Journal of Economic Behavior and Organization 235 (2025) 107074

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Why do some nudges work and others not?

Matej Lorko [a,*], Tomáš Miklánek [b], Maroš Servátka [c]

[a] Faculty of Economics and Finance, University of Economics in Bratislava, Slovakia, and Faculty of Business Administration, Prague University of Economics and Business, Czech Republic
[b] Faculty of Business Administration, Prague University of Economics and Business, Czech Republic
[c] MQBS Experimental Economics Laboratory, Macquarie Business School, Sydney, Australia, and Faculty of Economics and Finance, University of Economics in Bratislava, Slovakia

ARTICLE INFO

Keywords:
Nudge
Incentives
Tax compliance
Experiment

ABSTRACT

While nudges have recently gained popularity, many nudging interventions fail, and the effects of successful ones are often short-lived. We conjecture that the success of a nudge depends on how it interacts with the underlying economic incentives that determine the payoff-maximizing behavior of the decision-maker. For example, in the domain of tax compliance, a nudge is likely to be effective only if it is financially optimal for the taxpayer to pay the tax. To test our conjecture, we run a multi-period experiment in which we manipulate tax audit probability and nudge participants to report their income. In addition, we vary how often the nudge appears to test whether more frequent nudging increases long-run compliance. We observe that the first application of a nudge has a positive immediate effect on income reporting, irrespective of whether it is optimal to comply or not. However, subsequent nudges increase income reporting only if the nudge is aligned with the taxpayer's incentives. More frequent nudging in the direction opposite to incentives yields no effects on long-run compliance. Policy implications are discussed.

1. Introduction

Thaler and Sunstein (2008, p.6) define nudges as "aspects of the choices architecture that alter people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives". These inconspicuous and cost-efficient policy interventions have recently gained prominence, with governments around the world establishing over 200 behavioral insights teams (also known as "nudge units") to suggest, test, and apply behavioral nudges (DellaVigna and Linos, 2022; Hallsworth, 2023; Hubble and Varazzani, 2023).

Behavioral research has shown that for nudges to work, they must be easy, attractive, social, and timely (BIT Handbook, 2014). Yet many nudges across diverse domains (e.g., healthcare, environment, finance) are found to be ineffective despite satisfying these criteria. Therefore, it is natural to ask why some nudges work and others do not. Given the large number of implemented nudges and studies analyzing their outcomes, the main approach to answering this question has been conducting meta-analyses (e.g., Hummel and Maedche, 2019; Mertens et al., 2022; Bruns et al., 2025). While insightful about the context and type of interventions and their features playing a role, the approach is inconclusive about *why* this happens.

In this study, we propose a conjecture that a nudge will work only if it is aligned with the underlying economic incentives. This simple yet fundamental idea seems to have been ignored by policymakers and, to the best of our knowledge, has not received attention

* Corresponding author.
E-mail addresses: matej.lorko@gmail.com (M. Lorko), tomas.miklanek@vse.cz (T. Miklánek), maros.servatka@mgsm.edu.au (M. Servátka).

https://doi.org/10.1016/j.jebo.2025.107074
Received 10 January 2025; Received in revised form 4 May 2025; Accepted 15 May 2025
Available online 24 May 2025

in empirical research.[1,2] We conjecture that if a nudge goes against economic incentives, it will be ineffective as the economic forces correct behavior that is not in the decision-maker's best interest. On the other hand, a nudge aligned with economic incentives will be effective and persist over time as it reinforces optimal behavior. In repeated contexts, where nudges can be reapplied, a subsequent question emerges: Does more frequent nudging foster the desirable behavior?

We conduct a laboratory experiment designed to illustrate how incentives may interact with a nudging intervention. Our approach is complementary to randomized controlled trials, field experiments, and meta-analysis in that it allows us to carefully control the economic environment in which the data is collected. We explore our general research question about the interaction of nudges and economic incentives in the context of tax compliance, where nudges have been extensively applied and studied, and where we can easily manipulate whether it is in the best (expected) monetary interest for a decision-maker to report income and pay tax owed or not. Our findings have important implications for behavioral science and policy and highlight the relevance of existing incentives when designing interventions.

2. Nudging and tax compliance

Tax compliance nudges are eminently attractive to governments, complementing systematic policy instruments such as enforcement (Slemrod, 2019; Slemrod and Yitzhaki, 2002). Given the dramatic magnitudes of tax evasion (see e.g., Internal Revenue Service, 2022; Kukk et al., 2020), even if the nudge increases the compliance rate only slightly, it can add immense revenues to government budgets. A rich variety of nudges have been proposed and applied, ranging from moral appeals, appeals to social norms or peer examples (Hallsworth et al., 2017), highlighting public goods provision (Bott et al., 2020), to correcting procrastination and limited attention (Hernandez et al., 2017; Loewenstein and Wojtowicz, 2023; Mascagni and Nell, 2022).

However, the general literature reviewing nudging studies (Antinyan and Asatryan, 2025; Benartzi et al., 2017; DellaVigna and Linos, 2022; Mertens et al., 2022; Szaszi et al., 2018) reports mixed results regarding nudge efficacy, with some nudges having no effects (e.g., Gravert and Collentine, 2021) and some even backfiring (e.g., Holzmeister et al., 2022; Liu et al., 2016).[3] Mixed effects of nudges are also reported in the domain of income reporting and taxation. While in some cases tax nudges increase compliance (e.g., Hallsworth et al., 2017; Holz et al., 2020; Vainre et al., 2020) they can backfire as well (e.g., De Neve et al., 2021). A recent meta-analysis of 45 field experiments (Antinyan and Asatryan, 2025) concludes that while deterrence-style nudges can positively affect tax compliance, non-deterrence nudges are usually less effective. However, deterrence-style nudges are not a panacea as their effect decreases over a longer horizon.

In this paper, we explore a fundamental reason, from an economist's point of view, why nudges yield mixed results by examining the effects of (mis)alignment between the underlying economic incentives that determine optimal behavior and the implemented nudge.[4] Applying this general idea to the tax compliance context, we conjecture that even though nudges, by definition, bypass economic incentives, the interaction between the nudge and taxpayer incentives to comply (or not) determines whether the nudge will succeed or fail (see also Alm, 2019).[5] If the nudge is misaligned with the incentives and pushes the decision-maker in the opposite direction relative to what is optimal, it is deemed to fail. For example, if the controlling mechanisms and enforcing institutions are sufficiently weak, so there is a low probability of tax audit and/or low penalties for non-compliance, the decision to pay the tax is suboptimal for a risk-neutral (and in our setup, even for a realistically risk-averse) taxpayer. The underlying economic incentives restrain the impact of the applied nudge, which is unlikely to increase tax compliance as long as the incentives not to report income persist.[6]

From a policy-making perspective and for welfare evaluation, it is crucial to consider the effects of nudging over a longer horizon. However, a review by Szaszi et al. (2018) concludes that over 90 percent of nudging interventions implement only a one-off nudge, and the related studies usually evaluate only immediate effects. Notable exceptions include Manoli and Turner (2014), who provide

[1] One can argue that careful wording of Thaler & Sunstein (2008) that a nudge does not change economic incentives implicitly encompasses our conjecture. Sunstein (2017) posits that the presence of strong opposing preferences is one potential reason for the inefficiency of nudges. However, this aspect has not been investigated causally.

[2] For example, the existing meta-analyses do not systematically consider the incentive structure faced by the decision-makers who are subjected to nudging. In fact, as pointed out to us by an anonymous referee, Hummel & Maedche (2019) and Mertens et al., (2022) mainly use the words "economic" and "incentive" to characterize alternatives to nudges, namely traditional interventions based on providing economic/financial (dis) incentives whereas the incentives that are in place before the intervention are not considered.

[3] Maier et al. (2022) find that the prevalence of successful rather than unsuccessful nudges in the literature is driven by publication bias.

[4] In a similar vein, Holz et al. (2020) and Gil et al. (2023) argue that tax nudges result in different effect sizes for small vs. larger forms due to differences in their audit probability.

