Kenniscentrum
Psychologie en Economisch Gedrag

How choice architecture can promote and undermine tax compliance

Testing the effects of prepopulated tax returns and accuracy confirmation

Authors: Wilco van Dijk, Sjoerd Goslinga, Bart Terwel, & Eric van Dijk
No. 2 / 2020

The Knowledge Centre Psychology and Economic Behaviour is connected to the Department of Social, Economic, and Organisational Psychology of Leiden University

www.kcpeg.nl

How choice architecture can promote and undermine tax compliance

Testing the effects of prepopulated tax returns and accuracy confirmation

Authors: Wilco van Dijk, Sjoerd Goslinga, Bart Terwel, & Eric van Dijk

No. 2 / 2020

The Knowledge Centre Psychology and Economic Behaviour is connected to the Department of Social, Economic, and Organisational Psychology of Leiden University

Journal of Behavioral and Experimental Economics 87 (2020) 101574

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

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

How choice architecture can promote and undermine tax compliance: Testing the effects of prepopulated tax returns and accuracy confirmation

Wilco W. van Dijk[a,b,*], Sjoerd Goslinga[a,c], Bart W. Terwel[c], Eric van Dijk[a,b]

[a] Department of Social, Economic and Organisational Psychology, Leiden University, PO Box 9555, Leiden 2300 RB, the Netherlands
[b] Knowledge Centre Psychology and Economic Behaviour, Leiden University, the Netherlands
[c] Netherlands Tax and Customs Administration, the Netherlands

ARTICLE INFO

Keywords:
Default effects
Prepopulated tax returns
Tax compliance
Moral costs
Status quo bias

ABSTRACT

We tested the effects of prepopulated returns and accuracy confirmation on compliance. Participants were asked to report correct liabilities for different types of returns, whereby some had to confirm the accuracy of each reported liability and others not. Results showed that correctly prefilled returns yielded the highest rate of compliance, followed by returns that were not prefilled, followed by returns that overestimated liabilities, and with returns that underestimated liabilities displaying the lowest compliance. Moreover, accuracy confirmation increased compliance only for returns that overestimated liabilities. The present study indicates that both morality and defaults play a pivotal role in shaping the effects of prepopulated returns on compliance. Our findings imply that prepopulating tax returns should be done with care, because it can increase tax compliance when done correctly, but undermine it when done incorrectly.

One of the most significant innovations over the last twenty years in personal income tax systems, has been the development of prepopulated tax returns. Tax administrations use data from their own records and information that has been collected from third parties to prepare these returns. To establish accurate tax liabilities, taxpayers are usually required to check that prefilled information is complete and correct, and if this is not the case to self-report the correct and relevant information. Prepopulated returns increase administrative efficiency and make compliance with tax laws easier. Not surprisingly, these returns are becoming the norm—a notion corroborated by a recent survey showing that in 40 of the 58 surveyed advanced and emerging economies, personal income tax returns are (at least in part) prepopulated by tax administrations (OECD, 2019).

Prepopulated returns, however, may have a downside, as prefilling returns constitutes a change in choice architecture that may influence the perceived unethicality of tax evasion. In traditional returns, taxpayers need to self-report all relevant information—which can be done correctly or incorrectly. The choice architecture of prepopulated returns, however, requires taxpayers to review the accuracy of prefilled information—whereby they should retain accurate information, add missing information, and correct inaccurate information. We posit that especially this last requirement may impact how unethical taxpayers consider underreporting.

Traditional economic models explain dishonesty as an economic trade-off between the expected benefits of cheating and its perceived costs. According to this view, tax evasion would be a function of the expected financial gain of underreporting, the chances of being caught, and the imposed fines when being caught (Allingham & Sandmo, 1972; Becker, 1968). Modern perspectives on decisions in the moral domain have criticized the economic cost-benefit view as being too narrow. In particular, perspectives that consider morality as an underlying driver of decisions assume that people (also) derive utility from an internal standard of being able to see themselves as a moral person (Jacobsen, Fosgaard, & Pascual-Ezama, 2018; Mazar, Amir, & Ariely, 2008). According to these views, people are only prepared to cheat if they can maintain a self-concept of being honest; self-serving dishonesty would then be restrained by a concern to perceive oneself as moral. Any aspect that allows people to reduce ethical dissonance—the internal conflict between the temptation to profit from unethical behaviour and the desire to maintain a positive image of oneself—then constitutes a potential risk factor for cheating (Ayal, Gino, Barkan, & Ariely, 2015). We contend that prepopulated tax returns constitute such an aspect; a reasoning that fits with Jacobsen et al. (2018), who identified choice architecture as a potential situational risk factor for

This work was supported by The Netherlands Tax and Customs Administration and The Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS-KNAW). We thank Marco Burelli, Maike Grimberg, and Manon Schutter for assisting with the data collection.

