact, in cases where liquidity constraint is the main driver of non-compliance, it is even possible to conceive of scenarios in which BI would backfire and reduce social welfare. Other types of interventions, such as payment plans may be more appropriate in these situations, but such interventions are not BI and are thus outside the scope of this handbook.

Given the plethora of interacting factors influencing compliance, it is important to ensure that BI intervention development projects start with a thorough problem analysis. This is why a section of this handbook deals with the various drivers of non-compliance, and why we emphasise the need to work on a case-by-case basis to understand why a particular group of taxpayers at a particular time in a particular country or region is non-compliant. While it is useful to learn from the experience of other countries and authorities, BI interventions are often only successful if they are specifically designed for unique contexts. What works in one context may not necessarily produce the same results in another one[5]. Module 2 of this handbook demonstrates that many interventions have mixed records depending on the context in which they have been used.

[1] https://www.oecd.org/gov/regulatory-policy/behavioural-insights.htm
[2] Thaler, R. H., & Sunstein, C. R. (2021). Nudge: The Final Edition. Penguin.
[3] Kirchler, E., Hoelzl, E., & Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. Journal of Economic psychology, 29(2), 210-225.
[4] Kirchler, E., Muehlbacher, S., Kastlunger, B., & Wahl, I. (2010). Why pay taxes? A review of tax compliance decisions. Developing alternative frameworks for explaining tax compliance, 59. https://icepp.gsu.edu/files/2015/03/ispwp0730.pdf
[5] Antinyan, A., & Asatryan, Z. (2019). Nudging for tax compliance: A meta-analysis. ZEW-Centre for European Economic Research Discussion Paper, (19-055).

Case study 1: Ghostbusting in Detroit: Evidence on non-filers from a controlled field experiment[6]

Background: The City of Detroit collects a local income tax, and residents are obligated to file a local tax return. Non-filing, however, is a serious issue. In 2014, 48 percent of residents failed to file local taxes compared to only 6.1 percent for federal income taxes.

Possible explanations for not filing in this situation include:

- The limited administrative capacity of the city which leads to a perceived low probability of punishment.
- The required effort to file a return.
- Not being aware that the Detroit has an income tax, or that one must file a separate return for this tax.

Objective: In collaboration with the City of Detroit, researchers tested several interventions aimed at increasing the filing rate by sending letters and postcards with different messages to non-filers. The goal of the study was to establish which type of messaging would be the most effective at increasing filing rates.

Intervention design: The authors did not explicitly test which drivers where related to noncompliance in Detroit. Instead, they developed their letter interventions based on previous research of message types that were successful in other noncompliance situations. The mechanisms tested were:

1) *Penalty salience*. This message aims to appeal to fear of punishment. The message using this mechanism recalled that tax non-return is punishable with a fine of up to $500 and 90 days in jail.

2) *Punishment probability*. This letter informed recipients that the City of Detroit already knew the recipient's total federal income, which is information provided to the City by the IRS. The purpose of this was to make recipients believe that the City already had sufficient information to convict them if they failed to file, increasing the perceived risk of punishment.

3) *Compliance cost*. This group received a blank tax form and a return envelope, reducing their cost of complying.

4) *Civic pride*. This letter made a moral appeal, emphasising the importance of tax collection to aid the resurgence of Detroit.

5) *Penalty salience x punishment probability*. The final intervention combined the first two, with a message first stating the recipient's income and then the penalty for non-compliance.

Experimental design: 7142 suspected non-filers were randomly allocated to one of six groups. These groups received letters or postcards from the City, with each group receiving a different message based on the intervention design explained above.

The sixth group was a control group that received a generic letter without any appeal to comply. In addition, another 2400 non-filers were allocated into a no-contact control group. The researchers tracked the recipients for 90 days after the intervention.

Key results:

- Penalty salience was the measure that performed the best, with 10.1 percent of recipients filing tax returns compared to 3 percent in the contact-only control group and 0.3 percent in the no-contact control group. This was followed by penalty salience x punishment probability at 9.7 percent, compliance cost at 6.2 percent and punishment probability at 4.9 percent.
- In terms of net revenue, the difference was even starker: The penalty salience resulted in net revenues of $8 per recipient, far ahead of penalty salience x punishment at $0.15. No other treatment resulted in positive net revenue.
- The "Civic pride" treatment group's filing rate (3.8 percent) was not statistically significantly different from the contact-only control group.
- The study found no evidence of network effects; none of the interventions had any impact on the behaviour of the recipients' neighbours.

Objectives

This handbook is meant to be a toolbox for authorities and policymakers who are not behavioural scientists but who wish to use behavioural insights in their work. It is based on recent academic findings – such as described in case study 1 - and supports a methodical and evidence-based application of behavioural insights in tax policies.

This handbook aims to:

- Provide authorities and policymakers with a process which can be used to develop, test and implement behavioural interventions that improve tax compliance (module 1).
- Unpack the diversity of drivers of non-compliance in the context of taxation which have been identified in previous studies. This overview of drivers can be used as a starting point for authorities and policymakers to brainstorm about what drivers are relevant for the non-compliance issue(s) in their own countries. This identification is necessary before moving on to designing tailored measures. Furthermore, the handbook will provide an overview of previously tested behaviourally informed interventions by describing the mechanisms which were at their core. This overview is also expected to inspire users of this handbook when designing their interventions in their own countries (module 2).
- To help select the most appropriate research designs for each of the different stages involved in developing and evaluating measures using behavioural insights (module 3).

[6]Meiselman, B. S. (2018). Ghostbusting in Detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics*, *158*, 180-193. https://www.isid.ac.in/~epu/acegd2017/papers/BenShuchMeiselman.pdf

1. Module 1 – How to use the BASIC model to develop behavioural interventions to improve tax compliance

A key aspect of using BI for policy making is a methodical way of intervention development. To avoid the risk of "borrowing" unsuitable behavioural strategies from other contexts, interventions must be meticulously developed through a systematic process. This process guides policymakers from an initial understanding of the problem, through the design of interventions, to the rigorous testing of their effectiveness. There are various models and guidelines available that offer a stepwise approach from defining the problem behaviour to developing, testing and implementing solutions. In this handbook, the OECD BASIC model[7] will serve as the primary framework to illustrate a typical BI intervention development approach.

This module includes two main sections. First, we will use the BASIC model to illustrate an intervention development process. Second, we will discuss key ethical considerations within the context of using BI in the context of tax compliance.

1.1. Using the BASIC model to improve tax compliance

1.1.1. General introduction to The BASIC model

BASIC is a framework intended to guide policymakers on how to use behavioural interventions. It was developed in response to demand from policymakers and authorities around the world, who wished to standardise the process of using BI, allowing them to develop and implement BI in a more systematic, responsible, and less haphazard way. While BASIC is far from the only framework on behavioural interventions, it has a few key advantages over other frameworks, the most important being its comprehensiveness: BASIC is specifically designed with non-behavioural experts in mind and deals with every step of developing policy-related behavioural interventions. Second, because BASIC does not focus on interventions for any particular area of policy, it has a great degree of versatility. While this handbook deals with tax compliance, policy-makers familiar with BASIC can use the framework in other policy areas as well.

This module does not solely repeat the BASIC model as already presented in the publication by the OECD. For each of the five BASIC stages, it specifically addresses behavioural and ethical aspects that relate to the issue of improving tax compliance. This module also includes examples of case studies to illustrate certain steps of the development process.

Figure 1 The BASIC model.

As depicted in figure 1, the model consists of five subsequent stages that provide a structured approach to integrating behavioural insights into the policymaking process, from problem identification to the evaluation of interventions and their potential for broader implementation.

[7] Nakagawa, J., Papa, F., Cavassini, F., Drummond, J., Naru, F., & Esson, A. (Eds.). (2019). *The basic toolkit: Tools and ethics for applied behavioural insights* (Brochure). Organization for Economic Cooperation and Development (OECD). Developed in partnership with Dr Pelle Guldborg Hansen of Roskilde University. Retrieved from https://www.oecd.org/gov/regulatory-policy/BASIC-Toolkit-web.pdf

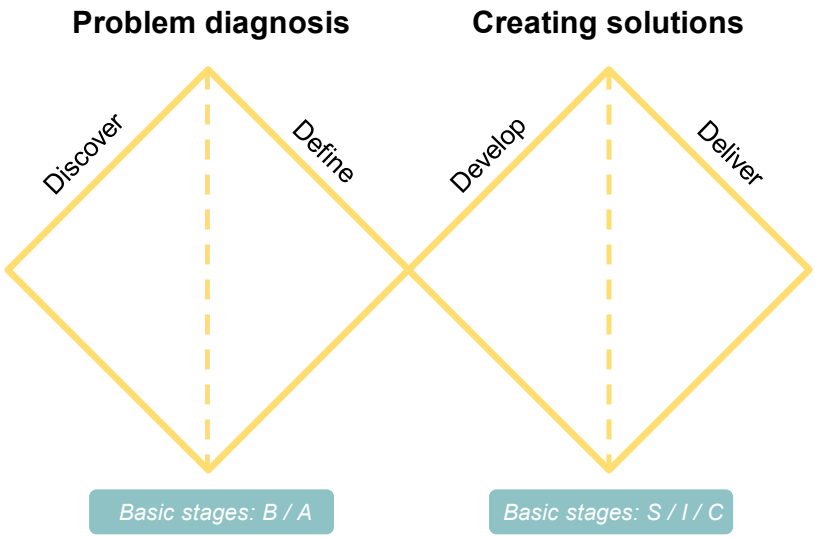
1.1.2. A 'double diamond' perspective on intervention design

In practice, intervention development is not a linear, but an iterative process, where each stage builds on the insights of the preceding stage. In the BASIC model, stages 1 and 2 focus on the diagnosis of the problem – these two stages are about gaining more specificity about what the problem is and what factors causes it. To arrive at the problem diagnosis, first the problem is explored by investigating the tax system, taxpayers' behaviours, and other related factors. When a comprehensive overview of the problem has been made, the second step of the problem diagnosis is to set priorities and determine the most important issues and taxpayer behaviour(s) in the system.

This is followed by a similar cycle when designing a solution for this problem. Stage 3 of the basic model concerns the generation of possible strategies to address the problem, whereas stage 4 and 5 involve developing, testing, and selecting strategies to identify the most optimal solution(s).

This iterative approach mirrors the process observed in generic intervention design models, emphasising the importance of exploration before convergence on a solution. Notably, it aligns with the principles of the Double Diamond model, which is widely utilised in intervention development across a diverse range of activities. Developed by the British Design Council[8], the Double Diamond model highlights the significance of an iterative approach of alternating cycles of diverging (i.e., generating ideas) and converging (i.e., prioritising and selecting ideas) activities to stimulate creativity through exploration and strategic decision-making. Although we will focus on the five BASIC stages in module 1, it is important to understand these iterative cycles behind the intervention development process. If, for example, a certain prototype intervention does not provide the expected outcomes, an extra iteration is likely to help improve the prototype, which is more cost-effective than completely starting over.

Figure 2 The Double Diamond model.



1.1.3. Stage 1: Behaviour

The precondition for designing a tailored behavioural intervention is having a clear understanding of the behaviour that the measure aims to change. Tax compliance in itself is not a behaviour, but rather a result of several more specific behaviours as will be shown below.

In this initial stage, the focus is on outlining the behaviours that relate to tax compliance and providing guidance on narrowing down and selecting suitable target behaviours for intervention. Additionally, the SMART framework is introduced, a framework that can be used to set objectives and determine the success or failure of the intervention that will be designed and tested in the later stages of BASIC. The main outcome of this stage is a description of the target behaviour(s) that will be the focus of the behavioural intervention.

[8] Design Council. (2023) The Double Diamond model. Retrieved from: https://www.designcouncil.org.uk/our-resources/the-double-diamond/

What are the possible target behaviours for tax compliance? Tax compliance consists of a sequence of interactions between taxpayers and the tax authority, where taxpayers are required respond in certain ways in order to be tax compliant (e.g., by submitting a tax return or by paying tax debt in time). One way to unpack what behaviours underpin tax compliance is to think about the journey of the taxpayer. Although these so-called tax journeys differ between countries (and within countries between individuals) they typically have a number of common elements. The journey often starts by filing a tax return, which requires taxpayers to learn about and pay attention to taxes, and to set aside time to fill out the return (which may require planning ahead). After filing a return, the taxpayer then receives a tax assessment notice which they must read and understand, and, if they owe taxes, they must arrange to pay them in a timely manner using the correct identification, requiring financial planning on their part to cover this expense. Different behavioural and non-behavioural drivers, such as fear of punishment, ease of filing and ease of payment affect different stages of this tax journey, something which we cover in depth in Module 2.

The tax journey of the taxpayer is likely to differ depending on the tax concerned. Thus, the first step of this process is for the authorities to select a type of tax for which they wish to enhance compliance.

The second step is to map the tax journey for this type of tax and to clearly define what is considered compliant and non-compliant for each taxpayer's action within this journey. When establishing these criteria, it is important to ensure that they can be effectively measured by the administration. For instance, if the administration experiences delays in processing mail of paper tax returns, these delays should be reflected in the compliance definition.

The third step is to identify the most important candidate behaviours for improvement. A good approach is to consider and discuss all interactions and behaviours a taxpayer has to perform to be tax compliant within the tax journey. The result of this exercise is the compilation of a comprehensive list of all candidate target behaviours.

Box 1 Example list of behaviours

What might a list of behaviours look like for personal income tax?

- Inform oneself about how to submit a tax return and the relevant deadlines.
- Choose the form in which to declare tax (online/ paper)
- File a tax return by the deadline.
- Report income accurately.
- Find out what constitutes legitimate deductions.
- Claim only legitimate deductions.
- Save necessary receipts or other documentation as evidence for entitlement to deductions.
- Co-operate with audits.
- Pay any taxes owed on time.

It is not feasible for an intervention to target all the behaviours concerned. Furthermore, not all the behaviours have equal weight in driving non-compliance. Some behaviours are likely to be more problematic for some target audiences. For example, persons who have to declare taxes for the first time are more likely to need access to information about the process. Furthermore, not all behaviours carry the same importance as root causes of non-compliance. Finally, not all behaviours are equally susceptible of being influenced by BI. Therefore, it is necessary to prioritise among all relevant behaviours in order to select the ones which are most problematic and where there is greatest scope for effectiveness in terms of compliance with the use of BI.

Selecting target behaviours. There are various criteria that can be used to prioritise the most problematic target behaviours within a tax journey. What criteria are used and how they are weighed will depend on the overall project objectives and organisational strategy. Since all humans are subject to biases, it is vital that a proper standard for evaluation is used during this step, so as to prevent the process from becoming arbitrary and coloured by pre-existing beliefs of the staff involved. The OECD suggests the following criteria for narrowing down behaviours:

Importance. Is a change in this behaviour an institutional priority? What percentage of the target population fails to perform the behaviour (correctly)? How central is it in the behavioural system, that is to say the tax journey? If the behaviour cannot convincingly be argued to be important, there is a risk that efforts to change the behaviour may be haphazard or may be abandoned prematurely. If the relevant institutions, such as the government and the tax authorities do not agree that targeting for example late filing is an important endeavour, it will be more difficult to carry out. In assessing the importance of targeting a behaviour, one must keep in mind both the frequency and cost of the behaviour (how common is late filing? How much does it cost the tax authorities?), as well as potential positive spillovers that may result from targeting a certain behaviour. For example, studies[9] have suggested that behavioural interventions targeting one tax may also have a positive impact on compliance with other un-related taxes (i.e. property tax and gross-sales tax for self-employed individuals).

Ethics. What are the ethical implications of targeting a particular behaviour? Given the positive relationship between income and compliance, many interventions targeting non-compliance inevitably end up disproportionately targeting low-income and vulnerable groups, which may contribute to a feeling of unfairness and resentment. There is also the potential for unintended side effects: While the target behaviour may be "claim only legitimate deductions", there is a risk that interventions will deter some taxpayers from claiming even those deductions that they *are* legally entitled to.

Impact. How great is the societal impact of altering this behaviour? For tax behaviour, the impact would include first and foremost the sheer amount of revenue raised from improving a behaviour, for example the revenue raised from more taxpayers correctly reporting their incomes or not claiming deductions they are not entitled to. In addition, interventions may also reduce administrative costs by, for example, increasing the number of taxpayers who file on time and whom the tax authorities do not have to send reminders to. On the other hand, the impact assessment must also include aspects such as fairness and the perception of the system: As mentioned in the previous section, a behaviour that mostly concerns lower-income individuals may by many be perceived as "punching down", even if it does raise revenue.

Feasibility. Is there political support for targeting the behaviour, and how controversial might it be to use BI to do so? In the past, initiatives by tax authorities to improve compliance have met with political controversy, such as the U.S. Internal Revenue Service hiring spree, which led to the agency losing some of the funding it had previously been promised[10]. Policymakers and authorities have to consider which behaviours may be more controversial to target, where targeting may be perceived as more intrusive or unjustified.

Data access. How much data is available on the behaviour, and is there an agreed-upon way of measuring it? While it is easy to estimate precisely how many taxpayers are filing late, and how many taxpayers do not pay taxes owed in time, it is harder to estimate exactly how many taxpayers are underreporting their incomes as any such measure relies on either self-reporting or on the use of proxies to estimate "true" incomes.

Frequency. For individual taxpayers, most tax-related behaviours take place only once a year during tax filing season, whereas self-employed individuals and firms may engage in tax-related behaviours throughout the year. This distinction is important as many behavioural interventions aim to establish and reinforce habits, which generally is not a good approach to a behaviour that takes place very infrequently.

There are multiple ways to prioritise the list of target behaviours on the selected criteria. The OECD developed a priority filter questionnaire [link] to help prioritise. Regardless of which way target behaviours are assessed, it is important to use all available evidence and involve available expertise at this stage. This diagnosis may be established based on analysing existing administrative data, using brainstorming techniques which allow the tax authority experts to unpack the behaviour and prioritise or using qualitative research. For more information about the use of methods which can be useful for this diagnosis stage of the process see module 3. When the target behaviour is selected, it is important to clearly define it. This ensures a common understanding of the

[9] López-Luzuriaga, A., & Scartascini, C. (2019). Compliance spillovers across taxes: The role of penalties and detection. *Journal of Economic Behavior & Organization*, *164*, 518-534. https://www.sciencedirect.com/science/article/pii/S0167268119302057

[10] https://thehill.com/business/4395737-republicans-win-faster-irs-cuts-in-funding-deal/

focus of the study and clarifies the involved people and timing of the behaviour. Box 2 contains guiding questions for this.

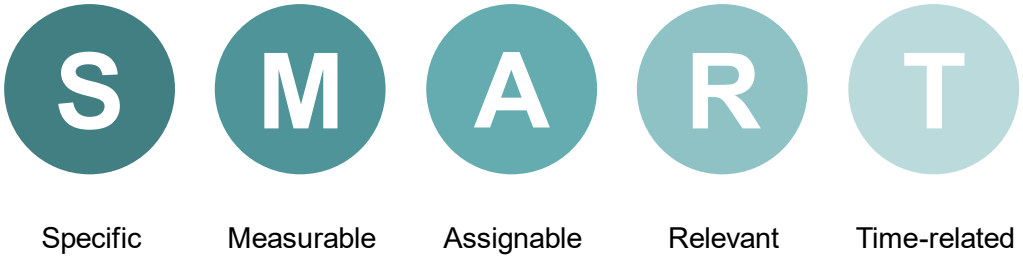
Box 2 Guiding questions for specifying the target behaviour.

Michie et al., (2014; p.48) propose to use the following questions to specify the selected target behaviour in detail:

- Who needs to perform the behaviour?
- What do people have to do differently to achieve the desired outcome?
- When will they perform the behaviour?
- Where will they perform the behaviour?
- How often are they required to perform the behaviour?
- With whom will they have to perform the behaviour?

Defining the objectives for the intervention which will be developed and trialled. The next step in the process is to formulate a clear set of SMART objectives for the intervention which will be tested. The purpose of these objectives is to allow authorities to make a judgement about the extent to which a given intervention was a success. Typically, a BI process of developing interventions will compare several options (see for example case study 1 above). These will most likely perform differently in addressing the problem behaviour. However, it is possible that none of them achieves a sufficiently large improvement compared to the scale of the initial problem. It is also possible that several of them perform well. The SMART objectives should define the threshold which will allow to judge the extent to which a given BI project was or was not effective.

Figure 3 Smart target behaviours



SMART desired policy outcomes should be:

Specific to the target behaviour. The policy outcome should be directly related to the behaviour and be quantifiable, i.e. “decrease in filings which are made after the due date” rather than “increase of timely filing”. The first example is more specific as it defines that is considered as timely everything that is filed before or on due date.

Measurable. As noted above, some behaviours such as late filing are easier to measure than others. What is important is that there is an agreed-upon way to measure the frequency of the target behaviour, and that the organisation responsible for carrying out the BI is able to collect this data (either themselves or through a third party). For example, it could be the case that is considered as “timely” any submission received up to five days after the due date provided such exact data is available.

***Assignable*.** to a specific group of individuals. In the case of tax compliance behaviours, this may be “all taxpayers” but could also be “late filers” or “taxpayers known to have cheated in the past”. It could also be those eligible for a specific tax first time (newly arriving persons to the country or young people entering labour market).

Realistic. Consulting with behavioural experts and reviewing historical data can help tax authorities and policymakers set realistic expectations for what kind of change can be achieved by targeting, for example, late filing. What is realistic will depend not only on the behaviour being targeted but the resources available, the timespan of the intervention, and on underlying factors such as demographics (see Module 2). This will also

depend on the resources available to implement the intervention. Depending on the resources available one can expect the reach of the intervention to vary. For example, interventions that are administered through online means (e.g. email remainders) are relatively cheap and can reach a large number of people. On the other hand, interventions that rely on being administered by advisors (for example providing assistance) are much more resource intensive and will reach substantially fewer people.

Time-related. This refers to setting a "deadline" at which point the desired outcome should have been met, i.e. during which tax year. How long an intervention should be allowed to run before being evaluated depends on the nature of the intervention. To determine a suitable timespan, we strongly recommend reviewing existing research on BI related to the target behaviour(s).

It must be noted that contextual factors can have a large influence on the policy outcomes. In Module 2, we introduce a number of demographic factors that research shows have an impact on tax compliance. While demographics cannot be manipulated through policy, these and other non-modifiable factors must be considered to help set realistic expectations for practitioners: A country or region with particularly, for tax compliance purposes, unfavourable demographics may perhaps not reach the level of compliance of another country or region with particularly favourable demographics, even with the help of BI. This is important to consider in order to set realistic expectations, and based on those expectations decide whether it is worthwhile to pursue BI and decide on what outcomes would justify ending an intervention.

Mapping the contextual factors of the target behaviour

When a target behaviour is selected it is important to gain a deeper understanding of the context in which the target behaviour occurs. This helps to identify potential targets for the behaviour intervention as well as potential positive or negative spillover effects of modifying the target behaviour.

Mapping the context is often done by drawing a process map - or by refining the initial tax journey - that includes all actions and decisions before, during and after the target behaviour. A process map creates an easy overview of processes and sequences and is particularly helpful when a process involves a great number of steps (as the tax process does) that may be easy to overlook in text. Process maps help create context when it would otherwise be difficult to understand how the target behaviour fits with actions and decisions before and after the behaviour. The initial tax journey is typically a good starting point for the process map. At this stage, it is important to further refine the steps and interactions proximate to the target behaviour (e.g. by including specific deadlines or detailed information how a certain action should be completed).

Case study 2. Interactive process map of a tax journey.

Background: The Taxpayer Advocate Service is an independent organisation within the US Internal Revenue Service. Its main objective is to ensure that every taxpayer is treated fairly by helping taxpayers with individual problems.

Intervention: To help all taxpayers navigate the complex system of personal income tax, the Taxpayer Advocate Service developed a digital interactive roadmap of a taxpayer's journey in the United States. It shows the main stages of the US tax system, depicted as a US subway map, and includes various stages of a taxpayer's journey, including direct links to resources at a particular stage.

Outcome: The tool is available **online** (and **pdf**) and aims to help taxpayers to find their way in the system. The Taxpayer's roadmap is an example of a process map to understand the interactions and context of taxpayer's behaviours within this journey[11].