[5] In addition to the underlying economic incentives induced by tax enforcement (Batrancea et al., 2019), tax compliance can be affected by tax morale encompassing social norms, trust and reciprocity to the government (e.g., Luttmer & Singhal, 2014), third-party reporting (e.g., Bagchi & Dušek, 2021; Kleven et al., 2011), or fear (Bergolo et al., 2023). To focus on the interaction of incentives with nudges, our laboratory approach minimizes the above-mentioned effects (and if present, their influence should be similar across all treatments).

[6] Relatedly, we speculate that the main reason why deterrence nudges are more successful than non-deterrence nudges is the possibility that deterrence nudges are more likely to change the "perceived" incentives. For example, if nudge significantly increases the taxpayer's subjective probability of being audited (in the spirit of Slemrod et al., 2001), the decision to pay taxes becomes optimal from the perspective of perceived economic incentives. We note that our paper is not intended to offer insights into changes in the perceived audit rate since in our experiment the likelihood of audit is always stated explicitly.

evidence that the effect of a one-time nudge sharply decreases over time and disappears after three years, which suggests that nudging may need to be applied repeatedly. Importantly, Antinyan et al. (2021) claim that after crossing some optimal threshold, more frequent nudging does not lead to further significant gains and decreases overall efficiency, because of its repetitive costs. Additionally, Dunning et al. (2017) show that a one-time nudging intervention may have unintended adverse effects in the long run by disrupting existing habits of desired behavior.

Our study contributes to the understanding of the effects of nudges on tax compliance in two ways. First, we make a distinction between optimal and non-optimal compliance. While the existing tax compliance experiments vary whether it is optimal to comply by changing crucial parameters (e.g., Alm and Malézieux, 2021; Malézieux, 2018), our study conceptually relates the underlying incentives to nudge effectiveness. Our second contribution is particularly important for policy implications. Rather than focusing on the immediate effect of a one-off nudge, lessons from which may be limited for environments featuring a periodically repeated activity like reporting one's income and paying taxes, we explore whether the conjectured effectiveness of nudges aligned with economic incentives persists every time the nudge is applied and whether it increases with a higher frequency of nudging.

We conduct a controlled multi-period laboratory experiment in which subjects execute an individual real-effort task to earn a taxable income.[7] In each period, a subject can report the earned income or not. By varying the audit probability, we manipulate whether it is optimal to report income and thus pay tax (i.e., the audit rate is high enough that a risk-neutral taxpayer maximizes his expected payoff by complying) or not (i.e., the audit rate is low enough that unless the taxpayer is extremely risk averse, she should not report income). As the subject is focused on completing the task to earn income, she may forget to report her income for tax purposes. Depending on the experimental condition, the subject may be exposed to a nudge in the form of a reminder to report her income in some periods. A crucial feature of our experimental environment is that income reporting decisions are repeated. We investigate the interaction of audit rate (determining the economic incentives) and the presence and frequency of a nudge, leading to the following testable hypotheses:

Hypothesis 1. The first nudge increases immediate tax compliance only if it is optimal to pay tax.

Hypothesis 2. Repeated nudges increase immediate tax compliance only if it is optimal to pay tax.

Hypothesis 3. If it is optimal to pay tax, more frequent nudges increase long-run tax compliance. If it is not optimal to pay tax, more frequent nudges do not affect long-run compliance.

Contrary to Hypothesis 1, we find that when people are nudged for the first time, they are significantly more likely to report their income, independently of whether the audit probability is low or high. However, and in line with Hypothesis 2, repeated nudges significantly increase immediate tax compliance (in the period when the nudge is applied) only under a high probability of audit, i.e., if the nudge is aligned with the underlying economic incentives of a taxpayer. Regarding Hypothesis 3, we find that more frequent nudges have a weakly statistically significant positive effect on long-run tax compliance under the high probability of audit, while more frequent nudges under low probability yield no effect.

3. Experimental design

Our experiment employs the canonical tax game introduced by Allingham and Sandmo (1972), and consists of 16 periods in which subjects execute an individual real-effort task originally used in Lorko et al. (2019). In this task, the computer screen displays an inequality between a pair of two-digit numbers and subjects indicate whether the inequality is true or false. Immediately after the answer is submitted, a new, randomly generated inequality appears. Each period lasts 120 s.

To successfully complete the task and earn income for the period, subjects must provide 50 correct answers while providing fewer than 10 incorrect answers. The task is easy to understand and requires no prior training. However, it is also cognitively demanding and captivating, which opens space for inattention towards income reporting. We set the threshold for correct answers to 50 so that most subjects would take between 60 and 90 s to finish the task in each period. We implement the limit of 10 incorrect answers to discourage subjects from random clicking.

3.1. Income reporting and the reminder nudge

In each period, the screen shows information about the current number of correct and incorrect answers, a counter displaying elapsed time, and a "Report your income" button (see a screenshot in the Appendix). Clicking the button reports the income for the current period, while not clicking the button results in the income not being reported. We chose to report the current period income

[7] While there is a growing number of field studies on tax nudges we conduct our experiment in a laboratory environment because it allows us to (1) control underlying economic incentives, (2) reduce the complexity of the decision-making environment, (3) control who is nudged and how often, (4) ensure that the nudge is acknowledged, (5) eliminate potentially confounding influences such as spillover effects from different nudges and/or nudges to which an individual was exposed indirectly, e.g., via communication with a different individual who was originally targeted (Mascagni, 2018), and (6) conduct a clean test of the long-term effect of repeated nudges without attrition of subjects. Many of these important necessary considerations for answering our research question are unattainable in the field. For a discussion about the external validity of laboratory tax experiments, see Alm et al. (2015).

instead of the past period income because reporting the current period income makes all periods identical, meaning there is no first period without reporting and no last period without the effort task.[8] We reduce income reporting into a binary decision (report all income or none) to make the incentive structure easy to understand and to limit additional cognitive load (e.g., Deck et al., 2021) from thinking about the optimal level of underreporting.[9]

Importantly, while the button is displayed on the screen for the entire duration of each period, it is only active between the 30th and 60th second from the start of the period (see Table 1 for a period overview). We limit the time when income reporting is possible to parallel the outside-of-the-lab environment where the tax return window usually lasts for approximately one-third to one-quarter of a year, depending on the country and the tax system. The setup allows for not reporting the income due to inattention (Bordalo et al., 2022; Taubinsky, 2013), imperfect memory (Ericson, 2017) or limited self-control (Mertens et al., 2022), and the nudge in the form of a reminder can help to overcome these barriers, as is often the case in the realm of various financial decisions (DellaVigna and Linos, 2022; Mertens et al., 2022).

The implemented nudge takes the form of a pop-up window that covers the entire task at the moment when the income reporting button becomes active. The message informs subjects that they can report their income in the next 30 s. To acknowledge the message and remove it from the screen, subjects must click the OK button. A screenshot of the nudge is provided in the Appendix.[10]

3.2. Earnings

In each period, subjects can earn 100 tokens if they provide 50 correct answers and fewer than 10 incorrect answers.[11] Not completing the task successfully yields zero income for the period. Reporting income automatically subtracts a fixed tax of 30 tokens from the 100 tokens earned in the given period.

In each period, subjects can be audited with an audit probability of either 10 % or 60 %, depending on the experimental condition. The audit probability is kept constant for each subject across all periods. Failure to report income, if audited, results in having to pay the tax owed of 30 tokens and a fine of additional 30 tokens. No tax or fine is imposed if income for the period is zero, regardless of whether it was reported or not. The tax and fine remain constant for all subjects across all periods.[12] We postpone feedback about whether the audit occurred until all 16 periods have been completed to eliminate the potential confounding effects of being audited on the ensuing reporting decisions (e.g., the "bomb-crater" effect observed by Mittone et al., 2017, or the “echo-effect” documented by Kastlunger et al., 2008).

While in everyday life nudges often appeal to competing interests and aim to change people’s behavior in a way that serves some purpose, often social, we refrain from using the collected tax revenue to finance any charities or public goods to prevent another set of potential confounding effects summarized by Alm and Malézieux (2021). For example, if the collected tax revenue is contributed to a public good, subjects could be more likely to respond to the nudge, increasing the chance of observing a positive effect in our data that would hinge on a specific procedural feature. Instead, we implement a conservative design that makes it harder to detect a positive effect. If we observe an effect under our harsh rudimentary conditions, the effect is likely to be magnified if the nudge results in additional benefits to others or oneself.