* Corresponding author.
E-mail address: dijkwvan@fsw.leidenuniv.nl (W.W. van Dijk).

https://doi.org/10.1016/j.socec.2020.101574
Received 17 February 2020; Received in revised form 15 June 2020; Accepted 15 June 2020
Available online 17 June 2020

unethical behaviour.

An example of how choice architecture may induce dishonesty was provided by Mazar & Hawkins (2015), who showed that cheating was more likely when it involved passively accepting an incorrect (advantageous) default than when it meant actively overwriting a correct default. The underlying notion that maintaining a positive self-concept is less difficult for passive than for active acts of dishonesty aligns with the broader literature on omission bias showing that immoral acts of omission are judged as less maliciously motivated and less morally reprehensible than immoral acts of commission (Spranca, Minsk, & Baron, 1991). While these studies did not explicitly address tax returns, the relevance seems clear: Whereas correctly prepopulated returns can promote compliance, incorrectly prepopulated returns can undermine it.

Prefilling tax returns with accurate information will likely lead to more compliance. In prepopulated returns, changing correct entries into incorrect ones requires more (immoral) action than 'just' filling out incorrect information in non-prepopulated returns. Both compliance decisions require filling out incorrect information, but the former also requires overwriting correctly prefilled information and thus is a more active—and hence more immoral—form of non-compliance. In contrast, incorrectly prepopulated returns may lead to more underreporting. Taxpayers may be especially tempted to be non-compliant if prepopulated returns contain inaccuracies that are advantageous, when left uncorrected—thereby providing financial benefits at little moral costs.

Inactions leading to non-compliance do not necessarily imply a self-serving motive. Prefilled information can serve as default options—certain courses of action that take effect if nothing is specified by the decision-maker (Thaler & Sunstein, 2008)—that nudge taxpayers toward accepting preset liabilities (status-quo bias).[1] Doing so would positively affect compliance when liabilities are preset correctly, but negatively when they are preset incorrectly. Default effects should equally affect compliance in prepopulated returns that overestimate liabilities and in those that underestimate liabilities. A morality perspective on compliance, however, predicts that compliance will be more prevalent in latter returns—those that yield financial benefits when left unchanged.

Arguing from a morality perspective, the above can be summarized in the overarching hypothesis that correctly prefilled returns (*correct returns*) yield more compliance than returns that are not prefilled (*blank returns*), incorrectly prefilled returns that overestimate liabilities (*higher returns*), and incorrectly prefilled returns that underestimate liabilities (*lower returns*), respectively. We tested this hypothesis in a study that resembled a tax setting, and in which relevant data fields were sometimes left blank, sometimes correctly prefilled, sometimes prefilled with too high liabilities, and sometimes prefilled with too low liabilities. Compliance was measured as the percentage of liabilities that were correctly reported.

Experimental research on the impact of prepopulated tax returns on compliance has been sparse and the results far from conclusive, sometimes even conflicting. To illustrate, Fonseca and Grimshaw (2017) found that compliance was the same in correct returns as in blank returns, and observed more compliance in those returns than in higher or lower returns—whereby the latter two did not differ in compliance. In contrast, both Doxey, Lawson, & Stinson (2019) and Fochmann, Müller, & Overesch (2018) found that compliance in correct returns was the same as in higher returns and also found more compliance in those returns than in blank and lower returns. Moreover, Doxey et al. found more compliance in blank returns than in lower returns, whereas in Fochman et al.'s study no difference in compliance was found between these types of returns. Thus, the only consistent finding in these studies was that correct returns yielded more compliance than lower returns; none of these studies supported a morality interpretation of compliance.[2]

Research on dishonesty indicates that most people, rather than cheating all the time, occasionally act dishonestly while being honest on other occasions. This combination helps them to maintain a self-concept of being a moral person (Ayal et al., 2015; Jacobsen et al., 2018; Mazar et al., 2008). The inconclusive findings of aforementioned studies might have been due to insufficient leeway given to participants to cheat in a self-perceived acceptable way. In these earlier studies, participants made only a few compliance decisions (four or six), which restricted their opportunities to cheat on some and be honest on other occasions.

In the present study, we aimed at getting a more conclusive answer to the question how prepopulated returns impact compliance. We conducted a high-powered, controlled experiment in which participants had to make 100 compliance decisions. This provided participants with ample opportunity to combine honest and dishonest decisions in a way that enabled them to cheat without necessarily seeing themselves as an immoral person. Additionally, we tested another form of choice architecture that could possibly affect tax compliance: As a between-participants manipulation, half of the participants were presented with tick boxes that required them to confirm the accuracy of each liability they reported. We introduced this intervention, because we expected that it would improve compliance—a hypothesis based on several arguments. First, theorizing suggests that subtle measures to heighten self-engagement may increase honesty (Ayal et al., 2015). Confirming the accuracy of liabilities may constitute such a measure and yield a similar effect on in reviewing and filing tax returns. In addition, for each type of return, accuracy confirmation makes non-compliance a more active, hence more immoral, decision that is taken less likely. Furthermore, the required confirmation increases attention to the prefilled information, which should heighten alertness for inaccuracies, making corrections more likely. To our knowledge, we are the first to test this intervention in a tax setting.

