Do Behavioral Nudges in Prepopulated Tax Forms Affect Compliance? Experimental Evidence with Real Taxpayers

Miguel A. Fonseca and Shaun B. Grimshaw

Defaults, in the form of prepopulated fields within the tax form, have been identified as potential mechanisms that tax authorities can use to reduce noncompliance. They achieve this by simplifying the process of filing taxes, thus reducing the scope for errors. However, defaults may increase the scope for evasion if set incorrectly. The authors report experimental data on the effect of correct and incorrect defaults. They find that prepopulating tax returns is a worthwhile policy only if it is done with highly reliable information. Setting default levels that underestimate taxpayers' true tax liability leads to significant drops in compliance and tax revenue. The authors also study whether nudges that contain messages with descriptive norms about compliance can mitigate the adverse effect of prepopulated returns with incorrect values. Nudges that react to inputs from the taxpayer effectively raise compliance, whereas static nudges do not. This result demonstrates the limits to the applicability of nudges in a public policy sphere as well as possible adverse effects resulting from poor implementation.

Keywords: defaults, tax compliance, prepopulation of tax returns, reactive nudges, experiments

Governments are turning to the marketer's toolkit to design the way in which they interact with their citizens (Dolan et al. 2010; Thaler and Sunstein 2008). While nudges such as default options were initially applied to insurance purchasing (Johnson et al. 1993) and other consumer choice domains, research has increasingly recognized the importance of choice framing for nudging public policy–relevant decisions, such as organ donation choices (Johnson and Goldstein 2003) or retirement pension choices (Madrian and Shea 2001; for a review, see Johnson et al. 2012).

In an important development in the public policy sphere, defaults are being introduced to the tax domain. The U.K. tax authority is now moving toward online tax filing (Her Majesty's Revenue & Customs [HMRC] 2015). Within that framework, it is using information about taxpayers' income and/or tax-deductible expenses from third parties such as employers, banks, or pension companies to prepopulate the tax form. This move follows an international trend: the state of California already prepopulates elements of its state tax returns with the Ready Return program; in addition, tax return prepopulation happens to varying degrees in more than ten European Union countries and Australia (Forum on Tax Administration 2006; Jensen and Wöhlbier 2012). A prepopulated field in a tax form is effectively a default.

This article reports the results of an online experiment designed to shed light on the impact on filing behavior of introducing defaults and norm-based nudges in online tax returns. Our experiment contributes to the literature on default options by exploring the potential compliance benefits from prepopulating fields in tax returns; in particular, what the potential pitfalls are if default values are set incorrectly. Our study also explores, for the first time in a tax context, the potential for nudges that invoke descriptive norms to change compliance behavior. We consider static nudges, which have been the focus of attention in the social norm messaging literature. We also examine a novel form of nudges that react to users' inputs, which are especially well-suited to online environments.

Taxes are an interesting and relatively unexplored domain of research in marketing. They are a ubiquitous payment by all consumers, yet they differ from most personal consumer payments along two important dimensions, as noted by Lamberton (2013): most consumer purchases have an intrinsic personal benefit and, for the most part, can be controlled by consumers. In contrast, taxes are compulsory payments, the benefits of which are not directly experienced by individuals but, rather, experienced indirectly through the provision of public goods such as roads, schools, and law enforcement. This might explain why

Miguel A. Fonseca is Associate Professor of Economics, University of Exeter, and Associate Researcher, Economic Policies Research Unit, University of Minho (e-mail: m.a.fonseca@exeter.ac.uk). *Shaun Grimshaw* is Postdoctoral Research Associate, Tax Administration Research Centre, University of Exeter (e-mail: sbg203@exeter.ac.uk). Funding from the Tax Administration Research Centre, ESRC/HMRC/HMT grant ES/K005944/1 is gratefully acknowledged. The authors thank the *JPPM* review team for their very helpful input. They also thank Joe Scarlett-Smith, Rossy Bailey, Adrian Haldane, Soma Chaudhury, and other HMRC researchers and staff for their feedback at various stages of development of this project. The views expressed in this article are the authors' own and do not reflect HMRC policy or the opinions of any HMRC personnel. Jerome Williams served as associate editor for this article.

consumers often view taxes as a loss of personal freedom (Kirchler 1998), and why consumers exhibit tax aversion, defined by Sussman and Olivola (2011, p. S92) as "a dislike of taxes per se that goes above and beyond any associated financial costs."

Filing a tax return is similar to financial consumer decisions such as choosing a pension plan. Both tasks are procedurally complex and cognitively demanding; they require a reasonable degree of financial literacy as well as knowledge about the regulatory framework. Just as there is ample evidence of individuals and households making errors in their financial decision making (Bernheim 1998; Beshears et al. 2013; Lusardi and Mitchell 2007, 2014), errors in filing decisions account for a significant portion of noncompliance: the U.K. government estimates that it loses £6.5 billion (19%) of its tax revenues due to filing errors (National Audit Office 2015); Andreoni, Erard and Feinstein (1998) estimate that 7% of U.S. taxpayers make mistakes when filing their tax returns. Governments are starting to recognize that the benefits of simplifying decision processes and helping people with their financial decision making also apply to tax filing (Government Accountability Office 2005; Reeson and Dunstall 2009). Moreover, the financial case for doing so is overwhelming.

However, tax filing differs from other types of financial decisions in important ways that are likely to have implications for the impact of defaults. When a company decides which of the 401(k) plans to introduce as a default, it knows its employees' income but not their plan preferences. When a government agency prepopulates tax returns, it assumes that taxpayers want to minimize their tax burden, but it only has potentially noisy third-party data about taxpayers' true taxable income. It may have accurate salary data from employers but only incomplete data on capital earnings, such as dividend payments (Bloomquist et al. 2012). In that sense, the tax domain presents an important dimension for the use of nudges in public policy: honesty.

The asymmetry of information between the tax authority and taxpayers about the latter's income introduces operational and ethical concerns when considering the introduction of prepopulation. In particular, the tax authority could inadvertently prepopulate tax forms incorrectly. This may lead to unintended consequences of a policy designed to improve compliance. One possible outcome is that the tax authority could underestimate taxpayers' liabilities. Taxpayers may simply accept the incorrectly prepopulated values because of status quo bias or behavioral inertia (Samuelson and Zeckhauser 1988), or because they trust the tax authority's assessment to be correct. A mistake of this nature would potentially leave taxpayers open to an audit and any associated penalties from their noncompliance, because the legal responsibility for correctly filing the tax return still lies with the taxpayer. Another possible outcome is that the tax agency prepopulates tax returns in a way that overestimates tax liabilities. Either over- or underestimating tax liabilities when prepopulating tax returns raises ethical considerations given the duty of care that tax administrations have toward taxpayers, and it may lead to a public relations blow. It could also lead to additional audits, the cost of which would offset the increase in overall revenue that prepopulation offers.

The policy could make misreporting more prevalent by making noncompliance a passive act, rather than an active one (Mazar and Hawkins 2015; Spranca, Minsk, and Baron 1991). Incorrect prepopulation also reveals to taxpayers what the tax agency knows (and, importantly, what it does not know) about their affairs; taxpayers may also interpret mistakes as incompetence on the part of the tax authority, thus extending the opportunity for deliberate evasion. In the context of defaults, our contribution to the literature is twofold. On the one hand, we contribute to the literature on tax compliance by examining the psychological determinants of compliance as a manifestation of honesty through the use of defaults. On the other hand, we also contribute to the understanding of acts of omission and acts of commission in the context of honesty.

Given the potential for unexpected noncompliance to emerge from the use of incorrect defaults, it is important to understand whether other types of nudges can be effective at mitigating any adverse effect of prepopulation. Recently, Smith, Goldstein, and Johnson (2013) advocated for the use of "smart nudges," which react to users' behavior in real time. One of the benefits they propose is the potential to correct mistakes users may make along the decision process. We are particularly interested in reactive nudges, as they are well-suited to online environments such as the one considered by HMRC; we are also interested in understanding the extent to which they can mitigate potential errors from incorrect defaults. To this end, we implemented a series of nudges, some of which included normative messages about compliance. Depending on the treatment, these messages appeared on the screen as a function of the amounts declared by participants on the tax form. To the best of our knowledge, this is the first study to test the effectiveness of reactive nudges on behavior.

In the following section, we develop the theoretical framework underpinning our experiment and the resulting hypotheses; we then present the experimental design. We outline the results and conclude by discussing the relevance of our findings to the academic and policy literature streams.

Theoretical Framework and Hypotheses

If a tax authority prepopulates a particular entry on a tax form, such as employment income or taxable expenses, it is effectively imposing a default action on taxpayers. When deciding the value to enter as the default, tax agencies must consider that the information they have about that field may be potentially unreliable or uncertain. This introduces a moral dimension to the use of defaults, because taxpayers have an incentive to misrepresent their tax liabilities. Will defaults and other forms of nudges be effective in promoting desirable behavior in this environment?