Table 2 shows per-period earnings for all possible combinations of income reporting and audit, provided that the task is completed successfully. Each subject is paid the cumulative sum of his/her payoffs from all 16 periods. We chose to pay for all periods instead of paying for one or more randomly chosen periods to establish a dominant strategy for almost all empirically relevant risk attitudes (Dave et al., 2010; Eckel and Grossman, 2008; Holt and Laury, 2002), in both the low-audit-probability condition and the high-audit-probability condition. Assuming Constant Relative Risk Aversion (CRRA), indifference between reporting and not reporting under 10 % and 60 % probability of audit occurs at the r coefficient values of 4.6 and -0.93, respectively.[13] In the risk attitude calibration by Holt and Laury (2002), 97 % of subjects fall between these two values. Our payoff protocol, in which earnings accumulate across all 16 periods, accentuates the differences across audit probabilities even more. In the case of the 10 % audit probability, never reporting the income (across all 16 periods) results in higher expected earnings compared to always reporting 99.99 % of the time. Thus, a payoff-maximizing agent with any risk aversion from a typically observed range will never report her income. If the probability of audit is 60 %, never reporting leads to higher expected earnings compared to always reporting only 14.2 % of the time. Thus, the audit rate is high enough that it is favorable for all risk-averse, all risk-neutral, and most risk-loving agents to always report. In addition, the implemented protocol of paying for all periods mimics the actual tax-paying practice.

[8] After first two periods, approximately 98% of subjects successfully completed the task in each subsequent period, resulting in identical per period earnings. Reporting the past instead of current period income would thus result in essentially identical conditions for subjects.

[9] A meta-study by Alm & Malézieux (2021) shows that even when subjects have the option to report any share of their income, approximately 64% of them report either full or zero income. The third most frequently used share - half of the income - occurs in fewer than 5% of the cases.

[10] Nudges that are transparent and require conscious decision-making and deliberate action are generally well accepted by individuals (e.g., Jung & Mellers, 2016; Sunstein, 2016). Therefore, it is unlikely that our nudge would trigger a strong negative reaction from subjects. Furthermore, van den Akker and Sunstein (2025) find that participants were slightly opposed to only two of the 36 proposed nudging interventions - those that were non-transparent and targeted System 1.

[11] In the experiment, one experimental token was worth one Eurocent. The exchange rate was announced in the subject instructions.

[12] For an overview of the effects of different parameter variations on tax compliance, see Blackwell et al. (2007), Malezieux (2018), or Alm & Malézieux (2021).

[13] The CRRA utility function is defined as $U(x) = \frac{x^{1-r}}{1-r}$ where x denotes a monetary payoff.

Table 1
Overview of a period with a nudge reminder.

Time from the start of the task	Event
Second 1	Task starts
Second 30	A nudge message appears
Second 30	The income-reporting button becomes active
Second 60	The income-reporting button becomes inactive
Second 120	Task ends

Table 2
Period earnings by income reporting and audit.

	Audit	No audit
Income reported	Income – Tax = 70 tokens	Income – Tax = 70 tokens
Income not reported	Income – Tax – Fine = 40 tokens	Income = 100 tokens

3.3. Experimental conditions

We employ a 2 × 4 factorial across-subject design (see Table 3 for an overview), in which we vary (1) the audit probability (10 % or 60 %) to identify how a nudge interacts with the underlying economic incentives and (2) how often a nudge appears on the screen to identify the effect of nudging frequency on reporting behavior. In the Baseline condition, the nudge is never displayed. In the other three conditions, the nudge appears for the first time in Period 5, as we allow subjects to become familiar with the real-effort task and income reporting in the first four periods. In the Single condition, the nudge is never displayed again after Period 5. In the Periodic condition, the nudge is displayed again in Periods 9 and 13, i.e., four periods apart. In the Regular condition, the nudge is displayed in every following period after Period 5 until and including Period 13. The last three periods (i.e., Periods 14 to 16) in the Regular condition do not feature the nudge so that we can investigate whether the nudge effect persists after the intervention is removed.

3.4. Procedures

In each experimental session, the same experimenter handed out printed instructions (all experiment-related materials are included in the Appendix) to subjects and read them aloud. Any subject questions were answered in private. After reading the instructions, each subject was required to complete a set of 10 non-incentivized comprehension questions. Subjects were allowed to use the printed instructions and ask for help from the experimenter. If a comprehension question was answered incorrectly, a message window identifying the question with an incorrect answer appeared. The first period began only after each subject in the session correctly answered all 10 questions.

After completing all 16 periods, subjects participated in an incentivized risk-elicitation task where they had to choose between 70 tokens with certainty and a lottery which, to a certain extent, resembled the payoffs and probabilities used in the income-reporting task. Subjects were informed that one of their choices would be randomly chosen and the resulting payoff would be added to their experimental earnings. The risk-elicitation task included two salience manipulations of contrast and prominence, to generate a proxy for being susceptible to nudging (Bordalo et al., 2022; see the Appendix for details).

Subjects then completed a three-item modified Cognitive Reflection Test (CRT, Primi et al., 2016) in which they could earn 50 tokens for each correctly answered question. The CRT was followed by unincentivized elicitation of dishonesty attitudes based on the World Values Survey (Inglehart et al., 2014), a set of basic demographics, and two open-ended questions. The first question asked subjects about how they made their reporting decisions and the second allowed them to leave any other comments about the experiment.

The final screen displayed per-period information about the subjects' individual performance, their reporting decision, whether an audit occurred, and what their resulting task earnings were. The screen also showed earnings from the randomly selected risk elicitation choice and the CRT. At the end of the experiment, the experimental earnings were converted from tokens into cash and paid privately to the subjects.

3.5. Subjects

A total of 393 subjects, all students at the University of Economics in Bratislava, participated in the experiment. Due to variance in show-up rates, the number of subjects in each condition ranged from 45 to 53. One of the subjects left the experiment after the third period and is, therefore, excluded from the analysis. The sample analyzed consists of 392 subjects (210 females) with a mean age of 21.5 (SD 2.0).

All experimental sessions were conducted in the Bratislava Behavioral and Experimental Economics Research Lab at the University of Economics in Bratislava between November 2022 and February 2023. The experiment used a computerized interface programmed in zTree (Fischbacher, 2007). Subjects were recruited using the Online Recruitment System for Economic Experiments (ORSEE; Greiner, 2015) from a subject database maintained by the lab. On average, an experimental session lasted approximately 90 min

Table 3
Experimental conditions.

Nudge frequency	Low audit probability (10 %)	High audit probability (60 %)
Baseline (never)	Baseline-Low	Baseline-High
Single (Period 5)	Single-Low	Single-High
Periodic (Periods 5, 9, 13)	Periodic-Low	Periodic-High
Regular (Periods 5–13)	Regular-Low	Regular-High

including the initial instructional period and payment of subjects. The subjects earned EUR 13 on average.[14]

4. Results

Fig. 1 displays summary statistics of task performance and income reporting. The upper-left panel shows that the task was completed successfully 96 % of the time. The success rate increased to 98 % after the initial four periods. The upper-right panel displays the average duration of task completion, which was 92 s in the first period and then gradually declined to 69 s in the last period. The sudden increase in task duration in Period 5 was caused by the nudge, which appeared in that period for the first time in all conditions except for the two Baseline conditions that did not feature a nudge in any period. Similar but less pronounced increases are detectable in Periods 9 and 13, in which the nudge reappeared in the two Periodic conditions.

The lower-left panel shows that the income reporting rate (which is our measure of subject tax compliance as reporting one's income automatically results in paying tax owed) was consistently higher under the high audit probability (59.9 % across all 16 periods and all conditions) than under the low audit probability (33.3 %). The income reporting rates increased and became relatively steady after the initial four periods (65.3 % and 36.3 %, for the high and low audit probability, respectively). Finally, the lower-right panel displays the average time, counted from the start of the period, at which income reporting took place and shows that whenever a nudge appeared, subjects who decided to report their income did so on average four seconds sooner compared to periods without the nudge (36.7 and 40.7 s, respectively).

Due to apparent learning effects occurring in the first four periods when subjects were familiarizing themselves with the experimental environment and reporting decisions, in what follows, we only analyze data from the fifth period onwards. Recall that this is when nudging started. We note that there were no statistical differences in income reporting rates across conditions given the audit probability in the initial four periods (Pearson's chi-square test yields p-values of 0.14 and 0.79 for the 10 % and 60 % audit probability, respectively).