1. Method

1.1. Design and participants

One hundred and two students from Leiden University were randomly assigned to one of the two conditions of a mixed design with Returns (correct, blank, lower, higher) as within-participants factor and Confirmation (non-confirmation, confirmation) as between-participants factor.[3]

[1] In the current study, we were not able to tease apart omission bias from status-quo bias. In our paradigm, inaction always leads to reporting the prefilled liabilities, whereas to make changes to prefilled liabilities, action is always needed. This feature that by retaining the prefilled numbers one is both inactive and retains the status quo is a correct description of how prefilled tax forms work. From a research standpoint one could argue that one cannot disentangle these effects in prefilled tax forms. This is not unique to tax studies but in fact often observed in decision-making research. Ritov and Baron (1992) are among the few who untangled these biases. In their decision-making research, they used scenarios in which change occurred unless action was taken, and they concluded that omission bias plays a major role in status-quo bias. In a tax setting, however, implementing a change unless one takes action seems unrealistic.

[2] Other noteworthy research on prepopulated tax forms and compliance concerns a study by Kotakorpi and Laamanen (2016). Using data from a Finnish policy experiment, they found that receiving a (partially) prefilled income tax return lead to a significant reduction in non-prefilled deductions and self-reported income, and an increase in deductions that were prefilled in the new system. Outside the tax context, Duncan and Li (2018) found in a context-free experiment that prefilled values increased honest reporting, whereas Morrison and Ruffle, (2020) in an insurance context found that prefilled values were only limited in their ability to reduce dishonesty in claim reports.

[3] We aimed to recruit 50 participants per between-participants condition, consistent with previous recommendations (Simmons, Nelson, & Simonsohn, 2013).

1.2. Procedure, task, and experimental manipulations

Upon arrival, participants received an envelope containing a form that listed the correct liabilities for the task (see below). Participants were seated in separate cubicles, and received instructions and completed the task on a computer. The study was approved by the institutional ethics board and informed consent was obtained from all participants. After completing the study, participants were fully debriefed, paid, and thanked.

The task consisted of reviewing and filing 100 (simplified) tax returns. Participants were asked to report, for each return, the correct liability that was listed on the earlier received form. There were 25 returns of each type of return and the presentation order for these 100 returns was randomized. During the task, four returns were presented simultaneously below each other on a screen, and participants were thus presented with 25 screens in total.[4]

All returns had an income field on the left and a liability field on the right. Each of the 100 returns included a different income, which was a random number between 90,000 and 110,000 points. The correct liability also differed for each of the 100 returns and was a random percentage between 48% and 52% of the shown income. Whereas the 100 income fields were always correctly prefilled, the 100 liability fields were not—these fields were correctly prefilled (*correct returns*: 25x), not prefilled (*blank returns*: 25x), incorrectly prefilled with liabilities that were 10% too low (*lower returns*: 25x), or incorrectly prefilled with liabilities that were 10% too high (*higher returns*: 25x).

Although participants were asked to report the correct liabilities, they could report any liability they wanted with the restriction that it could not be lower than zero. Participants were also not allowed to leave liability fields empty. After filing all four returns on a screen, participants continued to the next screen with again four returns; this continued until all 100 returns were presented and filed. The details of the returns and the order in which they were presented were the same for all participants.

Participants were informed that after filing all returns, reported liabilities would be subtracted from their total income, and remaining points would be converted into money (€0.05 per 100,000 points), rounded off to the nearest €0.10, and paid out. Instructions clarified that payoffs could vary between €0 and €5, and included examples of possible payoffs; these examples made apparent that lower (higher) reported liabilities increased (decreased) payoffs.

Participants were also informed that there was a 5% chance that, after they filed all their returns, these returns would be audited. After participants had filed all 100 returns, their returns were independently audited with a 5% probability. Payoffs were calculated as follows: (1) when there was no audit, payoffs were calculated as described before; (2) when there was an audit and participants' total reported liabilities were equal or higher than the correct total, payoffs were calculated as described before; and (3) when there was an audit and participants' total reported liabilities were lower than the correct total, the amount of underreporting was subtracted twice from their total income. This third calculation corresponds to a situation in which audited tax evaders have to pay the liabilities they still owed, increased with a fine of 100% of the underreported liabilities. With this payoff structure, full evasion (i.e., reporting a liability of 0 for all 100 returns) would yield the highest payoff possible (€5) when there was no audit, but the lowest (€0) when an audit did take place. Independent of whether there was an audit, full compliance (i.e., reporting the correct liability for all 100 returns) would result in a payoff of approximately €2.50, whereas overreporting would results in a payoff between €0 and €2.50.