Gigerenzer (2010) argues that the same heuristics that guide choices in nonmoral domains are also at play in moral decision making. The default heuristic, which states that "if there is a default, do nothing about it," has been offered as the prime explanation for cross-country discrepancies in participation rates of 401(k) savings plans with and without a default option (Madrian and Shea 2001) or signup rates for organ donation (Johnson and Goldstein 2003) and should also determine people's compliance behavior when filing a tax return with or without prepopulation.

Defaults should reduce the cognitive cost of making decisions, which should in turn help decision makers better understand and weigh information (Peters et al. 2006) and ultimately reduce the time spent performing the task. Kotakorpi and Laamanen (2015) study the effectiveness of prepopulation on tax filing behavior by examining Finnish tax filing data. Between 1995 and 2004, Finland's tax authority began prepopulating

sections of tax forms for a subset of the Finnish taxpayer population. Kotakorpi and Laamanen report that taxpayers who received a prepopulated return were more likely to report the items that were prepopulated and less likely to report deductions that were not prepopulated. The authors examined only changes in filing behavior in the income and deduction fields and did not have access to actual earnings data for those taxpayers. This is because the Finnish government did not audit any subset of taxpayers who were in the treatment and control groups. As such, their study cannot speak to whether prepopulation led to changes in compliance. Importantly, a large proportion of taxpayers simply accepted the prepopulated returns and chose not to file a modified form. Kotakorpi and Laamanen attribute this behavior to taxpayers avoiding the cognitive and temporal costs associated with the complex process of engaging with (and potentially modifying) their tax forms.

> H_1: Prepopulating income fields leads to quicker completion of tax filing decisions across all treatments.

An important idiosyncrasy of tax filing decisions is that the tax authority may not always know taxpayers' true level of income. Any third-party information it uses to prepopulate a tax form could be incorrect. As such, introducing defaults in a tax context means it is possible to set the default value at an incorrect level: for instance, the tax authority could either overestimate or underestimate a taxpayer's income for that year.

Setting a default at a level that underestimates a taxpayer's income intrinsically changes the nature of a potential misrepresentation of income by the taxpayer from an act of commission to an act of omission. The literature on moral psychology has found consistent evidence for an omission bias in decision making: morally reprehensible acts of commission are judged more harshly than acts of omission that carry the same negative consequence. Harmful acts of commission, unlike acts of omission, presumably signal malicious intent on the part of the decision maker (Baron and Ritov 1994; Sánchez-Pagés and Vorsatz 2009; Spranca, Minsk, and Baron 1991).

The action principle in moral psychology (Cushman, Young, and Hauser 2006) also suggests that it is easier to passively refrain from acting morally than to actively transgress a moral norm. Teper and Inzlicht (2011) show that this is indeed the case for behavior in both prescriptive and proscriptive domains: participants in their experiment were more likely to offer to help a fellow student with a disability if they were asked directly than if they were passively given the option to help. Likewise, participants were less likely to cheat in a math quiz when doing so involved an action rather than an omission. Therefore, it should be psychologically easier for participants to misreport their tax liabilities when doing so is a default action than when it is an active choice. Finally, another potential motive for noncompliance following incorrect prepopulation is behavioral inertia: taxpayers may simply accept the prepopulated values at face value and submit the tax return as it is (Madrian and Shea 2001; Samuelson and Zeckhauser 1988).

> H_2: Prepopulating fields in a way that underestimates tax liabilities leads to higher noncompliance.

Defaults can influence choice to the extent that decision makers may believe that defaults are a suggestion by the policy maker and, as such, imply a recommended course of action (Johnson and Goldstein 2003). In the tax context, defaults carry additional significance because the tax authority, an expert body and part of government, has a duty of care toward its taxpayers. If the tax authority chooses a default value that overestimates a taxpayer's liability, then taxpayers may interpret the incorrect prepopulation as a signal that the tax authority does not have their interests at heart (Wright [2002] and Brown and Krishna [2004] define this process as "marketplace meta-cognition") or as a signal of incompetence, leading to higher noncompliance.

> H_3: Prepopulating fields in a way that overestimates tax liabilities leads to higher noncompliance.

There are two approaches within the social sciences to conceptualize the determinants of honest behavior. On the one hand, there is the external incentives approach, anchored in the economics of crime literature (Allingham and Sandmo 1972; Becker 1968). In this theoretical framework, the decision to be dishonest revolves around the calculus of expected utility: people weigh the relative gains from being dishonest against the probability of being caught and the associated penalties. Therefore, increases in the expected benefit from dishonesty, either through changes in the penalties or probability of detection, should increase the extent of dishonesty.

Independently of our own work, Bruner et al. (2015) study the role of defaults in taxes using the external incentives approach. The authors develop an individual tax evasion experiment, in which participants are presented with a tax form in which, depending on the treatment, some fields are prepopulated. Participants have two types of income sources: "matched" income, which is verifiable through third-party data (e.g., salaries), and "unmatched" income (e.g., self-employment income). Bruner et al.'s experimental design incorporates treatments in which itemized deductions are possible to ascertain the impact of prepopulation within the context of the U.S. tax code. Their study focuses on how changes in the regulatory framework, such as the audit likelihood or the presence of itemized deductions, affect behavior with and without prepopulation. Noncompliance is measured in three ways: underreported taxes on unmatched income, underreported taxes on the deduction, and overall underreported taxes. The authors find that prepopulating tax returns in a way that implies a lower tax liability increases noncompliance.

On the other hand, there is the internal incentives approach, rooted in the social psychology literature. This theoretical framework is based on the idea that moral actions are driven by internal rewards, which are, to some extent, uncorrelated with the degree of financial reward at stake. People develop an understanding of socially normative behavior from an early age (Campbell 1964), as well as through group membership (Akerlof and Kranton 2002; Sherif and Sherif 1953). Norms in turn establish a set of prescribed behaviors from which we derive psychological well-being. Neuroeconomics research supports this account: people who cooperate or enforce cooperative behavior in social dilemmas exhibit similar brain activation patterns (De Quervain et al. 2004; Rilling et al. 2002) as those observed when people experience financial rewards or consume food (Knutson et al. 2001; O'Doherty et al. 2002).

The theory of self-concept maintenance, developed by Mazar and Ariely (2006) and Mazar, Amir, and Ariely (2008), provides a particularly useful conceptualization of how internal reward mechanisms determine honest behavior. It postulates that when people consider whether to break a social norm

Table 1.   Contents of the Taxpayer Profile Used in the Experiment

Field	Description	Value (in ECU)
Self-employment income	Income from contract with local authority	25,200
Self-employment income	Income for work done for ACS Ltd	27,100
Self-employment expenses	Cost of travel to work	2,500
Property income	Revenue from letting a flat	20,000
Property expenses	Cost of estate agent and legal fees for letting of flat	2,000 times the roll of a six-sided die

such as honesty, they trade off the financial gains from cheating against their positive self-concept as an honest person. According to this theory, people manage this problem by finding a balance between these competing psychological demands. They do so by engaging in an amount of dishonest behavior that brings in financial rewards but is not enough to force people to reassess their self-image.

In this context, descriptive norms —norms pertaining to what people in a group or population do—can be a powerful driver of self-concept maintenance by prescribing a particular mode of behavior (Cialdini et al. 2006). Descriptive norms can be a powerful driver of behavior, influencing littering (Cialdini et al. 1990; Reno, Cialdini, and Kallgren 1993), environmentally friendly behavior (Goldstein, Cialdini, and Griskevicius 2008), or energy consumption in the household (Schultz et al. 2007).

> H₄: Introducing nudges that remind participants of descriptively normative behavior reduces noncompliance.

Importantly, when developing the focus theory of normative conduct, Cialdini, Reno, and Kallgren (1990) and Cialdini, Kallgren, and Reno (1991) argue that the effectiveness of social norm messages is critically dependent on whether they are focal in the decision maker's attention and, therefore, consciously salient. This approach suggests that nudges containing descriptive norm information will be more effective at preventing noncompliance if they are reactive to users' behavior. Furthermore, because the displayed message reacts to user behavior, this should create a perception among users that the message is directly targeted at them. This "personalization" of the message content should increase the effectiveness of the reactive nudge, according to the literature on survey responses (Kanuk and Berenson 1975; Yu and Cooper, 1983), as compared with static nudges.

> H₅: Nudges that are activated by noncompliant filing behavior lead to more compliance than always-present static nudges.