4.1. *The first nudge*

We begin by testing Hypothesis 1 that the first exposure to a nudge increases immediate tax compliance only if the audit probability is high. We find that in Period 5, subjects report their income significantly more often in the three nudge conditions (Single, Periodic, Regular) compared to the Baseline condition under both low and high audit probability. The differences (see the left panel of Fig. 2) are 20 percentage points under the low audit probability and 18 percentage points under the high audit probability, with Pearson's chi-square tests yielding p-values of 0.02 and 0.01, respectively, rejecting Hypothesis 1. The observed increased compliance regardless of audit probability suggests that the first nudge is special, which could help explain why previous studies focused on one-off interventions did not detect the association with the underlying incentives.

Result 1. The first nudge increases immediate tax compliance regardless of whether it is optimal to pay tax or not.

4.2. *Repeated nudges*

Next, we test Hypothesis 2 that repeated nudges increase immediate tax compliance only if the nudge is aligned with the taxpayer's economic incentives, i.e., only if it is optimal to pay tax. Table 4 presents OLS regressions with clustered standard errors at the subject level. For clarity of interpretation, we run separate models for each probability of audit. In Models 1 and 2, we include Periods 5–16 and find that nudges increase the immediate income reporting by 5 and 10 percentage points for low and high audit probability, respectively. However, the effect of nudge is significant only for the high audit probability (see also the right panel of Fig. 2).

A higher cognitive reflection score is associated with a significantly higher probability of making optimal decisions not to report the income if the audit probability is low and to report if the audit probability is high. Higher risk aversion is associated with a higher income-reporting rate under both audit probabilities, while a higher dishonesty score is associated with a lower income-reporting rate.[15] We also find that females are less responsive to the audit probability than males. Inconsistency in the saliency manipulation choices is not significantly correlated with reporting decisions in any specification and its exclusion from the set of control variables

[14] For comparison, the hourly wage that students could earn at the time of the experiment as a research assistant or in manual jobs typically ranged from EUR 5 to EUR 7.

[15] We calculate the dishonesty score from the set of five questions eliciting acceptance of different forms of cheating using the principal component analysis (Jackson, 2005).

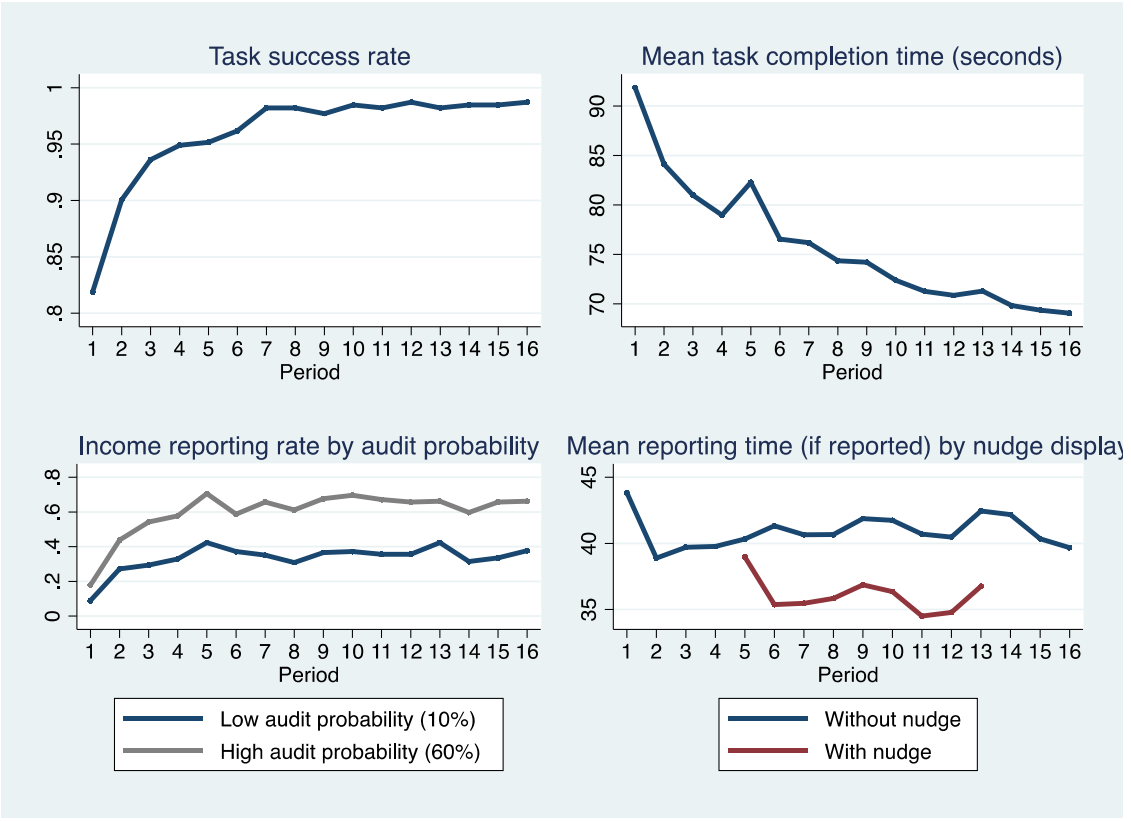

Fig. 1. Task performance and income reporting by period.

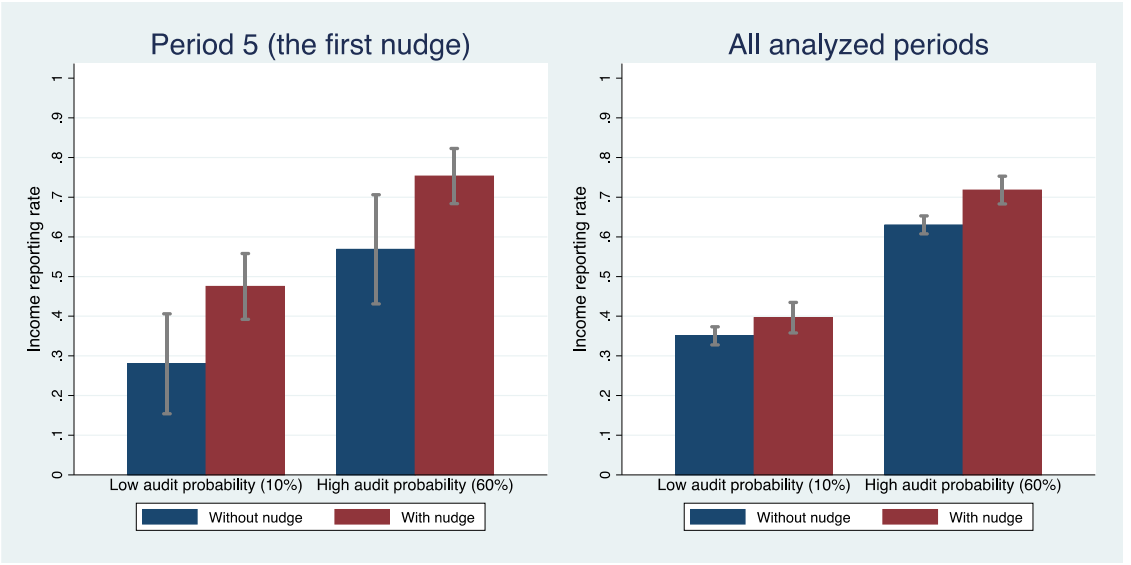

Fig. 2. Income reporting rates with and without a nudge.

does not qualitatively change any results. For brevity, we do not include it in the subsequent results. As a robustness check, in Models 3 and 4 we exclude Period 5 (i.e., the first nudge) and find qualitatively similar results, namely that nudges significantly increase immediate income reporting, by 8 percentage points, under high audit probability, while the increase of 2 percentage points under low audit probability is insignificant.

Table 4
Effect of nudge on income reporting.