1.1.4. Stage 2: Analysis - How to identify the main psychological drivers behind the target behaviour(s)?

Once the first stage has been completed and a target behaviour is selected, it is time to focus on the psychological and non-psychological drivers that influence the behaviour. This, in turn, will allow authorities to explore the behaviour from a BI perspective, which is the goal of this step. The drivers of the behaviour are what causes it. It is by tackling these

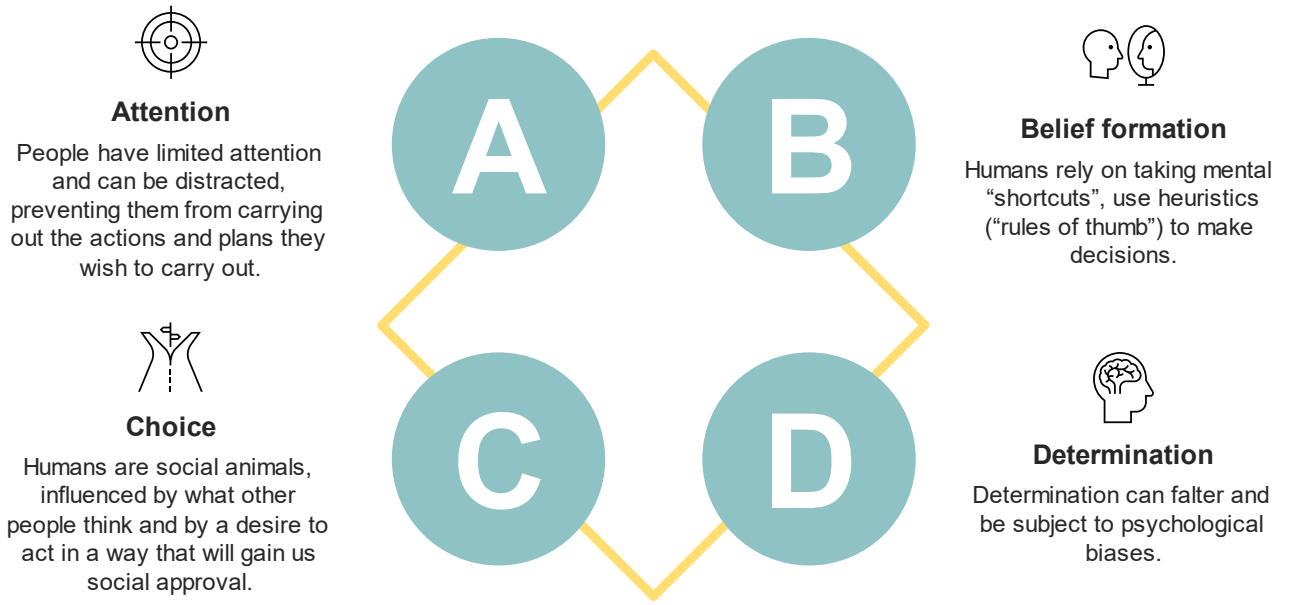
[11] Ciraolo, C. D. (2020). Showing Us How to Get It Done: Nina Olson. *Pitt. Tax Rev.*, *18*, 183. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4529551

drivers that behavioural interventions are expected to be effective in tackling the problem at stake. As said above, some drivers will be behavioural while others may be non-behavioural. For example, poverty may be a driver of non-payment of taxes, but it is one which is not going to be tackled through behavioural interventions. On the other hand, the fact that people do not create savings, or do not prioritise tax payment over other types of expenses which would allow them to pay the tax when it is due are drivers that can be tackled through behavioural interventions.

A first step of this analysis is to establish as many potential drivers as possible. Not all the drivers will be retained for the design of the intervention(s) as not all of them are equally important in causing the problematic behaviour. Mapping them systematically is important to avoid a situation where an important driver is omitted in the considerations about the interventions to be tested.

To ensure a systematic and comprehensive approach to this analysis, multiple frameworks exist that help to bridge the gap from behavioural science to practical applications[12]. The OECD uses what is known as the Attention, Belief formation, Choice and Determination (ABCD) framework, but other noteworthy examples are the COM-B framework, and the theoretical domains framework which have been extensively applied to public policy, sustainability and healthcare[13]. Which framework to use is ultimately a subjective matter, highly dependent on contextual factors including which framework the professionals involved have experience of using in the past. At a minimum, the chosen framework needs to include both automatic (implicit) processes and reflective processes, both of which play a role in virtually all behaviours. The framework must also consider both individual and contextual factors.

Figure 4 The ABCD framework[4]



The ABCD framework considers behaviour to stem from the following four distinct factors.

Attention. People have limited attention and can be distracted, preventing them from carrying out the actions and plans they wish to carry out. For a behaviour such as filing a tax return in time, or for correctly stating one's income, this step is crucial to consider. A taxpayer who intends to file in time when they receive a reminder letter, may be distracted and forget to do so at a later moment. If the process of filing demands too much attention (due to being long and complicated), there is also an increased risk that the taxpayer becomes distracted and forgets to report some income (such as from a side project) that they had.

Many countries have put in place pre-filled tax forms as a "default" based on the income and deductions the tax agency believes the taxpayer is entitled to. Due to lack of attention, individuals, while not desiring to commit tax

[12] Rhodes, RE, McEwan, D & Rebar, AL 2019, 'Theories of physical activity behaviour change: A history and synthesis of approaches', Psychology of Sport and Exercise, vol. 42, pp. 100-109. https://doi.org/10.1016/j.psychsport.2018.11.010
[13] Michie, S., Atkins, L., & West, R. (2014). The behaviour change wheel. *A guide to designing interventions. 1st ed. Great Britain: Silverback Publishing*, *1003*, 1010. Retrieved from https://www.researchgate.net/profile/Susan-Michie-2/publication/311857816_Changing_Behaviour_to_Improve_Clinical_Practice_and_Policy/links/588249b9aca272b7b4425460/Changing-Behaviour-to-Improve-Clinical-Practice-and-Policy

evasion, may simply submit the default and think no more of it. This default option bias is well documented in many areas of behaviour, including saving and organ donations[14].

Another frequently adopted strategy within this category is to remind public of tax filing through public campaigns at the right time. For example, a campaign evaluation in Pakistan revealed that exposure to newspaper advertisements providing information about tax eligibility significantly increased taxpayers' likelihood of income tax filing[15] (a 19.4% increase in tax filing probability).

Questions to consider: Is the targeted decision point well-timed; are people making the decision to engage in the behaviour (like filing taxes) when they are in a good state of mind and not stressed? What may be seizing the attention of taxpayers in the context where the behaviour is engaged (or not engaged) in? How can we design our online platforms and paper-based forms in such a way as to grab and maintain the attention of taxpayers?

Belief formation. Humans rely on taking mental "shortcuts", or in other words, use heuristics to make decisions. Heuristics simplify the decision-making process by allowing individuals to quickly draw conclusions and make judgments without the need for comprehensive information processing, especially in complex or uncertain situations. While they efficiently reduce the cognitive load during decision-making, the outcome frequently results in either over- or underestimating the probability of certain outcomes. The Cognitive Bias Codex (link) provides an overview of approximately 200 well documented cognitive biases that result from these mental shortcuts. This is a particularly important factor for tax compliance. For example, taxpayers tend to vastly overestimate the likelihood of being caught not complying, making them more likely to comply[16]. Belief formation can also have a negative impact on compliance. For example, when taxpayers allocate too little time to file their tax returns because they underestimate how long the filing process will take, or mistakenly believe that they are not obligated to file a tax return.

Choice. Humans are social animals, influenced by what other people think and by a desire to act in a way that will gain us social approval. If people believe that no-one in their social circle (neighbourhood, friends etc.) think that paying taxes is the morally right thing to do, there is a high likelihood that they are less likely to do so themselves. Beyond social approval, intrinsic motivation – what is known in literature as "tax morale" – plays an important role, and compliance frequently stems from a belief that, by complying, the taxpayer is "doing the right thing", which provides its own reward. Again, many interventions appeal to and attempt to strengthen tax morale. Other choices that may be targeted by interventions include the choice on whether or not to file online, which tax authorities typically prefer, and the choice of whether to pay taxes owed in a lump sum or in instalments.

It is also crucial for policymakers and authorities to avoid "choice overload", when too many options cause an individual to freeze, making them to either not make a choice at all, or choose more or less at random. A complicated tax code that offers dozens of similar-sounding categories for even the most common deductions could cause this to happen.

Determination. Determination can falter and be subject to psychological biases. Anyone who has ever tried to quit smoking, or attempted to lose weight can attest to this. The less mentally taxing it is to do the right thing, and the less determination is needed, the more likely it is that individuals choose to do it and carry through with their decision. Because of this, making compliance as seamless as possible is an important factor to encourage it. For example, taxpayers faced with paying a tax bill after submitting their tax return (due to having underpaid taxes over the past year) should be able to pay this bill in a fast and easy way. It is important that the filing and payment processes are easy and that taxpayers know that they are easy and do not require a great deal of determination. If the system is perceived to be complicated, even if this is not true, it will impact how many taxpayers comply in a timely manner.

[14] Steffel, M., Williams, E. F., & Tannenbaum, D. (2019). Does changing defaults save lives? Effects of presumed consent organ donation policies. *Behavioral Science & Policy*, *5*(1), 69-88. https://journals.sagepub.com/doi/pdf/10.1177/237946151900500106

[15] Koumpias, A. M., & Martínez-Vázquez, J. (2019). The impact of media campaigns on tax filing: quasi-experimental evidence from Pakistan. *Journal of Asian Economics*, *63*, 33-43. Retrieved from: https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1154&context=icepp

[16] Aliyev, G. R. o. (2019). *Taxpayers' Misperceptions and Two Novel Behavioral Interventions to Counter Tax Evasion* (Doctoral dissertation, Pardee RAND Graduate School). https://www.rand.org/content/dam/rand/pubs/rgs_dissertations/RGSD400/RGSD435/RAND_RGSD435.pdf

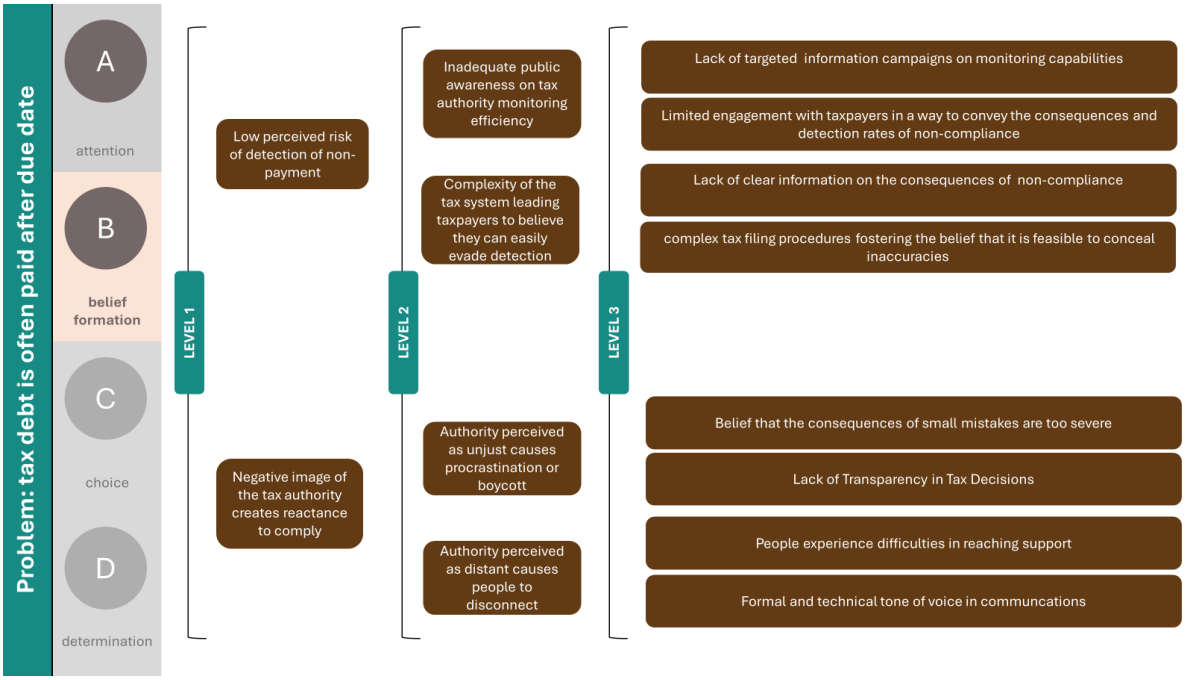
Methods for evidence-based BI analysis. This ABCD framework, as any framework, is first and foremost a general guide, and still leaves questions as to the specifics:

- How do we find out what might distract people while they are filing their taxes?
- How do we know what their pre-existing beliefs are that may impact their willingness to pay taxes?
- How do we identify whether the process of filing and paying taxes is too cumbersome?

An accurate behavioural diagnosis consists of the most important modifiable drivers that influence the target behaviour. Often, strong direct evidence on is absent, due to the specificity of the target behaviour within its unique context. Even historical data within the same country may prove of limited use, due to changing procedures or methods of filing. To generate ideas on possible modifiable drivers of the target behaviour, the four stems of the ABCD model serve as a starting point to generate ideas of potential drivers or to categorise existing information. Depending on the available information, explorative research methods – such as surveys, qualitative interviews or co-design sessions (further explained at module 3) need to be deployed to obtain a complete overview. Case study 3 provides an example of how stakeholder interviews led to the identification of various drivers for tax compliance in the construction industry.

A systematically approach to further narrowing down potential ideas to specific drivers is to use problem trees. Creating problem trees helps to organize, understand and refine the drivers influencing target behaviours. By breaking down the issue into its components using the ABCD model as a guide and by using existing insights on the problem based on exploratory research such as surveys or stakeholder interviews, one can systematically map out the modifiable drivers affecting the target behaviour. The figure below depicts a hypothetical example of a problem tree on tax payment non-compliance, specifically focussing on 'belief formation' drivers.

Figure 5 example of a problem tree to organize drivers.



The next step is to select the drivers that could be the target of the behavioural intervention. The main question at this step is 'what needs to change for the desired behaviour to occur?'[14], which can be done by charting a map that includes all drivers to illustrate how all drivers interact with each other and with the target behaviour and change over time. Based on this overview, the most important drivers should be selected as intervention targets.

Case study 3: Nudging towards tax compliance: A fieldwork-informed randomized controlled trial (Study 1)[17]

This study illustrates a methodical process to analysing and prioritising behavioural drivers within the context of tax compliance in the construction industry in Estonia. The authors used this approach to develop an e-mail message intervention tailored to the specific circumstances.

Background: The authors aimed to increase compliance in filing of payroll information. Because traditional deterrence-based approaches may not be sufficient to address tax compliance challenges in certain contexts, the authors suggest a nuanced approach. In the study, they describe the development and testing of so-called 'tailored compound interventions' (TCIs). In a TCI, multiple nudges that target different behavioural mechanisms behind tax compliance are employed at the same time.

Objective: This study aimed to identify and prioritise behavioural drivers to increase tax compliance in the construction industry by combining existing literature with a qualitative study of the target group.

Identifying behavioural drivers: The authors conducted semi-structured in-depth interviews with 16 employers in the construction industry. In these interviews they explored reasons for non-compliance. They analysed the transcripts to identify the major themes, which they then situated within psychological models of tax compliance and nudge interventions.

Results: Four major themes emerged from the interviews:

- ***Not paying pays off***. Employers view tax compliance through a cost-benefit analysis lens, driven by the need to survive in a competitive market. Under this perspective, evading taxes is considered a means of "saving" money, especially when the perceived risk of detection is low.
- ***Not paying is the norm***. The non-payment of taxes is perceived as the norm among subcontractors, with many considering undeclared salaries standard practice. Interviewees express a common sentiment that official payment of entire salaries is rare across companies, indicating a widespread acceptance of evading tax obligations.
- ***Unfair government***. Some interviewees criticised the government for being overly controlling towards entrepreneurs, contrasting it with what they perceive as irrational spending by politicians. Negative experiences with government agencies or a perceived lack of adequate public services contribute to a sense of unfairness among taxpayers.
- ***Red tape***. Employers in the construction industry highlight the burden of bureaucracy, leading to a preference for unofficial agreements and financial dealings. The complex tax system is cited as a reason to avoid official channels altogether, with the belief that some tasks are easier to handle "unofficially" due to the overwhelming red tape associated with the tax system.

These themes were classified into a working model of compliance behaviour, where eight potential prevention and promotion targets for an e-mail message intervention were identified. The selection process included the following criteria: the potential of an intervention target to increase compliance; the malleability of an intervention target through an e-mail message; and the likelihood of adverse effects. Based on these considerations, four targets were selected: strengthening perceived risk of non-compliance; weakening descriptive norms of non-compliance; weakening the adversarial construal and strengthening the collaborative construal of tax authority.

These four intervention targets were combined in a single e-mail. To strengthen the perceived risks of non-compliance, the e-mail announced new enforcement initiatives by the authority, such as upcoming audits. To weaken the descriptive norms of non-compliance, the e-mails also stated that 92% of employees receive their salaries with due taxes deducted. To weaken the adversarial construal of the authority, the e-mails also stated that assuring fair competition was a priority for the authority. Finally, to strengthen the collaborative construal of the authority, the letters also exemplified the investment value of tax revenues.

Progressive refinement of the problem scope

With the selection of the drivers as targets for a BI intervention, the problem diagnosis is completed. Through various steps within the Behaviour and Analysis stage the problem scope will be progressively refined from a broad, country-wide perspective to specific targeted, actionable issues. This refinement involves narrowing

down from general patterns observed across the entire country to the specific behaviours and areas that will benefit most from intervention.

At this juncture, it is also important to reflect on whether a behavioural intervention is the optimal path forward. It is important to consider if there might be underlying issues that require different strategies. This reflection ensures that the chosen approach addresses the root causes effectively and is not a superficial solution to more complex challenges. Box 3 includes guiding questions to help with this reflection.

Box 3. Stop and think: Is a behavioural intervention an appropriate way forward?

While the BASIC framework is a useful tool for developing interventions, it does not fully address the consideration of whether a BI intervention is the optimal approach in the first place. at this problem diagnosis stage, it therefore is crucial to reflect on the appropriateness of BI to solve the issue.

This box contains four guiding questions to help with a decision on whether a BI intervention approach is the best way forward:

1) Is the root cause of this problem fundamentally a behavioural issue?
2) Are the key drivers of the behaviour modifiable with a BI approach? Poverty, for instance, is not a driver that can be directly modifiable through behavioural insights.
3) Are there any ethical concerns associated with using a BI approach to influence this behaviour?
4) Is a BI intervention development the best way forward for this issue, or is it possible to refine existing procedures and approaches that are not specifically BI-focused?
5) Is the targeted magnitude of change realistically attainable with a behavioural intervention?
6) Do records indicate that BI have been successfully used to change a similar behaviour in the past?
7) Are there alternative strategies that might be more effective or appropriate for addressing this issue?
8) What resources (time, budget, expertise) are required to design and implement the BI intervention, and are they available?

1.1.5. Stage 3: Strategies

In this stage, we move from the problem identification and analysis towards generating solutions. The main goal is to generate behavioural strategies to target the drivers that have been identified in the previous stage. In this step, technical expertise of behavioural scientists and other experts is often required. By understanding the basics of the process, non-experts will be able to better interact with these experts and ultimately achieve a better outcome. In addition to the ABCD model, the overview of previous BI interventions (see Module 2) is also a useful source at this step.

It is important to emphasise that this stage requires creative thinking and expert discussions on possible solutions. A good start is to ensure a mutual understanding of the project scope, and the outcomes of the previous stages to ensure that all participating stakeholders in this stage have a comprehensive understanding of the targeted behaviour(s) and their root causes. This collective understanding fosters collaboration to ideate effective strategies that address the complexity of the problem at hand. Design thinking strategies (module 3) are well-suited methods that facilitate this creative process. Rushing through this stage poses a risk of prioritising quick solutions, potentially leading to a focus on past interventions or replicating strategies that may not consider the unique contextual factors of the current tax journey or prevailing zeitgeist.

[17] Vainre, M., Aaben, L., Paulus, A., Koppel, H., Tammsaar, H., Telve, K., ... & Uusberg, A. (2020). Nudging towards tax compliance: A fieldwork-informed randomised controlled trial. *Journal of Behavioral Public Administration*, *3*(1). https://journal-bpa.org/index.php/jbpa/article/download/84/67

Using the ABCD model to identify strategies for change. Returning to the ABCD framework, we will now suggest ways of turning the theory into practice. What kind of strategies might target each of these aspects of behaviour?

Attention. In interventions aimed at improving tax compliance, the limited human attention is crucial to consider. One way to do so is through increasing the salience of relevant information. This is especially pertinent to communication interventions, where the following features can be optimised for salience and attention-grabbing:

- Communication channel. An important distinction can be made between passive communication (such as through announcements on websites) and active communication (such as through sending reminders). Within active communication, the channel matters as well. For instance, in-person visits by the tax authorities are much more difficult to ignore than letters sent by post.
- Timing of the communication. Well-timed communication can grab attention. A reminder that is sent right before the deadline for filing taxes can be more salient than one sent a month before. Similarly, digital prompts can be used for precise targeting of on-going actions (for example, if they are displayed while the taxpayer is filling in their declaration), making the communication directly relevant and thus, salient.
- Source of the communication. An envelope of a reminder letter that clearly states that it is from a tax authority and contains relevant cues (such as the logo) can increase perceived importance of the communication, and thus stimulate paying attention.
- Message content. If information on potential penalty for non-compliance is clearly stated in a top part of the letter, taxpayers might be more inclined to read the rest of the letter.
- Complexity of the content. Simple communication can make the most important information more visible, as taxpayers do not have to navigate through the complexity to find it.
- Visual display. For instance, the use of colour can make the most important elements salient (e.g., the action of accessing the tax declaration portal in a blue frame contrasting with the rest of the letter)

When designing behavioural interventions beyond the realm of communication, attention also needs to be considered. For instance, interventions can plan for lack of attention, working around this issue to ensure that the intervention has at least some impact even without attention being paid. Changing the default option in a way that improves compliance is one example of an intervention that does not rely on attention, as taxpayers like consumers in general are prone to choose the default option without thinking or paying much attention to it.

Belief formation. Incorrect beliefs may both increase or decrease the frequency of a desirable behaviour. When pre-existing beliefs and expectations are harmful, well-placed disclaimers and targeted information or messages may be effective at dispelling them. Measures targeting belief formation are about the content of the message[18] that the target audience receives through the intervention and which, if any, short-cuts that message appeals to and tries to address.

An example is a disclaimer at the top of a tax return informing the taxpayer of how long they should set aside to fill out the return (this number should be higher than the average time needed for completion to ensure as many as possible do not run out of time). This may help reduce the number of late filers. Other examples include messages with the consequences for non-compliance and framing non-compliance in terms of an active choice rather than as a mere oversight (to induce moral guilt). On the other end, information provision can also focus on providing taxpayers with information that they may need to file the return.

Choice. Interventions addressing choice are about making the preferred choice (one that leads to the desired behaviour) easier or even the default without completely removing the choice (though certain situations may of course call for removing choices or options).

A typical choice issue in the tax domain is allowing taxpayers to decide between annual payment and monthly withholding at the source. From a ‘rational agent’ perspective withholding strategies initially suggest a

[18] While the content of the message is crucial for belief formation, it is important to note that there are other features that can facilitate changes in beliefs – for instance, if the information is coming from a highly credible (compared to non-credible) source, people are more likely to judge the content as valid, and update their beliefs (the „messenger effect”). Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of economic psychology*, *33*(1), 264-277.

disadvantage for taxpayers, as it leads to earlier tax payments. However, social science research reveals that individuals often prefer withholding or even over-withholding. Taxpayers, seeking a forced savings plan, are inclined to overpay throughout the year and receive a lump sum refund, which they may view as a windfall for essential purchases. In contrast, annual payments lack the psychological mechanisms that encourage saving, and taxpayers may find it challenging to allocate funds for significant expenses without the structure provided by withholding. Thus, withholding not only aligns with government interests but also caters to taxpayers' psychological preferences, making it a more favourable option than annual payments[19].

As noted earlier, the choice of whether to comply or not is informed by both extrinsic and intrinsic (tax morale) factors. One extrinsic factor mentioned earlier is the ease of choosing the right thing. Ensuring that tax return forms and online platforms are user-friendly and easy to navigate, with as clear language as is possible, will avoid non-compliance due to misunderstandings or due to taxpayers losing patience.

Choices can be framed as to appeal to loss aversion, for example by having invoices for unpaid taxes emphasise avoiding late payment fees (which would be a “loss”) instead of emphasising the amount the taxpayer is required to pay.

Determination. As noted earlier, ensuring that the desired behaviour requires as little determination as possible (i.e. that it is easy to pay a tax bill) is important to encourage the behaviour. Another way to encourage taxpayers to stick with compliant behaviours can be to leverage social pressure and show in what way the tax will benefit the community. Furthermore, it has been shown in literature that publicly committing to a goal (losing weight, quitting smoking) is a so-called “commitment device” as others come to expect the person who made the public commitment to stick to it, and may ask questions if they do not (“I thought you said you were going to lose weight?”). While it is unfeasible to expect taxpayers to publicly declare their intention not to commit tax evasion, a similar effect may be reached by making taxpayers feel that others are expecting them to comply. Peer pressure-oriented nudges informing non-compliant individuals that they are, in fact, the only (or one of a small number of) non-complier in their neighbourhood aim to create this feeling.

1.1.6. Stage 4: Intervention

The main objectives in this stage are to develop and test the intervention.