Materials and Methods

The Experimental Task

Participants in our study took on the role of a fictitious taxpayer. Their task was to complete a tax form based on a profile of income and expenses for that fictitious taxpayer. The profile detailed two sources of income and two corresponding expenses that could be used to reduce tax liabilities. Table 1 outlines the profile used in the experiment. Payoffs were denominated in experimental currency units (ECU); 1,000 ECU were worth £.50 (at the time, $.75).

The experimental instructions (for a copy of the instructions, see the Web Appendix) detailed that participants would be paid according to the income in their profile minus any tax or fines due from their tax declaration and any potential audit. The instructions

also detailed that after participants filed their tax return, the "experimental tax authority" could audit it. If a participant's tax return was audited, the computer compared the values in the tax return with the values in the profile. The probability with which the experimental tax authority carried out audits was a function of the actual declared tax liability on the return, but it could never exceed 10%: the probability of audit was 3.3% if the declared liability was greater than or equal to 45,200 ECU, 6.6% if the declared liability was between 22,600 ECU and 45,199 ECU, and 10% if the declared liability was less than 22,600 ECU. Participants did not know the actual probability of audit or how it changed as a function of declared tax liabilities; they knew only that the probability varied with the amount declared and was limited to 10%. We did not intend to mimic the audit policy of the Internal Revenue Service (IRS), HMRC, or any other tax authority.

Participants were required to submit a tax return based on the following fixed (and known) parameters: a tax rate of 40% and a penalty rate applied to unpaid tax of 50%. The values for the probability of audit, the tax rate, and the fine rate were set so that the optimal action for a risk-neutral, payoff-maximizing participant was to underreport his or her tax liability, matching the reality in the field (Allingham and Sandmo 1972; Andreoni, Erard, and Feinstein 1998).

Although the instructions did not tell them to do so, participants could increase their financial payment by either underdeclaring income or overdeclaring expenses. In either case, the most they could gain would be to declare a tax liability of zero. This translates into a possible gain relative to full compliance of £13.56 (at the time, US$20.34) for a task that took 22 minutes on average.[1]

After reading the instructions, participants were asked to complete a practice tax form based on a simple profile for which they were told they would not be paid. After completing the practice form, participants were informed about the payoffs they would have received had they been audited (vs. not been audited) on their practice tax declaration.

The majority of items in the tax return were verifiable if audited. It is essential that income amounts be verifiable because they form a direct part of the participant's payoff; as such, the experimenter must know the value to be able to pay it. Expenses, however, offer the experimenter the ability to set unverifiable items, in that the expenses reduce the tax paid, so participants can increase their payoff by raising expenses without the experimenter needing to know the true value. Unverifiable expenses potentially allow participants a greater opportunity to evade, a mechanism found to have an effect in empirical studies (Kleven et al. 2012). We allocated the value

---

[1]The task can be broken down as an average of seven minutes to read the instructions, two minutes to perform the practice round, three minutes for the tax filing, and ten minutes to complete the questionnaire.

of one of the expenses (i.e., property expenses) to be equal to the roll of a six-sided die multiplied by 2,000 ECU. Because a participant's die roll is unverifiable, it is rational for them to declare the maximum allowable value for the expenses field—that is, 12,000 ECU, equal to rolling a six. Although we can never verify whether a participant misreported that expense item, we can detect noncompliance at the sample level, because the distribution of die rolls (and therefore of declared values on that item in the tax return) should be uniform if participants are compliant (Fischbacher and Föllmi-Heusi 2013).

When participants had completed their tax form, they saw their tax calculation on the screen. This was done to remove any computational burden from the participants, and it is similar to tax calculators that are available online. They could then either repeat the process to change their details or submit their tax return. After submission, the computer randomly determined whether the participants were to be audited, and the participants saw their payoff from the experiment.

Participants then completed a questionnaire; it included two open-ended questions about their choices in the experiment, questions regarding participants' attitudes toward taxes using Likert scales, and questions regarding sociodemographic characteristics. They were informed that the questionnaire would not affect their payoff, and they could leave any question blank if they wished. Finally, participants were told they had completed the experiment and given details about how to opt out of having their responses included in the data set, had they wished to do so.

Participants' experimental balance was calculated at the end of the experiment as the total of the two income streams in the profile minus the tax payable on their declared liability and any fines occurred from the underpayment of tax due. It is important to note that overdeclaration of income could not raise participants' payoffs, and the experimental instructions were clear about this. Participants' earnings in ECU were converted to cash at a rate of 50 pence per 1,000 ECU; average earnings were £29.62 (at the time, US$44.43).

Experimental Design

The experiment consists of seven different treatments in a between-subjects design, summarized in Table 2. In our baseline treatment, Base, the tax form was not prepopulated. In the Corr treatment, the tax form had the self-employment income field prepopulated with the same total amount as in the profile, the sum of the two values given for self-employment income, and the tax form displayed that the information in the tax authority database was the two values corresponding to the two self-employment income streams in the profile. This corresponds to the case in which the tax authority has access to quality third-party reporting and therefore can correctly prepopulate the taxpayer's income (Gale and Holtzblatt 1997). In the United Kingdom, third-party reporting forms the basis of the Pay-As-You-Earn system, such that the correct tax is paid at source and many employees are not required to submit a year-end tax return.

In the Under treatment, the self-employment income field was prepopulated with an incorrect value equal to one of the two subitems of the self-employment income in the profile, and the tax form displayed that the information in the tax authority database was that single income stream. This captures the case in which the tax authority either has incomplete access to third-party data (e.g., an employer not providing this information) or is unaware of that stream of income. This error in prepopulation leads the tax authority to underestimate the tax liability of the participant. In the Over treatment, the tax form displayed that the information in the tax authority database consisted of three values, one of which was a double-counted entry. Thus, the value used to prepopulate the self-employment field of the tax form was greater than the actual income level in the participant's profile. This error in prepopulation leads the tax authority to *over*estimate the tax liability of the participant.

As argued in H_2, we expected a large incidence of noncompliance in the Under treatment, either because inertia leads people not to change their prepopulated entries or because people learn of the experimental tax authority's ignorance of the true profile values and engage in active noncompliance. To test whether behavioral nudges can mitigate the negative effects of incorrect prepopulation, we consider three additional versions of the Under treatment. The first was UnderGeneric, in which the prepopulated value was locked. To edit that field, participants had to first click a checkbox positioned next to it. In addition, participants also had to recheck that box to confirm the new value they entered before filing the tax return.

Table 2. Treatments Used in the Experiment

Treatment	Description
Base	No information reported and all four fields left blank
Corr	Correct self-employment income streams reported, correct self-employment income prepopulated
Over	Double counting of one income stream reported, incorrect (value too high) self-employment income prepopulated
Under	Omission of one income stream reported, incorrect (value too low) self-employment income prepopulated
UnderGeneric	Omission of one income stream reported, incorrect (value too low) self-employment income prepopulated, click of checkbox required to edit prepopulated field (and confirmation of edit)
UnderAlways	Omission of one income stream reported, incorrect (value too low) self-employment income prepopulated. Additional message on screen: "Most people in your circumstances enter an income value of more than 40,000. Values below this amount are more likely to be audited. Click the tickbox to confirm you wish to proceed."
UnderTrigger	Omission of one income stream reported, incorrect (value too low) self-employment income prepopulated. Same message as UnderAlways displayed only if participant files self-employment income value less than 40,000.

The second version was UnderAlways, which featured the following message: "Most people in your circumstances enter an income value of more than 40,000. Values below this amount are more likely to be audited. Click the tickbox to confirm you wish to proceed." This treatment was intended to trigger a descriptive norm of compliance and reminded participants of the nature of the audit rule. Social psychologists (e.g., Cialdini et al. 2006) have long argued for the effectiveness of descriptive norms as catalysts of behavior change (for a review of the evidence of norms applied to tax compliance, see Onu and Oats 2014).

Recently, a case has been made for the inclusion of reactive defaults. Smith, Goldstein, and Johnson (2013) propose using defaults that react to inputs by decision makers. We implement a treatment that approximates this recommendation: Under-Trigger, in which the same message as UnderAlways is featured, but only if the participant enters a total self-employment income amount lower than 40,000 ECU.

Our choice of nudge in the UnderAlways and UnderTrigger treatments was based on one of the mechanisms used by tax authorities to identify tax evaders, which is to target outliers from within a given group (e.g., based on industry). For example, the IRS's "DIF score" will produce "audit flags" for taxpayers who deviate from the average behavior of their group (Alm and McKee 2004). Because the probability of audit is endogenous with respect to the participant's declaration in the experiment, we can use a nudge to inform participants of the tax authority's operational process. We opted for the value for income displayed in the nudge to be below the actual value given in the profile, reflecting the process whereby outlying declarations are subject to higher probability of audit. We also decided that it should be above the value used for the prepopulation for the message to have some degree of saliency.