	Periods 5–16 (all analyzed periods)		Periods 6–16 (without the first nudge)	
	(1) Report rate Low audit P	(2) Report rate High audit P	(3) Report rate Low audit P	(4) Report rate High audit P
Nudged	0.05	0.10***	0.02	0.08**
	(0.04)	(0.03)	(0.04)	(0.04)
Cognitive reflection	−0.09***	0.05**	−0.09***	0.05**
	(0.02)	(0.02)	(0.02)	(0.02)
Female	0.17***	−0.09**	0.18***	−0.09**
	(0.04)	(0.04)	(0.04)	(0.04)
Risk aversion	0.07***	0.05***	0.07***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)
Dishonest	−0.02*	−0.03*	−0.02*	−0.03*
	(0.01)	(0.01)	(0.01)	(0.01)
Period	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.12*	0.35***	0.10	0.33***
	(0.07)	(0.08)	(0.07)	(0.08)
N	2292	2412	2101	2211
R^2	0.16	0.07	0.16	0.07

Notes: Standard errors are reported in parentheses. All regressions include clustering of standard errors at the subject level. *, **, and *** indicate significance at the 10 %, 5 %, and 1 %-level, respectively. Nudged is a dummy variable that equals 1 if the subject was nudged in the relevant period and 0 otherwise. The control variables include gender, cognitive reflection (Primi et al., 2016), dishonesty attitudes (Inglehart et al., 2014), and risk aversion (see the Appendix).

Result 2. Repeated nudges significantly increase immediate tax compliance only if the nudge is aligned with incentives, i.e., only if it is optimal to pay tax.

Result 2, together with Result 1, demonstrates that the first nudge that goes against economic incentives might temporarily perturb behavior but is ineffective if repeated (see also Fig. 3). While the nudge may work the first time it is applied, due to the element of surprise (in the spirit of Falk, 2007), saliency (Bordalo et al., 2022), or experimenter demand effect (Zizzo, 2010), nudging is unlikely to increase tax compliance in the long run as these factors dissipate over time, while the economic incentives persist. In other words, the economic forces correct behavior that is not in the decision-maker's best interest.

4.3. Nudging frequency

Our Hypothesis 3 states that more frequent nudges increase long-run tax compliance, but only if it is optimal to pay tax. In the OLS regressions presented in Table 5, we therefore estimate the effect of nudging once (Single condition), three times (Periodic condition), and nine times (Regular condition). We use the income reporting rate of individual subjects as the unit of observation and estimate the effects of nudging frequency across Periods 5–13 (i.e., periods in which the nudge appeared in at least one condition, Models 1 and 2), and across Periods 5–16 (Models 3 and 4). In line with Hypothesis 3, we find that under the high audit probability, the Periodic nudge condition and the Regular nudge condition weakly significantly increase the likelihood that a subject will report her income compared to the Baseline condition. On the other hand, nudging more frequently under the low audit probability yields no significant effects on the income reporting rate.

Result 3: If it is optimal to pay tax, nudging more frequently leads to (weakly significantly) higher long-run tax compliance. If it is not optimal to pay tax, more frequent nudging does not affect long-run compliance.

4.4. Nudge removal

Our design also allows us to investigate what happens after removing the nudge. To that effect, we analyze the likelihood that a subject reports income in three periods after the nudge appears for the first time (Periods 6–8) and three periods after the Regular nudge is removed (Periods 14–16). Table 6, Models 1 and 2 show that under high audit probability, subjects nudged in Period 5 continued to report income insignificantly more frequently in Periods 6–8 without a nudge. On the other hand, under low audit probability, we observe that nudge removal results in an insignificant decrease in the likelihood of reporting income in Periods 6–8. We find similar directional effects in Models 3 and 4, comparing Periods 14–16 in the Regular condition, where the nudge was always displayed in the nine preceding periods, to the Baseline condition. In summary, we find directional but statistically insignificant support for the intuition that once a nudge aligned with economic incentives is removed, its positive effects on compliance may persist, while removing a misaligned nudge may backfire into lower compliance.

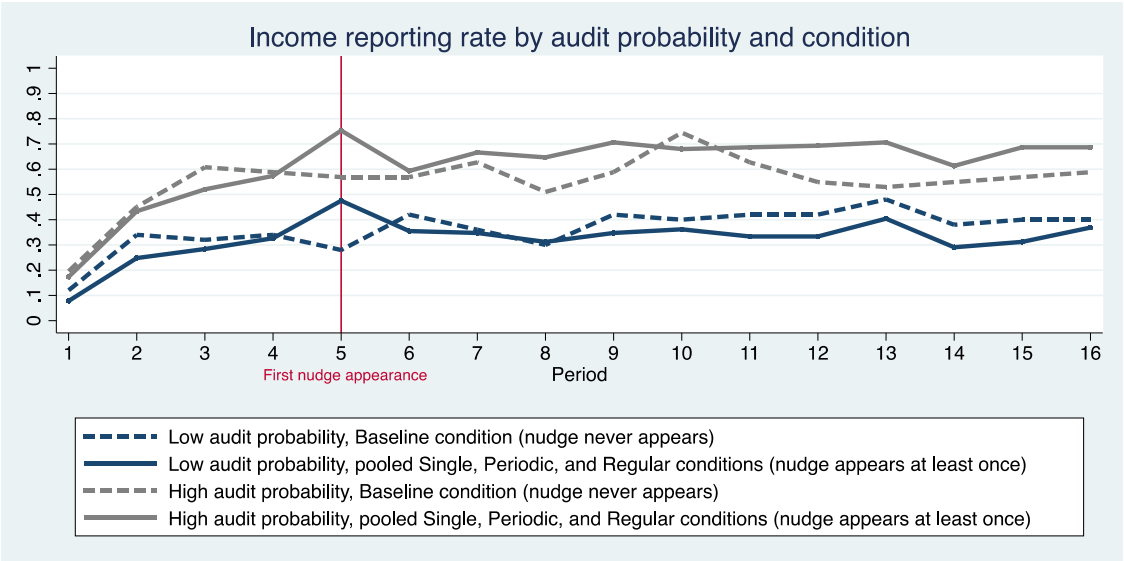

Fig. 3. Income reporting rate by audit probability and condition.

Table 5
Effect of nudging frequency on long-run tax compliance.

	Periods 5–13		Periods 5–16	
	(1) Report rate Low audit P	(2) Report rate High audit P	(3) Report rate Low audit P	(4) Report rate High audit P
Single nudge condition	−0.02	0.07	−0.02	0.08
	(0.06)	(0.06)	(0.06)	(0.06)
Periodic nudge condition	0.01	0.09	0.01	0.10*
	(0.06)	(0.06)	(0.06)	(0.06)
Regular nudge condition	−0.00	0.12**	−0.03	0.10*
	(0.06)	(0.06)	(0.06)	(0.06)
Control variables	YES	YES	YES	YES
Constant	0.15**	0.36***	0.15**	0.34***
	(0.07)	(0.09)	(0.07)	(0.09)
N	191	201	191	201
R^2	0.31	0.16	0.32	0.17

Notes: Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10 %, 5 %, and 1 %-level, respectively. The control variables include gender, cognitive reflection (Primi et al., 2016), dishonesty attitudes (Inglehart et al., 2014), and risk aversion (see the Appendix).

4.5. Auxiliary analyses

Finally, Table 7 presents our auxiliary regression analyses of behavior that our hypotheses are not explicit about. We find that when the nudge appears, income is reported more promptly and there are fewer late report attempts (i.e., trying to report the income after the reporting window is closed). Although the nudge also prolongs the task duration (because subjects must acknowledge it once it appears on the screen), it does not significantly affect whether the task is completed successfully. It thus does not reduce the period income, nor the period cash earnings after tax and audit.

While we find that the high audit probability (unsurprisingly) leads to lower period earnings, it also increases late reporting rates. A higher cognitive reflection is associated with more frequent successful task completion as well as with higher earnings due to a higher frequency of optimal reporting choices. Finally, we find that females take significantly longer to complete the task than males but report their income sooner.

5. Discussion

Due to their simplicity and inexpensive implementation, nudging interventions have been studied extensively in many domains of economic behavior, including tax reporting, on-time utility payments, healthy lifestyle, retirement saving, pro-environmental behavior, and others. Governments find nudges attractive because of their capability to shift behavior in the desired direction without a need for mandates, choice restrictions, and changes in economic incentives. However, a large fraction of nudging

Table 6
Effects of removing the nudge.