In the *confirmation condition*, participants needed to confirm the accuracy of each reported liability separately by ticking a box, placed to the right of each liability field and labelled 'The liability that I report is the correct liability'. Thus, in this condition, participants had to tick in total 100 boxes, one for each liability they reported. Participants in the *non-confirmation condition* were not presented with tick boxes and not required to confirm the accuracy of the liabilities they reported.

[4] In the current study, randomisation resulted in a maximum of three returns of the same type on one screen.

2. Results and discussion

Data from four outliers on age (> 7 *SD*s older) were excluded from the analyses.[5] During the task, 33 participants were mistakenly presented with an incorrect liability on their list for the fourth correct return, and their responses for this return were coded as missing values. Consequently, analyses were performed with data from 98 participants (75 women, 23 men; M_{age} = 21.89 years, *SD* = 3.19) and included 9,767 observations (i.e., 100 compliance decisions of 65 participants made and 99 compliance decisions of 33 participants). There were 51 participants in the non-confirmation condition and 47 in the confirmation condition.[6] On average, participants needed 21 minutes to complete the study. Returns of two participants were audited, and their mean earnings were €2.35 (*SD* = €0.21).[7] For the 96 participants whose returns were not audited, mean earnings were €2.97 (*SD* = €0.79). Whereas 12 participants showed full compliance (i.e., reported the correct liability for all 100 returns), 13 choose full evasion (i.e., reported a liability of 0 for all 100 returns).

2.1. Overview of the analyses

We first conducted an overall repeated-measures analysis of variance (ANOVA). As previous research has shown that people are more likely to cheat when they are tired or bored (see Jacobsen et al., 2018), we included the position of the returns in the task as an additional factor in the analysis. To assess this factor, termed Time, we divided the 100 returns into five blocks of 20 returns, whereby a block consisted of five consecutive returns of each type of return. That is, the first block consisted of the first five correct returns, the first five blank returns, the first five higher returns, and the first five lower returns. The second block consisted of the next five consecutive correct, blank, higher, and lower returns, etc. The overall analysis was thus performed with Returns (correct, blank, higher, lower) and Time (block 1, block 2, block 3, block 4, block 5) as repeated measures, Confirmation (non-confirmation, confirmation) as between-participants factor, and compliance as dependent variable.[8] As robustness test, we reran the overall analysis without data from participants who displayed either full compliance or full evasion. Below, we first report the findings of the overall analysis, and discuss the effect of Time on compliance. This is followed by an evaluation of the first hypothesis, including both planned and post-hoc comparisons. Next, we evaluate the second hypothesis, again with several follow-up analyses.

[5] Three outliers had been assigned to the confirmation condition and one to the non-confirmation condition (they were 46, 53, 59, and 61 years old, respectively).

[6] For the reported analyses, the confirmation (12 men and 35 women) and the non-confirmation condition (11 men and 40 women) did not differ in gender composition ($\chi^2[1] = 0.21$, $p = .81$), nor did participants in the confirmation condition (M_{age}= 21.89 years, *SD* = 2.94) differ in age from those in non-confirmation condition (M_{age}= 21.88 years, *SD* = 3.43), $t(98) = 0.17$, $p = .65$.

[7] A show-up fee of €1 is not included in the reported payments.

[8] An analysis with gender as an additional factor in the design showed the same pattern of results. No interactions with gender were found, and therefore data were collapsed across gender for the reported analyses. Results did show a main effect of gender, which indicated that female participants were more compliant (*M* = 67.2%, *SD* = 37.4) than male participants (*M* = 38.7%, *SD* = 39.1), $F(1, 94) = 10.66$, $p = .002$. This finding is consistent with previous research suggesting that, generally, males cheat more than females (Duncan and Li 2018; Jacobsen et al., 2018).

Table 1
Test statistics of main and interaction effects of the overall test including the total sample and the selected sample.

Test statistics	Total sample (n = 98)		Selected sample (n = 73)	
Returns (within-participants)	$F(2.23, 214.19) = 35.64$	$p < .001$	$F(2.38, 168.98) = 40.99$	$p < .001$
Time (within-participants)	$F(2.12, 203.47) = 4.49$	$p = .011$	$F(2.14, 152.10) = 4.47$	$p = .011$
Confirmation (between-participants)	$F(1, 96) = 2.55$	$p = .11$	$F(1, 71) = 5.13$	$p = .027$
Returns × Time	$F(7.67, 736.51) = 2.20$	$p = .028$	$F(7.69, 545.80) = 2.22$	$p = .026$
Returns × Confirmation	$F(2.23, 214.19) = 2.58$	$p = .072$	$F(2.38, 168.98) = 2.98$	$p = .045$
Time × Confirmation	$F(2.12, 203.47) = 0.34$	$p = .85$	$F(2.14, 152.10) = 0.28$	$p = .77$
Returns × Time × Confirmation	$F(7.67, 736.51) = 1.41$	$p = .19$	$F(7.69, 545.80) = 1.43$	$p = .18$

Note. The selected sample did not include the data from participants who displayed either full compliance (i.e., reported the correct liability for all 100 returns; n = 12) or full evasion (i.e., reported a liability of 0 for all 100 returns; n = 13).