Sample and Recruitment

Our sample consisted of a pool of participants who volunteered to take telephone and online surveys run by ICM, a market research company. ICM sent an invitation e-mail to its participant pool to take part in an online decision-making experiment. When registering their interest in the experiment, participants were asked to fill out a questionnaire comprising a series of standard demographic questions. ICM included only participants who stated that they were over 18 years old and were either self-employed or employed full-time; this meant that they were U.K. tax residents. ICM then invited at random 755 people from those who met our sampling criteria. Participants were required to have Internet access as well as a six-sided die; the invitation e-mail included several online links to simulated dice roll websites for those who did not have access to a physical die. Out of the 755 people invited, 554 completed the experiment.[2] Just over 60% of our participants were male; participants' age ranged from 18 to 78 years old; the average age was 44.5 years old for women and 47.3 years old for men. We found that 34% of participants were self-employed and

66% were employees; the gender distribution was roughly the same for both employment categories.

Experimental Procedures

The experiment was operationalized through a customized website designed by the experimenters and hosted by the University of Exeter. The experiment took place between February 9, 2015, and April 12, 2015. ICM provided each participant with a link to the experimental website and a unique login username and password. We could not match usernames to actual participant data, and ICM did not have access to participant decisions, making this a double-blind experimental design. This was made explicit to participants when they were invited to participate.

After logging in, each participant read an on-screen set of instructions that detailed the task they were required to perform. Participants were also told that they would be paid a fixed £5 (US$7.50) sum for completing the experiment and would have the opportunity to earn more in line with their decisions in the experiment. The instructions detailed several examples of the potential outcomes from various declaration choices (for the full set of screenshots, see the Web Appendix).

Results

We treat each individual decision in our analysis as an independent observation and make treatment comparisons using standard statistical tests. Unless stated otherwise, we report two-sided tests throughout. We complement these with econometric analysis, which incorporates individual characteristics as well as responses to the postexperimental questionnaire.

Our main measure of analysis is based on the fields in the tax form that are verifiable by the experimenters—therefore, we exclude property expenses. We define observable tax liability as self-employment income + property income – self-employment expenses. In line with this variable, we define compliance as (declared observable tax liability)/(actual observable tax liability). If an individual has a compliance value of 1, then (s)he is classified as compliant. This ratio defines two types of noncompliance. If the ratio is smaller than 1, that person is undercompliant; if that ratio is higher than 1, that person is overcompliant. The latter type is relevant because the prepopulated amount in the Over condition overestimated the participant's taxable income, and passive acceptance of the default value could lead people to overpay taxes.

Prepopulation and the Default Heuristic

To address H_1, we analyze data on completion times for the tax-filing component of the experiment. The median completion time in our sample was 155 seconds, and 90% of participants completed the filing task in 6 minutes or less. We do not include in the analysis of completion times four outlier participants who took more than 90 minutes to finish.

If participants passively accept defaults as part of their filing process, as the heuristics approach would suggest, we should observe shorter completion times in the treatments with defaults than the Base treatment. Figure 1 displays average completion times across all seven treatments. Compared with Base, we found either no significant difference in completion times (Corr: $z = .65$, $p = .514$; UnderGeneric: $z = 1.73$, $p = .083$,

---

Figure 1. **Average Completion Time for the Experimental Task**
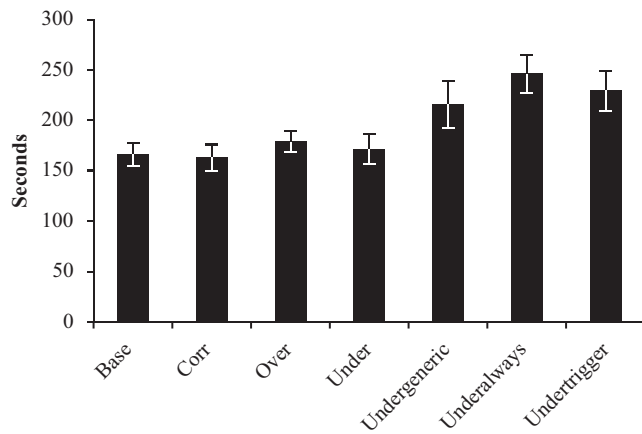
Mann–Whitney test [MWT]) or *longer* completion times (UnderAlways: z = 4.14, *p* < .001, MWT; UnderTrigger: z = 2.83, *p* = .005, MWT).

> Observation 1: The introduction of defaults does not lead to shorter average completion times.

This finding suggests that, at least for some participants, the presence of defaults led to greater deliberation with regard to what amounts to declare. As such, it is relevant to condition our analysis on compliance behavior.

Figure 2 shows the proportion of compliant, undercompliant, and overcompliant participants in each treatment; Figure 3 outlines the average completion time for each type of participant. We note that in all but the Over treatment, the number of overcompliant participants is extremely small, so there is no meaningful analysis to be done in those cases. Thus, we report completion times only for overcompliant participants in the Over treatment.

It is noteworthy that in the Base treatment, undercompliant types took, on average, 90 seconds longer to complete the tax return than compliant types (z = 3.05, *p* = .002, MWT). The same is true in Corr, in which the correct amount was already prepopulated; in fact, there is no statistically significant difference in completion times between Corr and Base for either compliant (z = .351, *p* = .726, MWT) or undercompliant (z = .126, *p* = .900, MWT) types.

We note that compliant participants took longer to file their return in Over than in Base (z = 2.497, *p* = .013, MWT) and Corr (z = 2.654, *p* = .008, MWT). In contrast, undercompliant types took as long in Over as in Base (z = 1.416, *p* = .157, MWT) or Corr (z = 1.187, *p* = .253, MWT). Notably, the 40% of overcompliant people in the Over condition took as long as the compliant (z = .835, *p* = .404, MWT) and undercompliant (z = .680, *p* = .497, MWT) types to complete the tax return.

> Observation 2: Undercompliance requires greater deliberation time than compliance when those behaviors are acts of commission.