	Periods 6–8, all conditions		Periods 14–16, Baseline and Regular condition	
	(1) Report rate Low audit P	(2) Report rate High audit P	(3) Report rate Low audit P	(4) Report rate High audit P
Nudge removed	−0.03	0.03	−0.11	0.04
	(0.06)	(0.06)	(0.07)	(0.07)
Nudge displayed	0.03	0.13**		
	(0.06)	(0.07)		
Control variables	YES	YES	YES	YES
Constant	0.07	0.30***	0.08	0.20*
	(0.08)	(0.10)	(0.12)	(0.12)
N	191	201	99	100
R^2	0.27	0.16	0.29	0.17

Notes: Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10 %, 5 %, and 1 %-level, respectively. Nudge removed is a dummy variable that equals 1 if the subject was nudged in the previous period(s) and 0 otherwise. Nudge displayed is a dummy variable that equals 1 if the subject continued to be nudged in the relevant periods and 0 otherwise. The control variables include gender, cognitive reflection (Primi et al., 2016), dishonesty attitudes (Inglehart et al., 2014), and risk aversion (see the Appendix).

Table 7
Auxiliary analysis.

	(1) Late report	(2) Reporting time	(3) Task duration	(4) Period income	(5) Period earnings
Nudged	−0.03***	−4.34***	2.11**	−2.21	−0.02
	(0.01)	(0.66)	(1.03)	(1.39)	(0.01)
Period	−0.01***	−0.10*	−0.90***	0.12	0.00
	(0.00)	(0.06)	(0.07)	(0.08)	(0.00)
High audit prob.	0.03**	−0.47	1.70	−1.92	−0.18***
	(0.01)	(0.92)	(1.16)	(1.50)	(0.01)
Cognitive reflection	−0.03***	−0.76*	−0.49	1.71**	0.03***
	(0.01)	(0.46)	(0.67)	(0.73)	(0.01)
Female	0.02*	−1.66**	6.67***	−0.48	−0.03**
	(0.01)	(0.81)	(1.21)	(1.50)	(0.01)
Risk averse	0.00	−0.04	0.05	1.26*	−0.00
	(0.00)	(0.28)	(0.39)	(0.72)	(0.01)
Dishonest	0.00	0.20	0.30	−0.60	−0.00
	(0.00)	(0.25)	(0.35)	(0.44)	(0.00)
Constant	0.13***	44.63***	78.10***	90.27***	0.80***
	(0.03)	(1.62)	(2.23)	(3.45)	(0.03)
N	4704	2408	4605	4704	4704

Notes: Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10 %, 5 %, and 1 %-level, respectively. Nudged is a dummy variable that equals 1 if the subject was nudged in the relevant period and 0 otherwise. High audit probability is a dummy variable that equals 1 if 60 % audit probability applied and 0 if 10 % audit probability applied. The control variables include gender, cognitive reflection (Primi et al., 2016), dishonesty attitudes (Inglehart et al., 2014), and risk aversion (see the Appendix).

experiments finds no or very weak effects. Based on our general conjecture, supported by our experimental findings, we speculate that a potential explanation behind many failed nudges is the possibility that the underlying economic incentives were ignored. While nudges are capable of changing behavior, they cannot overturn economic incentives. We observe that the first nudge is special as it temporarily perturbs behavior, but if it goes against economic incentives, its effect quickly dissipates. Our empirical findings have important implications for policymakers and the broader behavioral science community and highlight that understanding and considering existing incentives is crucial for designing successful interventions.

Our research resonates with recent work investigating factors that may interact with nudges. Löfgren and Nordblom (2020) conclude that the effect of nudging is stronger for relatively less important decisions.[16] This theoretical prediction is empirically supported by Bonander et al. (2023). de Ridder et al. (2022) extend it by the claim that agents cannot be nudged into decisions that are not aligned with their initial preferences, which is akin to our findings. On the other hand, Farhi and Gabaix (2020) allow for this possibility in their theoretical model of optimal taxation by assuming that susceptibility to nudges is an exogenous (and heterogeneous) parameter and that nudges can directly affect utility. Similarly, Bordalo et al. (2022) argue that while FAST agents (Forgetful

[16] Our design is unable (and was not intended to) test the predictions of their theory as it would require strict assumptions about the exact functional form and parametrization of individual utility functions to vary the intensity, but not the direction, of the underlying incentives.

and Salient Thinkers) could be nudged into decisions that are against their initial preferences, nudges can influence FBOR agents (Forgetful But Otherwise Rational) only if the nudging direction is aligned with their initial preferences. Finally, Goldin (2015) and Johnson and Goldstein (2003) claim that nudging could be particularly effective when preferences are unclear or agents are inconsistent in their choices. Our study does not consider such an environment.

We also explore the effects of repeated nudges. Interestingly, extant research differs in theoretical predictions when nudging is repeated over several periods. On the one hand, Taubinsky (2013) argues that nudges in the form of reminders lead to habit formation and therefore repeated nudges increase the probability that an individual will act again in the future, even without a nudge. The same reasoning is suggested by Mullainathan (2002) and empirically observed in the works of Tobias (2009) and Henderson et al. (2011). On the other hand, Sellen et al. (1997) claim that the increased probability of acting on a reminder might not be permanent, and Ericson (2017) argues that anticipated and unanticipated reminders yield different theoretical predictions. While unanticipated reminders increase the probability of task completion, anticipated reminders can both increase and decrease this probability.

We find that more frequent nudging may lead to higher long-run tax compliance if it is optimal for taxpayers to report income. While our analysis shows that the effect of more frequent nudging is relatively weak, we note that nudges generally do not overturn preferences but instead make it easier for an individual to act in accordance with her underlying preferences.[17] From this perspective, a weakly significant increase in long-run compliance can be seen as the upper limit of what nudges can achieve. In line with the earlier findings that nudges can sometimes backfire, we find that removing nudges yields an insignificantly negative effect on compliance when it is not optimal for the taxpayer to comply. In practice, this could mean that if the taxpayer perceives the nudge as just an instant call to action, she may not pay tax in the future unless she is called to action by a subsequent nudge again.

All in all, our study shows that economic incentives interact with nudging. While nudges can be effective if they are aligned with underlying economic incentives, policy-makers should not expect nudges to reliably increase tax compliance if the tax audit rates (or at least subjectively perceived audit probability) and/or penalties are too low, parallel to the observed null effects of honesty nudges in an environment with zero risk of penalty for not reporting truthfully (Dimant et al., 2020). This is particularly important if the cost of the nudging intervention is non-negligible. It is therefore crucial from the policy perspective to understand the perceived economic incentives of taxpayers.

A methodological afterthought about the source of our empirical evidence is warranted. A frequent objection to laboratory experiments is whether their results generalize to individual or firm behavior outside the lab. This is not the right question. No experimental results, whether laboratory or field, generalize per se to a different setting from the one in which the experiment was conducted, unless there is a particular theory about the generalization process and the conditions of that theory are satisfied.[18] The main value of experiments is in testing falsifiable theoretical ideas. Since economists strive for their theories to be general, these theories are often silent about institutional assumptions and define only the most fundamental environmental details (if at all). However, without specifying an institution and environment, one cannot run an experiment and observe behavior (Smith, 1982). An experiment is a specific example of the theory. Just as theory is not meant to capture all details of everyday life, neither is experiment. Instead, the suitability of an experimental design is evaluated based on how closely it captures the examined theoretical idea and whether the precepts of experimental economics, including parallelism, are satisfied. If a crucial element of the environment is missing, it ought to be implemented if there is a theoretical reason why its omission would influence behavior. However, that means the element should have been a part of the original theoretical idea or concept in the first place. If the omission does not change the behavioral prediction, the principle of parsimony suggests it is better to adopt the more straightforward explanation.

The virtue of both theory and experiment is simplicity and there is value in testing theoretical ideas under the most rudimentary conditions. If a laboratory test delivers proof of concept, the underlying theoretical idea that has received empirical support can be applied in the field and tested further for robustness. A laboratory test is thus a part of the process that expands our understanding of everyday life phenomena (Smith, 1976). Nevertheless, the external validity of laboratory experiments cannot be ignored. The world is complex, and multiple factors influence behavior in an uncontrolled manner. An advantage of laboratory experiments is that they can control the impact of these factors and systematically study them, one at a time (Falk and Fehr, 2003). Laboratory behavior can also be linked to the field setting by a series of treatments (List, 2006). In the case of the current study, future research should attempt to replicate our findings in a field trial, ideally over multiple reporting periods and controlling for the subjective probability of being tax audited. Other behavioral factors influencing tax compliance in the field that are worthwhile exploring with respect to nudge efficacy include beliefs about deterrence or tax morale. Given that they are outside the scope of our research question, we also leave them for future research.