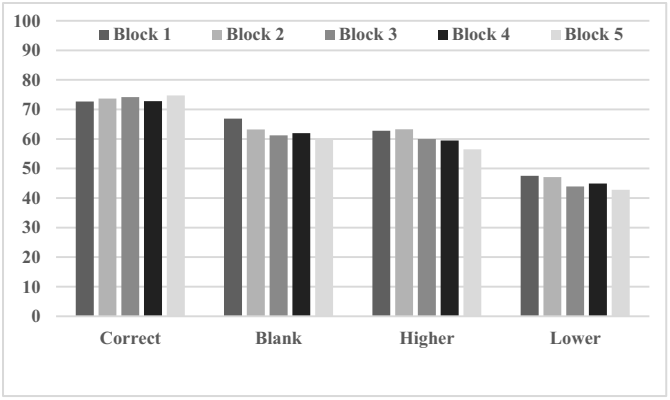

Fig. 1. Compliance per type of return for five consecutive blocks of 20 returns.

2.2. Overall analysis

Results of the overall analysis showed a main effect of Returns, a main effect of Time, but no main effect of Confirmation (the robustness test, however, did yield this main effect). We also found an interaction between Returns and Time and a (marginally significant) interaction between Returns and Confirmation, but neither an interaction between Time and Confirmation nor a three-way interaction between Returns, Time, and Confirmation (see Table 1, for test statistics).

2.3. The effect of time on compliance

The effect of Time indicated that, overall, participants became less compliant over time (62.5%, 61.8%, 59.9%, 59.8%, and 58.5%, for the five consecutive blocks). Moreover, post-hoc comparisons to interpret the Returns × Time interaction showed that compliance decreased over time in blank returns ($F[4, 267.79] = 4.48, p = .005$), higher returns ($F[4, 277.84] = 4.20, p = .007$), and lower returns ($F[4, 276.97] = 2.25, p = .059$), but not in correct returns ($F[4, 247.10] = 0.51, p = .64$) (see Fig. 1).[9] The decrease in compliance over time is consistent with aforementioned research (see Jacobsen et al., 2018). A compelling reason for the constantly high prevalence of compliance in correct returns is that in these returns inaction (e.g., due to tiredness or boredom) automatically results in compliance.

2.4. The effect of type of return on compliance

To test the first hypothesis, we examined the effect of Returns on compliance in the non-confirmation condition only. Results supported the hypothesis and showed more compliance in correct returns (69.9%) than in blank returns (56.2%), higher returns (50.2%), and lower returns (41.3%), respectively (see Table 2, upper row).[10,11] The result that compliance was more prevalent in accurately prefilled returns than in returns that were either incorrectly prefilled or not prefilled is consistent with both a morality perspective on compliance and a default effect. Our finding, however, that lower returns yielded less compliance than higher returns aligns with an ethical dissonance argument, but not with a default effect. With this observation we do not mean to imply

[9] Robustness tests, in which we reran the analyses without data from participants who displayed either full compliance or full evasion, yielded the same pattern of results.

[10] In the current research, the magnitude of the inaccuracy in lower and higher returns was set at 10%. From a morality perspective, not adjusting advantageous inaccuracies should yield higher moral costs, the larger the inaccuracies. Whereas the moral costs of non-adjustments of disadvantageous inaccuracies would not be affected by the size of the inaccuracies. This would imply that for lower returns, but not for higher returns, non-compliance decreases with increasing size of the inaccuracies. Future research could examine this by varying the size of inaccuracies in incorrectly prefilled returns. For example, by presenting participants with returns that are prefilled with liabilities that are 10% vs 50% vs 90% too low or too high.

[11] Additionally, we conducted a one-sample t-test in which the mean of advantageous inaccuracies (i.e., too low reported liabilities) in blank returns was tested against -10% (i.e., the percentage that was implemented in the incorrectly prefilled returns). In this analysis, only participants from the non-confirmation condition and who reported at least one incorrect liability were included (n = 32). Results showed that the inaccuracy for blank returns was greater (M = -26.2%, SD = 40.3%) than -10%, $t(31) = 2.27, p = .03$. A robustness test, in which the analysis was rerun excluding participants who displayed full evasion (n = 4), indicated that the mean inaccuracy in blank returns was not different from -10%, $t(27) = 0.97, p = .34$. This finding indicated that the difference in compliance between lower and blank returns was not due to a difference between these returns in the size of inaccuracies. Only 9 participants in the non-confirmation condition reported at least one disadvantageous liability (i.e., a too high reported liability) in the blank returns. Therefore, no analysis was conducted for disadvantageous inaccuracies.