From Strategies to intervention ideas

After the most important strategies have been identified, the next step is to develop actionable intervention ideas. Ideally, this stage is a creative and collaborative effort, drawing on the earlier identified strategies to design innovative and effective interventions. A possible method to further operationalize the intended strategies is to utilize the Behaviour Change Technique taxonomy[20], that includes 93 distinct behavioural change techniques. These are organized into 16 groups based on their function, such as goals and planning or feedback and monitoring. Utilizing this taxonomy helps in identifying which specific techniques could be effective in influencing the identified behavioural drivers and formulating targeted interventions. The output of this step are concrete descriptions or schematic representations of intervention ideas, combining selected behaviour change techniques with context-specific insights to create a tangible, testable version of each strategy.

Usability, feasibility and pilot-testing

A good intervention should achieve measurable improvements of the targeted behaviour. However, a good intervention must also be feasible: Some interventions may have to be rejected even if they might have “worked”, because the overall effort and required resources to develop, test and implement them are too high.

The OECD framework outlines multiple considerations at this stage that should be discussed between the behavioural experts that will conduct the experiment and the involved policymakers.

[19] Thomas, K. D. (2019). The Modern Case for Withholding. *UC Davis L. Rev.*, *53*, 81. https://scholarship.law.unc.edu/cgi/viewcontent.cgi?article=1475&context=faculty_publications

[20] Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., ... & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Annals of behavioral medicine, 46(1), 81-95.

- How is success defined? Going back to the SMART objectives covered earlier in this module, it is important to make sure that there is an agreed-upon definition and that everyone agrees that expectations for success are realistic.
- What research design works best? In Module 3, we will describe a number of quantitative and qualitative methodologies to choose from. There is no “correct” answer as to which methodology is best, but rather this will come down to context, including issues such as how much funding is available.

  On the same note, a sample size must be agreed on before research commences. A bigger sample size will yield more accurate findings but will also result in greater costs (both in terms of time and money).
- How soon will users be involved in testing the design? Pilot studies have an important role, as they help identify any issues with the proposed methodology at an early stage[21]. This should be considered in the project planning.
- What are the risks? Are there any potential unintended outcomes? A risk assessment should be conducted together with all experts and stakeholders, with particular attention paid to negative spillovers. A particular risk in this domain is ‘guinea pig’ criticism, where participants may feel they are being experimented on without clear benefits. To counteract this, it is crucial to develop and implement a clear communication strategy.
- How will data be stored and kept safe? What data is eligible to use for research purposes and can be shared with commercial or academic research partners and/or for external publications? Non-disclosure agreements are particularly important, as well as agreement on the use of aggregate results if you are working with academic researchers who may wish to publish any novel findings. A data protection officer should be consulted to ensure privacy regulations will be followed.

To answer these types of questions it is crucial to assess the usability and feasibility of the intervention idea(s). Usability refers to how easily and effectively a user can interact with a system or intervention. It can be assessed through methods such as:

- **Think-Aloud Protocols**: Users vocalize their ideas, associations and immediate responses while interacting with the intervention, providing real-time insights into their experience and challenges.
- **System Usability Scale (SUS):** A quick, standardized questionnaire that gives a numerical measure of the usability of the system.
- **Focus Groups**: Groups of users discuss the intervention, providing feedback on their collective experience, which can reveal usability issues and areas for improvement.

Feasibility refers to whether an intervention can be effectively implemented and sustained in practice. It encompasses several key areas that can be assessed through various methods, including focus groups, project team meetings or small-scale experiments:

- **Implementation**: This considers whether the intervention can be successfully delivered to the intended participants within the existing infrastructure.
- **Practicality**: Examines if the intervention can be carried out with available means, resources, and circumstances, often requiring a cost analysis to ensure affordability.
- **Integration**: Assesses how well the intervention can be assimilated within current systems, often involving stakeholder meetings to evaluate compatibility with existing infrastructure.
- **Expansion**: Looks at the potential for the intervention to be scaled up to serve the full target population, ensuring the design is suitable for broader application.

[21] John, P. (2017). *Field experiments in political science and public policy: Practical lessons in design and delivery*. Taylor & Francis. https://opac.units.it/sebina/repository/catalogazione/immagini/PDF/Field%20experiments%20in%20political%20science%20and%20public%20policy.pdf

- **Limited Efficacy Testing**: Preliminary testing, such as manipulation checks, is conducted to determine if the intervention is likely to be effective in practice, thereby showing promise before extensive resources are committed.

These feasibility and usability assessments do not only help to decide which intervention is most promising, but also lead to refinements that enhance user engagement and the overall effectiveness of the intervention. Moreover, it helps to obtain a realistic picture on what is needed to actual implement and test the intervention, including the resources and departments that are needed for support.

As a final preparatory step before testing the intervention, a pilot test is highly recommended. Pilot testing involves a small-scale preliminary study with a limited sample size to test the experimental protocol and intervention. A successful pilot test includes verifying the feasibility of institutional, technical, and systemic aspects; identifying any unforeseen challenges that could affect timelines; determining potential indicators of the intervention's effect size; assessing the feasibility of the data analysis procedures; and evaluating whether the current intervention requires further refinement based on initial outcomes and feedback.

Testing prototype interventions. As we will cover in module 3, there are many ways of testing interventions, but in this section we will focus on field Randomised Controlled Trials (RCTs). In an RCT, participants from a target population are randomly allocated between one or several treatment groups, and a control group, and are then tracked over time in their natural environment. These types of studies are a powerful tool to test behavioural interventions. The random allocation of participant to treatment and control groups ensures that any differences in outcomes can be attributed to the intervention itself, rather than other factors. This method allows policymakers to observe the real-world effects of behavioural strategies within the natural environment of taxpayers, providing robust evidence on the effectiveness of such interventions. Module 3 includes a more comprehensive description of this method.

Regardless of the chosen research design, it is crucial that a “quitting threshold” is set before the intervention starts. This may include minimum effect size targets throughout a specified period of time, and the number of times the experiment will be repeated before the intervention is abandoned. Setting those criteria before implementation of the intervention is crucial to ensure high credibility of the results and to prevent research waste. If an intervention has failed to meet those pre-defined criteria, it should be strongly considered to discontinue the project. It may be tempting for practitioners to continue to pursue an ineffective intervention due to how much time has been spent developing it, but to do so would be to fall victim to the sunk cost fallacy.

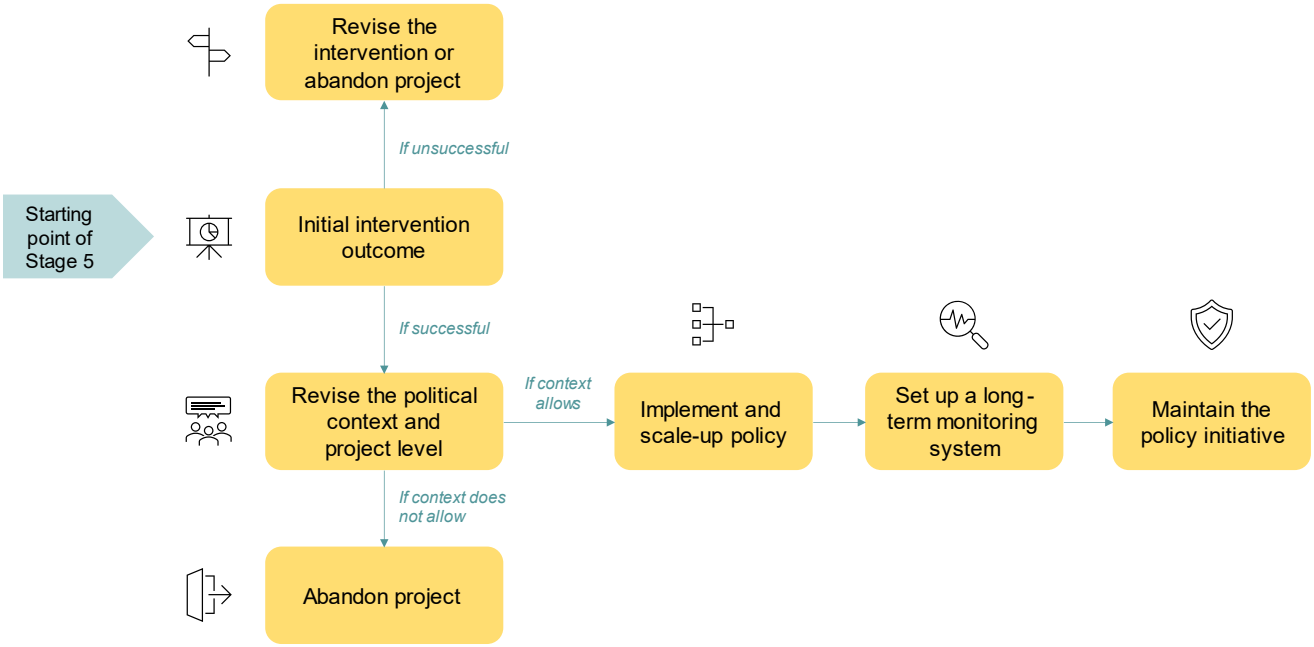
1.1.7. Stage 5: Change

In this final stage, the results and conclusions are shared with the public and with relevant stakeholders. If the intervention developed in the last stage worked, our next step is to consider how to scale up the intervention so that it can be applied to the general population.

If, on the other hand, the intervention was unsuccessful as judged by the pre-defined criteria, there are two options: Abandon the project and start over from scratch or iterate and work to improve the intervention based on the collected data. If choosing the latter option, practitioners must be careful to base any changes on the data collected or on existing literature, rather than make random changes again and again which could result in false positives.

Challenges with scaling up will be discussed.

Figure 6 Considerations in Stage 5



Revisiting the political context and project level. Earlier in this module we mentioned that political feasibility is an important consideration. Given the time it can take to design and test an intervention, it may be years into the future by the time this stage is reached. Thus, it is appropriate to go back and consider whether the political environment has changed in any considerable way since the project first began. Are the findings controversial, and are they convincing enough to motivate change given the political and legal landscape and current priorities?

Implementing and scaling behaviourally informed policies. Depending on the nature of the intervention, scaling up may require elected officials to pass new laws or regulations, or government authorities to issue new guidelines. Practitioners must think carefully about whether there are any reasons to believe the results may not be generalisable, in other words may not be the same once the intervention is scaled up. This is particularly important if lab experiments were used in Stage 4. Carefully analyse whether your data suggests any demographic group is adversely affected by your intervention, and if this is the case, how you might mitigate this issue (i.e. whether it is possible to exempt this group from the intervention).

Setting up a long-term monitoring system. Whether a lab experiment, an RCT or another research design is used in Stage 4, the research you have conducted will have been limited in time span. As part of implementing your findings, you must also implement built-in measures for *ex post* data collection to ensure the intervention is working as you had intended. This data collection must also cover spillover effects (both positive and negative).

Maintaining the policy initiative. Due to staff turnover, the employees involved with launching the intervention may not work at the relevant tax authority forever. To avoid the intervention being watered down or even discontinued by future employees who were not part of the development process and may not understand why the intervention is necessary, instructions must be given for the proper maintenance of the intervention. Those involved with developing and implementing the intervention need to consider what parts of the organisation (what roles/departments) should be involved in maintenance and think carefully about the information that future employees will need to understand and carry on with the intervention.

Dissemination of the results. As a rule, research conducted by policymakers and public authorities should be shared with the public. Beyond that, however, it may be useful to share the research findings with other authorities that may deal with similar or related behaviours. Sharing results with tax authorities in other countries can both help establish good reciprocal relationships where they may share their own future research with you, and it is also a way to receive external feedback on your research design and your findings. Ideally, findings should also be translated to other languages to maximise their reach.

Please note that even *negative* results – i.e., unsuccessful interventions – should be shared to avoid the intervention being tested again in the future (which would be a waste of time and money).

1.1.8. Ethical considerations during intervention development, testing and implementation

Behavioural interventions – nudges in particular - have in recent years been subject to ethical controversy and criticism. Critics allege, among other things, that BI, violates principles of transparency and informed consent of the citizenry. Furthermore, there are questions as to the long-term effects of behavioural interventions and any unexpected side effects that these interventions may have. Furthermore, critics have alleged that BI effectively interfere with individual autonomy, even in case of nudges that technically do not remove any choices from the individual. BI has also been criticised for an inappropriately strong focus on the individual, allowing policymakers to ignore underlying system-level problems causing behaviours (like non-compliance) in favour of attempting to manipulate behaviour with the use of behavioural sciences[22]. Related to this, there is a potential for behavioural interventions to reduce welfare in the case where a taxpayer is not able comply with the behaviour that the intervention is attempting to induce. We will now address these objections in turn and describe how these issues may be resolved or mitigated.

First, it is important to note that transparent nudges, in which participants are informed that they are being subject to a behavioural intervention and the purpose of that intervention, routinely perform just as well as hidden interventions. This may seem counterintuitive but may be explained by most nudges influencing people towards an action that they agree that they should do, even if they are not currently doing it (whether due to limited willpower, not enough time or because they haven't thought about it). In the context of tax compliance, most taxpayers do agree that they should be filing and paying their taxes and that they should do so on time and in a correct manner. Nudges therefore may be interpreted by the vast majority as a convenient help towards making the choice they either were or wish they were already making. Another reason why transparent nudges may work as good or better as hidden nudges is that some taxpayers may realise that the nudge is present even if it is hidden, and the lack of transparency may make them feel that they are being unjustly manipulated by government authorities, causing some of them to lash out and do the opposite of whatever the nudge intended for them to do as a way of asserting their independence. This allows for the opportunity for using disclaimers – either on the intervention itself when feasible (such as on a mailing) or on the tax agency's website – to mitigate any discussion on transparency.

Second, with respect to the long-term effectiveness, many behavioural interventions, particularly those stemming from lab experiments, have been implemented without a long-term follow-up. Throughout this handbook, we mainly rely on RCTs which, being conducted over a longer time period, have a greater potential of capturing adverse side effects or interventions that whose impact taper off over time. This handbook also includes steps to explore potential negative side effects and to set up long-term monitoring to ensure that they continue to work as intended.

Finally, policymakers should be aware of the limits of BI solutions and also investigate non-behavioural methods to increase tax compliance. This goes in particular for issues like poverty and corruption. Nevertheless, we disagree that behavioural issues cannot be described as "underlying" issues. If a user-unfriendly tax return design is driving non-compliance, then a behavioural approach will be able to address this.

As with any research that collects personal data, encryption and safe storage are necessary ethical precautions when conducting RCTs and other forms of research on BI that uses such data.

In module 3, we added ethical guidelines for each of the described research methods, to further promote ethical conduct during intervention development and testing.

[22] Chater, N., & Loewenstein, G. (2023). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, *46*, e147.

2. Module 2 – How to build on existing findings: a literature overview of behavioural insights into tax compliance

2.1. Introduction

This module offers an overview of recent literature on behavioural insights for tax compliance. By exploring existing research, new projects can leverage the knowledge and experiences from prior studies. This module helps to narrow down suitable target behaviours and to guide the choice of methodology, a topic we will cover to a greater extent in Module 3. The module contains the following sections:

- **Measuring tax compliance.** In examining the effectiveness of behavioural interventions in tax compliance, it is essential to employ clear and objective outcome measures. In this section, we present various compliance measures that we identified in the literature review. The measures are organised into four categories, providing a wide range of options to investigate compliance: understanding the tax process, attitude towards taxes, declaring taxes, and paying taxes.
- **Drivers of tax compliance.** This section includes studies that investigate the relationship between drivers and tax compliance. These studies explore how cognitive, emotional, and situational factors impact individuals' decisions and actions regarding their tax obligations. We provide an overview of all identified drivers, summarising the findings of studies that have explored the relationship between each driver and tax compliance. This overview serves as a reference for behavioural intervention design projects.
- **Experimental studies on behavioural interventions.** Experimental studies on behavioural interventions to improve tax compliance provide an important source of information on what mechanisms and intervention channels may be effective and feasible. This section includes an overview of all recent field trials on tax compliance, organised by their main intervention mechanism.
- **Benchmarking.** This section contains various effect sizes from successful randomised controlled trials that can be used as a point of reference for future studies. The overview includes intervention details, effectiveness, and additional taxes collected and helps to determine realistic objectives for behavioural interventions to improve tax compliance.
- **Narrative summary meta-analysis and systematic reviews**. This section contains an overview of selected literature reviews on tax compliance. These documents are summarised with a focus on how they can inform the development of behavioural interventions.

2.2. Literature review strategy

In December 2023, we conducted a literature search in Scopus, Web of Science, PsycINFO and Google Scholar, restricting it to studies from 2015 to December 2023. The search strategy was centred around two themes: ‘tax compliance’ and ‘behavioural insights’ (see appendix for the search strategy). In addition, we added the collection of studies on this topic from the Joint Research Centre of the European Commission and by the Belgian Service Public Fédéral Finances. For all studies of interest, we extracted relevant information using a standardised data extraction form, and this knowledge then formed the base for this chapter.

2.3. How to measure tax compliance

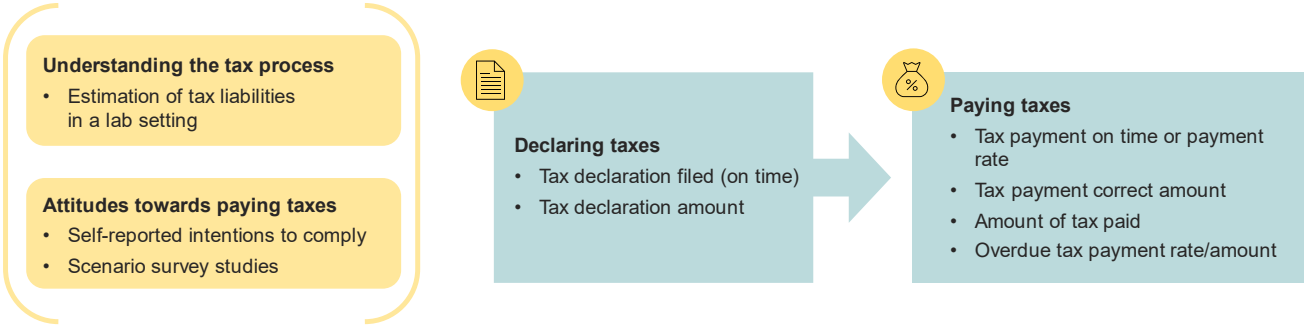
When examining the effectiveness of behavioural interventions, it is crucial to use clear and objective indicators of tax compliance. In that way, the targets for improvement become concrete and allow for quantitative analysis of the effectiveness of the intervention. Understanding the different indicators of tax compliance can also be helpful with making appropriate comparisons between different studies.

Importantly, there is not a universally optimal compliance indicator; it varies depending on the exact behaviours that are targeted. Furthermore, researchers tend to use multiple indicators to measure the impact of single interventions to demonstrate the robustness of their results. This so-called triangulation between different

indicators of compliance can demonstrate the validity of the results better than a conclusion drawn from a single measure. We elaborate on the importance of triangulation in Module 3.

We compiled a list of the measures of compliance that we found through the literature review and organised these measures into four categories: understanding the tax process, attitude towards taxes, declaring taxes, and paying taxes. 'Understanding the tax process' relates to taxpayers' actions to familiarise themselves with the intricacies of the tax system, gaining knowledge about filing requirements, deadlines, and overall procedures. 'Attitude towards taxes' refers to individual predisposition towards taxes, which may encompass people's feelings, opinions, and evaluations about taxation, including their willingness to comply with tax regulations, perceptions of fairness in the tax system, or overall views on the role of taxes in society. 'Declaring taxes' involves the formal submission of tax declarations, where taxpayers provide accurate and comprehensive information about their income, deductions, and other relevant financial details to tax authorities. 'Paying taxes', encompasses fulfilling the financial obligations outlined in the tax declaration by submitting the owed taxes within the stipulated timeframe, completing the taxpayers' journey through the tax process. This category also involves payment of overdue tax, in case of failure to pay on time. It is important to note that, although all outcome measures that we identified can be categorised in one of the four categories, they do not encompass every conceivable action and decision in the realm of tax compliance. For instance, in situations where taxpayers owe revenue to tax authorities, there payment information needs to be provided, enabling authorities to transfer the money.

Figure 7 Categories of measures of tax compliance



Understanding the tax process. Many intervention studies aim to improve compliance by targeting people's understanding of the tax process. However, virtually none of these studies directly measure if their understanding improves by the intervention. A possible reason may be the methodological challenge of designing and implementing a measure that accurately detects a meaningful change in understanding of how a certain tax process works in real life field studies. This is more feasible in a controlled laboratory setting, where researchers can control the presentation of information and implement controlled experiments to explore understanding of particular aspects of the tax system. For example, one study evaluated how accurate taxpayers were able to comprehend nonlinear incentive schemes within the US income tax system. To measure this, researchers measured how accurately participants were able to forecast the amount of due income tax for sixteen different income scenarios[23]. Based on these estimates it was possible to investigate the hypothesised structural overestimation of low-income filers and underestimation of high-income filers.

Attitude towards taxes. Attitude towards taxes is primarily measured with a self-reported intention to comply, or by using fictional scenarios where an opportunity for non-compliance (or compliance) is presented to participants. An example of such a scenario is a description of a small business owner who is preparing their income tax return. Subsequently, individuals were confronted with a decision to engage in tax evasion by deliberately inflating their automobile expense deduction through fraudulent overstatement of the business

[23] Rees-Jones, A., & Taubinsky, D. (2016). *Heuristic perceptions of the income tax: Evidence and implications for debiasing*. National Bureau of Economic Research. Retrieved from https://faculty.wharton.upenn.edu/wp-content/uploads/2015/08/Heuristic-Perceptions-of-the-Income-Tax-1.pdf

usage percentage[24]. A more general way of evaluating attitudes towards compliance is by including a question such as: "How justifiable do you think it is to evade taxes if an easy opportunity to do so presents itself?"[25]

Declaring taxes. Tax declaration outcomes are frequently assessed using binary measures, such as whether the taxpayer filed their declaration on time or at all. These binary indicators measure the proportion of taxpayers who declared taxes within specified deadlines and can be used to evaluate the effectiveness of behavioural interventions to improve filing compliance such as reminder letters[26]. Additionally, tax declaration can be measured in terms of the monetary amount declared, for example to estimate compliance on the intensive margin (i.e., the variations in specific components of taxpayers' declarations)[27].

Paying taxes. Various measures can be employed to assess tax payment compliance, and the choice often depends on the specific focus of the intervention. The binary measure of tax payment on time or the payment rate allows for the calculation of the proportion of taxpayers who make payments before a set deadline, irrespective of the accuracy of the payment amount[28]. Another binary measure assesses the correctness of the payment amount, calculating the share of taxpayers who pay the accurate amount before the deadline[29]. The 'amount of tax paid' quantifies the monetary value of the taxes paid, providing an indicator of financial compliance[30]. For interventions targeting overdue payments or debt repayment, both the rate and amount of overdue tax payments can be measured, offering insights into participants' compliance to repayment obligations[31,32].

One distinction between the measurements is that compliance can be measured on the extensive and intensive margin. Extensive margin measures concern binary outcomes, such as whether taxpayers filed their declaration or paid taxes on time, providing a broad assessment of participation. In contrast, intensive compliance measures focus on specific details of tax declarations, evaluating changes in the quantity or amount of reported income or deductions to gauge the depth and accuracy of compliance on a more detailed level. While extensive measures indicate overall engagement, intensive measures offer a finer-grained analysis of the adjustments made within taxpayers' declarations.

Finally, it is important to note that the measures of compliance summarised in this handbook are primarily from academic literature, where the authors are often constrained in terms of data access. When tax authorities run interventions directly, they may be able to construct more fine-grained measures of compliance, such as likelihood of repeated failure of filing (or paying) taxes.

2.4. Drivers of tax compliance

2.4.1. Introduction to drivers of compliance

Drivers refer to determinants that contribute to shaping individuals' decisions and actions related to tax compliance. These factors encompass a wide range of influences, including cognitive, emotional situational, external, and demographic elements. Often, multiple interacting drivers play a role in influencing taxpayers'

[24] Shafer, W. E., & Wang, Z. (2018). Machiavellianism, social norms, and taxpayer compliance. *Business Ethics: A European Review*, *27*(1), 42-55. Retrieved from https://www.researchgate.net/profile/William-Shafer/publication/319913133_Machiavellianism_social_norms_and_taxpayer_compliance/links/5c7a67b5299bf1268d31150f/Machiavellianism-social-norms-and-taxpayer-compliance

[25] Doerrenberg, P., & Schmitz, J. (2015). Tax compliance and information provision–A field experiment with small firms. *ZEW-Centre for European Economic Research Discussion Paper*, (15-028). Retrieved from https://www.econstor.eu/bitstream/10419/110751/1/dp9013.pdf

[26] Jamison, J. C., Mazar, N., & Sen, I. (2021). Applying behavioural insights to tax compliance: experimental evidence from Latvia. Retrieved from https://ore.exeter.ac.uk/repository/bitstream/handle/10871/121984/document.pdf?sequence=4

[27] Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2020). You've got mail: A randomized field experiment on tax evasion. *Management science*, *66*(7), 2801-2819. Retrieved from https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.2019.3390

[28] Chirico, M., Inman, R., Loeffler, C., MacDonald, J., & Sieg, H. (2019). Deterring property tax delinquency in Philadelphia: An experimental evaluation of nudge strategies. *National Tax Journal*, *72*(3), 479-506. Retrieved from https://www.nber.org/system/files/working_papers/w23243/w23243.pdf

[29] Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2020). You've got mail: A randomized field experiment on tax evasion. *Management science*, *66*(7), 2801-2819. Retrieved from https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.2019.3390

[30] Chirico, M., Inman, R., Loeffler, C., MacDonald, J., & Sieg, H. (2019). Deterring property tax delinquency in Philadelphia: An experimental evaluation of nudge strategies. *National Tax Journal*, *72*(3), 479-506. Retrieved from https://www.nber.org/system/files/working_papers/w23243/w23243.pdf

[31] Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behaviouralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of public economics*, *148*, 14-31. Retrieved from https://www.nber.org/system/files/working_papers/w20007/w20007.pdf

[32] Perez-Truglia, R., & Troiano, U. (2018). Shaming tax delinquents. *Journal of Public Economics*, *167*, 120-137. Retrieved from https://www.nber.org/system/files/working_papers/w21264/w21264.pdf

compliance behaviour, either promoting or hindering compliance. As covered in Module 1, Identifying the most important set of drivers for a specific tax compliance issue is a crucial step at the beginning of an intervention development process.