Declaration of competing interest

We, Matej Lorko, Tomáš Miklánek, and Maroš Servátka declare that we have no relevant material or financial interests that relate to the research described in this paper.

[17] Our results parallel the findings from literature studying the impact of nudges to decrease household water and energy consumption. Assuming that customers prefer to pay lower bills, nudges to save water or electricity are aligned with economic incentives, and the success of these nudges is therefore consistent with our overarching conjecture. The implications of our findings regarding the long-run impact of nudges are particularly important for the utility sector that lacks an enforcement mechanism and warrant further experimentation with saving nudges in field settings.

[18] For an example, see the model by Levitt & List (2007). For counterarguments, see Camerer (2015) and Falk & Heckman (2009).

Acknowledgements

This project was funded by the Czech Science Foundation (GAČR, project no 23–07136S) and by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic (VEGA, project no 1/0167/21). The project has been reviewed and approved by the Ethical Committee of the University of Economics in Bratislava (no. EKEUBA-VEGA1/0167/21–2/2022). We thank the editor, Daniela Puzello, and two anonymous referees for their constructive insights and suggestions that helped us improve the paper. We would also like to thank Armenak Antinyan, Lubomír Cingl, Paolo Crosetto, Klára Kalíšková, John List, Philippos Louis, Luigi Mittone, Johanna Mollerstrom, Jean Robert Tyran, and the audiences of the Economic Science Association meetings in Lyon and Cape Town for their helpful comments.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2025.107074.

Data availability

Data will be made available on request.

References

Allingham, M.G., Sandmo, A., 1972. Income tax evasion: a theoretical analysis. J. Public Econ. 1 (3–4), 323–338.
Alm, J., 2019. What motivates tax compliance? J. Econ. Surv. 33 (2), 353–388. https://doi.org/10.1111/joes.12272.
Alm, J., Bloomquist, K.M., McKee, M., 2015. On the external validity of laboratory tax compliance experiments. Econ. Inq. 53 (2), 1170–1186.
Alm, J., Malézieux, A., 2021. 40 years of tax evasion games: a meta-analysis. Experim. Econ. 24 (3), 699–750.
Antinyan, A., Asatryan, Z., 2025. Nudging for tax compliance: a meta-analysis. Econ. J. 135 (668), 1033–1068. https://doi.org/10.1093/ej/ueae088.
Antinyan, A., Asatryan, Z., Dai, Z., Wang, K., 2021. Does the frequency of reminders matter for their effectiveness? A randomized controlled trial. J. Econ. Behav. Organ. 191, 752–764.
Bagchi, S., Dušek, L., 2021. The effects of introducing withholding and third-party reporting on tax collections: evidence from the U.S. state personal income tax. J. Public. Econ. 204, 104537. https://doi.org/10.1016/J.JPUBECO.2021.104537.
Batrancea, L., Nichita, A., Olsen, J., Kogler, C., Kirchler, E., Hoelzl, E., Weiss, A., Torgler, B., Fooken, J., Fuller, J., Schaffner, M., Banuri, S., Hassanein, M., Alarcón-García, G., Aldemir, C., Apostol, O., Bank Weinberg, D., Batrancea, I., Belianin, A., Zukauskas, S., 2019. Trust and power as determinants of tax compliance across 44 nations. J. Econ. Psychol. 74, 102191. https://doi.org/10.1016/J.JOEP.2019.102191.
Benartzi, S., Beshears, J., Milkman, K.L., Sunstein, C.R., Thaler, R.H., Shankar, M., Tucker-Ray, W., Congdon, W.J., Galing, S., 2017. Should governments invest more in nudging? Psychol. Sci. 28 (8), 1041–1055.
Bergolo, M., Ceni, R., Cruces, G., Giaccobasso, M., Perez-Truglia, R., 2023. Tax audits as scarecrows: evidence from a large-scale field experiment. Am. Econ. J.: Econ. Policy 15 (1), 110–153. https://doi.org/10.1257/POL.20200321.
BIT Handbook, 2014. Four Simple Ways to Apply EAST Framework to Behavioural Insights | The Behavioural Insights Team. https://www.bi.team/publications/east-four-simple-ways-to-apply-behavioural-insights/.
Blackwell, C., 2007. A meta-analysis of tax compliance experiments. Int. Studies Program Work. Paper 7, 24.
Bonander, C., Ekman, M., Jakobsson, N., 2023. When do default nudges work? ArXiv Preprint. *ArXiv:2301.08797*.
Bordalo, P., Gennaioli, N., Shleifer, A., 2022. Salience. Annu. Rev. Econ. 14, 521–544.
Bott, K.M., Cappelen, A.W., Sørensen, E.Ø., Tungodden, B., 2020. You've got mail: a randomized field experiment on tax evasion. Manage. Sci. 66 (7), 2801–2819.
Bruns, H., Fillon, A., Maniadis, Z., Paunov, Y., 2025. Comparing transparent and covert nudges: a meta-analysis calling for more diversity in nudge transparency research. J. Behav. Experim. Econ. 116, 102350.
Camerer, C.F., 2015. The promise and success of lab–Field generalizability in experimental economics: a critical reply to Levitt and List. Handbook Experim. Econ. Methodol. 249–295. https://doi.org/10.1093/ACPROF:OSO/9780195328325.003.0016.
Dave, C., Eckel, C.C., Johnson, C.A., Rojas, C., 2010. Eliciting risk preferences: when is simple better? J. Risk. Uncertain 41, 219–243.
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. J. Polit. Econ. 129 (5), 1425–1463.
de Ridder, D., Kroese, F., van Gestel, L., 2022. Nudgeability: mapping conditions of susceptibility to nudge influence. Perspect. Psychol. Sci. 17 (2), 346–359. https://doi.org/10.1177/1745691621995183.
Deck, C., Jahedi, S., Sheremeta, R., 2021. On the consistency of cognitive load. Eur Econ Rev 134, 103695.
DellaVigna, S., Linos, E., 2022. RCTs to scale: comprehensive evidence from two nudge units. Econometrica 90 (1), 81–116.
Dimant, E., van Kleef, G.A., Shalvi, S., 2020. Requiem for a nudge: framing effects in nudging honesty. J. Econ. Behav. Organ. 172, 247–266. https://doi.org/10.1016/j.jebo.2020.02.015.
Dunning, T., Monestier, F., Pineiro, R., Rosenblatt, F., Tuñón, G., 2017. Is Paying Taxes Habit forming? Theory and Evidence from Uruguay. University of California-Berkeley Working Paper.
Eckel, C.C., Grossman, P.J., 2008. Forecasting risk attitudes: an experimental study using actual and forecast gamble choices. J. Econ. Behav. Organ. 68 (1), 1–17.
Ericson, K.M., 2017. On the interaction of memory and procrastination: implications for reminders, deadlines, and empirical estimation. J. Eur. Econ. Assoc. 15 (3), 692–719.
Falk, A., 2007. Gift exchange in the field. Econometrica 75 (5), 1501–1511.
Falk, A., Fehr, E., 2003. Why labour market experiments? Labour Econ. 10 (4), 399–406. https://doi.org/10.1016/S0927-5371(03)00050-2.
Falk, A., Heckman, J.J., 2009. Lab experiments are a major source of knowledge in the social sciences. Science 326 (5952), 535–538. https://doi.org/10.1126/SCIENCE.1168244.
Farhi, E., Gabaix, X., 2020. Optimal taxation with behavioral agents. Am. Econ. Rev. 110 (1), 298–336.
Fischbacher, U., 2007. z-Tree: zurich toolbox for ready-made economic experiments. Experim. Econ. 10 (2), 171–178.
Gil, P., Holz, J.E., List, J.A., Simon, A., Zentner, A., Andres, L., Carney, K., Chen, H., Elias, J., Gars, J., Livingston, J., Londono-Velez, J., Metalfe, R., Novgorodsky, D., Riley, E., Rodriguez, A., Scartascini, C., & All, A.Z. (2023). *Toward an understanding of tax amnesties: theory and evidence from a natural field experiment.* https://doi.org/10.3386/W31210.
Goldin, J., 2015. Which way to nudge: uncovering preferences in the behavioral age. Yale LJ 125, 226.