Table 2
Mean percentages of compliance, under-compliance, and over-compliance per type of return and confirmation condition (standard deviations between parentheses).

	Returns Correct	Blank	Higher	Lower
Condition				
Non-confirmation				
Compliance	69.9[a]	56.2[b]	50.2[c]	41.3[d]
	(39.8)	(44.4)	(45.9)	(45.8)
Under-compliance	25.9[a]	34.6[b]	27.1[a]	53.0[c]
	(36.8)	(40.0)	(35.7)	(44.5)
Over-compliance	4.2[a]	9.2[b]	22.7[c]	5.6[a,b]
	(16.1)	(20.1)	(32.3)	(16.9)
Confirmation				
Compliance	77.8[a]	69.6[b]	71.5[b]	49.5[c]
	(37.8)	(41.4)	(40.6)	(44.3)
Under-compliance	22.0[a]	28.9[a]	22.6[a]	49.5[b]
	(37.7)	(40.0)	(35.7)	(44.5)
Over-compliance	0.3[a]	1.4[a]	6.0[a]	0.9[a]
	(1.0)	(20.1)	(32.3)	(16.9)

Note. Compliance refers to correctly reported liabilities, under-compliance refers to liabilities that are incorrectly reported and lower than correct liabilities, and over-compliance refers to liabilities that are incorrectly reported and higher than correct liabilities. Means per row with different superscripts differed significantly ($p < .05$, with Holm-Bonferroni correction).

that the default effect did not play any role in our study. After all, we did observe that at least some participants retained incorrectly overestimated liabilities (in the case of higher returns), while it would have been in their interest to correct these errors.

Results of a follow-up analysis also indicated that the reported liabilities are, at least in part, driven by a self-serving motive. According to an ethical dissonance argument, incorrectly prefilled returns should be less often adjusted if they contain liabilities that are too low—and hence can provide financial benefits at little moral costs (i.e., by inaction)—than if they are prefilled with liabilities that are too high. Corroborating this notion, we found that lower returns were left unchanged more often than higher returns (29.6% vs 15.6%; $t[50] = 3.67$, $p = .001$; see Table 3). The finding that 15.6% of higher returns—returns with disadvantageous inaccuracies—were not adjusted, however, indicates that compliance was also affected by defaults settings.

Results of an additional follow-up analysis also indicated that that compliance can be driven by both a self-serving motive and a preference for sticking to defaults. We found that of the 51 participants in the non-confirmation condition, 24 left the same number of underestimated and overestimated liabilities unchanged. Whereas 21 participants left underestimated liabilities more often unchanged than overestimated liabilities, and only 6 participants left overestimated liabilities more often unchanged than underestimated liabilities. The finding that nearly half of the participants did not differentiate between advantageous and disadvantageous defaults fits with a preference to retain the default. Whereas a morality perspective is supported by the finding that there were statistically significant more participants who reported advantageous defaults more often than disadvantageous defaults ($n = 21$), as compared to participants who reported disadvantageous defaults more often than advantageous defaults ($n = 6$), $\chi(1) = 7.26$, $p = .01$.

Thus, the evaluation of the first hypothesis and follow-up analyses suggest that prepopulated returns are a potential risk for incorrect reporting. It does not necessarily imply a self-serving motive, however, as prefilled liabilities also serve as default options that take effect when insufficient attention is paid.

Table 3
Percentages of different types of adjustments made in higher and lower returns per condition (standard deviations between parentheses).

Condition	Non-confirmation		Confirmation	
Returns	Higher	Lower	Higher	Lower
No adjustments	15.6[a, x]	29.6[a, y]	4.8[b, x]	28.9[a, y]
	(27.9)	(35.8)	(13.8)	(38.2)
Adjustments: Correct	50.2[a, x]	41.3[a, y]	71.5[b, x]	49.5[a, y]
	(27.9)	(35.8)	(13.8)	(38.2
Adjustments: Too little downward	5.8[a]	—	0.8[b]	—
	(13.2)		(3.2)	
Adjustments: Too much downward	27.1[a, x]	19.9[a, y]	22.6[a, x]	19.3[a, x]
	(35.7)	(32.7)	(38.6)	(36.2)
Adjustments: Too little upward	—	3.4[a]	—	1.3[a]
		(7.1)		(4.6)
Adjustments: Too much upward	1.2[a, x]	5.6[a, y]	0.3[a, x]	0.9[a, y]
	(3.8)	(16.9)	(0.1)	(1.2)
Adjustments: Total	84.4[a, x]	70.4[a, y]	95.2[b, x]	71.1[a, y]
	(27.9)	(35.8)	(13.8)	(38.2)

Note. Correct adjustments refer to changes made in either higher or lower returns that resulted in the report of a correct liability. Too little downward adjustments refer to reported liabilities in higher returns that were lower than the prefilled liabilities, but still higher than the correct liabilities. Too much downward adjustments refer to reported liabilities in higher returns that were lower than the correct liabilities, whereas they refer to reported liabilities in lower returns that were lower than the incorrectly prefilled liabilities. Too little upward adjustments refer to reported liabilities in lower returns that were higher than the incorrectly prefilled liabilities, but still lower than the correct liabilities. Too much upward adjustments refer to reported liabilities in either higher or lower returns that were higher than the correct liabilities. Means per row with different first superscript (a or b) differed significantly between conditions, means per row with different second superscript (x or y) differed significantly within condition ($p < .05$).