This section supports this step by providing an overview of drivers that have been related to tax compliance in previous studies. Although each tax compliance issue is unique and requires its own exploration, this overview helps to understand the range of potential influencing factors and to focus on what type of variables need to be explored. The drivers discussed in this section were identified through the literature review. In cases where drivers are disputed, with some papers showing them as significant and others as insignificant, we have made qualified assessments based on the number of papers on each side and the quality of the studies.

We have divided the drivers for individual taxpayer compliance into two categories: modifiable and non-modifiable. The non-modifiable drivers are mainly demographic variables that cannot be changed with behavioural interventions, but that need to be considered to establish baselines for compliance. For example, if low-income status predicts tax payment non-compliance on income tax in a certain country, this variable needs to be considered when evaluating the effect of compliance interventions over time or between groups. Furthermore, non-modifiable drivers are useful to identify specific subgroups within the taxpayer population that may be more susceptible to compliance challenges. By recognising these inherent characteristics, policymakers can design tailored behavioural interventions that address the unique needs and circumstances of these subgroups. Modifiable drivers, on the other hand, are potential targets that policymakers and authorities may aim to change through behavioural intervention. In addition to drivers for individual taxpayers, we also added a list of drivers that specifically relate to owners of small businesses, as these were recurrently highlighted in various studies.

It is important to note that not all drivers may have a causal relationship with compliance; as we note throughout the text, some drivers are likely merely highly correlated with other drivers that, in turn, have a causal relationship with tax compliance.

2.4.2. Individual taxpayers: Non-Modifiable drivers

Table 1 provides an overview of all identified non-modifiable drivers for individual taxpayers. Annex 1 contains a more detailed information about the studies from which the drivers were extracted, including the references, study type, population and direction of the effects.

Table 1 An overview of all identified non-modifiable drivers for individual taxpayers.

Non-modifiable drivers – Individual taxpayers		
Driver	**Number of studies**	**Relationship with compliance**
Age	7	Most studies report a positive relationship between age and compliance, indicating that older people are more compliant. One study reports that age has an inverted relationship with compliance, suggesting that compliance peaks in middle-age and then drops (i.e., an inverted-U relationship).
Gender	7	Most studies (five out of seven) report women are more likely to be tax compliant than men.
Education	3	All studies indicate a positive relationship between educational attainment and tax compliance.
Income	3	All three studies show a positive relationship between income and compliance.
Employment	1	One study shows a link between being employed (as opposed to unemployed) and self-reported tax compliance intentions.
Parenthood	1	One study indicates that having children is linked to higher tax compliance.
Homeownership	1	One study finds a positive relationship between homeownership and tax compliance.

Level of corruption (country level)	1	One study shows a negative relationship between corruption and compliance
Level of income inequality (country level)	1	One study finds a negative relationship between income inequality and compliance.
Level of public expenditure (country level)	1	One study indicates that a positive relationship exists between tax compliance and public expenditure, and more broadly between tax compliance and taxation as a % of GDP.
Strength of legal rights (country level)	1	One study indicates that a positive relationship exists between tax compliance and the strength of individual legal rights in a country.

Age is positively correlated with tax compliance. This may be attributed to several factors. Firstly, age is also positively correlated with income up until retirement age, potentially leading to increased compliance (as described below, income is positively correlated with compliance). Moreover, older taxpayers have had more experience with filing taxes. Given that some non-compliance is due to mistakes and lack of knowledge of the system, it makes sense that this would be less common among experienced (older) taxpayers. Additionally, personal norms - that been related to compliance[33] - may explain the differences between younger and older taxpayers, with older generations potentially raised with norms placing a greater relative emphasis on social responsibility over individualism and personal achievements. If personal norms are the main channel through which age influence compliance, it may not be the case that the younger generation today, who are raised with different norms, will automatically become more compliant as they age.

Women tend to exhibit higher compliance rates compared to men, despite earning less on average. Few papers discuss why this may be the case, but it could be attributed to social conditioning, which may discourage women from evading taxes.

Married couples are *not* less likely to offend than unmarried individuals, but **parenthood** is associated with lower likelihood of tax offending. This could be due to parents feeling that penalties they may suffer would also hurt their children. Parenthood may also reduce short-term orientation[34], contributing to higher compliance rates.

Higher education is positively linked with self-reported tax compliance intentions. This could be because higher education is associated with higher income, and a better understanding of the tax system, making it easier to comply. It is also conceivable that those with higher education feel a greater social obligation to pay taxes, as they would in most Western countries have received their education for free (or highly subsidised).

Income and employment status also influence tax compliance. Employed taxpayers and taxpayers with higher incomes are more likely to comply, while debt to tax authorities is negatively correlated with timely payments. Tax non-compliance is in some cases due to liquidity constraint, an issue that would not affect those employed and with higher incomes to the same extent as people who are unemployed. Also, while those with higher incomes are obligated to pay more taxes, low-income taxpayers may suffer a greater loss of utility from paying taxes (due to diminishing marginal returns).

Homeownership is linked to stronger intentions to comply, potentially due to homeowners in general having higher incomes, as well as having an asset (their home) that can be seized in the event of non-payment.

State level expenditure. Research suggests a positive link between state level expenditure and tax compliance, which may at first seem contradictory as a higher state level expenditure would naturally be associated with higher taxes that fewer taxpayers would want to or be able to afford to pay. However, higher

[33] McKerchar, M., Bloomquist, K., & Pope, J. (2013). Indicators of tax morale: an exploratory study. *eJTR*, *11*, 5. Retrieved from https://www.unsw.edu.au/business/sites/default/files/documents/eJTR_Vol_11_No1_2013.pdf#page=5
[34] Li, Y. J., Haws, K. L., & Griskevicius, V. (2019). Parenting motivation and consumer decision-making. *Journal of Consumer Research*, *45*(5), 1117-1137. https://carlsonschool.umn.edu/sites/carlsonschool.umn.edu/files/2018-10/parenting_jcr_2018_1.pdf

state level expenditure is also associated with better public services and safety nets, which may help increase tax morale.

Income inequality is associated with less compliance, likely due to more individuals being unable to pay their taxes for financial reasons, as well as less equal countries lacking the social cohesion and social norm of solidarity that may be found in their more equal counterparts.

Corruption. Countries that struggle with corruption tend to also struggle to encourage tax compliance, possibly due to taxpayers not wishing to contribute to public institutions they perceive as corrupt, or because they worry that the money they pay will be misappropriated.

Strength of legal rights. Just as corruption is linked to lower compliance, strong legal rights (typically associated with transparent and democratic nations) are linked to higher compliance.

2.4.3. Discussion and implications

We classify these drivers as non-modifiable since they cannot be altered by behavioural interventions. It is possible for governments to implement policies that may impact the levels of homeownership, education, income and employment in a country or region, as well as the country-level variables such as the level of public expenditure. While those variables have been widely investigated in the economic literature, they are not directly related to BI and therefore beyond the scope of this handbook.

The non-modifiable drivers, do, however, create a baseline for compliance. A country with very unfavourable demographics may perhaps never reach the levels of compliance seen in countries with favourable demographics, even with the use of BI.

The other reason why it is important to understand non-modifiable drivers is because they may shift over long periods of time. This may cause a problem when monitoring and evaluating the persistence of behavioural interventions: It may appear to practitioners as if an intervention that initially caused tax compliance to improve has stopped working if compliance falls back to its previous levels, when the reason may be that demographics have changed since the intervention was first introduced.

It is important to note that many studies on non-modifiable drivers are establishing a link between the self-reported intention to comply or self-reported compliance. It is conceivable that some of these factors may influence what taxpayers report that they intend to do, rather than what they do in the end. Caution is thus advised.

2.4.4. Individual taxpayers: Modifiable drivers

Table 2 contains an overview of all identified modifiable drivers for individual taxpayers. Annex 1 contains a more detailed information about the studies from which the drivers were extracted, including the references, study type, population and direction of the effects.

Table 2. An overview of all identified modifiable drivers for individual taxpayers.

Driver	Number of studies	Relationship with compliance
Trust	2	Both studies indicate a positive relationship between trust and compliance.
Subjective norms	1	One study in our review found a link between subjective norms and compliance.
Personal norms	3	Three studies have found that personal norms, such as believing that paying taxes is a moral imperative or that wealth carries with it responsibilities, are correlated with tax compliance.
Injunctive norms	1	One study has found a positive relationship between injunctive norms and tax compliance.
Higher detection rate	2	Real or perceived higher detection rates have a positive relationship with tax compliance according to two studies.

Higher punishment	1	One study has found that higher punishments for non-compliance are linked to greater tax compliance.
Machiavellian personality	1	One study has linked Machiavellian personality traits to reduced compliance.
Views on immigration and globalism	2	One study has found a negative relationship between individual levels of globalism and tax compliance, whereas another found individuals with negative attitudes towards immigration tend to be less likely to comply.
Tax knowledge	1	One study found taxpayers with a greater understanding of the tax system are more likely to comply.
Perception of fairness	1	One study found taxpayers who perceive the system to be fair are more likely to comply.
Life satisfaction	1	One study has found a link between life satisfaction and individual tax compliance.

Institutional trust, alongside **general trust**, significantly correlates with tax compliance. Taxpayers who perceive (whether rightly or not) governmental institutions as being ineffective or corrupt are more inclined to evade taxes. As we saw in the previous section, corruption is linked to lower levels of tax compliance. Individuals who lack institutional trust may feel sense of disillusionment with the system which may justify a belief that they are entitled to keep as much of their money as they can rather than paying taxes.

Norms (personal, subjective and injunctive) and beliefs also influence the choices that are part of complying. Individuals who view wealth as carrying moral responsibilities are more likely to comply with tax obligations, reflecting a collectivist mindset. Conversely, negative attitudes towards taxes often accompany non-compliance, highlighting the importance of dispelling misconceptions about tax utilisation and fostering a sense of civic duty.

Related to this, **Attachment to one's country** has emerged as another determinant. Research suggests that individuals with higher individual levels of globalism are more prone to tax evasion. This may be due to a weaker attachment to one’s country creating a weaker sense of obligation to contribute to the public services of said country. Conversely, **negative sentiments towards immigration**, which are often intertwined with lower institutional trust, can fuel non-compliance, signalling the complex interplay between societal attitudes and tax behaviour.

Personality traits and personal knowledge are also important drivers. **Machiavellian personalities**, which are characterised by moral indifference, tend to exhibit lower tax compliance intentions, reflecting the intrinsic link between morality and tax adherence. Personal knowledge of the tax system also has a positive impact on compliance, likely due to more knowledgeable taxpayers making fewer mistakes when filing. Life satisfaction also has a positive relationship with compliance, possibly due to people positive with their lives not feeling the need to evade taxes or feeling that they have more to lose from doing so because they are in a good place in their lives.

Finally, the spectre of **punishment and detection** plays a significant role. A higher risk of detection is strongly associated with greater compliance, underscoring the deterrent effect of enforcement measures. Likewise, the **severity of punishment** for non-compliance has also been shown to influence tax behaviour. These two drivers of compliance have been extensively researched in standard economic literature and are included in the seminal model of tax compliance by Alligham and Sandmo[35].

2.4.5. Self-employed taxpayers: Drivers

Table 3 contains an overview of all identified drivers for self-employed taxpayers. Annex 1 contains a more detailed information about the studies from which the drivers were extracted, including the references, study type, population, and direction of the effects.

[35] Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of public economics*, *1*(3-4), 323-338.

Table 3. An overview of all identified drivers for self-employed taxpayers.

Driver	Number of studies	Relationship with compliance
Trust	2	Two studies have found that trust drives tax compliance among self-employed, as it does among those who are employed.
Positive experience with tax authorities	1	One study has found that a positive experience with tax authorities and feeling treated respectfully by auditors is linked to greater tax compliance among self-employed, whereas those self-employed who experience a feeling of self-blame after contact with tax authorities are less likely to comply.
Financial scarcity	1	Self-employed who are struggling financially are less likely to comply according to one study.
Perception of fairness	1	Self-employed who perceive the fiscal system to be fair are more likely to comply according to one study
Short-term orientation	1	One study has found self-employed individuals with greater short-term orientation are less likely to comply.
Mental segregation	1	Self-employed who mentally segregate the money they have to pay towards taxes from their overall revenue are more likely to have a positive attitude towards taxation according to one study.
Lower conscientiousness	1	Self-employed who exhibit lower conscientiousness are less likely to comply.
Attitude to taxes	1	Self-employed who display a negative attitude to taxes are also less likely to be tax compliant.
Personal norms	2	Two studies have shown that self-employed whose personal norms support tax compliance and who feel that paying taxes is a moral obligation are more likely to comply.
Subjective norms	1	One study has indicated that subjective norms play a role in improving tax compliance among self-employed.

Tax compliance among the self-employed, is greatly affected by norms and by the level of perceived institutional trust. Notably, self-employed who mentally segregate the taxes they owe from their overall revenue are more likely to comply, most likely due to not suffering the same loss of utility from paying as they never considered the money, they pay in taxes to be theirs in the first place. Conscientiousness, the trait of being diligent and careful, also impacts compliance among self-employed. While we cannot rule out that conscientiousness may also play a role for other taxpayers, generally, filing taxes as a self-employed person is more difficult and requires greater attention to detail. Financial scarcity may also play a more prominent role for tax compliance among the self-employed, as their taxes are not automatically deducted from their paycheques by their employers.

Having a positive experience with the tax authorities is important for compliance, and, like individual taxpayers, so is feeling that the tax system is fair. Finally, short-term orientation, meaning a tendency to choose smaller-sooner over larger-later rewards, is associated with non-compliance.

2.5. Behavioural interventions: The options.

2.5.1. Introduction

This section includes all experimental studies that tested one or more behavioural interventions in a real-life setting. The interventions are thematically organised by the type of behavioural intervention (i.e., intervention mechanism category). In Appendix A, we included a table, where the studies have been classified according to these categories.

While these are categories of singular mechanisms, it is crucial to note that they are often combined together to yield the largest effects. This was done in some of those studies and is sensible from a practical point of view, as it is likely to optimise the intervention effect for a specific tax compliance issue. However, the drawback of such an approach is the increased difficulty of understanding which mechanism had what effect. Behind a successful intervention with multiple mechanisms, it could be hidden that one of the mechanisms had a high positive effect, while the other backfired. Therefore, next to testing combined mechanisms, some studies test the separate mechanisms as well in other experimental conditions. As an example, Hernandez et al. (2017) tested the effects of commission/omission framing and of deterrence separately, but also interacted the two in one letter by including the following message: “So far, we have thought of your payment delay to be accidental. However, if you disregard this notice, we will consider it an intentional choice of yours and we will treat you as a dishonest taxpayer. As part of the execution procedures, we can, for example, block your bank account, salary, and, in addition, you will have to cover all execution expenses that arise.” By analysing the differences between these conditions, it is possible to evaluate the individual effects as well as the potential added value of combining the two mechanisms.

Furthermore, interventions may generate spillover effects as well, which are indirect or unintended impacts. For instance, because taxpayers are embedded in social networks, Eerola et al., (2019) looked at how sending letters with information on a general increase in enforcement intensity of landlord taxes in Finland, impacted compliance of the household members of the targeted taxpayers, as well as of the people in their neighbourhood.

2.5.2. Intervention mechanism versus intervention channel

Intervention mechanisms pertain to the psychological or behavioural principles underlying the approach, such as framing or deterrence, while intervention channels involve the specific means or platforms through which the intervention is delivered, such as letters, emails, or social media. It is important to make a clear distinction between these two aspects, as the effectiveness of an intervention may vary depending on both the chosen mechanism and channel. Therefore, when an intervention proves ineffective, it is important to check whether the lack of impact is attributed to the chosen channel rather than the underlying mechanism. In such cases, testing how the intervention's delivery medium, such as letters or emails, was perceived by taxpayers (for example by using interviews or focus groups) could uncover insights into the potential role of the communication channel.

2.5.3. Overview of behavioural interventions

Table 4 contains all intervention mechanisms of field RCTs that were identified in the literature review, categorised according to the ABCD framework. Below, we summarised the mechanisms for each of these categories.

Table 4. Categories of intervention mechanisms according to ABCD framework[36]

Attention	Belief formation	Choice
Simplification	Information provision: deterrence	Social norms: descriptive or injunctive norms
Reminders	Information provision: non-deterrence	Social norms: minority norms
In-Person Engagement	Commission/omission framing	Shaming
	Positive incentives	Value of public goods
		Simplification

Simplification. Because of the limited ability of the human mind to process large amounts of information, we sometimes experience cognitive overload. This can be a reason for non-compliance, given the relative complex

[36] While interventions can also work through the Determination mechanism (the ‘D’ in ABCD), none of the studies that we reviewed utilised this mechanism.

task of declaring and paying taxes. For instance, in a study using US data from 1979 to 1990, people forego as much as $600[37] on average, just to avoid the additional procedure of specifying deductions[38],[39].

A frequently used simplification intervention is to rewrite the reminder letter to improve understanding and leave out unnecessary and complicating information. Below, we have included an example such a letter:

- The letter begins with a brief description of the purpose of the letter, immediately followed by a call-to-action (i.e., what must a taxpayer do) and a deadline: “Please pay your income tax by June 3, 2016.”
- The letter minimises the use of formal legal language to improve overall understanding.
- The letter specifies concrete next steps regarding what to do and how to do it. In a separate table, the letter includes the interest due on each day until the deadline so that taxpayers do not need to calculate it themselves and have additional incentive to pay the liability before the deadline.

Another way to increase compliance is to simplify the taxpayers’ task. For instance, a study in the United States, improved this by attaching a tax form and return envelope to an information letter, making it easier to file taxes physically[40].

We reviewed ten studies that provided different forms of interventions using this mechanism, out of which nine showed a positive and significant treatment effect. In one case, the positive effect was only sustained over a short period of time.

Reminders. As taxpayers may fail to comply due to simple forgetting, this mechanism entails sending reminders at appropriate times. While many field studies have included a simple reminder letter as a control group, some also explicitly studied the effects of the frequency of reminders. For example, one study examined the effects of the frequency of SMS reminders on compliance in the context of property tax in China, sending them either only once, once a week for 4 weeks, or twice a week for 4 weeks[41].

It is also important to note that sending any letter regarding the taxes, as is the case in many of the studies in this review, is a form of reminder. This is also the reason that a ‘simple reminder’ letter it is often used as a control group. In this way, any effect can be attributed to the intended mechanism instead of only a reminder-effect of the letter.

Of the six studies reviewed with interventions that used this mechanism, five showed it to be effective at improving compliance. In the case of tax debts, one study showed that reminders are effective at improving payment, but only when the size of the tax debt is low.

In-Person Engagement. Only two studies included an in-person component in communication with taxpayers[42]. In both studies representatives from the tax authority contacted taxpayers who missed the due date by phone and encouraged them to commit to paying taxes. Also, taxpayers were offered instalment plans or a physical meeting at the tax authority office to solve any disputes[43]. One study found a positive effect of the intervention on compliance, whereas the other study only found a positive effect on the intention to comply. In this latter study, subsequent analysis showed no increased payment rates compared to a control group.

Information provision: deterrence. According to standard economic theory, people respond to penalty amount and probability of being caught in case they do not comply. While empirical evidence points to these factors as drivers of compliance, people may not know about them when filing taxes, due to factors such as

[37] Inflation-adjusted amount, expressed in 2014 dollars.
[38] Hernandez, M., Jamison, J., Korczyc, E., Mazar, N., & Sormani, R. (2017). Applying behavioral insights to improve tax collection. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/27528/116046-WP-Tax-Collection-PUBLIC.pdf;sequence=1
[39] Benzarti, Y. (2015). How taxing is tax filing? Leaving money on the table because of hassle costs. Retrieved from https://www.austaxpolicy.com/wp-content/uploads/2016/09/Benzarti.pdf
[40] Meiselman, B. S. (2018). Ghostbusting in Detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics*, *158*, 180-193. Retrieved from https://www.isid.ac.in/~epu/acegd2017/papers/BenShuchMeiselman.pdf
[41] Antinyan, A., Asatryan, Z., Dai, Z., & Wang, K. (2021). Does the frequency of reminders matter for their effectiveness? A randomized controlled trial. *Journal of Economic Behavior & Organization*, *191*, 752-764. https://www.econstor.eu/bitstream/10419/250341/1/E2021-17.pdf
[42] Gemmell, N., & Ratto, M. (2018). The effects of penalty information on tax compliance: Evidence from a New Zealand field experiment. *National Tax Journal*, *71*(3), 547-588. Retrieved from https://ir.wgtn.ac.nz/bitstream/handle/123456789/20293/Working%20Paper.pdf?sequence=1&isAllowed=y
[43] Mogollon, M., Ortega, D., & Scartascini, C. (2021). Who’s calling? The effect of phone calls and personal interaction on tax compliance. *International Tax and Public Finance*, 1-27. Retrieved from https://link.springer.com/article/10.1007/s10797-021-09655-4

lack of information or a limited understanding of how the tax system works. Therefore, behavioural interventions have been developed to increase the salience of the penalty amount or probability of being caught.

The typical way in field experiments is to state the potential penalty in a message, such as the amount of interest charges given the owed debt[44]. The penalty stated does not have to be only monetary – for instance: "As part of the execution procedures, we can, for example, block your bank account, salary, and, in addition, you will have to cover all execution expenses that arise[45]." Deterrence information is not only provided through letters. An example of a different channel for this intervention mechanism is by phoning taxpayers who are in debt to discuss the consequences of non- compliance[46].

In total our review included Seventeen studies providing information meant to deter from non-compliance. Out of these, 13 studies suggest the mechanism is effective, with one of the remaining four noting the mechanism was effective but only for small taxpayers, and another stating it was effective for taxpayers with small debts (the study looked at collection of tax debt owed).

Information provision: non-deterrence. This type of intervention mechanism entails presenting helpful information on how to comply, instead of focusing on punishment or incentives. For instance, by providing step by step information to a specific group of taxpayers who do not yet have experience with the system. This was done in Australia for first-time taxpayers who missed the due date[47]. The tax authority sent "a newly designed non-deterrence letter ('welcome letter') to encourage lodgement of a taxpayer's first income tax return. The letter also promoted the use of a 'free and easy' online tool for filing taxes. The welcome letter sought to create a 'positive and supportive first experience' with the tax authority.

Our literature review included six studies with interventions that used this mechanism. Three of these found it to be effective, whereas three did not (relative to a reminder or a standard behavioural letter).

Commission/omission framing. This mechanism concerns framing non-compliance of as an intentional choice rather than omission, for example through forgetting. In general, individuals are often less comfortable with unethical behaviour when it is described as being an intentional action rather than a failure to take action[48,49]. An example of this mechanism is the following message that was included in a letter to individuals who were overpaid by a government benefit and did not respond to a prior message: "Previously, we treated your lack of response as an oversight. Now, if you do not call [phone number], we will treat it as an active choice."[50]

Our review included four studies that investigated the efficacy of this type of framing, and all four found it to be effective. One study indicated the framing works particularly well when combined with deterrence (see below).

Positive incentives. This mechanism involves encouraging compliance through ways other than punishment, such as through personal rewards, or rewards for the community. This approach has been implemented most often through lotteries, leveraging the behavioural insight that people tend to overestimate low-probability outcomes. For instance, one study aimed to increase timely payment of by giving taxpayers the opportunity to enter a lottery, with either financial (1000CHF) or non-financial (a wellness weekend) prizes, in exchange for a formal promise to pay taxes on time[51].