Gravert, C., Collentine, L.O., 2021. When nudges aren't enough: norms, incentives and habit formation in public transport usage. J. Econ. Behav. Organ. 190, 1–14.
Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Associat. 1 (1), 114–125.
Hallsworth, M., 2023. nature human behaviour A manifesto for applying behavioural science. Nature Human Behav. 7, 310–322. https://doi.org/10.1038/s41562-023-01555-3.
Hallsworth, M., List, J.A., Metcalfe, R.D., Vlaev, I., 2017. The behavioralist as tax collector: using natural field experiments to enhance tax compliance. J. Public Econ. 148, 14–31.
Henderson, C.M., Beck, J.T., Palmatier, R.W., 2011. Review of the theoretical underpinnings of loyalty programs. J. Consum. Psychol. 21 (3), 256–276. https://doi.org/10.1016/j.jcps.2011.02.007.
Hernandez, M., Jamison, J., Korczyc, E., Mazar, N., Sormani, R., 2017. Applying Behavioral Insights to Improve Tax Collection.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. Am. Econ. Rev. 92 (5), 1644–1655.
Holz, J.E., List, J.A., Zentner, A., Cardoza, M., Zentner, J., 2020. The 100 million nudge: Increasing tax compliance of businesses and the self-employed using a natural field experiment.
Holzmeister, F., Huber, J., Kirchler, M., Schwaiger, R., 2022. Nudging debtors to pay their debt: two randomized controlled trials. J Econ Behav Organ 198, 535–551.
Hubble, C., Varazzani, C., 2023. Mapping the global behavioural insights community - Observatory of Public Sector Innovation. https://oecd-opsi.org/blog/mapping-behavioural-insights/.
Hummel, D., Maedche, A., 2019. How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. Journal of Behavioral and Experimental Economics 80, 47–58.
Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., Puranen, B., 2014. World Values survey: Round six-Country-Pooled Datafile Version, 12. JD Systems Institute, Madrid.
Internal Revenue Service, 2022. Federal tax compliance research: tax gap estimates for tax years 2014-2016. www.irs.gov.
Jackson, J.E., 2005. A User's Guide to Principal Components. John Wiley & Sons.
Johnson, E.J., Goldstein, D., 2003. Do defaults save lives?. In: Science, 302 American Association for the Advancement of Science, pp. 1338–1339.
Jung, J.Y., Mellers, B.A., 2016. American attitudes toward nudges. Judgm. Decis. Mak. 11 (1), 62–74.
Kastlunger, B., Kirchler, E., Mittone, L., & Pitters, J. (2008). *Sequences of audits, tax compliance, and taxpaying strategies*. https://doi.org/10.1016/j.joep.2008.10.004.
Kleven, H.J., Knudsen, M.B., Kreiner, C.T., Pedersen, S., Saez, E., 2011. Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. Econometrica 79 (3), 651–692.
Kukk, M., Paulus, A., Staehr, K., 2020. Cheating in Europe: underreporting of self-employment income in comparative perspective. Int. Tax Public Finance 27 (2), 363–390.
Levitt, S.D., List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? J. Econ. Perspect. 21 (2), 153–174. https://doi.org/10.1257/JEP.21.2.153.
List, J.A., 2006. The behavioralist meets the market: measuring social preferences and reputation effects in actual transactions. J. Polit. Econ. 114 (1), 1–37. https://doi.org/10.1086/498587.
Liu, C.-W., Agarwal, R., Gao, G., 2016. The dark side of positive social influence.
Loewenstein, G., Wojtowicz, Z., 2023. The Economics of Attention. *Available at SSRN 4368304*.
Löfgren, Å., Nordblom, K., 2020. A theoretical framework of decision making explaining the mechanisms of nudging. J. Econ. Behav. Organ. 174, 1–12. https://doi.org/10.1016/j.jebo.2020.03.021.
Lorko, M., Servátka, M., Zhang, L., 2019. Anchoring in project duration estimation. J. Econ. Behav. Organ. 162, 49–65.
Luttmer, E.F.P., Singhal, M., 2014. Tax morale. J. Econ. Perspect. 28 (4), 149–168. https://doi.org/10.1257/JEP.28.4.149.
Maier, M., Bartoš, F., Stanley, T.D., Shanks, D.R., Harris, A.J.L., Wagenmakers, E.J., 2022. No evidence for nudging after adjusting for publication bias. Proc. Natl. Acad. Sci. U.S.A. 119 (31). https://doi.org/10.1073/PNAS.2200300119.
Malezieux, A.E., 2018. A practical guide to setting up your tax evasion game.
Manoli, D.S., Turner, N., 2014. Nudges and learning: evidence from informational interventions for low-income taxpayers.
Mascagni, G., 2018. From the lab to the field: a review of tax experiments. J. Econ. Surv. 32 (2), 273–301.
Mascagni, G., Nell, C., 2022. Tax compliance in Rwanda: evidence from a message field experiment. Econ. Dev. Cult. Change 70 (2), 587–623.
Mertens, S., Herberz, M., Hahnel, U.J.J., Brosch, T., 2022. The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains. Proc. Natl. Acad. Sci. U.S.A. 119 (1). https://doi.org/10.1073/PNAS.2107346118.
Mittone, L., Panebianco, F., Santoro, A., 2017. The bomb-crater effect of tax audits: beyond the misperception of chance. J. Econ. Psychol. 61, 225–243.
Mullainathan, S., 2002. A memory-based model of bounded rationality. Q. J. Econ. 117 (3), 735–774.
Primi, C., Morsanyi, K., Chiesi, F., Donati, M.A., Hamilton, J., 2016. The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). J. Behav. Decis. Mak 29 (5), 453–469.
Sellen, A.J., Louie, G., Harris, J.E., Wilkins, A.J., 1997. What brings intentions to mind? An In situ study of prospective memory. Memory 5 (4), 483–507. https://doi.org/10.1080/741941433.
Slemrod, J., 2019. Tax compliance and enforcement. J. Econ. Lit. 57 (4), 904–954.
Slemrod, J., Blumenthal, M., Christian, C., 2001. Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. J Public Econ 79 (3), 455–483.
Slemrod, J., & Yitzhaki, S. (2002). Tax avoidance, evasion, and administration. In *Handbook of Public Economics* (Vol. 3, pp. 1423–1470). Elsevier.
Smith, V., 1976. Experimental economics: induced value theory. Am. Econ. Rev. 66 (2), 274–279. https://www.jstor.org/stable/1817233.
Smith, V.L., 1982. Microeconomic Systems as an Experimental Science 72 (5), 923–955.
Sunstein, C.R., 2016. Do people like nudges. Admin. L. Rev. 68, 177.
Sunstein, C.R., 2017. Nudges that fail. Behavioural Public Policy 1 (1), 4–25. https://doi.org/10.1017/BPP.2016.3.
Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., Aczel, B., 2018. A systematic scoping review of the choice architecture movement: toward understanding when and why nudges work. J. Behav. Decis. Mak. 31 (3), 355–366.
Taubinsky, D., 2013. From intentions to actions: a model and experimental evidence of inattentive choice. Work Pap.
Thaler, R.H., & Sunstein, C.R. (2008). *Nudge: Improving decisions About Health, wealth, and Happiness*. Penguin.
Tobias, R., 2009. Changing behavior by memory aids: a social psychological model of prospective memory and habit development tested with dynamic field data. Psychol. Rev. 116 (2), 408–438. https://doi.org/10.1037/a0015512.
Vainre, M., Aaben, L., Paulus, A., Koppel, H., Tammsaar, H., Telve, K., Koppel, K., Beilmann, K., Uusberg, A., 2020. Nudging towards tax compliance: a fieldwork-informed randomised controlled trial. J. Behav. Public Administr. 3 (1).
van den Akker, M., Sunstein, C.R., 2025. Do people like financial nudges? Judgm. Decis. Mak. 20, e2.
Zizzo, D.J., 2010. Experimenter demand effects in economic experiments. Experim. Econ. 13, 75–98.