Figure 2. **Percentage of Compliant, Undercompliant, and Overcompliant Participants**
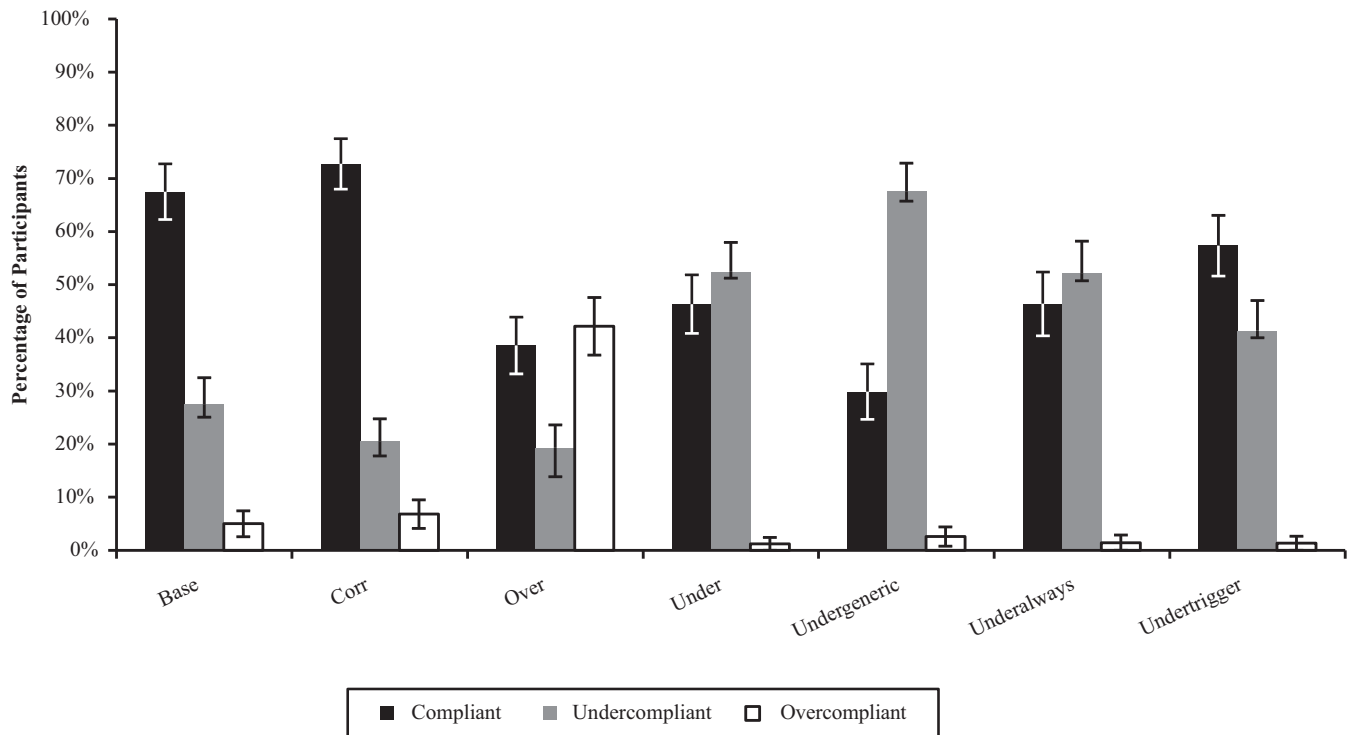
Figure 3.   **Average Completion Time for the Experimental Task Conditional on Participant Type**
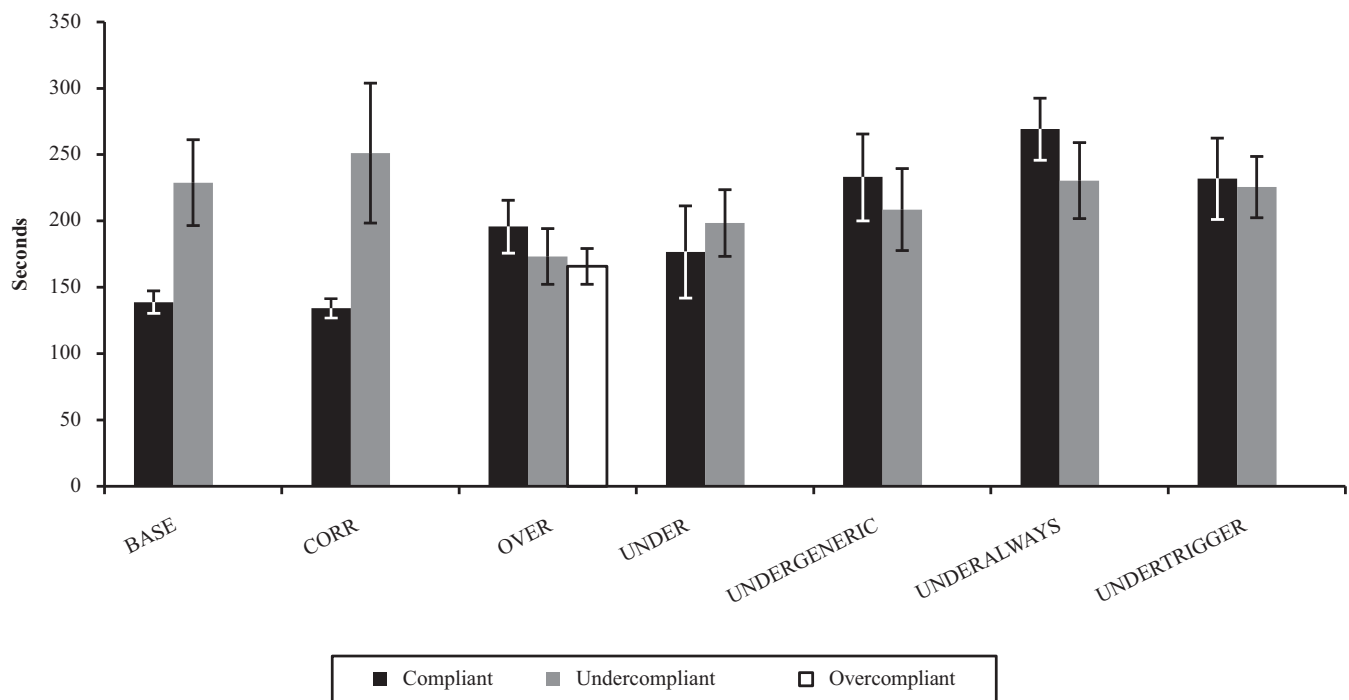
The average completion times by compliant participants in the Under and Over conditions are not significantly different from those of their counterparts in Base and Corr. The same is true for undercompliant participants. However, when we divide the group of overcompliant participants in Over into those who did not alter the prepopulated amount ($N = 28$) and those who did ($N = 7$), the average completion time of the former subgroup is 139 seconds, while the completion time of the latter is 274 seconds. The average completion time of those who accepted the incorrect default is not significantly different from the completion time by compliant types in Corr ($z = .142$, $p = .887$, MWT). This suggests that the behavior of some participants may be driven by inertia.

Likewise, if we split the sample of undercompliant participants in the Under condition into those who left the default value unchanged ($N = 27$) and those who did not ($N = 16$), a similar pattern emerges to that observed in Over. Those who accepted the default value and underreported their tax liabilities took, on average, 142 seconds to complete the filing task; this is not significantly different from the average completion time by compliant types in Base ($z = .801$, $p = .423$, MWT) or Corr ($z = .672$, $p = .502$, MWT). This suggests that the behavior of some participants in both treatments is also driven by the default heuristic.

We can therefore classify undercompliant people in the Under condition into two categories: 15 participants (35%) are passively undercompliant—that is, they accept the default value and are honest otherwise. The remainder are actively undercompliant—some keep the incorrect default value and evade in other fields in the tax return, and others change all fields, including the prepopulated one.

Observation 3: A minority of people passively accept incorrect defaults; their compliance behavior is primarily a function of the prepopulated value. However, most undercompliance is driven by active choice.

In short, our data broadly rejects H_1. Only a few participants' behavior is consistent with the default heuristic. Most deviations in completion time are explained by deliberate actions.

The Effect of Prepopulating Tax Forms

We begin by examining the impact of correctly prepopulating income fields in tax forms on compliance. Figure 2 indicates that in the Base condition, just over two-thirds of participants were compliant, and 28% were undercompliant. There are relatively more compliant participants (73%) and relatively fewer undercompliant participants (20%) in Corr, although the distributions of types are not significantly different in the two treatments ($p = .574$, Fisher's exact test [FET]). However, the proportion of participants who entered the correct value in the income field that is subject to prepopulation went from 75% in the Base treatment to 99% in Corr ($z = 4.356$, $p < .001$, MWT).

Observation 4: Correctly prepopulating a field in the tax form leads to higher compliance in that field, although not to higher compliance overall.

In contrast, prepopulating the income field incorrectly leads to a decrease in the proportion of compliant participants. In the Over treatment, that is due to an increase in the proportion of overcompliant people, while the proportion of undercompliant participants is unchanged. As a result, we observe a significant

difference in the distributions of types in the Base and Over conditions ($p < .001$, FET). In the Under treatment, the drop in compliant participants relative to Base is due to an increase in undercompliant types, while the proportion of overcompliant types remains unchanged. The type distributions in Base and Under are significantly different ($p = .002$, FET).

> Observation 5: Introducing incorrectly prepopulated fields results in fewer compliant people.

In short, our data provide strong support for H_2 and H_3.

The Effectiveness of Norm-Conveying Nudges

The next question is to what extent can nudges imbedded within the tax return mitigate the adverse effects of incorrect prepopulation. To answer this, we use the Under treatment as the de facto baseline condition and investigate whether nudges can "recover" compliance levels back to those observed in the original Base treatment (or even higher).

Figure 2 shows that the effect of nudges on compliance is rather mixed: in the UnderGeneric treatment (which featured a checkbox that participants had to uncheck before altering the content of the prepopulated field), the proportion of fully compliant participants is significantly lower than in the Under treatment ($p = .049$, FET). Figure 3 shows that although average completion time for compliant types was significantly higher in UnderGeneric than in Under ($z = 3.103$, $p = .002$, MWT), there was no significant difference in average completion time among undercompliant types in both treatments ($z = .030$, $p = .976$, MWT). In other words, the introduction of a checkbox increased undercompliance by introducing a physical barrier to changing the prepopulated field. This is manifested in the extra 56 seconds it took compliant participants to file their returns in the UnderGeneric treatment compared with compliant participants in Under.

The introduction of a descriptive norm message plus a confirmation checkbox (UnderAlways) had no discernible effect on the proportion of compliant participants ($p = 1.000$; FET). It increased the average length of time that compliant participants took to complete the tax return by over 90 seconds ($z = 4.592$, $p < .001$, MWT) compared with those in the Under treatment. There was no significant difference in the completion time of undercompliant participants in UnderAlways ($z = 1.555$, $p = .120$, MWT).

It was more difficult to be passively undercompliant (as described in relation to the Under treatment) in the UnderAlways treatment, because participants were forced to acknowledge the statement of the norm by checking a box before they could submit their return. In other words, despite creating a psychological barrier to filing an incorrect tax return, which manifested itself in longer average completion times for the compliant participants, there was little effect in terms of dissuading undercompliance.

Finally, in the UnderTrigger treatment, the same message was triggered by the participant's filing behavior, which led to a greater proportion of compliant than undercompliant people, although the distributions of participant types in Under and UnderTrigger were not significantly different ($p = .200$, FET). Again, the effect of the norm trigger on the compliant types' behavior is apparent: they took, on average, 45 seconds longer in UnderTrigger than in Under ($z = 2.446$, $p = .014$, MWT).