---

[44] Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of public economics*, *148*, 14-31. Retrieved from https://www.nber.org/system/files/working_papers/w20007/w20007.pdf

[45] Hernandez, M., Jamison, J., Korczyc, E., Mazar, N., & Sormani, R. (2017). Applying behavioral insights to improve tax collection. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/27528/116046-WP-Tax-Collection-PUBLIC.pdf;sequence=1

[46] Gemmell, N., & Ratto, M. (2018). The effects of penalty information on tax compliance: Evidence from a New Zealand field experiment. *National Tax Journal*, *71*(3), 547-588. Retrieved from https://ir.wgtn.ac.nz/bitstream/handle/123456789/20293/Working%20Paper.pdf?sequence=1&isAllowed=y

[47] Dong, S. X., & Sinning, M. (2022). Trying to make a good first impression: A natural field experiment to engage new entrants to the tax system. *Journal of behavioral and experimental economics*, *100*, 101900. Retrieved from https://www.econstor.eu/bitstream/10419/233884/1/1757182004.pdf

[48] Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of behavioral decision making*, *3*(4), 263-277. Retrieved from https://www.sas.upenn.edu/~baron/papers/vac1990.pdf

[49] Jamison, J. C., Mazar, N., & Sen, I. (2021). Applying behavioral insights to tax compliance: experimental evidence from Latvia. Retrieved from https://ore.exeter.ac.uk/repository/bitstream/handle/10871/121984/document.pdf?sequence=4

[50] Hallsworth, M., List, J. A., Metcalfe, R. D., Rotaru, K., & Vlaev, I. (2015). The making of homo honoratus: From omission to commission. *Journal of Consumer Psychology*. Retrieved from https://myscp.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/jcpy.1392

[51] Koessler, A. K., Torgler, B., Feld, L. P., & Frey, B. S. (2019). Commitment to pay taxes: Results from field and laboratory experiments. *European Economic Review*, *115*, 78-98. Retrieved from https://www.econstor.eu/bitstream/10419/177848/1/1019357142.pdf

Out of the three studies identified by our review, only one suggested it to be effective, and even in that study only non-financial rewards were able to improve compliance.

Social norms: descriptive or injunctive norms. This mechanism relies on people's proclivity to do what the others are doing. This is particularly important when people overestimate the prevalence of undesirable behaviours (such as tax non-compliance) and underestimate the prevalence of desirable behaviours (such as compliance), which is often the case[52]. The theory of normative conduct[53] identifies 'descriptive' norms, which communicate the behaviour of others, and 'injunctive' norms, which communicate the opinions of others. Put differently, descriptive norms say what others do; injunctive norms say what others believe: including what behaviours they approve of."[54]

For instance, a recent study included both of these mechanisms inside the letters that were sent to taxpayers who were late with paying their declared income tax. For a descriptive norm, it was stated that "Nine out of ten people pay their tax on time." and for an injunctive norm, it was stated that "The great majority of people agree that everyone in the UK should pay their tax on time". The latter norm was empirically determined, through a survey they ran beforehand, where they found that indeed, 88% of taxpayers agreed with that statement[55]. Social norms can also be communicated in a dynamic way, for example by highlighting the descriptive social norm that an increasing number of taxpayers file their tax by the deadline each year[56].

Out of the twelve studies that we reviewed, only three found this mechanism to effective at altering compliance behaviour, with two studies even finding that an intervention utilising this mechanism reduced compliance. Out of the three that did find a positive effect, one noted that the effect was not persistent, while another noted that while the intervention outperformed the traditional letter that taxpayers would otherwise have received, it did not outperform a behaviourally informed control letter.

Social norms: minority norms. This is a way to frame the social norm information, emphasising that the undesirable behaviour of the respondent is uncommon. On top of further emphasising that the targeted individual is not doing what the majority of others is doing, it has been hypothesised to evoke feelings of being 'on the radar' of the tax authority to a higher extent than merely stating the descriptive norms[57]. An example of a minority norm message is: "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."[58]

Two out of five studies that were part of our review show a positive impact from this type of mechanism on compliance.

Shaming. Shaming interventions include a type of penalty, where taxpayers' non-compliance behaviour is exposed publicly. We found this mechanism tested in one study[59], where a list of non-compliant taxpayers was published online. Although the researchers found that the intervention did increase debt repayment, shaming is generally considered controversial, not least from an ethical viewpoint, and this is the only study in our review that incorporated this mechanism.

Value of public goods. This mechanism involves making it clear that the taxes paid by individuals are used to finance public goods and services, in hope to increase the motivation to pay taxes (Antinyan & Asatryan, 2019).

---

[52] Schultz, P. W. (2022). Secret agents of influence: leveraging social norms for good. *Current Directions in Psychological Science*, *31*(5), 443-450. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/09637214221109572

[53] Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in experimental social psychology* (Vol. 24, pp. 201-234). Academic Press. Retrieved from http://www.influenceatwork.com/wp-content/uploads/2015/05/A-Focus-Theory-of-Normative-Conduct.pdf

[54] Hernandez, M., Jamison, J., Korczyc, E., Mazar, N., & Sormani, R. (2017). Applying behavioral insights to improve tax collection. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/27528/116046-WP-Tax-Collection-PUBLIC.pdf;sequence=1

[55] Hernandez, M., Jamison, J., Korczyc, E., Mazar, N., & Sormani, R. (2017). Applying behavioral insights to improve tax collection. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/27528/116046-WP-Tax-Collection-PUBLIC.pdf;sequence=1

[56] Jamison, J. C., Mazar, N., & Sen, I. (2021). Applying behavioral insights to tax compliance: experimental evidence from Latvia. Retrieved from https://ore.exeter.ac.uk/repository/bitstream/handle/10871/121984/document.pdf?sequence=4

[57] Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, *57*(4), 904-954. Retrieved from https://www.nber.org/system/files/working_papers/w24799/w24799.pdf

[58] Hernandez, M., Jamison, J., Korczyc, E., Mazar, N., & Sormani, R. (2017). Applying behavioral insights to improve tax collection. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/27528/116046-WP-Tax-Collection-PUBLIC.pdf;sequence=1

[59] Perez-Truglia, R., & Troiano, U. (2018). Shaming tax delinquents. *Journal of Public Economics*, *167*, 120-137. Retrieved from https://www.nber.org/system/files/working_papers/w21264/w21264.pdf

For example: “Paying tax means we all gain from vital public services like the NHS [National Health Service], roads, and schools.”[60]

Of the eight studies that were part of our review and that included this mechanism, only two showed a positive effect. Three studies that showed the mechanism to be ineffective compared it either to a standard, behaviourally informed letter or to a reminder and found it to be less or not significantly more effective. On the other hand, for the two studies that did show an effect, the control groups either received no letter or a baseline letter.

2.6. Benchmarking

This section contains an overview of studies that have been selected for benchmarking purposes. Based on the findings of our literature review, we identified exemplary studies and included for each study the effect size of the most successful experimental treatment condition. In the table, we provide benchmarks for different Contexts and distinguish between types of taxpayers (individual vs self-employed), types of tax (income vs other), and stage of declaration (filing vs payment vs recovery)[61].

The table also includes a short description of the behavioural intervention, the effectiveness (in terms of percentage points changes in a given compliance indicator) in comparison to control groups, as well as an estimate of additional amount collected revenue given the cost of the intervention (if included by the authors)[62]. The effect sizes are displayed in percentage points. This is an absolute change in percentage between an intervention and a control condition.

In addition to the example studies in the benchmarking table, a recent meta-analysis also provides an important point of reference for new behavioural interventions[63]. The study includes a comprehensive analysis of the effectiveness of nudging treatments from 45 randomised controlled trials on tax compliance, with a total of 1000 treatment effects. Included studies had to be RCTs with interventions that targeted taxpayers (i.e., individuals or firms rather than, clusters or regions). Furthermore, they had to contain a nudging intervention and a dependent variable that measured tax payment behaviour. The researchers concluded that non-deterrence interventions (such as those evoking social norms, or the value of public goods) were not effective at increasing compliance both at the intensive and extensive margin compared to control groups that received neutral communication from the tax authority. On the other hand, the interventions emphasising traditional determinants of compliance such as audit probabilities and penalty rates, increased the probability to comply by 1.5 – 2.5 p.p. compared to non-deterrence nudges, which was equivalent to 5.9 - 12.6 p.p. increases on the intensive margin.

Table 5. Example studies for benchmarking

Context	First author (year)	Intervention description	Effectiveness	Cost-effectiveness	Measure of compliance
		Basic information about what the intervention entailed	*Increased compliance by the most successful intervention in each study, in percentage points (in terms of correctness,*	*Intervention cost:yield ratio (OECD 2013) OR increased amount of taxes collected and (if reported) cost of the interventions*	*Method of how tax compliance was measured in the study*

[60] Hernandez, M., Jamison, J., Korczyc, E., Mazar, N., & Sormani, R. (2017). Applying behavioral insights to improve tax collection. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/27528/116046-WP-Tax-Collection-PUBLIC.pdf;sequence=1
[61] Note that for some of the combinations of those categories, there are no studies included in this literature review.
[62] Note that we aimed to report any adverse spillover effects and differentiate between digital and paper filing; however, we excluded these aspects due to insufficient available information for review.
[63] Antinyan, A., & Asatryan, Z. (2019). Nudging for tax compliance: A meta-analysis. *ZEW-Centre for European Economic Research Discussion Paper*, (19-055). Retrieved from https://www.econstor.eu/bitstream/10419/223572/1/cesifo1_wp8500.pdf

			payment before due date, or payment of due amount).		
Individual citizen; Personal Income Tax; Payment	Hallsworth (2017)[64]	Physical letters sent to individual taxpayers who have declared but not yet paid the correct amount	5.1 p.p. increase in payment rate compared to control group with standard government letter	Most successful message generated £1.8 million in collections that would otherwise not have been present at the 23rd day after the intervention started[1]	Likelihood of paying declared tax liabilities within 23 days of receiving letter.
Individual citizen; Personal Income Tax; Payment	Hernandez (2017)[65]	Physical letters sent to individual taxpayers who failed to pay before the first deadline	8.4 p.p. increase in payment rate compared to the standard government letter	increased average payment amount by ~ €80 compared to the previous version of the letter. Estimated revenues are 28 times higher than the cost of sending the letters, including staff time	Payment rate, payment amount, outstanding tax liability, payment delay after 12 weeks.
Individual citizen; Personal Income Taxes; payment	Luts (2019)[66]	Physical letters sent to individual income taxpayers who have not paid on time	10.9 p.p. increase in payment compared to standard control letter	€22 million in earlier payment of personal income tax for the 2015 assessment year	Number of outstanding debts paid in part or in full within 14 days of receiving letter.
Individual citizen; Personal Income Tax; Filing	Dong (2021)[67]	Physical letters sent to first-time taxpayers who missed the due date	15 p.p. increase in lodgement rate compared to the control group that did not receive any communication	NA	Lodgement of tax declaration, use of MyTax government portal, new MyGov registrations, number of days to lodgement, number of inbound calls after the intervention, observed change in address within the 1-2 month time span that data

[64] Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, *148*, 14–31. https://doi.org/10.1016/j.jpubeco.2017.02.003
[65] Hernandez, M., Jamison, J., Korczyc, E., Mazar, N., & Sormani, R. (2017). *Applying Behavioral Insights to Improve Tax Collection: Experimental Evidence from Poland*. https://doi.org/10.1596/27528
[66] Roy, M. V., & Luts, M. (2019). *Nudging in the context of taxation*. https://www.iota-tax.org/sites/default/files/documents/iota_paper_belgium_nudging_final.pdf
[67] Dong, S. X., & Sinning, M. (2022). Trying to Make a Good First Impression: A Natural Field Experiment to Engage New Entrants to the Tax System. *Journal of Behavioral and Experimental Economics*, *100*, 101900. https://doi.org/10.1016/j.socec.2022.101900

					collection took place.
Individual citizen; Personal Income Tax; Filing	Persian (2022)[68]	E-mails sent to individual taxpayers	increased overall and early filing rates by 1.10 and 2.07 p.p., respectively, compared to the 'no-e-mail' control	NA	Proportion of annual tax returns filed at least two weeks before deadline, overall filing rate.
Individual citizen; Property tax; Payment	Antinyan (2021)[69]	SMS messages to taxpayers who have not paid on time	the probability tax payments is 12–14 p.p. higher than in a 'no message' control	$97 cost of messages yielded $42.840 tax revenue. The tax revenue was $10,500 in the control treatment (at no cost)[1]	Probability of payment within one month of treatment ending.
Self-employed individuals; Personal Income Tax; Filing	Jamison (2021)[70]	E-mails sent to self-employed individuals with history of non-compliance	Number of declaration submissions by deadline was higher by 4.2 p.p. compared to the control group without a message	NA	Submission of tax return by deadline, submission of tax return, number of days until submission measured 46 days after treatment.

2.7. Meta-analyses and literature reviews

In addition to the studies we have reviewed in this module, there exists several meta-analyses and literature reviews over the existing literature on tax compliance. In this final section, we will summarise the most recent of these studies.

A meta-analysis[71], using the data from 70 experiments, provides a comprehensive overview of the current evidence from laboratory experiments. The findings confirm that the two crucial factors from the standard economic theory – the probability of getting caught and the fine size, are indeed related to compliance – but also provide a more fine-grained overview on these relationships. Particularly, the likelihood of detection is more important to deter evasion than the size of the fines. This is in line with other research on crime, which has found that the likelihood of being punished has a greater influence on behaviour than the punishment itself. Furthermore, it was also found that tax amnesties, which allow tax evaders to pay their tax debts without being fined, to have a negative effect on compliance. That said, it must be noted that lab experiments are by design unable to capture the full consequences that a taxpayer may face if caught evading, which includes real punishments and accompanying social stigma that obviously cannot be replicated in an experimental setting.

One meta-analysis[72] covering 20 RCTs on tax compliance found that, on average, the most effective interventions tend to focus on deterrence, personalisation and simplification and salience. On the other hand, interventions that appealed to social norms had an on average negative effect on the target behaviour. These

[68] Persian, R., Prastuti, G., Bogiatzis-Gibbons, D., Kurniawan, M., Subroto, G., Mustakim, M., Scheunemann, L., Gandy, K., & Sutherland, A. (2022). Behavioural prompts to increase early filing of tax returns: A population-level randomised controlled trial of 11.2 million taxpayers in Indonesia. In *Behavioural Public Policy* (Vol. 7, p. 20). https://doi.org/10.1017/bpp.2022.25

[69] Antinyan, A., Asatryan, Z., Dai, Z., & Wang, K. (2021). *Does the Frequency of Reminders Matter for their Effectiveness? A Randomized Controlled Trial*.

[70] Jamison, J. C., Mazar, N., & Sen, I. (2021). *Applying behavioral insights to tax compliance: Experimental evidence from Latvia*. https://ore.exeter.ac.uk/repository/handle/10871/121984

[71] Alm, J., & Malézieux, A. (2021). 40 years of tax evasion games: a meta-analysis. *Experimental Economics*, *24*, 699-750. Retrieved from https://www.researchgate.net/profile/Antoine-Malezieux-2/publication/344042683_40_Years_of_Tax_Evasion_Games_A_Meta-Analysis/links/5f5238c1a6fdcc9879ca5861/40-Years-of-Tax-Evasion-Games-A-Meta-Analysis.pdf

[72] Retrieved from https://www.revenue.ie/en/corporate/documents/research/applying-behavioural-science.pdf

findings are broadly in line with our own review, which found interventions targeting social norms to be less effective (or in some studies counterproductive) to induce tax compliance compared to deterrence and many other options. This meta-analysis also found evidence for positive interactions between interventions, where two interventions combined produce greater results than either one its own. Finally, the paper notes the importance of designing interventions with the population segment being targeted in mind; deterrence appears to work better among younger than older taxpayers, who instead respond better to a "soft" tone.

In contrast, another review[73] suggested that, while historically the "social norms" and other moral suasion messages had proven ineffective, this may be due to them having been too passive. More recent studies have employed harsher language ("Don't be a cheater" instead of "Tax cheating hurts the whole community") and seen better outcomes, though it remains to be seen whether tax authorities are willing to use such language on a large scale.

[73] Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, *57*(4), 904-954. Retrieved from https://www.nber.org/system/files/working_papers/w24799/w24799.pdf

3. Module 3 – How to choose the right study methodology

The aim of this module is to support selecting the right research methods at key points in developing behavioural interventions. The first section of the module contains a discussion of multiple core principles of study methodology. These principles help to understand the interlinkages between study designs and the benefits of combining various methods. This section is followed by an outline of design thinking principles and strategies that can be applied throughout the intervention development process. These strategies help to empathise with the target population, to explore the full scope of the intervention context and to adopt an iterative approach to intervention design by learning from past experiences and updating the interventions along the way. The final section of the module expands on the most frequently used qualitative and quantitative research designs for behavioural intervention development.

3.1. Choosing a study design: Main principles

Utilising diverse study methodologies at various stages of intervention development is essential for a comprehensive understanding of the problem and creating an effective intervention. In diverging phases, exploratory methods help grasp the complexities of the target population and intervention context. In converging phases, quantitative methods provide empirical insights that help selecting the best possible strategy, while qualitative approaches offer nuanced perspectives and user feedback. Employing a mix of methodologies ensures a well-rounded comprehension, enabling researchers to iteratively refine interventions based on evolving insights and evidence.

3.1.1. Trade-offs

When it comes to selecting study designs, there is no fixed hierarchy in terms of overall quality. The best research design is first and foremost one that can answer the research question. A second factor to consider is the feasibility of the design given the available time, financial resources, and expertise. For example, the average duration of a systematic review is between 6 months and 2 years[74] from inception until submission, with an additional delay for publication. While a systematic review may in many cases be the most thorough way to identify effective existing strategies, it may not be feasible to wait multiple years for a comprehensive overview. Finally, the best design will depend on the stage of intervention development. Generally, qualitative methods are used early in the development process, whereas quantitative, experimental methods are employed later. In order to support the selection of designs and the appraisal of outcomes, we provide a description of several quantitative and qualitative methods and discuss their main advantages and weaknesses.

3.1.2. Triangulation

Often, combining various study designs strengthens the ability to apply findings more broadly (i.e., generalizability). Moreover, the increased richness of the data generally enables a meaningful analysis and interpretation of study findings. Triangulation refers to using more than one approach to addressing a research question as a strategy to obtain a deeper understanding of the topic under research[75]. There are four types of triangulation, that each provide their own benefits and drawbacks to a research project. [76] [77]

Methodological triangulation. Whenever possible, the development and testing of an intervention should involve the use of mixed methodologies. An example of this is to combine an experimental evaluation with focus groups to evaluate the effectiveness of an intervention and to identify how taxpayers feel about certain interventions after being exposed to them. This is important as even interventions that “work” may prove unsustainable if they elicit a strong negative reaction in the target population. For example, a reminder to pay one’s taxes could be perceived as demeaning, infantilising, or lacking in clarity. These feelings may not be

[74] Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., & Moher, D. (2012). Evidence summaries: the evolution of a rapid review approach. *Systematic reviews*, *1*(1), 1-9. https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/2046-4053-1-10
[75] Schippling, A. (2017). Investigator triangulation in the data interpretation process. An almost untouched research area. La Critica Sociologica. 51 (203), 87-100. https://repositorio.iscte-iul.pt/bitstream/10071/16880/1/text_Schippling.pdf
[76] Denzin, N. K. (1978). The research act: A theoretical introduction to sociological methods. New York: McGraw-Hill
[77] Patton, M.Q. (1999). Enhancing the quality and credibility of qualitative analysis. Health Sciences Research, 34, 1189–1208. ncbi.nlm.nih.gov/pmc/articles/PMC1089059/pdf/hsresearch00022-0112.pdf

captured by a traditional, quantitative design, but may become apparent in the context of a focus group. For example, box 4 includes citizen responses to a BI informed payment reminder that were obtained through a qualitative study. This demonstrates how undesirable negative spillover effects of interventions can be uncovered or even prevented with qualitative methods. It is important to note that for technical and privacy reasons, linking administrative data from the tax administrations with other sources such as surveys and focus groups is often not possible.

Box 4. Examples of citizen responses to BI informed late payment reminders[78]

- 'I do not like it at all…It's just so demeaning.'
- 'I was in a panic because I didn't know what to do. I didn't know where to turn. And I just kind of ignored it.'

Another example is to conduct one-on-one interviews to identify suitable target behaviours, using case studies to develop appropriate strategies, conducting an online or lab experiment as a pilot study to understand whether the intervention has potential and to sort out any design issues, and finally implementing an RCT to determine the intervention's potential in a real-world setting. Before introducing the intervention to a broader population (stage 5), focus groups or expert interviews may be appropriate to help deciding the best possible implementation strategy.

An example of this type of methodological mixing can be found in a 2019 paper from Enachescu et al[79], in which two studies are described: In the first study, focus groups were used to identify emotions elicited over the course of the taxation process, whilst in the second study, an experimental survey was used to both quantitatively measure the prevalence of these emotions and manipulate them using different scenarios posed to participants. Mixed methodologies are also increasingly used to understand how and why nudges, particularly reciprocity nudges, work to alter behaviour[80].

Caution must be taken to ensure conclusions are not muddled. When two methodologies are used to answer the same question ("Does this intervention work?"), researchers may struggle if these methodologies return different answers. To mitigate the issue of reconciling contradictory evidence from different methodologies, researchers should provide clear minimum objectives and thresholds for what the intervention is supposed to achieve. An example of such objectives may be an information nudge that has to improve tax compliance by a certain predefined percentage, while not eliciting negative emotional reactions or confusion from the target population (something that qualitative methods such as focus groups may measure).

Investigator triangulation. This refers to having more than one researcher (investigator) involved in the research process. This is a necessary (but not sufficient) condition to avoid tunnel-vision, misinterpretations, and logical errors. Multiple researchers can be involved in both the design, collection and interpretation of data, to ensure multiple perspectives are heard. Qualitative research in particular may benefit from interpretation groups to mitigate the issue of the data (such as answers to an interview) being interpreted according to the pre-conceived beliefs and biases of the researcher that collected the data. If multiple methodologies are used as in the example above, there should ideally be more than one researcher familiar with each methodology used (i.e., at least two researchers familiar with RCTs, two researchers familiar with focus groups, etc.).

Data triangulation. With data triangulation, multiple data sources are used to answer the research question, by varying the data collection between different people or different points in time. One example may be a survey used in combination with public records to contrast and/or verify people's responses; if survey respondents respond in ways that public records suggest they should not (e.g., if younger respondents report

[78] Merrick, W. (2022). But does the nudge fit? Institutional structure and behavioural insights. *Behavioural Public Policy*, 1-17. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/16167E76CF1B18FBC9D238FCC01F3CD4/S2398063X22000367a.pdf/but_does_the_nudge_fit_institutional_structure_and_behavioural_insights.pdf

[79] Enachescu, J., Olsen, J., Kogler, C., Zeelenberg, M., Breugelmans, S. M., & Kirchler, E. (2019). The role of emotions in tax compliance behavior: A mixed-methods approach. *Journal of Economic Psychology*, *74*, 102194. https://research.vu.nl/ws/portalfiles/portal/159685721/The_role_of_emotions_in_tax_compliance_behavior.pdf

[80] Monageng, N. (2023). Using Mixed Methods to Understand Tax Compliance Behaviour. *Electronic Journal of Business Research Methods*, *21*(1), 43-53. https://academic-publishing.org/index.php/ejbrm/article/download/2903/2162

having on average higher incomes than older respondents) this may suggest the sample is not representative and puts the quality of the data collected into question.

Theory triangulation. To triangulate theory means using two or more theoretical frameworks in your research instead of approaching your research question from only one theoretical perspective. This is especially useful to understand why an intervention has or hasn't worked (Stage 4 of BASIC), as the mechanisms behind how some interventions work are frequently contested. An example of this type of triangulation can be found in a 2021 study on tax compliance[81] that explored the factors driving tax compliance among small- and medium-sized enterprises in Indonesia, utilising both the Moral Tax theory and the Theory of Planned Behaviour. One advantage of the ABCD framework discussed in Module 1 is its versatility in incorporating different fundamental behavioural models, allowing for a fuller understanding of the behavioural drivers behind a target behaviour.

3.1.3. Design-thinking principles

Design-thinking is an interdisciplinary approach to problem-solving and innovation. It involves an iterative process that focuses on understanding user needs, defining problems, generating ideas, prototyping solutions, and testing and refining those solutions[82][83]. Although design-thinking originated in product development, it recently found its way into the domains of public policy and behaviour change intervention development[84]. Main reasons for this popularity include a **human-centred focus** that helps to understand, empathise, and collaborate with stakeholders; the use of **generative techniques** to elicit ideas and solutions to a given problem; and the **iterative approach** to intervention development with frequent prototyping and refinement. Below, we have highlighted these principles and explained how they can be incorporated into the intervention development approach.

Human Centred focus: How to empathise with key stakeholders. When it comes to stakeholder participation there is an important distinction between including stakeholders 'as subject' and including stakeholders 'as partner'. In the former situation, researchers study or interview people from the target population and use their input to inform intervention development. In the case of 'stakeholders as partner' people from the target population have an active role in the intervention development process. Instead of only providing input through on a predetermined topic list or questionnaire, stakeholders 'as partners' are empowered to collaborate during design activities, ensuring a more inclusive and participatory approach. For example, by participating in brainstorming activities, taxpayers may actively collaborate with policymakers, tax administrators and researchers to identify the most important set of behavioural drivers for the target behaviour[85]. Both types of stakeholder involvement are considered to improve the overall quality of the intervention design. Two design-thinking methods that stimulate a human centred focus during the development of behavioural interventions are *open to learn conversations* and the creation of *personas*.