2.5. The effect of accuracy confirmation on compliance

To test the second hypothesis—accuracy confirmation increases compliance—we examined the effect of Confirmation on compliance. Results showed only the hypothesized main effect of Confirmation when data of participants who showed full compliance or full evasion were not included in the analysis (see Table 1). The Returns × Confirmation interaction, however, suggests that the effect of Confirmation was moderated by type of return (see Table 1). To interpret this interaction, we conducted post-hoc comparisons between both conditions, separately for each type of return. Results yielded only a significant difference for higher returns: more compliance was found in the confirmation condition than in the non-confirmation condition (71.5% vs 50.2%; $t[95.85] = 2.44$, $p = .017$; see Table 2).[12]

In the introduction, we argued that accuracy confirmation can increase compliance for different reasons. First, confirming the accuracy of reported liabilities makes non-compliance a more active, hence more immoral, decision that is taken less likely. Consistent with the principle of self-engagement (Ayal et al., 2015), we hypothesized that needing to confirm the accuracy of each reported liability by ticking a box labelled 'The liability that I report is the correct liability' would increase compliance. The absence of a reliable overall effect of accuracy confirmation, however, did not fully support the hypothesis. It could be that our intervention did not establish a strong enough relationship between ticking an 'honesty' box and a more general perception of morality. In other words, our intervention might not have been enough morally self-engaging to yield more honest reporting.

Second, we argued that accuracy confirmation increases attention to prefilled liabilities and thereby heightens alertness for inaccuracies,

[12] The robustness test showed, in addition to a significant difference for higher returns (79.4% vs 50.3%; $t[66.88] = 3.19$, $p = .002$), also a significant difference for blank returns (77.0% vs 58.6%; $t[69.57] = 2.04$, $p = .045$).

which, in turn, makes corrections of inaccurate prepopulated returns more likely, and hence increases compliance. This argument was supported by the obtained effect of accuracy confirmation for higher returns. A finding, however, that also indicated that the intervention only resulted in more adjustments to prepopulated returns if not changing the prefilled liabilities would be financially costly. Results of several post-hoc comparisons collaborated this notion. First, we found that, overall, over-compliance was less prevalent in the confirmation condition than in the non-confirmation condition (10.4% vs 2.1%; t[57.26] = 3.15, p = .003), whereas this was not the case for under-compliance (30.7% vs 35.2%; t[96] = 0.60, p = .55).[13] Moreover, the difference between conditions in over-compliance was larger for higher returns (22.7% vs 6.0%; t[71.52] = 3.33, p = .001) than for correct returns (4.2% vs 0.3%, t[50.44] = 1.72, p = .09), blank returns (9.2% vs 1.4%, t[58.39] = 2.64, p = .011), and lower returns (5.6% vs 0.9%, t[51.91] = 1.97, p = .054; see Table 2).[14] For none of the types of return, under-compliance differed between conditions (ts < 1, ps > .53).[15] Together, results of these follow-up analyses comparisons provide further support for the notion that our intervention increased compliance mainly through decreasing over-compliance in higher returns—that is, in incorrectly prefilled returns that would yield financial costs if not adjusted (see Table 2).

The notion that our intervention seems more effective in heightening alertness to inaccuracies than in increasing honest reporting was also corroborated by results of several post-hoc comparisons of (non-) adjustments in higher and lower returns, specifically. First, we found that for higher returns, the prevalence of making no adjustments was lower in the confirmation condition than in the non-confirmation condition (4.8% vs 15.6%; t[74.25] = 2.44, p < .001). Whereas for lower returns, the percentage of no adjustments was the same in both conditions (28.9% vs 29.6%; t[96] = 0.10, p = .92; see Table 3).[16] This indicates that our intervention reduced default effects for returns that overestimated liabilities, but not for returns that underestimated liabilities. This notion was supported by our finding that higher returns were more often correctly adjusted in the confirmation condition than in the non-confirmation condition (71.5% vs 50.2%, t[95.85] = 2.44, p = .017), whereas for lower returns no such difference was found (49.5% vs 41.3%, t[96] = 0.90, p = .37).[17] Thus, the evaluation of the second hypothesis and follow-up analyses suggest that our intervention did not reduce self-serving dishonest reporting, but it did increase compliance through counteracting defaults effects in incorrectly prefilled returns that overestimated liabilities.