> Observation 6: Nudges that take the form of physical barriers to changing default entries in tax forms compounded the undercompliance that exists with incorrect prepopulation. Messages with descriptive norms did not achieve significant increases in the proportion of compliant types.

Observable Characteristics and Self-Reported Measures

As part of the postexperimental questionnaire, we collected data on several sociodemographic variables: age, gender, employment status, and self-reported annual income. We also asked participants a question related to their attitudes toward taxes in general: "Do you think cheating on taxes if you have a chance is justifiable? Please state 1 if it is never justifiable, 10 if it is always justifiable, or a value in between." We also asked participants to comment on how they filled in the income and expenses fields. We considered four categories when classifying responses: Rule Following, Honesty, Strategic/Evader, and Other. Although the overwhelming majority of responses fit only one category, some responses fit two categories—often Rule Following and Honesty. The majority of responses (57%) were classified as Rule Following, while the second-most-coded category was Honesty (27%); 21% of responses were classified as Strategic/Evader, and 13% were coded as Other. The proportion of each of the four response categories was approximately constant in all treatments—this means we cannot use these variables to explain treatment level differences in compliance. However, they may still be useful to explain choices at the individual level.

Table 3 reports the results of a series of logit models estimating the probability of being a compliant type. Model 1 considers only the relevant treatments as dummy variables; the omitted treatment is Base. The basic findings reported previously are confirmed: the average probability of being compliant is higher in Corr, though the difference is not significant. The reverse is true of the Over treatment (note that our dependent variable equals zero if a participant either underreports or overreports tax liabilities), although again, the difference is not statistically significant. The average likelihood of being compliant is significantly lower in Under; that likelihood drops significantly further in the UnderGeneric treatment ($\chi^2(1) = 4.50$, $p = .034$); the likelihood of being compliant in UnderAlways is not statistically significantly different from that in Under ($\chi^2(1) = .00$, $p = .997$). The likelihood of being compliant in the UnderTrigger treatment is not significantly different from Under ($\chi^2(1) = 1.89$, $p = .170$) but also not significantly different from Base.

Model 2 incorporates participants' observable characteristics—in particular, age, gender (through a male dummy variable), employment (through a dummy variable for self-employment [SelfEmpl]), self-reported annual income (Income), and age. None of these variables are significant, and the sign and significance level of the treatment dummies are unchanged. Model 3 incorporates a variable measuring people's attitudes toward paying taxes (TaxAtt) as well as coded responses to the open-ended questions about how they approached their filing decision (RuleFollower, Honest, Evader/Strategic, and Other). Introducing these variables does not change the sign or significance of the coefficients on the other regressors, except for the case of the coefficient on Under, which is now significant at the 5% level ($p = .037$).

Table 3.   Probit Estimates of the Determinants of Compliant Types

DV: Compliant	1	2	3
Constant	.731** (.238)	.409 (.512)	.509 (.736)
Corr	.250 (.338)	.128 (.348)	.388 (.435)
Over	−1.197** (.328)	−1.201** (.339)	−1.468** (.414)
Under	−.877** (.326)	−.942** (.336)	−.864* (.414)
UnderGeneric	−1.584** (.345)	−1.669** (.358)	−1.731** (.432)
UnderAlways	−.876* (.339)	−.982** (.357)	−.679 (.442)
UnderTrigger	.435 (.334)	−.526 (.344)	−.315 (.423)
Male		−.232 (.188)	−.145 (.229)
SelfEmpl		.258 (.208)	.060 (.250)
Income		.010 (.049)	.040 (.060)
Age		.009 (.008)	.00003 (.010)
TaxAtt			−.118* (.054)
RuleFollower			1.018** (.390)
Honest			1.407** (.382)
Evader/Strategic			−2.242** (.359)
Other			−1.328** (.482)
N	554	548	543
Pseudo R^2	.06	.07	.31
LL	−359.7	−354.2	−260.0

*$p < .5$.
**$p < .01$.
Notes: Compliant is a dummy variable equal to 1 if participant declared correct verifiable liability in tax form.

The coefficient on TaxAtt is negative and significant: participants who feel strongly that evading taxes is justifiable are more likely not to report their true tax liability in the experiment. In terms of the coded free-form responses, participants who described their behavior in terms of following rules or instructions, or who described their actions as a function of honesty, were more likely to be compliant types. In contrast, those who described their actions as evasive or strategically grounded were less likely to comply. Finally, those who did not have a clear description of their actions, or who reported not having a clear-cut strategy during the experiment, were less likely to be a compliant type. We conjecture that the last category captures some of the participants who were undercompliant through error, as opposed to premeditated evasion.

Observation 7: Attitudes toward taxes are strongly correlated with undercompliance.

Revenue

We conclude our analysis by looking at the revenue consequences of defaults and nudges in terms of the level of reported tax liability. Our current analysis has centered on the proportion of people who underreport, overreport, or correctly report their tax liabilities. It is possible that although defaults and other types of nudges we consider in our design might not significantly change the proportion of undercompliant participants, they might change the amount these participants underreport.[3] We report on the level

[3]We exclude from this analysis six participants who declared either a negative tax liability or a tax liability that meant they would make negative payoffs. In the previous analysis, these people would have been classified as undercompliant or overcompliant. Because they are so few, including them did not affect our analysis; however, because tax revenues are continuous, they constitute true outliers, and including them would skew means and standard errors.

of liability declared by participants (i.e., the factor that determines revenue before the application of any audit and payment of unpaid taxes or fines) as a proxy for revenue because the figure of actual revenue may be biased by the outcomes of particular audits.

Figure 4 displays the average reported tax liabilities in each of the seven treatments. The average tax liability in the Base treatment was 58,629 ECU. Liability is significantly higher in the Corr treatment (61,936 ECU; z = 2.051, p = .040, MWT). The average liability in Over is 67,711 ECU, which is significantly higher than Base (z = 4.620, p < .001, MWT) and driven by the high proportion of people who retained the incorrectly prepopulated entry in their tax form. The average liability in Under is equal to 49,456 ECU, significantly less than in Base (z = 3.257, p = .001).

The introduction of nudges in the context of the Under treatment led to a mixed result with regard to liability. The introduction of a checkbox in UnderGeneric led to a further reduction in revenues (44,547 ECU, z = 1.985, p = .047, MWT), while the static descriptive norm message (UnderAlways) resulted in larger average liabilities (52,190 ECU), although the difference relative to Under is not significant (z = .456, p = .648, MWT). However, the reactive nudge (UnderTrigger) led to significantly larger average revenues (55,660 ECU) relative to Under (z = 2.270, .023; MWT), to the extent that there were no significant differences relative to Base (z = 1.312, p = .189). This suggests that the reactive norm message, while unable to affect the number of undercompliant participants, did effectively reduce the amount of evasion they engaged in.

Observation 8: Reactive nudges reduced the amount of revenue lost from evasion on tax forms with incorrect prepopulation.

In short, our data provide only partial support for H_4 and H_5. While nudges with descriptive norms were generally ineffective at

Figure 4.   Average Declared Liability
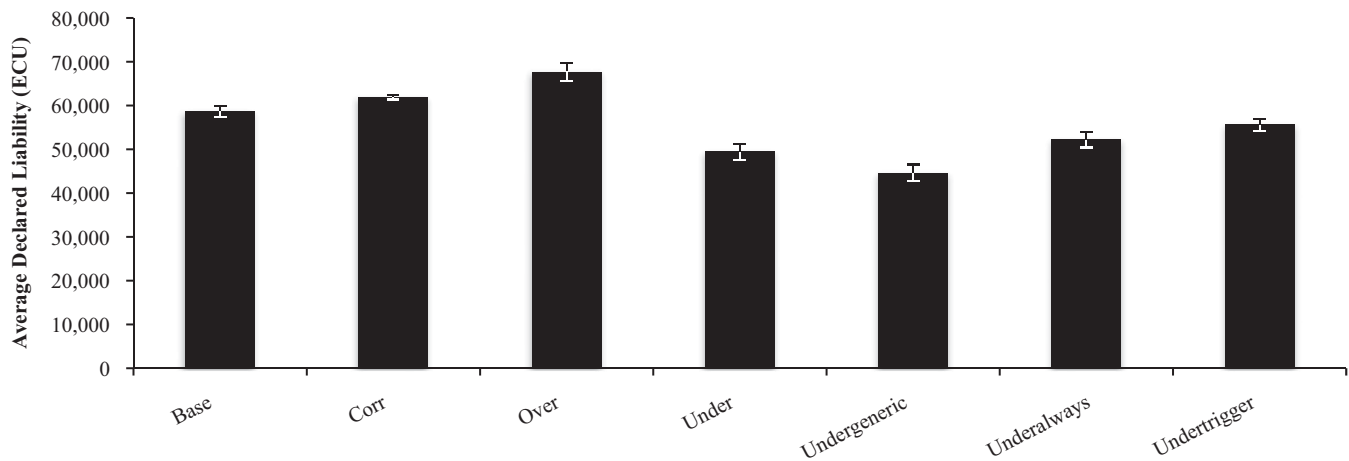
changing the proportion of undercompliant people, reactive norms did reduce the extent to which undercompliant people evaded.