Open to learn conversations. One way of increasing stakeholder partnership is to have open to learn conversations instead of traditional interviews. In these conversations, the goal is to question the fundamental basis of the status quo (e.g., the current policy measures to target non-compliance) and to encourage participants to think about new options or alternative solutions (in contrast to providing a set of initial ideas). To promote a solution-based conversation, 'how might we...?' questions are frequently adopted in these conversations, assuming that there is a solution and emphasising collaboration[86].

[81] Supriyono, A., Utami, I., & Muktiyanto, A. (2021). Exploration of tax compliance determination on micro, small and medium enterprise. *Jurnal Akuntansi*, *11*(1), 33-52. https://ejournal.unib.ac.id/index.php/JurnalAkuntansi/article/download/14012/7178
[82] Cash, P. J., Hartlev, C. G., & Durazo, C. B. (2017). Behavioural design: A process for integrating behaviour change and design. *Design Studies*, *48*, 96-128. https://www.sciencedirect.com/science/article/abs/pii/S0142694X16300667
[83] Mintrom, M., & Luetjens, J. (2016). Design thinking in policymaking processes: Opportunities and challenges. *Australian Journal of Public Administration*, *75*(3), 391-402. https://ojs.unbc.ca/index.php/design/article/download/1475/1217
[84] Cash, P. J., Hartlev, C. G., & Durazo, C. B. (2017). Behavioural design: A process for integrating behaviour change and design. *Design Studies*, *48*, 96-128. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0142694X16300667
[85] Sanders, E. B. N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *Co-design*, *4*(1), 5-18. https://www.tandfonline.com/doi/full/10.1080/15710880701875068
[86] Mintrom, M., & Luetjens, J. (2016). Design thinking in policymaking processes: Opportunities and challenges. *Australian Journal of Public Administration*, *75*(3), 391-402. Retrieved from https://onlinelibrary.wiley.com/doi/epdf/10.1111/1467-8500.12211

Personas. Personas are fictional characters or representations created to embody the characteristics, needs, and behaviours of specific user groups or stakeholders. These personas are based on real user research and data, providing a human face to the individuals who will be affected by the design process, whether it's for products, services, or policies. The purpose of creating personas is to make the target population such as taxpayers more relatable and understandable. A persona is often a brief description that typically includes details such as demographic information, goals, challenges, preferences, and other relevant attributes. By referring to these personas throughout the intervention development process, the design team can keep the needs and perspectives of the intended users at the forefront, ensuring that the final outcome is more responsive to diverse user needs. Personas help humanise the design process and guide decision-making by focusing on the individuals who will ultimately interact with and be impacted by the designed solutions.

Generative Techniques. Generative techniques in co-design refer to a diverse set of activities and methods aimed at sparking creative thinking, idea generation, and collaborative problem-solving among participants. These techniques encourage active engagement, exploration, and the generation of a wide array of solutions to address specific challenges. These activities – including brainstorming, mind mapping, and prototyping – help to achieve a rich pool of ideas that can be refined and developed collaboratively.

Sensitisers. A useful generative method to stimulate ideation in brainstorm activities are *sensitisers*. This concept refers to activities used to heighten participants' awareness and understanding of certain issues, perspectives, or challenges. Sensitisers help to foster empathy and encourage a deeper connection with the experiences and needs of the users or stakeholders involved in the co-design process. An example of this method is to ask taxpayers – one week prior to a brainstorm or interview session - to engage in a "Taxpayer Diary" activity at home. The diary entries may include details on the tax declaration and payment actions of the previous year, including any interactions with tax-related documentation. Participants might also reflect on their emotional responses or challenges faced during this process. This assignment helps to proactively engage participants in the ideation process at an earlier stage, thereby enhancing the overall quality of subsequent design activities. Sharing these reflections during a brainstorming session is likely to provide a rich foundation for discussions and creative problem-solving.

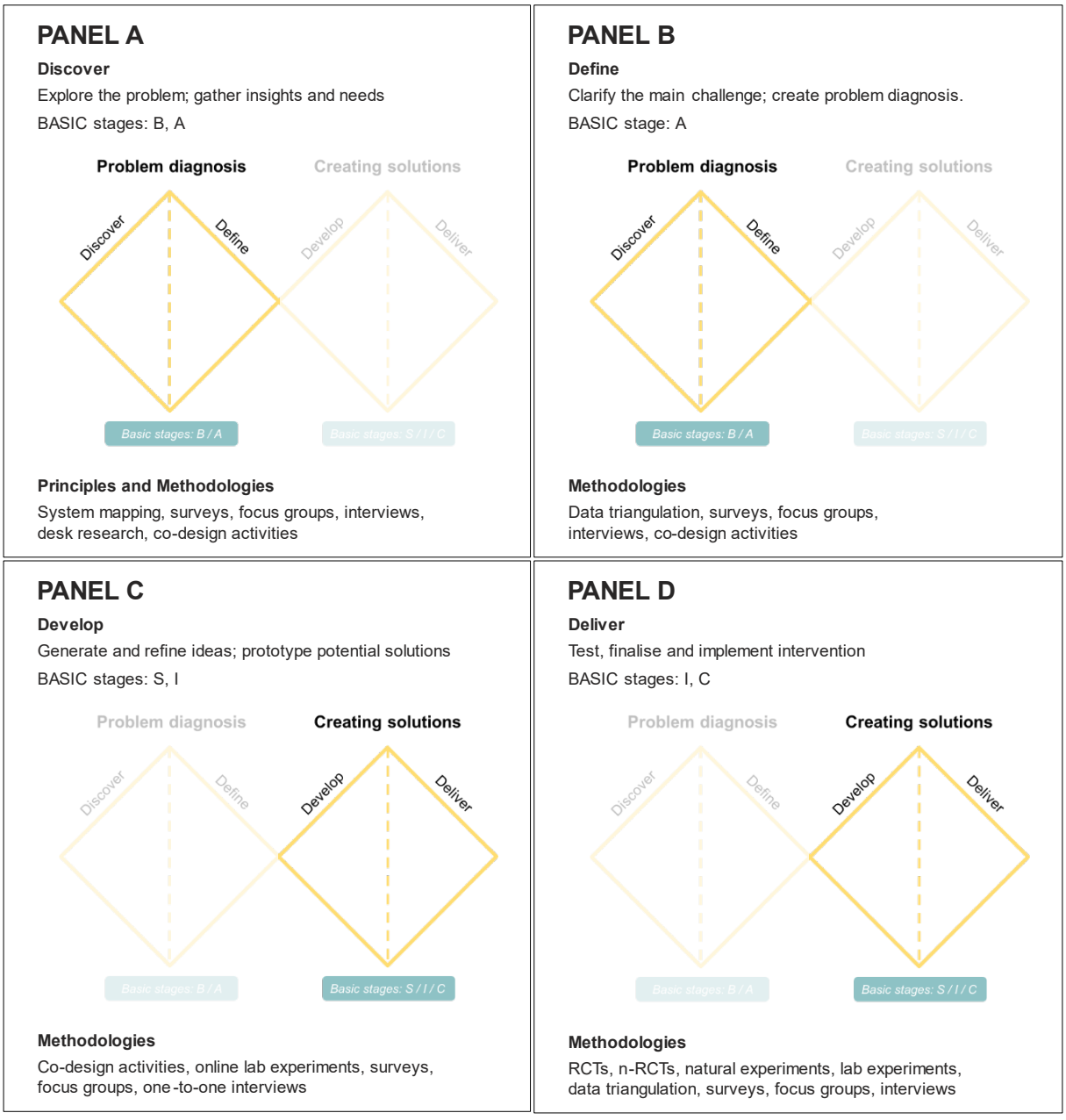
System mapping. System mapping involves visually representing the interconnected elements, relationships, and dynamics within a complex system relevant to the design context. This technique helps to provide a holistic understanding of the system's components and their interactions, supporting participants to identify key influencers and points of intervention. By creating visual maps, often using diagrams or charts, co-design teams can collaboratively analyse and explore the complexities of the system, informing strategic decision-making and facilitating a comprehensive approach to intervention design. System mapping serves as a valuable tool for uncovering insights, fostering shared understanding, and guiding collaborative efforts toward effective and sustainable design solutions. In the context of improving tax compliance, a system map could visually represent the various components such as tax regulations, taxpayer behaviours, government enforcement, and public perceptions. Relationships and feedback loops between these elements would be depicted, offering a holistic view of the complex dynamics at play. This map aids co-design teams in identifying influential factors and knowledge gaps, as well as understanding the interactions between these factors in relation to compliance. An example of a system map in the domain of tax compliance is the IRS subway map described in case study 2[12], or the 'tax compliance as a system model'[87].

Iterative intervention design. Although the BASIC model offers a stepwise approach to intervention development, the practice of intervention design is often nonlinear, with researchers and policymakers having to go back and forth between steps. For instance, if it becomes evident in the strategy stage that the initially compiled list of drivers for a specific target behaviour is incomplete, rather than rigidly adhering to the established process, design-thinking encourages a flexible approach to accommodate emerging insights and prompts the updating of ideas and concepts.

[87] Randlane, K. (2016). Tax compliance as a system: Mapping the field. International Journal of Public Administration, 39(7), 515-525.

This process is often visually illustrated by the double diamond model that is describe in module 1. This model represents a general process of problem solving that begins by defining the scope of the problem to be addressed, before moving to developing and testing solutions. The diamond shapes show how for each stage, first choice options are explored (diverging) and priorities are set (converging), with each stage requiring different types of research activities and objectives. Although the double diamond model is not explicitly included in the BASIC framework, alternating diverging and converging phases are included, particularly in the early stages. The figure below depicts how various study methodologies can be adopted in various stages of the development process. Panels A and B form the first diamond that starts with exploring and defining the problem and ends with a 'behavioural diagnosis' on what key drivers should be targeted with a behavioural intervention to increase compliance for that particular issue. Panels C and D form the second diamond that focus on generating, testing and implementing behavioural interventions as solutions. For each of the study methodologies that are covered in the section below we have indicated where they are most appropriate.

Figure 8. Suggested methodologies for each stage of the Double Diamond Model.



3.2. Descriptions of study methodologies

Below, we will describe each study type, discuss how it might be used in a behavioural intervention development project, and provide an overview of general strengths and limitations. In appendix B, a table

summary of these methodologies can be found. We also provide references to relevant quality or ethical guidelines, example studies in the domain of tax compliance and other sources of interest. For the purpose of this section, randomised control trials refer strictly to trials in the field.

3.2.1. Qualitative methodologies

In quantitative research, there is focus on objective truths through systematic observation and measurement. This approach strives to establish generalisable laws and causal relationships, prioritising statistical analysis and control. On the other hand, qualitative research acknowledges that much of knowledge is socially constructed and context dependent. The focus of qualitative research lies in exploring diverse dimensions of human experiences and meanings, employing flexible methods such as interviews and observations to capture the depth and nuance of the research question. Qualitative methods may help us understand the contexts in, and the mechanisms through which an intervention works or does not work.

Sample sizes in qualitative studies. Given the focus on exploring diverse dimensions of human experiences and meanings, qualitative researchers work with smaller, purposefully selected samples compared to quantitative study designs. Saturation, a crucial concept in qualitative research, relates to the generalizability and validity of findings in relation to the sample size. Saturation signifies the point at which researchers believe they have gathered enough data to thoroughly understand the phenomenon under investigation. It involves reaching a saturation point where new information no longer provides additional insights or perspectives. In qualitative studies, saturation is often assessed by comparing new data with data that has already been collected during the course of the study. When no additional information or perspectives emerge, the research has reached a saturation point.

Mitigating researcher bias. Researcher bias, also known as researcher subjectivity, occurs when the researcher's own perspectives, beliefs, or preconceptions influence the interpretation of qualitative data, potentially leading to subjective or biased conclusions. It is crucial that the public sector officials who are conducting behavioural research try and mitigate researcher bias to improve the study quality. One method to mitigate this bias is to involve participants in member checking. This is a collaborative approach where participants review and validate the researcher's interpretations, ensuring that the perspectives of those being studied are accurately represented.

Case study 4: A qualitative study of the impact of the UK 'bedroom tax[88]'

Background: The study focused on the impact of the UK government's "Removal of the Spare Room Subsidy," commonly known as the 'bedroom tax,' implemented in April 2013. This policy aims to reduce public spending on social housing by cutting housing benefits for tenants living in accommodation deemed too large for their needs. The policy affects a significant number of working-age social housing tenants, especially those with disabilities, who are among the poorest and most vulnerable in society.

Objective: The qualitative study aimed to investigate the effects of the bedroom tax on social housing tenants in a socio-economically deprived urban area in Northeast England. Specifically, the study examined its impact on health and well-being, social relationships, and the wider community.

Design: The study adopted a qualitative approach, using semi-structured interviews with tenants (38 participants), a focus group (7 participants), and interviews with service providers (5 participants). The sample comprised of individuals affected by the bedroom tax, identified through a social housing provider in the locality. Data collection involved exploring the financial impact, physical and mental health, social relationships, and service use through interviews and focus groups.

Outcome: The study revealed multifaceted impacts of the bedroom tax on affected individuals and communities. Participants experienced financial strain due to reduced income, leading to cutbacks in essential expenses such as food and utilities, and many resorted to food banks and struggled with utility bills, leading to increased debt and reliance on high-interest loans. The "bedroom tax" policy also contributed to stress, anxiety, and depression among participants, exacerbated by worries about re-location, inability to afford healthy food, and living in inadequately heated homes. These mental health issues were also linked to physical health problems and sleep disturbances. Participants in the study expressed strong emotional

[88] Moffatt, S., Lawson, S., Patterson, R., Holding, E., Dennison, A., Sowden, S., & Brown, J. (2016). A qualitative study of the impact of the UK 'bedroom tax'. *Journal of Public Health*, *38*(2), 197-205. Retrieved from: https://academic.oup.com/jpubhealth/article-pdf/38/2/197/6867779/fdv031.pdf

attachments to their homes and communities, which provided a sense of belonging and support. Moving to smaller properties was often seen as disruptive and detrimental to family life and community ties.

3.2.2. Focus groups

Description: Focus groups are a qualitative methodology in which small groups of individuals are gathered together to answer questions and discuss a certain topic in a conversation led by a researcher. Focus groups are interactive and involve free discussion and conversation, compared to surveys which ask a set of narrow questions that may not fully allow participants to explain their reasoning behind their behaviours. Using a focus group can allow researchers to explore attitudes and perceptions towards taxation and tax compliance and may also allow them to identify new barriers to compliance that they may not have thought of. Because of this, focus groups are also useful for generating new hypotheses and ideas for interventions.

Beyond the verbal responses, researchers leading a focus group may also record important non-verbal cues that may suggest, for example, that a participant is or isn't uncomfortable admitting something (such as admitting non-compliance).

Participants to a focus group may be recruited for example through advertisement (such as on social media), by reaching out to community organisations, or by reaching out to academic institutions who may help with recruiting students, faculty members and alumni.

Stage of BASIC: 1-2. Focus groups are mainly helpful to collect thoughts and input from stakeholders and experts at the initial stages when attempting to identify behaviours, understand what drives them and decide on a suitable target behaviour. They can, however, also be used at a debriefing stage to better understand the mechanics behind why an intervention worked or didn't work.

Strengths:

- ***Allows for in-depth discussions between participants***. Researchers may observe how participants react to each other's statements and arguments and through that way gain a better understanding. Moreover, responses of others often help trigger associations or responses, leading to a deeper discussion compared to individual interviews.
- ***Candid responses***. In a well-designed focus group, participants feel comfortable to share not just their opinions but elaborate on their beliefs and underlying motivations to an extent that they may not do in an ordinary survey (i.e. underlying motivations behind evasion). Unlike in surveys with multiple-choice, a focus group allows participants to phrase their responses in whatever way they wish and adds additional nuance.
- ***Real-time feedback***. Whereas in a traditional survey, any feedback from participants is received once they have completed the survey, a focus group provides researchers with real-time feedback, allowing them to steer and adapt the questions asked.
- ***Follow-up questions***. Another advantage of focus groups is the ability to ask follow-up questions that directly relate to participants' previous answers.

Weaknesses:

- ***Risk of bias***. Researchers may unconsciously influence the discussion and responses of the focus group. Interpretation is also subject not least to confirmation bias.
- ***Unpredictable in-group dynamics***. It is often hard for researchers to predict how the group will interact with one another, whether some participants may talk more and not let others have their say. Keeping discussion on-topic can also be difficult. Various methods exist to moderate the group dynamics and keep the discussion on-topic. An example is the nominal group technique, where participants are asked to first individually generate ideas or responses to a specific question before sharing them with the group.
- ***Self-selection bias***. Since focus groups take up a significant share of time, participants who sign up for them may not be representative of the overall population. Incentives may be used to mitigate this issue.

Sources

The Equality Challenge Unit has developed and published ethical guidelines for qualitative research [link]

3.2.3. One-on-one interviews

Description: Related to focus groups, one-on-one interviews (with taxpayers, other stakeholders or experts) can be used to collect in-depth data allowing to understand not just the decisions made by a taxpayer, but the reasoning behind the decisions. Although data collection in interviews is less controlled than in quantitative research, researchers often employ a degree of structure to guide discussions and ensure alignment with study objectives. Methods such as semi-structured interviews strike a balance, offering a framework of predetermined questions while allowing flexibility for participants to express their experiences and perspectives. The level of structure can vary, from loosely guided conversations to more rigid formats, depending on the research goals and the need to explore participant narratives in-depth. For example, in a study on financial literacy in Portugal, the topic list consisted of two topics, The first addressed the financial challenges faced by long-term low-income households, asking participants to detail their toughest financial management hurdles. The second topic explored strategies these households employ to navigate financial difficulties, soliciting both practical actions taken to manage budgets and professional observations on effective management tactics[89].

Stage of BASIC: 1-2. Like focus groups, interviews are not conducted to test interventions, but rather to help design them, though they can also be used for debriefing purposes. Experts may also be interviewed ahead of implementing a successful intervention on a large scale (stage 5).

Strengths:

- ***Individual autonomy.*** Unlike focus groups, there are no in-group dynamics that may adversely affect the outcomes.
- ***Equal individual focus.*** One-on-one interviews allow researchers to easier pay the same amount of attention to each participant and ensure each participant gets the same speaking time and questions.
- ***Tailoring.*** The flow and topics can be tailored to each participant with no concern as to how the question may influence the answers of other participants.

Weaknesses:

- ***Resource-intensive****.* With only one participant per session, interviews require more work relative to data output than most other research methods.
- ***Social desirability****.* Participants in one-to-one interviews may feel a stronger pressure to provide socially desirable responses due to the direct interaction with the researcher. This may impact the authenticity of the data as participants may tailor their answers to align with perceived expectations.

The Equality Challenge Unit has developed and published ethical guidelines for qualitative research [link]

3.2.4. Case studies

Description: Case studies involve researchers deep-diving into a particular case, such as a tax authority that have implemented an intervention. As part of the case study researchers would usually talk to the people who were involved in developing and implementing the intervention, asking questions about any issues they faced along the way, and also talk to ordinary taxpayers and other stakeholders who were affected (whether positively or negatively) by the intervention. By means of triangulation, these various data sources are analysed to gain insights into the complexities of the case under investigation. The case study method is beneficial for examining complex social phenomena, retaining the real-life aspects of events, and developing knowledge that is closely tied to specific contexts. It is crucial to note that case studies may involve both successful and unsuccessful interventions.

Stage of BASIC: 1-3. Case studies can help inform both behaviours suitable for targeting and strategies that may be used to do so.

[89] Rodrigues, S. P., Sousa, L., & Alarcão, M. M. (2016). Financial management in long-term low-income households: comparing perspectives of professionals and families in Portugal. European Journal of Social Work, 19(6), 977-991.

Strengths:

- ***Rich contextual understanding***. With the ability to deep-dive into a particular case, such as a behavioural intervention program previously implemented by an authority, more comprehensive information can be gleaned about the case of interest.
- ***Holistic examination***. Without the quantitative focus on particular outcomes or data points, case studies can provide a holistic understanding of why, for example, a particular intervention worked or did not work, that can incorporate both qualitative and quantitative elements.
- ***Tailored recommendations***. Related to the previous two points, the nature of a case study allows researchers to provide tailored recommendations for policymakers that take into account not just data but "soft" factors such as stakeholder perspectives.

Weaknesses:

- ***External validity***. Due to their focus on in-depth exploration of specific cases, it is challenging to generalise findings of case studies to broader populations or contexts. The unique nature of each case, potential selection bias, and small sample sizes contribute to the difficulty in applying case study results to diverse settings.
- ***Difficult to establish causation***. An in-depth look at one particular case, without the benefit of data to compare it to other cases, makes it difficult to establish causation (i.e. whether an implemented intervention is the reason tax compliance improved). Researchers must be aware of any potential confounding variables.

Sources:

The Equality Challenge Unit has developed and published ethical guidelines for qualitative research [link]

3.2.5. Quantitative surveys

Stage of BASIC: 1-2.

Description: Quantitative surveys may ask participants questions about their income, how much time they typically spend on preparing their tax return, or how they would rate their interactions with the tax authorities. Although Likert scales are frequently utilised for ranking questions, various question types exist that help to elicit accurate responses. Closed questions are often alternated with open questions, to further explore a specific choice or decision. Surveys can take place online, through mailings or, less frequently, in-person. The best way of conducting a survey will depend upon circumstances including the target group. An important feature is the possibility of combining quantitative surveys with online experiments, obtaining not only a measure of effectiveness of a (prototype) intervention, but also an extensive insight into the participants' motives and experiences.

Strengths:

- ***Larger sample sizes.*** Since quantitative data is generally easier to collect and does not rely on participants spending a lot of time outlining their thoughts and reasons behind a decision the way they might do in a focus group, quantitative surveys allow for larger sample sizes.
- ***Easier to interpret data.*** Whereas qualitative data relies on the participants using words to mean the same thing they mean to the researcher, quantitative data is relatively easy to analyse.

Weaknesses:

- ***Inability to determine causation***. Since surveys are not experiments, they cannot be used to determine a possible causal link between two variables. This is why surveys are typically used in the early stages of intervention development to identify possible targets.
- ***Sensitivity to wording***. Survey questions may be unclear, leaving participants to interpret the questions differently which would impact their answer.
- ***Limited depth***. While quantitative surveys are good for collecting surface-level information such as attitude towards taxes, their depth is limited. Qualitative methodologies tend to perform better at understanding why someone has a certain attitude towards taxes.

3.2.6. Online experiments

Description: Online experiments are the newest tool in the toolbox of experimental and behavioural researchers. Utilising digital platforms, participants engage with study materials, respond to stimuli, or complete tasks online, allowing the researchers to isolate the impact of this (treatment) variable. Depending on the research question, these experiments may contain one or several experimental and control conditions. As such, online experiments provide timely and cost-effective insights into behaviours, attitudes, and responses to online prototypes of behavioural interventions. Universities, research organizations, and market research companies have developed highly customizable platforms, enabling the simulation of specific tax processes targeted for testing, accommodating a wide range of experimental needs.

Stage of BASIC: 3-4, though online (non-experimental) surveys can be used at the earlier stages as well to gather information.

Strengths:

- ***Lower cost***. Compared to RCTs, online experiments are cheaper, mainly because they do not require an implementation process in tax administration systems or require long-term follow-up.
- ***Reduced self-selection bias***. Given that online experiments are easier to take part in, typically requiring less time commitment and not forcing participants to travel to a physical location, participants who agree to take part may be more likely to represent the overall population than in offline lab experiments. This is especially the case when a representative sample is drawn from large online panels.
- ***Rapid Data Collection.*** Online experiments allow for fast data collection and analysis, as participants can be reached and responses gathered in a shorter time frame compared to physical experiments.
- ***Diverse sample***. Due to not being constrained by physical space, online experiments can have virtually unlimited sample sizes, a great advantage when an experimenter seeks for a broad representation of demographics, cultures, and backgrounds or to target certain subgroups.

Weaknesses:

- ***Greater risk of unserious or inattentive participants***. Participants in online experiments make a relatively small commitment. While this has the effect of potentially recruiting a more representative sample (see above), it may also lead to more participants who are not taking the survey seriously. This is particularly an issue if real incentives are used, and participants thus stand to make money or other rewards from their participation. And while taking part of an experiment from the comfort of one's own home may induce more natural behaviour; this environment would also contain more distractions (including from notifications received on the phone or computer). Multiple methodological and statistical techniques exist to mitigate this and filter out inattentive participants, including attention-testing questions, straight lining detection and response time analysis.
- ***Certain interventions cannot be (easily) replicated online***. Recreating the complex aspects of real-life tax situations in online experiments can be hard. Although online setups become increasingly accurate, they do not capture real consequences of non-compliance making it tough to accurately test certain interventions.
- ***Difficulty reaching less-online populations***. While all demographic groups use the internet to some extent, online experiments may struggle to recruit from those groups that use it less. Given the increasing digitalization of the tax process, this problem becomes increasingly relevant. Specialized recruitment strategies, such as partnering with community organizations or with specialized panels may help engage less-online populations in these types of experiments.