3. Conclusions

Our study showed that choice architecture in tax returns can induce both correct and incorrect reporting: Whereas correctly prepopulated returns promoted compliance, incorrectly prepopulated returns undermined it. Moreover, both moral costs and default effects played a pivotal role in shaping the effects of prepopulated tax returns on compliance. Our finding that underreporting was more likely when it needed less effort, supports a morality interpretation of compliance. Whereas the observation that incorrectly prepopulated returns were often left unchanged—even if this had negative financial consequences—indicates that default effects also impact compliance behaviour. Results further suggest that needing to confirm the accuracy of reported liabilities can make taxpayers more attentive to inaccuracies in prepopulated returns, but also that it only nudges them into action when correcting inaccuracies yields financial benefits. Such an intervention may thus be more effective in reducing mindless overpayment of taxes than (more) mindful tax evasion. As the present study indicates that the effects of prepopulated tax returns on compliance are contingent upon the accuracy of prefilling information, it implies that to reap the intended positive effects of this form of choice architecture, tax administrations should handle prefilling in tax returns with great care.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.socec.2020.101574.

References

Allingham, M.G., Sandmo, A., 1972. Income tax evasion: a theoretical analysis. Journal of Public Economics 1, 323–338.

Ayal, S., Gino, F., Barkan, R., Ariely, D., 2015. Three principles of REVISE people's unethical behavior. Perspectives on Psychological Science 10, 738–741.

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

Doxey, M., Lawson, J., Shane Stinson, S, 2019. The effects of prefilled tax returns on taxpayer compliance. Unpublished manuscript.

Duncan, D., Li, D., 2018. Liar Liar: Experimental evidence of the effect of confirmation-reports on dishonesty. Southern Economic Journal 84, 742–770.

Fochmann, M., Müller, N., Overesch, M., 2018. Less cheating? The effects of prefilled forms on compliance behavior. Arbeitskreis Quantative Steuerlehre (arqus), Berlin Arqus Discussion Paper, No. 227.

Fonseca, M.A., Grimshaw, S.B., 2017. Do behavioral nudges in prepopulated tax forms affect compliance? Experimental evidence with real taxpayers. Journal of Public Policy & Marketing 36, 213–226.

Jacobsen, C., Fosgaard, T.R., Pascual-Ezema, D., 2018. Why do we lie? A practical guide to the dishonesty literature. Journal of Economic Surveys 32, 357–387.

Kotakorpi, K., Laamanen, J-P, 2016. Prefilled income tax returns and tax compliance: Evidence from a natural experiment. University of Tampere, Finland Tampere Economics Working Papers, No 104.

Mazar, N., Amir, O., Ariely, D., 2008. The dishonesty of honest people; A theory of self-concept maintenance. Journal of Marketing Research 45, 633–644.

Mazar, N., Hawkins, S.A., 2015. Choice architecture in conflicts of interest: Defaults as physical and psychological barriers to (dis)honesty. Journal of Experimental Social Psychology 59, 113–117.

Morrison, W.G., Ruffle, B.J., 2020. Insurable losses, pre-filled claims forms and honesty in reporting. Unpublished manuscript.

OECD, 2019. Tax administration 2019: comparative information on OECD and other advanced and emerging economies. OECD Publishing, Paris.

Ritov, I., Baron, J., 1992. Status-quo and omission bias. Journal of Risk and Uncertainty 5, 49–61.

Simmons, J.P., Nelson, L.D., Simonsohn, U, 2013. Life after p-hacking. In: Paper presented at the fourteenth annual meeting of the society for personality and social psychology. New Orleans, LA.

Spranca, M., Minsk, E., Baron, J., 1991. Omission and commission in judgment and choice. Journal of Experimental Social Psychology 27, 76–105.

Thaler, R.H., Sunstein, C., 2008. Nudge: improving decisions about health, wealth, and happiness. Yale University Press, New Haven, CT.

[13] Robustness tests yielded similar results for both over-compliance (12.5% vs 2.8%; t[41.71] = 3.04, p = .004) and under-compliance (31.4% vs 23.5%; t[71] = 1.19, p = .24).

[14] The robustness test yielded similar results (1.40 < t < 3.37, .002 < p < .18).

[15] The robustness test yielded similar results (ts < 1, ps > .53).

[16] Robustness tests yielded similar results for both higher returns (20.6% vs 6.3%; t[54.53] = 2.59, p = .012) and lower returns (29.6% vs 28.9%; t[71] = 0.15, p = .88).

[17] Robustness tests yielded similar results for both higher returns (79.4% vs 50.3%; t[66.88] = 3.19, p = .002) and lower returns (50.8% vs 38.0%; t[71] = 1.26, p = .21).

Kenniscentrum
Psychologie en Economisch Gedrag

Authors: Wilco van Dijk, Sjoerd Goslinga, Bart Terwel, & Eric van Dijk

No. 2 / 2020

www.kcpeg.nl