Discussion

Our study aims to understand how defaults and nudges containing descriptive norm messages shape tax-paying behavior. As such, one of the primary contributions of this article is to understand the behavioral drivers of honesty in the tax context. In this sense, our work builds on Mazar and Hawkins (2015), who study the extent to which people engage in deceptive behavior when it is financially beneficial. That study examines how lying changes as a function of whether the deceptive action was preset by the experimenter. Mazar and Hawkins find that deceptive behavior is indeed more prevalent when it is an act of omission. The authors propose that rejecting a correct default is psychologically difficult: it involves not only lying but also rejecting a preexisting truthful statement.

Our analysis offers limited support to Mazar and Hawkins's (2015) interpretation of deceptive behavior as being based on self-concept maintenance. The proportion of compliant participants in the treatments with correct default values is the same as in our baseline treatment. The time undercompliant participants spent completing the form when the prepopulated value was correct was no different than that spent by undercompliant participants in the baseline condition. However, collected tax revenues in the correct prepopulation treatment were significantly higher than in the baseline treatment. This is only possible if those evading do so by a smaller extent. Importantly, this primarily occurred in the nonprepopulated fields. Nudges that displayed a normative message about compliance in response to people's inputs raised average declared tax liabilities close to baseline levels.

We provide evidence of different motivations for why and when defaults are difficult to override, especially when they are incorrect. For a small subset of people in our sample, defaults reduced the cognitive cost of engagement with the filing task. Thus, those people will be compliant if the default is correct but will be under- or overcompliant if the default is incorrect. This behavior is therefore consistent with the heuristics approach in decision making

and supports the claim by Kotakorpi and Laamanen (2015) that prepopulated tax forms reduce the cognitive costs of tax filing.

A large proportion of participants, when faced with an incorrect default (either under- or overestimating tax liabilities), responded by evading even more. Participants may have interpreted incorrect prepopulation as incompetence, which could signal greater opportunities for evasion, which is a key determinant of noncompliance in the field (Kleven et al. 2012).

Defaults are extremely powerful: their effect dominates the power of normative messages, which have been shown to be particularly effective in other policy contexts (Cialdini et al. 2006). Only a normative message that was responsive to actual behavior was able to mitigate the adverse effect of an incorrect default, thus providing strong support to Smith, Goldstein and Johnson's (2013) proposal for using reactive defaults. Even then, the effectiveness of reactive nudges was primarily on the extent to which participants evaded, rather than on the proportion of evaders.

One important dimension that we could not explore in the present study is the role of taxpayer trust in the tax agency and government in general. It is plausible that people have low self-efficacy in the tax domain and choose to trust the values presented in the tax form. This makes it difficult to distinguish between inertia and self-efficacy explanations of behavior. Self-efficacy could also be at the heart of our revenue result: although the number of undercompliant people was not reduced, the amount of evasion was reduced. Future work could determine whether the reduction in evasion occurs because those who would have accepted the defaults actively changed information, or because those who were actively undercompliant actually became more conservative in their evasion.

One important avenue for future research is to further understand how descriptive and injunctive norm information interacts with defaults and other forms of nudges in an honesty and/or tax context. Norms and abidance to them are an integral part of one's identity, whether in a social or individual sense. Therefore, by manipulating the extent to which people violate their sense of group identity by evading, we could construct choice environments that deter dishonest behavior.

Finally, our data consist of decisions by U.K. resident taxpayers. It is possible that attitudes toward government and taxes may differ across countries. As such, further research should investigate the extent to which our results hold in different cultural contexts.

Policy Implications

The experiment detailed in this article reflects potential differences in the design of online tax forms in the United Kingdom today and those that may be used in the near future under recent proposals for change. The treatments reflect situations that might arise under the new filing system relating to the nature and quality of third-party reported data used to prepopulate tax forms.

When reviewing the results on an experiment designed to test the effects of defaults on honesty, Mazar and Hawkins (2015, p. 117) suggest that "it might be even more effective to ... have tax software automatically pre-fill key fields with available information and require applicants to actively override them rather than typing amounts into blank fields." We argue that this should only be the case for fields for which the tax authority has extremely reliable information. One such case could be employment income: in the United Kingdom, there is already a well-developed system of Pay-As-You-Earn tax reporting, which could be used to prepopulate tax returns of people who have multiple sources of income. Bloomquist et al. (2012) report that the quality of such third-party data in the United States is high, unlike data capital earnings. Our evidence suggests that tax-return entries for which government tax agencies have low-quality data should not be prepopulated, as this could lead to increases in noncompliance.

We considered measures that the tax authority can implement to mitigate the potentially increased levels of noncompliance, but they should be carefully considered. A lock on the prepopulated field with a nudge for honesty actually caused compliance to worsen if the prepopulated value was below the true level of income. In addition, a static nudge containing a descriptive norm message on compliance that was always present had no discernible effect on compliance or revenue.

A reactive nudge reminding users that a lower declaration of income leads to a higher probability of audit was much more effective in increasing compliance, particularly in relation to the major income item in participants' profiles. We note that the message we used in this experiment is highly specific to the profile used. Generating an equivalent message in a real tax system would be nontrivial for a tax authority.

A noteworthy effect of the reactive nudge is that it increases response times by compliant types while having a limited effect on those who underreport their tax liabilities, both in terms of reducing their number or changing their deliberation time. This introduces a welfare question: in a more complex environment where the filing task could be expected to take days rather than minutes, this nudge introduces an extra burden on compliant types while not necessarily reducing the number and behavior of noncompliant types.

More generally, our findings indicate an important scope for nudges that react to users' behavior. These types of nudges are particularly well-suited to online environments, in which we expect most of the interactions between people and their government or private service will take place in the near future.

The fact that these nudges react to user behavior may lead users to perceive that the nudge is directed at them, thus further increasing their potential. This is a promising area for future policy implementation.

References

Akerlof, George A., and Rachel E. Kranton (2002), "Economics and Identity," *Quarterly Journal of Economics*, 115 (3), 715–53.

Allingham, Michael G., and Agnar Sandmo (1972), "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1 (3/4), 323–38.

Alm, James, and Michael McKee (2004), "Tax Compliance as a Coordination Game," *Journal of Economic Behavior & Organization*, 54 (3), 297–312.

Andreoni, James, Brian Erard, and Jonathan Feinstein (1998), "Tax Compliance," *Journal of Economic Literature*, 36 (2), 818–60.

Baron, Jonathan, and Ilana Ritov (1994), "Reference Points and Omission Bias," *Organizational Behavior and Human Decision Processes*, 59 (3), 475–98.

Becker, Gary (1968), "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76 (2), 169–217.

Bernheim, B. Douglas (1998), "Financial Illiteracy, Education, and Retirement Saving," in *Living with Defined Contribution Pensions*, Olivia S. Mitchell and Sylvester J. Schieber, eds. Philadelphia: University of Pennsylvania Press, 38–68.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian (2013), "Simplification and Saving," *Journal of Economic Behavior & Organization*, 95, 130–45.

Bloomquist, Kim, Ed Emblom, Drew Johns, and Patrick Langtieg (2012), "Estimates of the Tax Year 2006 Individual Income Tax Underreporting Gap," *IRS-TPC Research Conference, Publication*, https://www.irs.gov/pub/irs-soi/12resconEstimates.pdf.

Brown, Christina L., and Aradhna Krishna (2004), "The Skeptical Shopper: A Metacognitive Account for the Effects of Default Options on Choice," *Journal of Consumer Research*, 31 (3), 529–39.

Bruner, David, Michael Jones, Michael McKee, and Christian A. Vossler (2015), "Tax Reporting Behavior: Underreporting Opportunities and Pre-Populated Tax Returns," Working Paper No. 15-11, Department of Economics, Appalachian State University.

Campbell, Ernest Q. (1964), "The Internalization of Moral Norms," *Sociometry*, 27 (4), 391–412.

Cialdini, Robert B., Linda J. Demaine, Brad J. Sagarin, Daniel W. Barrett, Kelton Rhoads, and Patricia L. Winter (2006), "Managing Social Norms for Persuasive Impact," *Social Influence*, 1 (1), 3–15.

Cialdini, Robert B., Carl A. Kallgren, and Raymond R. Reno (1991), "A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior," *Advances in Experimental Social Psychology*, 24, 201–34.

Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren (1990), "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places," *Journal of Personality and Social Psychology*, 58 (6), 1015–26.

Cushman, Fiery, Liane Young, and Marc Hauser (2006), "The Role of Conscious Reasoning and Intuition in Moral Judgment," *Psychological Science*, 17 (12), 1082–89.

De Quervain, Dominique J.-F., Urs Fischbacher, Valerie Treyer, Melanie Schelthammer, Ulrich Schnyder, Alfred Buck, et al. (2004), "The Neural Basis of Altruistic Punishment," *Science*, 305 (5688), 1254–58.

Dolan, Paul, Michael Hallsworth, David Halpern, Dominic King, and Ivo Vlaev (2010), *MINDSPACE: Influencing Behaviour Through Public Policy*. London: Institute of Government.

Fischbacher, Urs, and Franziska Föllmi-Heusi (2013), "Lies in Disguise: An Experimental Study on Cheating," *Journal of the European Economic Association*, 11 (3), 525–47.

Forum on Tax Administration (2006), "Using Third Party Information Reports to Assist Taxpayers Meet their Return Filing Obligations— Country Experiences With the Use of Pre-populated Personal Tax Returns," Organisation for Economic Co-operation and Development (March), http://www.oecd.org/tax/administration/36280368.pdf.

Gale, William G., and Janet Holtzblatt (1997), "On the Possibility of a No-Return Tax System," *National Tax Journal*, 50 (3), 475–85.

Government Accountability Office (2005), "Highlights of a GAO Forum: The Federal Government's Role in Improving Financial Literacy (GAO-05-93SP)," (November), http://www.gao.gov/assets/210/202486.pdf.

Gigerenzer, Gerd (2010), "Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality," *Topics in Cognitive Science*, 2 (3), 528–54.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius (2008), "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels," *Journal of Consumer Research*, 25 (3), 472–82.

HMRC (2015), "Making Tax Easier: The End of the Tax Return," https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/413975/making-tax-easier.pdf.

Jensen, Jonas, and Florian Wöhlbier (2012), "Improving Tax Governance in EU Member States: Criteria for Successful Policies," Occasional Papers 114 (August), Occasional Papers 114, European Commission, http://ec.europa.eu/economy_finance/publications/occasional_paper/2012/pdf/ocp114_en.pdf.

Johnson, Eric J., and Daniel G. Goldstein (2003), "Do Defaults Save Lives?" *Science*, 302 (5649), 1338–39.

Johnson, Eric J., John Hershey, Jacqueline Meszaros, and Howard Kunreuther (1993), "Framing, Probability Distortions, and Insurance Decisions," *Journal of Risk and Uncertainty*, 7 (1), 35–51.

Johnson, Eric J., Suzanne B. Shu, Benedict G.C. Dellaert, Craig Fox, Daniel G. Goldstein, Gerald Häubl, et al. (2012), "Beyond Nudges: Tools of a Choice Architecture," *Marketing Letters*, 23 (2), 487–504.

Kanuk, Leslie, and Conrad Berenson (1975), "Mail Surveys and Response Rates: A Literature Review," *Journal of Marketing Research*, 12 (4), 440–53.

Kirchler, Erich (1998), "Differential Representations of Taxes: Analysis of Free Associations and Judgments of Five Employment Groups," *Journal of Socio-Economics*, 27 (1), 427–37.

Kleven, Henrik J., Martin B. Knudsen, Claus T. Kreiner, Soren Pedersen, and Emmanuel Saez (2012), "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark," *Econometrica*, 79 (3), 651–92.

Knutson, Brian, Charles M. Adams, Grace M. Fong, and Daniel Hommer (2001), "Anticipation of Increasing Monetary Reward Selectively Recruits Nucleus Accumbens," *Journal of Neuroscience*, 21 (16), RC159.

Kotakorpi, Kaisa, and Jani-Pettri Laamanen (2015), "Complexity, Salience and Income Tax Reporting Behaviour: Evidence from a Natural Experiment," working paper mimeo.

Lamberton, Cait (2013), "A Spoonful of Choice: How Allocation Increases Satisfaction with Tax Payments," *Journal of Public Policy & Marketing*, 32 (Fall), 223–38.

Lusardi, Annamaria, and Olivia Mitchell (2007), "Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education," *Business Economics*, 42 (1), 35–44.

Lusardi, Annamaria, and Olivia Mitchell (2014), "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, 52 (1), 5–44.

Madrian, Brigitte C., and Dennis F. Shea (2001), "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116 (4), 1149–87.

Mazar, Nina, On Amir, and Dan Ariely (2008), "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance," *Journal of Marketing Research*, 45 (6), 633–44.

Mazar, Nina, and Dan Ariely (2006), "Dishonesty in Everyday Life and Its Policy Implications," *Journal of Public Policy & Marketing*, 25 (Spring), 117–26.

Mazar, Nina, and Scott A. Hawkins (2015), "Choice Architecture in Conflicts of Interest: Defaults as Physical and Psychological Barriers to (Dis)honesty," *Journal of Experimental Social Psychology*, 59, 113–17.

National Audit Office (2015), "Tackling Tax Fraud: How HMRC Responds to Tax Evasion, the Hidden Economy and Criminal Attacks," NAO Report https://www.nao.org.uk/wp-content/uploads/2015/12/Tackling-tax-fraud-how-HMRC-responds-to-tax-evasion-the-hidden-economy-and-criminal-attacks.pdf.

O'Doherty, John P., Rald Deichmann, Hugo D. Critchley, and Raymond J. Dolan (2002), "Neural Responses During Anticipation of a Primary Taste Reward," *Neuron*, 33 (5), 815–26.

Onu, Diana, and Lynne Oats (2014), "Social Norms and Tax Compliance," Discussion Paper 006-14, Tax Administration Research Centre, University of Exeter, https://tarc.exeter.ac.uk/media/universityofexeter/businessschool/documents/centres/tarc/publications/discussionpapers/Social_norms_and_tax_compliance_May_14.pdf.

Peters, Ellen, Daniel Västfjäll, Paul Slovic, C.K. Mertz, Ketti Mazzocco, and Stephan Dickert (2006), "Numeracy and Decision Making," *Psychological Science*, 17 (5), 407–13.

Reeson, Andrew, and Simon Dunstall (2009), "Behavioural Economics and Complex Decision-Making: Implications for the Australian Tax and Transfer System," Report No. 09/110 (August 7), CSIRO Mathematical and Information Sciences, http://www.taxreview.treasury.gov.au/content/html/commissioned_work/downloads/CSIRO_AFTS_Behavioural_economics_paper.pdf.

Reno, Raymond R., Robert B. Cialdini, and Carl A. Kallgren (1993), "The Transsituational Influence of Social Norms," *Journal of Personality and Social Psychology*, 64 (1), 104–12.

Rilling, James K., David A. Gutman, Thorsten R. Zeh, Giuseppe Pagnoni, Gregory S. Berns, and Clinton D. Kilts (2002), "A Neural Basis for Social Cooperation," *Neuron*, 35 (2), 395–405.

Samuelson, William, and Richard Zeckhauser (1988), "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, 1 (1), 7–59.

Sánchez-Pagés, Santiago, and Marc Vorsatz (2009), "Enjoy the Silence: An Experiment on Truth-Telling," *Experimental Economics*, 12 (2), 220–41.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius (2007), "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science*, 18 (5), 429–34.

Sherif, Muzafer, and Carolyn W. Sherif (1953), *Groups in Harmony and Tension*. New York: Harper.

Smith, N. Craig, Daniel G. Goldstein, and Eric J. Johnson (2013), "Choice Without Awareness: Ethical and Policy Implications of Defaults," *Journal of Public Policy & Marketing*, 32 (Fall), 159–72.

Spranca, Mark, Elisa Minsk, and Jonathan Baron (1991), "Omission and Commission in Judgment and Choice," *Journal of Experimental Social Psychology*, 27 (1), 76–105.

Sussman, Abigail B., and Christopher Y. Olivola (2011), "Axe the Tax: Taxes Are Disliked More Than Equivalent Costs," *Journal of Marketing Research*, 48 (Special Issue), S91–S101.

Teper, Rimma, and Michael Inzlicht (2011), "Active Transgressions and Moral Elusions: Action Framing Influences Moral Behavior," *Social Psychological & Personality Science*, 2 (3), 284–88.

Thaler, Richard H., and Cass R. Sunstein (2008), *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Wright, Peter (2002), "Marketplace Metacognition and Social Intelligence," *Journal of Consumer Research*, 28 (4), 677–82.

Yu, Julie, and Harris Cooper (1983), "A Quantitative Review of Research Design Effects on Response Rates to Questionnaires," *Journal of Marketing Research*, 20 (1), 36–41.