Case study 5: "Do Behavioural Nudges in Prepopulated Tax Forms Affect Compliance? Experimental Evidence with Real Taxpayers"[90]

Background. For decades, research on nudging has suggested that default options affect consumer behaviour. In the context of tax compliance, default options on tax return forms can simplify the filing process and reduce errors. However, incorrectly set defaults can lead to increased tax evasion. This experiment investigated the impact of correct and incorrect defaults on tax compliance and whether nudges can mitigate the adverse effects of incorrect defaults.

Objective. The study sought to examine the effects of prepopulating tax returns with correct and incorrect default values on tax compliance. Additionally, the study aimed to explore whether nudges, specifically those invoking descriptive norms about compliance, are able counteract the negative impact of incorrectly set defaults.

Design. The study employed an online experiment where participants acted as fictitious taxpayers completing a tax form based on a given income and expense profile. The experiment included seven treatments:

1) Base: No default, all fields left blank.
2) Corr: Correct default of self-employment income.
3) Over: Default with an overestimated self-employment income.
4) Under: Default with an underestimated self-employment income.
5) UnderGeneric: Similar to Under but required participants to actively edit prepopulated fields.
6) UnderAlways: Similar to Under but included a message about compliance norms.
7) UnderTrigger: Similar to UnderAlways but the message was displayed only if the declared income was below a certain threshold.

Participants completed the tax form, and their decisions were then compared across treatments to assess compliance behaviour.

Outcome: The study found that correct default options on tax return forms are beneficial for compliance only when based on highly reliable information. Defaults with incorrect values, whether underestimating or overestimating tax liability, lead to significant drops in compliance and tax revenue. Additionally, the study showed that nudges containing messages about compliance norms can effectively increase compliance when they react to user inputs (as in UnderTrigger when the message was only displayed if the declared income was below a threshold). However, static (i.e. the UnderAlways treatment) nudges are ineffective. These findings highlight the importance of accurate defaults in tax policy and the limited effectiveness of nudges without real-time responsiveness.

Sources

Bachard and Williams formulated ethical guidelines for conducting online experiments [link].

3.2.7. Lab experiments

Description: Lab experiments take place within a physical space where participants are randomly allocated into control and treatment groups, and then typically given a number of tasks (i.e. filling out a tax return) to complete. These tasks may be done on paper or may be digital. For tax compliance experiments, it is common to use simulations where the participants are digitally "filing" their taxes using platforms made to look identical to those used by the tax authorities in the country where the experiment takes place, in order to increase immersion and realism. In some experiments, participants interact with one another, but in most cases the tasks are completed separately. The treatment group(s) receive the same tasks as the control group, with only one difference (the intervention). For example, the treatment and control groups may both have a tax return that they must fill out, but with the distinction that the treatment group receives a disclaimer on top of the form

[90] Fonseca, M. A., & Grimshaw, S. B. (2017). Do behavioral nudges in prepopulated tax forms affect compliance? Experimental evidence with real taxpayers. *Journal of Public Policy & Marketing*, *36*(2), 213-226. https://ore.exeter.ac.uk/repository/bitstream/handle/10871/29935/jppm.15.128.pdf?sequence=1

informing them of the penalties for tax evasion. **When to use? BASIC stage 3-4.** Lab experiments excel at isolating and quantitatively estimating the impact of a treatment variable and are thus best used at the later stages. Depending on the project, lab experiments may be used in isolation or as a pilot study before launching an RCT.

Strengths:

- ***Offers maximum degree of control.*** In a traditional lab experiment, the environment and all the tasks carried out as part of the experiment are designed and chosen by the researcher.
- ***Ability to infer causation****.* Due to the high degree of control by the researcher over the environment, lab experiments allow for causal inferences to a greater degree than any other methodology. Furthermore, lab experiments may have more than one control group or use other designs that allow for more than one treatment to be tried at any given time without substantially increasing the cost.
- ***Isolation of factors****.* Through complete control of the tasks and environment faced by participants, researchers can isolate the impact of a single factor (the treatment) with a far greater precision relative to real world data collection which by nature is messier with confounding factors.
- ***Relatively low cost****.* Relative to RCTs, lab experiments are typically lower in cost as they usually involve one-off interactions.
- ***Ease of replicability****.* Due to the detailed nature of a lab experiment's design, attempting to replicate the findings is easier than for other types of experiments.

Weaknesses:

- ***External validity****.* The controlled nature of lab experiments limits the ability to generalize findings to real-world situations, as the artificial setting does not represent the complexities and variability present in everyday contexts.
- ***Limited sample size****.* Lab experiments typically require physical rooms, limiting the sample size compared to online methods. To ensure that the capacity is sufficient for the study objectives, it is important to conduct a sample size calculation on beforehand to determine the required number of participants for the planned statistical analyses.
- ***Self-selection bias****.* Participants who agree to take part in a lab experiment may not always resemble the population as a whole, even if they are demographically representative. They may, for example, have a greater interest in psychology or research in general than the average person. This is particularly the case with student-only samples, as most students have never paid taxes in their lives and thus have no experience with it. Consequently, there are more prone to make decisions in a manner unlike what experienced taxpayers would, simply due to their lack of experience.
- ***Hawthorne effect****.* Those who participate in lab experiments may act differently due to knowing that they are being observed. In the context of tax-related experiments, participants may choose the ethical option of complying, whereas in real life they would not have done so. By making lab experiments as anonymous as possible and using real incentives, researchers may mitigate this issue.
- ***Earned vs unearned endowments****.* Most lab experiments do not feature earned endowments, despite evidence suggesting participants act differently with money that they have earned compared to money that they have been given. In some experiments, participants first undertake a number of tasks (such as solving mathematical equations) and then use the money earned in those tasks in the subsequent task(s) (i.e., they file a tax return and pay taxes based on the amount they earned). This, however, is still not a universal rule. While it is not clear whether earned endowments make a difference in the context of tax compliance, a difference has been established in many other settings. In any case, these experimental procedures do not fully reflect the complexities of real life situations.
- ***Lack of real incentives****.* Lab experiments do not always feature real incentives, meaning participants are taking part in entirely hypothetical tasks that have no outcome for them in real life.
- ***The need for a specialized laboratory.*** Conducting lab experiments requires access to facilities equipped to simulate real-world tax processes, enhancing the study's realism and immersion. This necessity often leads to setting up collaborations between tax authorities and academic institutions,

Case study 6: “Tax policy and the news: An empirical analysis of taxpayers’ perceptions of tax-related media coverage and its impact on tax compliance”[91]

Background: Tax compliance is linked to trust in authorities and in the perception of the power of those authorities to enforce their rules (i.e. ability to enforce tax compliance). Both these variables can be impacted by media coverage, which suggests media coverage may have an impact on tax compliance.

Objective: The experiment investigated the influence of specific tax related information in the news on intended tax compliance, aiming to determine whether exposure to different types of media coverage influence taxpayers' perceptions and compliance intentions, even in the absence of changes in the overall tax policy environment.

Design: 487 employees in Austria took part in a questionnaire-based experiment. Participants were randomly assigned to one of four scenarios presenting news stories about Austrian tax authorities, stories that were aimed at manipulating levels of trust and perception of the authority’s power in a 2x2 design:

1) High trust, high power
2) High trust, low power
3) Low trust, high power
4) Low trust, low power

In the high trust condition, participants were given news stories to read that painted the political situation in the country as stable, with authorities trustworthy and legislation transparent, whereas the opposite was true of the news stories provided to those in the low trust condition. The news stories used in the high power condition described authorities as efficient and determined to effectively and severely punish tax evaders, whereas the stories in the low power condition gave participants the opposite image. Participants responded to Likert-type scale items measuring trust in authorities, perceived power of tax authorities, and intended tax compliance.

Outcome: Exposure to media coverage presenting tax authorities as trustworthy led to higher trust among participants compared to scenarios suggesting untrustworthiness. In the same way, exposure to media coverage depicting tax authorities as powerful resulted in higher perceived power among participants. Both higher trust and perceived power were associated with greater intended tax compliance. This experiment highlights the potential role of media coverage in shaping taxpayers' perceptions of trust and power, which in turn may influence compliance intentions. These findings underscore the importance of considering media influence in tax policy efforts aimed at promoting compliance.

Sources

The Economic Science Association (ESA) has developed ethical guidelines for the conduct of lab experiments [link].

3.2.8. Randomised control trials

Randomised control trials are a form of scientific experiment that measures the effectiveness of interventions. In an RCT, participants are randomly assigned to either the treatment group, which receives the intervention, or the control group, which does not. This randomization helps ensure that any observed differences in outcomes can be confidently attributed to the intervention rather than other factors. RCTs have the unique ability to establish causal relationships between interventions and outcomes. In the analysis, statistical methods, such as regressions, and generalized linear models, are commonly employed to compare outcomes between the treatment and control groups[92]. Conducting RCTs can be resource-intensive, and as such, pilot and feasibility studies play a crucial role in the preparatory phase. These preliminary studies help to refine protocols, identify

[91] Kasper, M., Kogler, C., & Kirchler, E. (2015). Tax policy and the news: An empirical analysis of taxpayers’ perceptions of tax-related media coverage and its impact on tax compliance. *Journal of Behavioral and Experimental Economics*, *54*, 58–63. https://doi.org/10.1016/j.socec.2014.11.001

[92] Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: an international journal of obstetrics and gynaecology*, *125*(13), 1716. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6235704/pdf/nihms966617.pdf

potential challenges, and estimate necessary resources, ultimately enhancing the efficiency and cost-effectiveness[93].

Step of BASIC: 4 - Intervention. RCTs are expensive and are best employed as the final test of an intervention prior to widespread implementation, frequently after a pilot evaluation has already been conducted.

Strengths:

Establishing causation*.* The purpose of an RCT is to be able to infer causation between two variables. With a random sample tracked over time, with only one variable changing (the intervention for the treatment group), it is easier to infer causation. Therefore, it is a key method to determine if an intervention actually increases compliance rates in the target population.

External validity*.* Relative to lab experiments, RCTs benefit from taking place in participants' natural environment. These real-world conditions allow for a better generalization of study findings to diverse and everyday situations.

Reduced selection bias in sampling for the study*.* Selection bias arises when study participants are not accurately reflective of the target population, which does not necessarily mean they should represent the broader population. When selection bias occurs, observed effects may be influenced more by the characteristics of the selected participants than the actual intervention or treatment, leading to inaccurate generalisability of the study findings. RCTs are effective in preventing selection bias primarily through randomization. This helps to balance both known and unknown factors across the groups at the start of the trial, reducing the risk that the composition of the groups will influence the outcome, thus maintaining the study's validity and generalizability to the target population.

Successful randomization. Randomisation in RCTs helps to ensure that each participant has an equal chance of being in one of the study groups, thus distributing potential confounding variables evenly between the groups. This process mitigates the impact of confounding variables by ensuring these are equally distributed across all groups, which enhances the integrity and reliability of the trial results. Furthermore, successful randomization facilitates a more unbiased comparison between intervention outcomes. Conducting an RCT does not automatically result in a successful randomization. For example, if the randomization process is flawed or if certain groups of participants refuse to participate this will affect the generalizability of the trial's results.

Ability to track long-term effects. Unlike lab experiments, which tend to be one-time interactions, RCTs are usually conducted in real-world settings over long stretches of time, making it easier to observe how interventions perform over extended periods in diverse and complex environments. Long-term effects may be tracked by including follow-up assessments over the course of the RCT. Establishing a well-structured and consistent follow-up protocol allows researchers to capture the durability of effects and provides valuable insights into the long-term efficacy of the intervention under investigation in the RCT.

Weaknesses:

Expensive*.* There are several reasons why RCTs are often expensive. First, introducing a study in a real-world setting requires ethical and regulatory compliance that require planning, approval and monitoring procedures. Second, the intervention needs to be integrated in the actual tax journey, requiring changes in existing procedures and IT-systems. This often involves alignment between multiple departments within the tax authority. Third, given that RCTs often involve long-term follow-up rather than the one-time interaction associated with lab experiments, they generally involve the project teams working over a longer period of time.

Time-consuming*.* Due to the nature of RCTs involving following a group of participants over the course of a relatively long period of time, this type of study is naturally time-consuming. Shorter RCTs mitigate this problem, but at the expense of not being able to capture potential long-term effects and side-effects.

Sensitivity to external events. At the beginning of an RCT, the researchers cannot know for certain what may happen over the course of the study period, which in some cases may be years. Many RCTs were outright interrupted by the Covid-19 pandemic or had their external validity affected by the global financial crisis of 2008

[93] Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., ... & Moore, L. (2021). A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *bmj*, *374*. https://www.bmj.com/content/bmj/374/bmj.n2061.full.pdf

which temporarily altered a lot of consumer behaviour. This is particularly a concern when an RCT takes place in a volatile environment and/or over a long period of time which could see external events interfere with the treatment, and could also cause participants to drop out or have their characteristics change in such a way that they are no longer the target for the study (i.e. losing income leading to no longer paying taxes making someone uninteresting to most tax compliance RCTs)

Drop out. Depending on the nature of the study, participants dropping out or not adhering to the "treatment protocol" that they have been assigned to can be a major disadvantage with RCTs. If participants drop out or become unreachable, this may affect the demographic make-up such that the control group no longer resemble the treatment group(s). This weakness is only relevant for interventions that require some form of cooperation or active involvement of the participants (e.g., following an e-learning on financial budgeting).

Ethical considerations. Even though the aim of the experimenters is to improve compliance behaviour, this is not always what happens in practice. If an intervention backfires and harms participants, this is an ethical concern particularly if participants did not consent or did not know that they were part of a study. For example, participants may be randomly assigned to a treatment group that receives a letter intended to improve tax compliance by appealing to peer pressure. For participants who are non-compliant due to liquidity constraints or cognitive disorders, such a letter may cause emotional distress without improving compliance. To anticipate these potential adverse events, adherence to ethical guidelines is imperative.

Sources.

Evans formulated ethical guidelines for conducting an RCT [Link]

Case study 7: "The behaviouralist as tax collector: Using natural field experiments to enhance tax compliance"[94]

Background: This study focuses on behavioural interventions to improve the payment of outstanding tax liabilities. An estimated 16 percent of the gross tax gap (in the U.S.) stem from late or enforced payments, often costing tax authorities significant amounts of resources.

Objective: The experiments aimed to increase payment of tax liabilities with the use of targeted messages towards individuals who had not yet paid taxes that the tax authorities had determined that they owed.

Design: In the first experiment, 100 000 taxpayers from England, Wales and Northern Ireland who had declared their income but not yet paid their tax liabilities despite receiving a reminder received a letter from the tax authorities with one out of five messages. Which message each taxpayer received was random, with the randomisation process utilising the self-assessment reference numbers that each taxpayer is randomly assigned by the tax authorities. The messages were developed based on prior empirical findings and psychological theories, resulting in three messages that were norm-based and two that focused on public goods:

1) "Nine out of ten people pay their tax on time" (basic norm)
2) "Nine out of ten people in the UK pay their tax on time" (country norm)
3) "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" (minority norm)
4) "Paying tax means we all gain from vital public services like the NHS, roads and schools" (public goods, gain)
5) "Not paying tax means we all lose out on vital public services like the NHS, roads, and schools" (public goods, loss).

A control group received only a standard letter with basic information.

In the second experiment, 119 527 taxpayers received one of thirteen messages either appealing to minority norm, which the first experiment had shown to be the most effective, or to descriptive (what others do) or injunctive (what others think should be done) norms. Two of the groups received purely financial messages with warnings that interest rates would accrue on the amount and explanations of how they could pay.

[94] Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of public economics*, *148*, 14-31. https://www.nber.org/system/files/working_papers/w20007/w20007.pdf

Recipients' payments were tracked over 23 days. The main outcome measure was the percentage of tax debt that was paid during that period.

Outcome: All messages in the first experiment served to increase payment rates, which resulted in an estimated £3 million in additional revenue collected over the course of the 23 days. Notably, the minority norm appeal had the strongest effect, nearly three times greater than the second most effective appeal (country norm). The public goods appeals had a statistically significant effect, but surprisingly, the "loss" framing that appealed to loss aversion had no greater impact than the "gain" framing.

In the second experiment, appeals to norms again prove effective, replicating the findings of the first study, with appeals to descriptive norms outperforming messages appealing to injunctive norms. Like in the first experiment, minority norm appeals prove highly effective. The third category of treatment, the financial information, also increased payment rates.

3.2.9. Non-randomised control trials

Description: Non-randomised controlled trials (NRCTs) are research studies that investigate the effectiveness of interventions, treatments, or policies without using random assignment to allocate participants to different groups. This is a broad category, encompassing both studies with only a historical control, as well as studies where a control group is used but participants are not randomly allocated into the control and treatment groups. N-RCTs originated in medical research where the use of a control group can be considered ethically troublesome but has since spread to other areas where researchers feel a control group is unnecessary either for ethical or practical reasons (i.e. where researchers believe they already know how a hypothetical control group would act based on previous data).

Step of BASIC: 3 – Strategy and 4 - Intervention. A non-randomised trial could take place in the third step as an alternative to, for example, a lab experiment, but would more typically take place in step 4 as a cost-effective or acceptable alternative to an RCT.

Strengths

- ***Cost-efficiency***. In studies using a historical control, fewer participants are needed as no participants have to be recruited into a control group.
- ***Feasibility***. Due to the smaller number of participants, and lower costs, an N-RCT can be a more feasible option when budget restraints do not allow for conducting an RCT.
- ***Real world applications***. Like RCTs, N-RCTs also take place in the real world and typically involve follow-up over relatively long periods of time. As such, their findings are easier to apply to a real-world setting than, for example, the findings from lab experiments.
- ***Ethics.*** In cases where interventions have a potentially significant positive impact on a group, for example an extensive intervention helping low-income taxpayers avoid falling into arrears, it can be argued that using a control group is ethically questionable when the researchers could offer the intervention to all participants instead.

Weaknesses:

- ***Selection bias***. Without randomisation, recruitment, or allocation to a group may be influenced by non-random factors such as taxpayer preference, or tax inspector choice. This can distort the observed treatment effects, making it challenging to attribute the outcomes solely to the intervention
- ***Difficulty establishing causation***. Interpreting results from Non-Randomized Controlled Trials (NRCTs) requires caution, particularly regarding causality. Due to the treatment and control groups not being randomly drawn from the same group of people (or due to the absence of a control group), it is harder to establish whether any changes in the treatment group's behaviour (i.e. increased tax compliance) is due to the treatment or due to the treatment group being different from the control group in some regard. When interpreting the findings, people must carefully consider other potential explanations for the results, relying on a thorough understanding of the study context and existing literature to draw conclusions.

- ***Data analysis challenges***. Data analysis is more complicated and sensitive to possible confounding factors, requiring sophisticated statistical techniques to handle. If confounding variables are not dealt with properly, results may be biased.

Sources.

Evans formulated ethical guidelines for conducting an RCT, guidelines that also apply to N-RCTs [Link]

3.2.10. Natural experiments

Description: A natural experiment is not per se an experiment, but rather a type of research that allows for casual inferences from observational data. In a natural experiment, researchers identify a case where a policy, such as an intervention to enhance compliance, is implemented in one country or region. Subsequently, they seek out a comparable country or region that did not adopt the policy, serving as a control group. It is imperative that this control group exhibits similarities, including comparable levels of compliance, similar economic and demographic characteristics, and a parallel compliance trend leading up to the policy intervention. Researchers would then look for a structural break, i.e., a sudden improvement in compliance after the policy was implemented that did not occur in the country or region that did not implement the policy.

Strengths:

- ***Real-world context and data***. Even more so than RCTs, natural experiments rely on data from the real world, providing a solid basis for external validity.
- ***Low cost***. In particular relative to RCTs, natural experiments, since they rely on already-collected data, are a budget-friendly option.
- ***Ability to infer causality***. Relative to non-experimental empirical studies, natural experiments have an advantage in establishing causal relationships due to focusing on similar regions before and after a policy was introduced and using statistical methods to identify structural breaks. However, establishing causality in natural experiments is contingent on the assumption that the treatment and control groups are similar in all relevant aspects except for the intervention, and that there are no other confounding factors influencing the observed outcomes.
- ***Ethics***. As the researchers are not involved in manipulating the behaviour of any taxpayer, there are few if any ethical downsides to natural experiments.
- ***Ability to track long term effects***. Without the heavy cost associated with conducting RCTs, natural experiments can theoretically study the impact of a policy indefinitely.

Weaknesses:

- ***Lack of control***. Since natural experiments can only be used for policies and interventions already applied in the real world and where there is a similar region that did not apply the policy and thus can act as a control group, this limits the control of the researcher. For newly developed policies that have never been applied before, natural experiments are not an option.
- ***Identifying an appropriate "control group" is often difficult***. If a difference between the treatment and control group is overlooked, this will affect the credibility of the findings of the experiment. And while natural experiments in theory allows for long-term inferences of a policy's impact, this assumes that the control and treatment group do not diverge in some fundamental way during that time period.
- ***Causal inferences may be muddled***. It is rare that policies are implemented in isolation, and a number of policies may be implemented in either the treatment or control group that may have an impact on the dependent variable.
- ***Limited replicability***. Natural experiments are frequently one-off events and cannot be easily replicated unlike lab experiments and RCTs.

Case study 7: “Nudges and threats: soft versus hard incentives for tax compliance[95] (Study 1)”

Background: In Sweden, individual taxpayers with tax liabilities are treated differently depending on the size of their debt. At levels below SEK 2,000 (around EUR 200), unpaid tax debts do accrue interest, but no further action is taken by the tax authorities to enforce payments. At levels above 2,000, unpaid debts are handed over to the national Enforcement Authority (EA), which in turn may seize assets and/or garnish the wages of the noncompliant taxpayer.

Objective: In the paper, the researchers use observational data from 2016 and 2017 to estimate the effects of the standard enforcement strategy, focusing on whether individuals end up on either side of the SEK 2,000 threshold for enforcement. The study aims to interpret the estimated effects as causal effects and assesses the impact of the threat of enforcement on taxpayer behaviour.

Design: In the analysis, they compared the compliance rate of individuals subjected to recovery measures (those just above the SEK 2,000 threshold) with individuals in the natural control group (those just below the SEK 2,000 threshold, and who did not undergo any measures).

Outcome: This study suggests that the mere threat of enforcement has a significant impact on taxpayer behaviour. The analysis reveals distinct effects of the threat of being transferred to the EA, particularly around the SEK 2,000 threshold, with an even greater effect from actually being transferred to the EA.

[95] Andersson, H., Engström, P., Nordblom, K., & Wanander, S. (2023). Nudges and threats: soft versus hard incentives for tax compliance. *Economic Policy*, *38*(116), 771–819. https://doi.org/10.1093/epolic/eiad017

Annex: Clarification of Terms

Behavioural Insights refers to applying knowledge from the behavioural sciences to understand how people make decisions and to design interventions that encourage desirable behaviours.

Choice Architecture refers to the designing of environments and/or systems that influence people's decisions by presenting options in a way that nudges them towards particular choices (i.e. paying their taxes) while preserving their freedom of choice.

Drivers are the underlying variables that predict individual tax compliance. Some drivers (i.e. trust) are modifiable and may be targeted by behavioural interventions, whereas others (i.e. age) are not.

Extensive margin measures binary outcomes, such as whether an intervention made taxpayers file on time or pay any of the tax debt they owed.

Intensive margin by contrast evaluates the total amount of tax raised or change in deductions claimed from an intervention.

Spillover Effects are effects on behaviour stemming from an intervention other than those effects that the intervention aimed at (i.e. an income tax intervention affecting real estate tax compliance). Spillover effects can be both positive and negative and should be thoroughly considered when evaluating the intervention.

Tax Journey is a term for the process or path that individuals go through when fulfilling their tax obligations, including activities such as filing tax returns, paying taxes, potentially seeking advice from tax professionals, and responding to tax audits or inquiries.

Triangulation refers to using multiple methods, data sources, or researchers to corroborate findings and enhance the validity and reliability of the research outcomes.

BASIC Framework. The acronym is short for Behaviour, Analysis, Strategies, Intervention and Change. A framework developed by the OECD to guide the implementation of behavioural insights into policy.

Nudge. An intervention that guides individuals towards making better decisions without restricting their freedom of choice.

ABCD Model. The acronym is short for Attention, Belief formation, Choice and Determination, and is a way of compartmentalising and analysing behavioural problems.

Appendix A: Field RCTs extraction table

First author, year	Country	Sample	Study objective	Mechanism categories	Intervention channel	Study results
Hallsworth, 2017	UK	200.000	Increase payment rate of overdue income tax	Descriptive norms, Injunctive norms, Minority norms, Value of public goods, Information provision: deterrence, other mechanisms	Physical letters sent to taxpayers	Minority norm framing increased payment rates the most (the highest effectiveness in both Experiment 1 and 2); Social norm and public goods messages both increased the payment rate; Descriptive norms have a significantly larger effect than injunctive norms on payment rates; Loss/gain framing of the public services did not change the effects; Financial messages also improved payment rate
Luts, 2019	Belgium	225.877	Increase tax payment with behaviourally informed letters	Simplification, Social norms, Value of public goods, Commission/omission framing, Information provision: deterrence, other mechanisms	Physical letters sent to taxpayers	A 9.2 p.p. overall increase in the rate of payment for the behaviourally informed letters compared to the former reminder letter; 7.4 p.p. increase in fully paid items; 10.9 p.p. increase in payment by the most successful treatment, which was the Explicit Penalty treatment (whether or not in one letter combined with the 'Active Choice'); the increase in payment for the various soft message treatments (e.g., public goods) were roughly the same as that of the 'new control' letter which was behaviourally informed but did not include those persuasive messages
Hernandez, 2017	Poland	149.925	Increase payment rates with behaviourally informed letters	Social norms, Minority norms, Value of public goods, Commission/omission framing, Information provision: deterrence, Information provision: non-deterrence, Other mechanisms	Physical letters sent to taxpayers	All the letters performed better than the traditional letter; 8.4 p.p. increase in payment rate compared to the standard government letter by the most successful letter (which was omission/commission combined with deterrence); all other behaviourally informed letters performed either the same, or worse than the standard behaviourally informed letter, specifically Public goods (positive), Pubic goods (negative), Social norms, and Deterrence performed significantly worse; the most successful letter (omission/commission), increased average payment amount by 347 PLN (roughly 80 EUR)compared to the old letter; 6.7 p.p. higher payment rate for the 'hard-tone' messages compared to the 'soft-tone messages' (this was based on grouping the messages together, e.g., public goods positive being a

						soft tone message); sending through registered (vs non-registered) mail did not make a difference
Bott, 2020	Norway	15.000	Increase self-reporting of foreign income	Information provision: non-deterrence, Value of public goods, other mechanisms	Physical letters sent to taxpayers	Moral suasion has a large effect on self-reported foreign income; For the follow-up year (right after the letter was sent), the average self-reported foreign income by the taxpayers who received one of the moral letters was almost double the amount self-reported by those who received the base letter; large effect of the detection letter, but the moral letters and the detection letter affect different margins of the taxpayer behaviour: the detection letter has a large effect on the extensive margin (the share of individuals who reported), whereas the moral letters only have a large effect on the intensive margin (the amount reported); the base letter (with information) itself had some effect on self-reported foreign income, but, overall, the underreporting seems to not primarily be driven by a lack of knowledge about how to report foreign income
Dong, 2021	Australia	18.000	Increase lodgement rate through letters	Information provision: deterrence, Information provision: non-deterrence	Physical letters sent to taxpayers	Both letters have large effects of about 14-15 percentage points on lodgement rates when compared to a control group that does not receive any communication from the tax authority over the duration of the experiment; The effect of the deterrence letter is not statistically different from that of the non-deterrence letter, suggesting that the primary mechanism of the interventions is provision of information; Australian citizens were more responsive to the letters
Perez-Truglia, 2018	US	34.334	Increase payment of owed tax through shaming letters	Information provision: deterrence, Shaming, Other mechanisms	Physical letters sent to taxpayers	Increasing the visibility of delinquency status increased compliance (2.1 p.p.) by individuals who have debts below $2500 (1st quartile), but had no significant effect on individuals with larger debt amounts; Financial reminders had a positive effect on payment rates independent of the size of the debt (0.7p.p.), while information about the delinquency of neighbours had no effect on payment rates.

John, 2018	UK	66.000	Examine the effectiveness of simplification of documentation, and social norm nudges on tax compliance	Simplification, Social norms	Tax bills with varied information	Simplification increased the number of people paying on time by four percentage points; the social norm did not change behaviour; In wave two of the study, which was carried out across all households, the descriptive social norm backfired, reducing the rate of payment
Kettle, 2017	Guatemala	627.242	Increase people's honest filling in of tax declarations, to increase non-zero declarations and the amounts declared	Other mechanisms	Variations of CAPTCHA	None of the interventions were effective at influencing taxpayer behaviour (the authors speculate that it might be because the prompts might have not invited thought about their content, and were too separate from the form itself, or that the decision about dishonesty might usually be made before already)
Koessler, 2019	Switzerland	2.201	Increase timely payment of taxes through asking taxpayers for commitment beforehand, and giving the opportunity to enter a lottery	Positive Incentives, Other mechanisms	Physical letters sent to taxpayers	Compliant taxpayers (from the previous year) were more likely to make a pledge (selection effect); the lottery with a non-financial reward (wellness weekend) was more effective than control at increasing timely payment, but the financial one was not (both were combined with an honesty pledge); solely giving the opportunity to make the promise does not increase timely payment compared to control
Schächtele, 2022	Brazil	163.260	Increase taxpayer registrations	Information provision: non-deterrence, Positive incentives	Physical letters sent to taxpayers	The payment rate increased for the group that was informed; however, the lottery backfired with lower compliance rate (even compared to the control), particularly among the owners of expensive properties (this is explained by the signalling of low enforcement which might have been generated by the lottery)
Vainre, 2020	Estonia	4.770	Increase compliance in filling in payroll information through an e-mail intervention containing multiple nudges tailored to the context	Information provision: deterrence, Information provision: non-deterrence, other mechanisms	E-mails sent to taxpayers	The e-mails improved tax declarations per employee, on average by 5.1-6.1% across the two intervention groups compared to not receiving an e-mail
Andersson, 2023	Sweden	57.000	Compare the effectiveness of various BI letters on	Reminders, Positive Incentives, Social norms	Physical letters sent to taxpayers	For the group that was below enforcement threshold, the effect of the pure reminder nudge on payment was 7 p.p. (roughly the same as stating the potential consequences

			payment of overdue taxes			of enforcement); for the group that was above the enforcement threshold, the same nudge had no effect (the authors suspect it might have been less salient in comparison to the information on enforcement the payers received anyway - but this is not very clear in the paper); social norm messages had a small effect for both the groups below enforcement threshold and above it (around 2 p.p. increase in probability of paying); the effects were pronounced only immediately after sending the letters, but not in the two months that followed
Zhang, 2023	U.S.	50.000	Test the effectiveness of various behaviourally informed letters on timely and accurate reporting of wages of those eligible for the benefits programme	Reminders, Positive Incentives, Information provision: non-deterrence, Descriptive norms	Physical letters sent to taxpayers	The letter reminding SSI recipients of their wage reporting responsibilities significantly increased both the likelihood of reporting any earnings (0.34 p.p. increase over the 0.97 % of the control group, which corresponds to 11.1%) and the total earnings reported (55.1% which corresponded to 4.89 $ higher income on average, than the control group)
Antinyan, 2021	China	1.742	Examine the effects of the frequency of SMS reminders on compliance in the context of property tax	Reminders, Information provision: deterrence	SMS messages to taxpayers	The probability of overdue tax payments in medium frequency (once a week) and high frequency (twice a week) is around 12–14 percentage points higher than in Control (more than 300% increase in compliance) and 5–7 percentage points higher than in low frequency (one SMS only), which amounted to around 40–60% increase in compliance.; The most effective reminders were those once a week (twice a week might have led to habituation, as the authors suggest).
Eerola, 2019	Finland	45.000	Investigate the effectiveness of various letters to potential landlords on reporting of rental income	Simplification, Information provision: deterrence, Other mechanisms	Physical letters sent to taxpayers	The strongest effects are found for the treatment letter that notified potential landlords of the use of third-party information on the ownership of apartments in tax enforcement; largest effects for individuals who had reported no rental income in the year prior to treatment; no spillover effects were found between landlords in the local rental markets; positive spillover effects where found within households

Castro, 2019	Argentina	54.000	Evaluate the effect of simplification of a notice on tax amnesty on tax payment, and also evaluate the effects of such treatments on future compliance (therefore, capturing the effects of amnesty as a policy)	Simplification, Positive Incentives	Physical letters sent to taxpayers	On average, 5% higher amount of tax paid (due to having entered the amnesty) among the tax payers who received the simplified letters compared to the control group; in the following year, those who were in the treatment group in the prior year, reduced their compliance; negative spillover effect was found: Increasing the share of people living in the same geographical block that received a new notice by 10%, increases the probability of non-payment for the previously compliant taxpayer by about 0.6pp
Jamison, 2021	Latvia	4.324	Investigate the effects of behaviourally informed e-mail messages on tax declaration	Reminders, Commission/Omission framing, Positive Incentives, Descriptive norms	E-mails sent to taxpayers	Number of declaration submissions was significantly higher (4.2p.p.) in the commission/omission group compared to the control group (without a message); simple reminder resulted in an insignificant increase; social norms message resulted in an insignificant increase.
Persian, 2022	Indonesia	11.157.069	Investigate the effect of behaviourally informed prompts on timely filing of individual income tax	Simplification, Information provision: non-deterrence, Other mechanisms	E-mails sent to taxpayers	All treatments performed better than the 'no e-mail control'; the simplified email that included planning prompts and asked people to sign up for further automatic email reminders performed better than all other treatment and control emails for the early filing outcome, and for all treatments except the simplified email with guidance for the overall filing outcome. Specifically, the planning email increased overall and early filing rates by 1.10 and 2.07pp, respectively, compared to the 'no-e-mail' control; email appealing to taxpayers' national pride by graphically showing what public goods the taxes are spent on, performed worse than all treatments except for the two control conditions; secondary analysis showed that the e-mails had less effect on the self-employed, and the authors hypothesise that this might have been due to the larger complexity of that filing system and the preparation of filing by accountants; overall, the treatments did not have an effect on the net amount of tax paid
Hallsworth, 2023	UK	40.000	Evaluate the effects of e-mails with various framing on individuals paying back	Commission/Omission framing, Simplification, Other mechanisms	E-mails sent to taxpayers	"The "Reciprocity," "Planning," and "Salience" treatments do not have a significant effect on repayment rates; The "Opportunity" treatment produced a 2.2% increase in payments. The effect in comparison to the control group

			government benefits that were overpaid (typically due to them not reporting changes in their eligibility)			is an 18.5% (0.06 SD) effect size (but the effect disappeared when adjusting for covariates); In terms of the omission to commission messages, the “Collective Omission to Commission” group created an 11.2% point increase in payment rates, which is equivalent to a 94.1% (0.33 SD) treatment effect size; the “Individual Omission to Commission” group produced a 10.9% point increase, which is equivalent to a 91.6% (0.32 SD) treatment effect size."
Chirico, 2019	US	19.039	Evaluate the effect of behavioural interventions on timely payment and payment rate of a property tax	Information provision: deterrence, Value of public goods, Descriptive social norms, Other mechanisms	Physical letters sent to taxpayers	Receiving the 'reminder-only' letter increases the rate of compliance after one month for an ever-paid tax payment by 3.7 percent above the holdout’s rate of compliance and by 3.9 % after three months (for the total amount paid the results were 2.2% after one month and 3% after three months); the effects of emphasising public goods provision, descriptive social norms, civic duty, had effects very similar to those of the reminder-only letter; the economic deterrence letters were more effective than the reminder, with 8.8% or 9.2% after three months on the probability on any payment being made, and 6.7% or 7.3% compared to the 'no letter' group, respectively, for the letter about a lien potentially being placed on the property, and the letter about a potential sale of the property by the sheriff if taxes are not paid (more comprehensive results in Table 2); in the next tax year, there was no effect that the reminders sent in the previous year made any difference in compliance
Cranor, 2020	US	90.000	Evaluate the effect of letters with messages framed in behaviourally informed ways	Information provision: deterrence, Descriptive norms, Minority norms	Physical letters sent to taxpayers	Detailed penalty notice significantly increased the fraction of taxpayers making a full payment before the statutory deadline or creating a payment plan by 1.6 p.p. (a 4.1 percent increase) relative to the control notice; the effect of the non-detailed information on the penalty was half the effect of the detailed penalty notice (and not significant); the effect of social norms treatment was even smaller and also not significant relative to control; an analysis of the heterogeneity of the effect revealed that only the tax payers with small-to-medium debt size were responsive to the threats of penalty, while those with large debt were not; follow-up measures show that part of the effect

						remains even 100 days after receiving the letters (but is smaller than after the 45 days)
Meiselman, 2018	USA	9.524	Increase filing rates among (suspected) non-filers through different types of letters	Information provision: deterrence, Simplification, Other mechanisms	Physical letters sent to taxpayers; Postcards sent to taxpayers; Tax forms sent to taxpayers	Penalty salience message is the most effective at increasing filing rates within every demographic subgroup; penalty salience message is also the most effective at increasing amount of taxes paid; only appeal to civic pride treatment group did not show significant improvement relative to control group that received only a basic information letter; all interventions had a net negative effect on social welfare though this depends greatly on assumption about social value of public spending, the cost of compliance, and possible future deterrence effects; no network effects or social learning was detected from any of the treatments
Guyton, 2016	United States	400.000	Investigate the impact of reminders on individuals' tax filing behaviour, particularly in the context of low-frequency decisions related to filing tax returns. The study aims to analyse the effectiveness of one-time and repeated reminders in reducing inattention, understanding the persistence of reminder effects, and exploring their implications for policy interventions in the realm of tax compliance.	Reminders	Physical postcards sent to taxpayers	Reminders effectively reduce inattention and increase tax filing, particularly among lower-income nonfilers. While reminders impact filing in the current year only, repeated reminders reduce recidivism, and individuals' responsiveness to reminders is influenced by past experiences.

Boyer, 2016	Germany	39.788	Investigating through a field experiment the potential crowding out of intrinsic motivation towards tax compliance. Study focuses on the German church tax and the role of taxes as externally imposed norms on contribution behaviour. Study investigates the motivational effect of externally imposed norms in a setting where we are able to obtain a good proxy for individuals' initial strength of intrinsic motivation.	Other mechanisms	Physical letters sent to taxpayers	Individuals with regular baseline contributions (the strongly intrinsically motivated) did not show any response to the information that the church levy is a tax; individuals who contributed only sometimes in the past (those with weaker intrinsic motivation) on the other hand, individuals in this group reduce their payments in response to the voluntary tax letter while the compulsory tax letter has no effect, suggesting that imposing norms on contribution behaviour does crowd out intrinsic motivation, but a sufficiently binding tax norm can compensate for this; baseline non-contributors increase their payments by a significant amount if they receive the compulsory tax letter but do NOT respond to the voluntary tax letter that communicates the existence of a non-binding legal norm
Lopez-Luzuriaga, 2019	Argentina	700	Investigating the spillover effects on compliance with one tax of an intervention that targeted another tax.	Other mechanisms	Physical letters sent to taxpayers	Taxpayers who received the treatment in the property tax declared more and were more likely to pay their gross-sales taxes than those in the control group, suggesting a positive spillover effect
De Neve[96], 2021	Belgium	1.216.317	investigate the effectiveness of behaviourally informed tax bills (physical letters) at increasing compliance in payment	Simplification, Reminders, Social norms, Minority norms, Descriptive norms, Information provision: Deterrence, Value of public goods, other	Physical letters sent to taxpayers	Simplifying the tax bill had a positive effect on the probability of paying on time, increasing it by 0.5 percentage point; Adding a deterrence message increased the probability of paying on time by a further 0.5 percentage point (thus, combined effect of simplification and deterrence is 1 p.p.). These effects are

[96] Four studies from the same publication are included in the following order: 1) Tax Payment experiment; 2) Reminder about Tax Payment experiment; 3) Tax filing experiment; 4) Tax filing reminder experiment.

			of taxpayers who declared a positive liability.			relatively small but significant: the combined effect of simplification and deterrence messages is 1.4% of the control mean (72.8%); The tax morale messages had no effect on compliance compared to simplification only
De Neve, 2021	Belgium	432.481	investigate the effects of behaviourally informed reminder letters on compliance across two tax years	Simplification, Reminder, Social norms, Minority norms, Descriptive norms, Information provision: Deterrence, Value of public goods, other	Physical letters sent to taxpayers	Simplifying the reminder letters increased the probability of payment by 10 percentage points (22.8% of the control mean); deterrence messages had an additional positive effect of 1.2 percentage points (2.7% of the control mean); Tax morale messages, however, had an opposite effect, slightly reducing tax compliance (−0.7 percentage point, or 1.6% of the control mean); The probability of paying taxes on time in the next tax year increased 1.3 percentage point in the simplification treatment (this long-term effect of simplification of the reminder letter is twice as large as the short-run effect of the simplification of the tax bill itself (0.5 percentage point increase in the probability of meeting the deadline); no evidence that simplification was less effective for those who had received a simplified letter in the previous year (sending the letter twice does not diminish effectiveness)
De Neve, 2021	Belgium	1.500.000	investigate the effectiveness of presenting the breakdown of government spending on public goods before individuals file their taxes online	Value of public goods, social norms, Other	Physical letters sent to taxpayers	The tax morale treatment (presenting the pie chart) in the tax filing experiment had no effect on declared taxable income; the survey results show that while the pie chart treatment was effective in improving taxpayers' knowledge and appreciation of how their taxes were spent, it fell short of improving their tax morale.
De Neve, 2021	Belgium	148.925	investigate the effects of behaviourally informed reminder letters on the filing taxes	Information Provision: Deterrence	Physical letters sent to taxpayers	Simplification and deterrence had a large positive effect on tax compliance among late filers.; Those who received a simplified letter were 2.6 percentage points more likely to file on time.; This probability increased by an additional 2.8 percentage points for those who received a simplified letter with a deterrence message, making them 17% more likely to file on time than the control group.

References

Andersson, H., Engström, P., Nordblom, K., & Wanander, S. (2023). Nudges and threats: soft versus hard incentives for tax compliance. *Economic Policy*, eiad017.

Antinyan, A., Asatryan, Z., Dai, Z., & Wang, K. (2021). Does the frequency of reminders matter for their effectiveness? A randomized controlled trial. *Journal of Economic Behavior & Organization*, *191*, 752-764.

Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2020). You've got mail: A randomized field experiment on tax evasion. *Management science*, *66*(7), 2801-2819.

Boyer, P. C., Dwenger, N., & Rincke, J. (2016). Do norms on contribution behavior affect intrinsic motivation? field-experimental evidence from germany. *Journal of Public Economics*, *144*, 140-153.

Castro, E., & Scartascini, C. (2019). Imperfect attention in public policy: A field experiment during a tax amnesty in Argentina.

Chirico, M., Inman, R., Loeffler, C., MacDonald, J., & Sieg, H. (2019). Deterring property tax delinquency in philadelphia: An experimental evaluation of nudge strategies. *National Tax Journal*, *72*(3), 479-506.

Cranor, T., Goldin, J., Homonoff, T., & Moore, L. (2020). Communicating tax penalties to delinquent taxpayers: Evidence from a field experiment. *National Tax Journal*, *73*(2), 331-360.

De Neve, J. E., Imbert, C., Spinnewijn, J., Tsankova, T., & Luts, M. (2021). How to improve tax compliance? Evidence from population-wide experiments in Belgium. *Journal of Political Economy*, *129*(5), 1425-1463.

Dong, S. X., & Sinning, M. (2021). Trying to make a good first impression: A natural field experiment to engage new entrants to the tax system. *Tax and Transfer Policy Institute-working paper*, *6*.

Eerola, E., Kosonen, T., Kotakorpi, K., Lyytikäinen, T., & Tuimala, J. (2019). Tax compliance in the rental housing market: Evidence from a field experiment. *VATT Institute for Economic Research Working Papers*, *122*.

Guyton, J., Manoli, D. S., Schafer, B., & Sebastiani, M. (2016). *Reminders & recidivism: evidence from tax filing & eitc participation among low-income nonfilers* (No. w21904). National Bureau of Economic Research.

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of public economics*, *148*, 14-31.

Hallsworth, M., List, J. A., Metcalfe, R. D., Rotaru, K., & Vlaev, I. (2015). The making of homo honoratus: From omission to commission. *Journal of Consumer Psychology*.

Hernandez, M. et al. 2017. Applying Behavioral Insights to Improve Tax Collection: Experimental Evidence from Poland. Washington, DC: World Bank. url : https://doi.org/10.1596/27528 .

Jamison, J. C., Mazar, N., & Sen, I. (2021). Applying behavioral insights to tax compliance: experimental evidence from Latvia.

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

Kettle, S., Oré, M. A. H., Ruda, S., & Sanders, M. (2017). Promoting Tax Compliance in Guatemala Using Behavioral Economics: Evidence from Two Randomized Trials. *Behavioral Insights for Development: Cases from Central America*.

Koessler, A. K., Torgler, B., Feld, L. P., & Frey, B. S. (2019). Commitment to pay taxes: Results from field and laboratory experiments. *European Economic Review*, *115*, 78-98.

López-Luzuriaga, A., & Scartascini, C. (2019). Compliance spillovers across taxes: The role of penalties and detection. *Journal of Economic Behavior & Organization*, *164*, 518-534.

Meiselman, B. S. (2018). Ghostbusting in Detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics*, *158*, 180-193.

Perez-Truglia, R., & Troiano, U. (2018). Shaming tax delinquents. *Journal of Public Economics*, *167*, 120-137.

Persian, R., Prastuti, G., Bogiatzis-Gibbons, D., Kurniawan, M. H., Subroto, G., Mustakim, M., ... & Sutherland, A. (2023). Behavioural prompts to increase early filing of tax returns: a population-level randomised controlled trial of 11.2 million taxpayers in Indonesia. *Behavioural Public Policy*, *7*(3), 701-720.

Roy, M. V., & Luts, M. (2019). Nudging in the context of taxation. https://www.iota-tax.org/sites/default/files/documents/iota_paper_belgium_nudging_final.pdf

Schächtele, S., Eguino, H., & Roman, S. (2022). Improving taxpayer registration through nudging? Field experimental evidence from Brazil. *World Development*, *154*, 105887.

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

Zhang, C. Y., Hemmeter, J., Kessler, J. B., Metcalfe, R. D., & Weathers, R. (2023). Nudging Timely Wage Reporting: Field Experimental Evidence from the US Supplemental Security Income Program. *Management Science*, *69*(3), 1341-1353.

Appendix B: Table comparing study methodologies described in Module 3.

Study methodology	Description	Strengths	Weaknesses
Field Randomised Control Trials	Quantitative method in which participants are randomly assigned to either the treatment group, which receives the intervention, or the control group, which does not.	• Establishing causation. • External validity. • Reduced selection bias. • Ability to track long-term effects.	• Expensive. • Time-consuming. • Sensitivity to external events. • Drop out. • Ethical considerations.
Non-randomised Control Trials	Quantitative method in which the effectiveness of interventions, treatments, or policies is assessed without using random assignment to allocate participants to different groups.	• Cost-efficiency. • Feasibility. • Real world applications. • Ethics.	• Selection bias. • Difficulty establishing causation. • Data analysis challenges
Online experiments	Participants engage with study materials on online platforms.	• Lower cost. • Reduced self-selection bias. • Rapid Data Collection. • Diverse sample.	• Greater risk of unserious or inattentive participants. • Certain interventions cannot be (easily) replicated online. • Difficulty reaching less-online populations.
Lab experiments	Quantitative method which takes place within a physical space where participants are randomly allocated into control and treatment groups, and then typically given a number of tasks.	• Offers maximum degree of control. • Ability to infer causation. • Isolation of factors. • Relatively low cost. • Ease of replicability.	• External validity. • Limited sample size. • Self-selection bias. • Hawthorne effect. • Earned vs unearned endowments. • Lack of real incentives.

Quantitative surveys	Method employing structured questionnaires to gather numerical data, aiming to analyse relationships between the relevant variables.	• Larger sample sizes. • Easier to interpret data.	• Inability to determine causation. • Sensitivity to wording. • Limited depth.
Natural experiments	Quantitative research that allows for casual inferences from observational data.	• Real-world context and data. • Low cost • Ability to infer causality. • Ethics. • Ability to track long term effects.	• Lack of control. • Identifying an appropriate "control group" is often difficult. • Causal inferences may be muddled. • Limited replicability.
Focus groups	Qualitative methodology in which small groups of individuals are gathered together to answer questions and discuss a certain topic in a conversation led by a researcher.	• Allows for in-depth discussions between participants. • Candid responses. • Real-time feedback. • Follow-up questions.	• Risk of bias. • Unpredictable in-group dynamics. • Self-selection bias.
One-on-one interviews	Qualitative methodology in which a researcher Interviews individual stakeholders with various degrees of structure.	• Individual autonomy. • Equal individual focus. • Tailoring.	• Resource-intensive. • Social desirability.
Case studies	Researchers deep-diving into a particular case.	• Rich contextual understanding. • Holistic examination. • Tailored recommendations.	• External validity. • Difficult to establish causation.

Appendix C: Index of case studies, tables, boxes and figures

verian

Funded by
the European Union

Visit our website:

Find out more
about the Technical
Support Instrument: