Cardiff Economics Working Papers

Working Paper No. E2021/17

Does the Frequency of Reminders Matter for their Effectiveness? A Randomized Controlled Trial

Armenak Antinyan, Zareh Asatryan, Zhixin Dai, Kezhi Wang

July 2021

ISSN 1749-6010

Cardiff Business School
Cardiff University
Colum Drive
Cardiff CF10 3EU
United Kingdom
t: +44 (0)29 2087 4000
f: +44 (0)29 2087 4419
business.cardiff.ac.uk

Does the Frequency of Reminders Matter for their Effectiveness? A Randomized Controlled Trial

Armenak Antinyan,[1,2] Zareh Asatryan,[3] Zhixin Dai,[4] Kezhi Wang [5]

Abstract

We assess the impact of reminder frequency on the probability of paying overdue property taxes in a randomized controlled trial in China. One reminder a week (sent as a text message) considerably increases the probability of tax compliance and results in tangible fiscal gains compared to a one-off reminder. However, increasing the frequency of reminders to two text messages a week diminishes their effectiveness. The takeaway of our study is that frequent reminders are an important trigger for human behavior, nonetheless, beyond a certain frequency the effectiveness of additional reminders seems to decline.

JEL: C93; H24; H26.
Keywords: Reminder Frequency; Randomized Controlled Trial; Tax Compliance.

[1] Wenlan School of Business, Zhongnan University of Economics and Law. antinyan.armenak@gmail.com.
[2] Cardiff Business School, Cardiff University.
[3] ZEW Mannheim. asatryan@zew.de
[4] *Corresponding Author.* China Financial Policy Research Center, School of Finance, Renmin University of China. dai@ruc.edu.cn.
[5] School of Public Economics and Administration, Shanghai University of Finance and Economics. kezhiwang@126.com.

1. Introduction

Reminders for overdue bills, approaching obligations, and appointments represent one of the most popular and effective interventions to steer individuals in a certain direction (Sunstein, 2014). The positive impact of reminders has been documented in such economically relevant settings as health decisions (Altmann and Traxler, 2014; Antinyan et al., 2021), financial behavior (Cadena and Schoar, 2011; Karlan et al., 2016), rule and tax compliance (Apesteguia et al., 2013; Gillitzer and Sinning, 2020), among others.

While the impact of reminder frame, content, and timing on individual decisions has been studied in the literature, little is known about the extent to which the reminders of different frequencies matter. To the best of our knowledge, it is largely unknown whether high frequency reminders will be more effective in nudging individuals to undertake a certain action (e.g., a message is sent every week reminding an individual to pay an overdue fine) compared to low frequency reminders (e.g., only one message is sent reminding an individual to pay an overdue fine).

On the one hand it is reasonable to expect that the positive impact of high frequency reminders may surpass that of low frequency reminders. Indeed, unlike low frequency reminders (or in the most extreme case one-off reminders), high frequency reminders may continually drive the individual's attention to the pending task and may not allow her to forget about it. Alternatively, high frequency reminders may create a payment pressure inducing those individuals who do not want to accomplish the pending task and repeatedly postpone it to act.

On the other hand, high frequency reminders may not have the expected large and positive impact compared to low frequency reminders. Since the treatment stimulus is implemented frequently over time, individuals can exhibit habituation, whereby the repeated presentation of a stimulus might decrease the reaction to the stimulus (Thompson and Spencer, 1966; Groves and Thompson, 1970; Rankin et al., 2009; Ito et al., 2018). In other words, individuals can become accustomed both to the fact of being regularly communicated and to the content of the communication, which may mitigate the effect of high frequency reminders on the probability of engaging in a certain conduct.

The present study reports the results of a field experiment that we carried out in September-November 2019 in the People's Republic of China in collaboration with the Baoshan Tax Administration in order to assess the impact of reminder frequency on the payment of overdue taxes. More specifically, we focus on 1,742 late property taxpayers in Baoshan district in Shanghai for whom the compliance due date was December 31, 2018 and who did not fulfill their obligations as of September 17, 2019 (i.e., there was roughly a nine-month delay). Our sample constitutes about 50% of the late property taxpayers.[6] In Shanghai, individuals pay property taxes for the second housing unit purchased if they are Shanghai residents, and for the first housing unit purchased if they are Shanghai non-residents.

The overdue tax obligations of the taxpayers we study cannot be explained by their unawareness of the property tax since taxpayers are usually notified about their tax obligations when purchasing a house. Thus, most likely, the majority of the late taxpayers have either forgotten or are unwilling to pay their overdue property tax obligations. Indeed, liquidity constraints may also play a role (e.g., Moulton et al., 2019; Brockmeyer et al., 2021), especially for taxpayers with high tax arrears. Nonetheless, given the large discrepancy between the housing prices in Shanghai and the average salary (Rapoza, 2017) individuals in our sample should be wealthy enough to afford either one (for Shanghai non-residents) or two (for Shanghai residents) houses in Shanghai. Thus, we believe that only a fraction of our sample should be affected by liquidity constraints.[7]

Keeping the content of the communication fixed, our experimental design manipulates the frequency of the reminders across four different treatments. In the *Control* treatment, no communication between tax authorities and taxpayers takes place. In the *Low frequency (LF)* treatment, only one digital message is sent on behalf of the tax administration in the beginning of the trial. The design of this treatment is aligned with that of many field experiments in taxation, whereby the tax administration contacts the taxpayers only once (see Mascagni, 2018; Slemrod, 2019; Antinyan and Asatryan, 2020 for reviews). The main contribution of our paper stems from

[6] The vast majority of the remaining late taxpayers paid the property tax in January 1, 2019-September 16, 2019 time interval. We dropped a small fraction of late taxpayers, who did not make a payment, because of imprecise or missing contact and ID information. Please refer to sub-section 2.3 for more details.

[7] According to the statistical yearbook of Baoshan district, the average cost of property in 2018 and 2019 was around 29,000 RMB (4,060 USD) per square meter. At the time of reminder dispatch, the exchange rate was 1 RMB=0.14 USD. Also note that the average overdue property tax equals 4412.506 RMB (617.751 USD) in our sample, which constitutes a tiny fraction of the house price.

Medium frequency (MF) and *High frequency (HF)* treatments. In the former, the same digital message is sent once every week for four weeks, whereas in the latter, it is sent twice every week for four weeks. A distinctive feature of the Chinese context is the fact that the communication between taxpayers and tax authorities is fully digital. Thus, increasing the frequency of reminders comes at little (financial and administrative) cost.[8]

The reminders were disbursed from September 17 to October 11 (inclusive), 2019 and the impact of the intervention was evaluated as of November 18, 2019. Thus, the trial lasted for about nine weeks, which is an acceptable trial length in the literature (e.g., Kettle et al., 2016; Hallsworth et al., 2017; Hernandez et al., 2017; Brockmeyer et al., 2018; Gillitzer and Sinning, 2020). We terminated the trial on November 18, since Baoshan tax administration requested relatively quick results to understand the most effective interventions. In case of positive results, the tax authorities intended to expose the taxpayers who remained non-compliant by the end of our trial to the most effective interventions.

The binary outcome variable under our scrutiny captures whether each taxpayer paid the overdue tax between September 17 and November 18, 2019, or not. The reminders dispatched during the experiment contained the following information: 1) the amount of the property tax due and the overdue fines; 2) a notice about restricting the house from trading until the tax obligation is fulfilled; 3) friendly tips for paying the tax and a phone number for questions.

Our main results are as follows. First, the increased frequency of reminders can have a remarkable influence on individual decisions. More specifically, the probability of overdue tax payments in *MF* and *HF* is around 12-14 percentage points higher than in *Control* (more than 300% increase in compliance) and 5-7 percentage points higher than in *LF* (around 40-60% percent increase in compliance). The impact of recurrent reminders is immediate, as we evidence (relatively) big drops in the share of non-compliant taxpayers on reminder days.

Second, increasing the number of reminders beyond a certain frequency may reduce the impact of the additional reminders on human behavior. Indeed, while the probability of tax compliance in *MF* increases by around 5 percentage points (or roughly 40%) compared to *LF*, the probability of tax compliance in *HF* increases by around 2 percentage points (or roughly 20%) compared to *MF*.

[8] The price of one SMS was 0.4 RMB (0.056 USD).

Third, the past behavior of individuals can characterize their sensitivity to the reminders irrespective of the reminder frequency. More precisely, those individuals who were late paying their property taxes in the past years were not sensitive to reminders of various frequency during the experiment. In our view, this may serve as an interesting finding for policymakers, since it illustrates that nudges may have limits for inducing taxpayers to comply.

This paper makes two contributions. First, the paper contributes to the literature that uncovers the impact of recurrent reminders on individual decisions (Altmann and Traxler, 2014; Calzolari and Nardotto, 2016; Karlan et al., 2016). While the mentioned field experiments juxtapose decisions in recurrent reminder treatments with that in a control treatment with no communication, our experimental design allows us to compare treatments with different frequency of reminders.

Second, the paper also adds to the literature that studies the impact of notifications sent by tax authorities on tax compliance (Mascagni, 2018; Antinyan and Asatryan, 2020; Slemrod, 2019). To this date various aspects of the notifications have been manipulated, such as the delivery method (e.g., Doerrenberg and Schmitz, 2017; Mogollón et al., 2021; Ortega and Scartascini, 2020), the content and the framing of the notification (e.g., Hallsworth et al., 2017; Hernandez et al., 2017; Bott et al., 2019; De Neve et al., 2019), and the timing of the communication (e.g., Gillitzer and Sinning, 2020). To the best of our knowledge, the impact of notification frequency on tax compliance is understudied so far.

2. The experiment

2.1. Background information

In order to curb the soaring housing prices and cool down the overheated housing market the Chinese authorities implemented a series of interventions in the last decade. First, *quota restrictions* – that set the maximum number of houses a household can purchase – and *credit restrictions* – that specify the minimum proportion of the total housing price to be paid in cash – were introduced starting from 2010 (Alm et al., 2021). In general, quota restrictions are implemented in a similar fashion in different cities in China and allow households that already hold one house buy a second one if the household is a resident in a given city (e.g., Shanghai). Households that own two or more houses cannot purchase any more (Du and Zhang, 2015; Alm et al., 2021). Non-resident households living in a given city are allowed to buy only one house

conditional on providing a proof of local tax receipts or social security records (Du and Zhang, 2015).

The Chinese government piloted property tax regulation in January 2011 in Shanghai and Chongqing with an expectation that the quota restriction regulation will be replaced with the property tax regulation in the long run (Du and Zhang, 2015). Unlike the quota restriction regulation, the property tax regulation is implemented differently in each city. In Shanghai, the property taxes for resident and non-resident families are calculated as follows (Cho and Choi, 2014; Du and Zhang, 2015):

(i) Shanghai residents pay property taxes for the second housing unit purchased after January 28, 2011 (the tax applies both to second-hand and newly built housing units);

(ii) Shanghai non-residents pay property taxes for the first housing unit purchased after January 28, 2011 (the tax applies both to second-hand and newly built housing units).

The due date for fulfilling the yearly property tax obligation is December 31 of each year. For instance, the property taxes for 2015 tax year should be paid by December 31, 2015. Partial payments are not possible and the tax should be paid in full. In case of a failure to do so, the taxpaying household is charged an additional 0.05 percent of the tax liability daily (or 18.25% of the tax liability yearly) starting from January 1 of the following year. Hence, those who fail to pay their 2015 property tax by December 31, 2015 are fined starting from January 1, 2016. Furthermore, in case of unpaid tax obligations, the property is restricted from trading.

Regarding the tax structure, for each household that pays taxes, 60 square meters per family member are tax-exempt. The tax rate is set to 0.6 percent and the tax base equals to 70% of the house price. The communication between tax authorities in Shanghai and property taxpayers unfolds as follows. First, the households that are eligible for paying taxes are notified about their property tax obligations when purchasing the house. Hence, the taxpayers who own a house are aware of the fact that they have annual property tax obligations. Second, every year, a notification is usually sent before the due date either at the end of November or at the beginning of December to all taxpayers requesting to pay the property taxes of the given year. Third, if the taxpayer does

not fulfill the yearly property tax obligation, she receives a reminder that she has an overdue property tax in the beginning of the next year. In other words, during our intervention, no other notifications requesting to pay the property tax (except for the ones we designed) were sent to the taxpayers.

The randomized controlled trial was run in Baoshan District – a suburban district of Shanghai – which covers an area of 424.58 km^2 with a population of about 2 million at the beginning of 2018 according to the data from Shanghai Statistics Bureau.[9] In 2018, the district had 21,093 taxpaying households for whom the due date was December 31, 2018. According to the administrative records, out of these taxpayers, 1,742 taxpayers did not meet their 2018 tax obligations as of September 17, 2019.[10] The average tax obligation was around 4,412 RMB (618 USD). All these households were included in the study sample.

2.2.Treatments

We designed the field experiment to test the effect of the frequency of reminders on individual decisions. In our specific setting, the decision is a binary outcome variable indicating whether each taxpayer paid the overdue tax between September 17 and November 18, 2019, or not. Keeping the content of the reminders fixed, our experimental design manipulates the frequency of the reminders sent by tax authorities across four different treatments:

(i) *Control*: no communication between tax authorities and taxpayers takes place.

(ii) *Low frequency (LF)*: only one digital message is sent on behalf of the tax administration in the beginning of the trial.

(iii) *Medium frequency (MF)*: the same digital message is sent on behalf of the tax administration once every week for four weeks. Overall, in this treatment, we sent 4 reminders during the experiment.

(iv) *High frequency (HF)*: the same digital message is sent on behalf of the tax administration twice every week for four weeks. Overall, in this treatment, we sent 8 reminders during the experiment.

[9] https://www.shbsq.gov.cn/shbs/en/aboutbaoshan/20190710/177055.html (retrieved on 07.07.2020)

[10] As noted in footnote 7 and discussed in sub-section 2.3, we exclude a small number of taxpayers with imprecise contact and ID information. In other words, the excluded taxpayers are not counted in the reported number.

We reached out to the late taxpayers using SMS and the text message read as follows:

"*This is a friendly reminder from the Baoshan Taxation Bureau: there is an unpaid housing property tax of (**) RMBs at (** Road) of Baoshan District. If you are the actual taxpayer, please pay as soon as possible. An overdue fine will be generated at 0.05% per day or 18.25% per year on your unpaid tax obligations. Furthermore, the house in arrears will be restricted from trading. It is recommended to pay the tax through "FuFeiTong payment and check" app (WeChat ID: **). Please call xxx-xxxxx or xxx-xxxxx for consulting other tax payment channels or residence permit exemption policies. If you have already paid the tax, please ignore this reminder. Thank you for your cooperation*."[11]

2.3.Procedures

In the beginning of September 2019, 1,779 taxpayers from Baoshan district were randomized into 3 treatment (N=445*3=1,335) and one control (444) arms.[12] To enhance balancing, we stratified randomization by gender and age.[13] The individual randomization exposes us to the threat of downward biases in our treatment effects due to potential spillovers across subjects in different treatment groups. Nevertheless, we believe that the likelihood of spillovers is extremely low in our scenario for fewfold reasons.

First, Baoshan district has around 2 million residents, out of which only a sub-sample of 21,093 individuals were eligible for paying the property tax in 2018 (i.e., around 1 %). Please refer to sub-section 2.1 for the eligibility criteria. Given the area that Baoshan district covers (424.58 km^2) and the overall number of property taxpayers (21,093), we have on average 50 taxpayers per 1 km^2, which is equivalent to 140 soccer fields. Given such geographical dispersion of the taxpayers, we believe that it is highly unlikely that they actively interact with each other resulting in strong spillovers across the trial arms. Please also note that we focus our attention on a final sample of

[11] For instance, an individual with a 1,000 RMB (140 USD) overdue tax obligation for year *t* will be required to pay a daily fine that equals 1,000 RMB *0.05%=0.5 RMB (0.07 USD) starting from day 1 in year *t+1* until the tax obligation is fulfilled. If the taxes due at the end of year *t* are not paid in year *t+1* the taxpayer has to pay a 18.25% yearly fine on the unpaid tax obligation (1,000RMB *0.05% *365=1,000 RMB *18.25%).

[12] At the time of the randomization overall 1,997 individuals did not pay the property tax. We dropped taxpayers with imprecise or missing contact and ID information.

[13] The user-written command randtreat (version 1.4) was used to perform the randomization (Carril, 2017).

1,742 taxpayers, which comprises around 0.09% of the residents in the district (later in section we explain why we drop 37 observations).

Second, in case of large spillovers, we should observe a substantial number of overdue tax payments in the *Control* during the trial (in other words those who receive treatment stimulus should induce those who do not receive treatment stimulus to fulfill the overdue tax obligations). The results in section 4 illustrate that this is not the case. Indeed, the treatment effects analysis suggests that the probability of compliance is much lower in the control group than in all of the treatment groups. Furthermore, Kaplan-Meier survival curves depicting daily compliance data show virtually no payments on the message days in the control group, despite payments in the treatment groups. In case of spillovers, one would expect the payment pattern to be similar across the control and treatment groups. Since there is virtually no spillover from the treatment groups to the *Control*, there is also no reason to believe that there can be considerable spillovers across *HF, MF, LF* treatments given that the allocation to treatment arms is random (thus, the geolocations of the subjects in the *Control* should be similar to that of the subjects in *HF*, *MF*, and *LF*).

Once the randomization was carried out, we verified that the treatment arms are well-balanced in terms of the observable characteristics under our disposal: the tax debt, the age, the gender of the dwelling owner and the number of months she paid a property tax for the properties she owns by the start of the reminder dispatch.[14]

We dispatched the reminders from September 17 to October 11 (inclusive), 2019 and evaluated the impact of the intervention as of November 18, 2019. The first reminder in all the treatments was sent on the same date (i.e., September 17, 2019). Starting from Week 2, the first reminder sent in *HF* always coincided with the weekly reminder sent in *MF*. In other words, if in Week 2 the first reminder was sent on Monday in *HF*, the weekly reminder was sent on Monday in *MF* as well. Table 1 illustrates the schedule of the reminders sent during the experiment. Following a recent theoretical literature about anticipated and unanticipated reminders (Ericson, 2017), we sent the reminders on different days each week, since surprise reminders can be more effective than anticipated ones. In our view, sending the frequent reminders following exactly the same schedule each week could undermine their effectiveness and somewhat reduce the "surprise effect" of these

[14] The results of this initial balance checks are available upon request.

reminders. In other words, following this design choice, we strive to make the frequent reminder less anticipated.

Table 1: The schedule of the reminders

	LF	MF	HF
Week 1	17.09.2019	17.09.2019	17.09.2019
			20.09.2019
Week 2		23.09.2019	23.09.2019
			26.09.2019
Week 3		29.09.2019	29.09.2019
			03.10.2019
Week 4		08.10.2019	08.10.2019
			11.10.2019

Note: The schedule of the reminders sent in the experiment.

We removed thirty-seven taxpayers included in the sample who paid their tax obligations before September 17 (i.e., the first day we dispatched the reminders). Furthermore, for four other taxpayers the age was indicated to be younger than 18 years old, which was obviously a mistake. We replaced these four values with the average age in the sample. In sum, we ended up with 1,742 taxpayers, which is the final sample used in the analysis. Table 2 illustrates that the final sample is well-balanced in terms of the observable characteristics mentioned earlier in this section.

Table 2: Balancing tests

	(1) Age	(2) Gender	(3) Tax Debt	(4) Number of Months Taxes Paid
LF	-0.018	0.004	13.051	-0.173
	(0.753)	(0.033)	(306.000)	(0.834)
MF	0.608	0.008	23.880	-0.614
	(0.783)	(0.033)	(316.194)	(0.846)
HF	-0.779	0.006	-8.078	0.542
	(0.732)	(0.033)	(290.786)	(0.820)
Constant	43.207***	0.616***	4405.304***	36.361***
	(0.546)	(0.023)	(212.628)	(0.575)
F stat.	1.20	0.02	0.01	0.63
P>F	0.308	0.996	0.999	0.595
R^2	0.002	0.000	0.000	0.001
N	1742	1742	1742	1742

Note: Note: Results from an ordinary least square (OLS) model (robust standard errors in parentheses). Dependent variables: *Age*- An integer, indicating the age of the respondent; *Gender*-A binary variable which equals one for males and 0 otherwise; *Tax Debt*- An integer indicating the property tax debt of taxpayer *i* (excluding the fines); *Number of Months Taxes Paid*- An integer indicating the number of months a taxpayer paid taxes for the property she owns by the start of the experiment. Independent Variables: *LF- A* dummy variable which equals one in low frequency treatment and 0 otherwise; *MF*- A dummy variable which equals one in medium frequency treatment and 0 otherwise; *HF*- A dummy variable which equals one in high frequency treatment and 0 otherwise; Significance levels: * p<0.1; ** p<0.05; *** p<0.01.

3. Empirical methodology

3.1.Treatments effect analysis

Given our research question, we estimate the following regression model:

$$Y_i = \beta_0 + \sum_{l=1}^{3} \beta_l T_{il} + \beta_2 X_i + \varepsilon_i, \qquad (1)$$

where Y_i is the tax compliance measure of individual *i*. T_{il} is an indicator variable denoting whether individual i belongs to treatment l. X_i is a vector of control variables, which includes individual *i*'s age, gender, the 2018 property tax amount, and the number of months the individual paid taxes for the property she owns by the start of the experiment. In the analysis that follows, we report specifications both with and without the control variables. In all specifications, the treatment in which no reminders are sent serves as the omitted category. Thus, the coefficient β_l depicts the causal impact of treatment T_l relative to the control treatment without communication.

Since the outcome measure is a binary variable, indicating whether individual *i* paid her overdue property tax during the experiment or not, we utilize a linear probability model (LPM) to estimate the equation (1). To examine the sensitivity of our results to model selection, we also utilize a probit model for our estimations. The marginal effects of the probit model are very similar to the LPM estimates.[15] In all regressions, we control for the heteroscedasticity of the residuals, introducing White robust standard errors.

3.2.Survival analysis

To assess the impact of the frequency of reminders on tax payments over time, we conduct survival analysis and estimate Kaplan-Meier survivor functions that report the probability of surviving after time *t* (Cleves et al., 2016). In our framework, the probability of surviving past time *t* represents the probability of not paying the overdue tax. Given the failure times $t_1, t_2, \dots, t_m$ (m is the number of observed failure times), the Kaplan-Meier (1958) estimate of the survivor function at time *t* is expressed by the following formula:

$$\hat{S}(t)=\prod_{j|t_j\le t}\left(\frac{n_j-d_j}{d_j}\right). \qquad (2)$$

In our framework, n_j is the number of individuals who have to pay their taxes at time t_j and d_j is the number of taxpayers who eventually did so. Time *t* is measured in daily intervals.

In line with Kalfbleisch and Prentice (2011), the confidence intervals are calculated using the asymptotic variance $\ln(-\ln(\hat{S}(t)))$, which equals

$$\hat{\sigma}^2(t)=\frac{\sum_{j|t_j\le t}\frac{d_j}{n_j(n_j-d_j)}}{\left(\sum_{j|t_j\le t}\ln\left(\frac{n_j-d_j}{d_j}\right)\right)^2}. \qquad (3)$$

The confidence bounds are determined by raising $\hat{S}(t)$ to the power of $\exp\left(\pm z_{\alpha/2}\hat{\sigma}(t)\right)$, where $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution and $\hat{\sigma}(t)$ is represented by (3). For more details about survival analysis an interested reader can refer to Cleves et al. (2016).

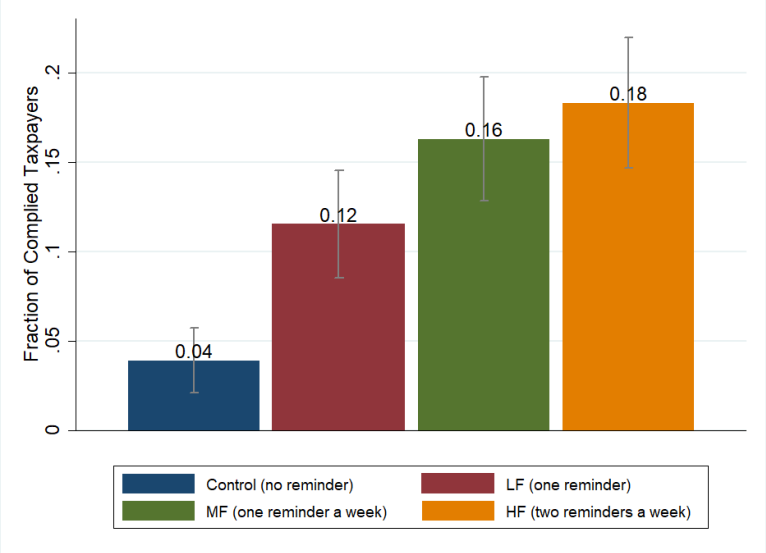
[15] The results are available upon request.

4. Results

4.1.Treatment effects analysis

Figure 1 reports the fraction of complied taxpayers by treatment groups with 95 percent confidence intervals. We see that tax compliance is substantially higher in each treatment than in the control. Furthermore, tax compliance in *MF* and *HF* is higher than in *LF*, while the difference between *MF* and *HF* smaller than the difference between *MF* and *LF*.

Figure 1. Tax compliance by treatment groups



Note: The figure reports the fraction of taxpayers who paid the 2018 overdue property tax from September 17 to November 18, 2019 by treatment with the associated 95% confidence intervals. The reminders were sent from September 17 to October 11 (inclusive), 2019.

The descriptive results depicted in Figure 1 are confirmed by formal statistical analysis. More specifically, Table 3 reports the estimates of the treatment effects analysis obtained from a linear probability model both with and without controls.[16] The table leads to three important conclusions.

First, the reminder significantly increased compliance in all treatments compared to the baseline with no communication. As captured by the positive and highly significant treatment dummies, the probability of compliance in *MF* and *HF* is approximately 12–14 percentage points higher

[16] The analysis reported in this section (and in the paper in general) focuses on the compliance throughout the entire experiment. As an additional robustness check we implement identical treatment effects analyses by focusing on compliance 25 and 40 days after the start of the experiment. The results are qualitatively similar and available upon request.

(more than 300% increase) as compared to the *Control*. In the same vein, as shown by the positive and highly significant coefficient of the *LF* dummy, the probability of compliance in this treatment is around 8 percentage points higher (roughly 200% increase) as compared to the *Control*. The positive impact of reminders is a standard result documented by the literature (Altman and Traxler, 2014; Apesteguia et al., 2013; Calzolari and Nardotto, 2016; Karlan et al., 2016).

Second, the frequency of reminders does matter for the decision to comply with taxes. According to a Wald test for equality of coefficients the twofold difference in the probability of compliance between *MF* and *LF* (F=3.89, p-value=0.049) as well as between *HF* and *LF* (F=9.3, p-value=0.002) is statistically significant.[17]

Third, beyond a certain reminder frequency, the probability of tax compliance does not increase considerably in the number of reminders sent. While the probability of tax compliance in *MF* increases by around 5 percentage points (or roughly 40%) compared to *LF* and this difference is statistically significant (F=3.89, p-value=0.049), the probability of tax compliance in *HF* increases by around 2.8 percentage points (or roughly 18%) compared to *MF* and this difference turns out to be statistically non-significant (F=1.26, p-value=0.261). Indeed, the statistically non-significant difference between *HF* and *MF* can stem from the relatively low number of observations in the trial. Nevertheless, even if this difference were statistically significant, we would evidence a twofold decline when comparing the difference between *MF* and *LF* with that of *MF* and *HF*.[18]

[17] For the Wald tests we use the point estimates of model 2. For the analysis with point estimates in model 1 which leads to similar conclusions please refer to Table 3.

[18] As a robustness check, we run additional analysis to account for multiple hypothesis testing. We test six null hypotheses corresponding to all pairwise comparisons across the three treatments and the control group. We apply the procedure developed by List et al. (2019) to control for the familywise error rate, which is the probability of making any type 1 error. The results are qualitatively similar to those reported in the main text, especially when focusing on the improved p-values described in Remark 3.7 of List et al.'s paper. These p-values are *"obtained by exploiting the logical restrictions among null hypotheses when there are multiple treatments",* which allows to improve power. The multiplicity adjusted p-values are reported in Table A1 in Appendix A.

Table 3: Treatment effects analysis

LPM	(1)	(2)
LF	0.076***	0.075***
	(0.018)	(0.018)
MF	0.124***	0.120***
	(0.020)	(0.019)
HF	0.144***	0.148***
	(0.021)	(0.021)
Age		0.001
		(0.001)
Gender		-0.010
		(0.016)
Tax Debt		-0.000*
		(0.000)
Number of Months Taxes Paid		-0.005***
		(0.001)
Constant	0.039***	0.208***
	(0.009)	(0.043)
Treatment Arm Comparisons (Wald tests)		
MF *vs* LF	F=4.13, p=0.042	F=3.89, p=0.049
HF *vs* LF	F=7.96, p=0.005	F=9.3, p=0.002
HF *vs* MF	F=0.62, p=0.430	F=1.26, p=0.261
Control Group Mean	3.9%	3.9%
F stat.	24.941	15.715
P>F	0.000	0.000
Adj. R^2	0.026	0.069
N	1,742	1,742

Note: Results from a linear probability model (robust standard errors in parentheses). Dependent variable: tax compliance measure of taxpayer *i*, which equals one if she paid her overdue property tax during the experiment and 0 otherwise. Independent Variables: *LF- A* dummy variable which equals one in low frequency treatment and 0 otherwise; *MF*- A dummy variable which equals one in medium frequency treatment and 0 otherwise; *HF*- A dummy variable which equals one in high frequency treatment and 0 otherwise; *Age*- An integer, indicating the age of the respondent; *Gender*-A dummy variable which equals one for males and 0 otherwise; *Tax Debt*- An integer indicating the property tax debt of taxpayer *i* (excluding the fines); *Number of Months Taxes Paid*- An integer indicating the number of months a taxpayer paid taxes for the property she owns by the start of the experiment. Significance levels: * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

4.2. Comparison with the literature

In this sub-section, we assess how the treatment effects of the current intervention compare to those estimated in the available literature. The underlying metadata comes from Antinyan and Asatryan (2020) and covers 133 treatment effect estimates retrieved from 8 papers that study the impact of notifications (or nudges) on property tax payment.[19] The list of papers and the number of estimates is detailed in Table 4 below.

[19] We only borrow those treatment effect estimates that are estimated for the entire sample (i.e., main treatment effects). If the paper reports regression models both with and without controls, the estimates of both models are included in the

Table 4: Papers and estimates

Paper	Number of Estimates	Communication Mode	Number of Contacts During the RCT	Country
Del Carpio (2014)	8	Letter	1	Peru
Castro and Scartascini (2015)	24	Tax Bill	1	Argentina
John and Blume (2018)	16	Tax Bill	1	UK
Chirico et al (2019)	42	Letter	1	USA
Eguino and Schaechtele (2020)	20	Tax Bill	1	Argentina
Pfeifer and Pacheco (2020)	5	Letter	1	Brazil
Brockmeyer et al. (2021)	6	Letter	1	Mexico
Collin et al. (2021)	12	Text Message	1	Tanzania
Total	133			

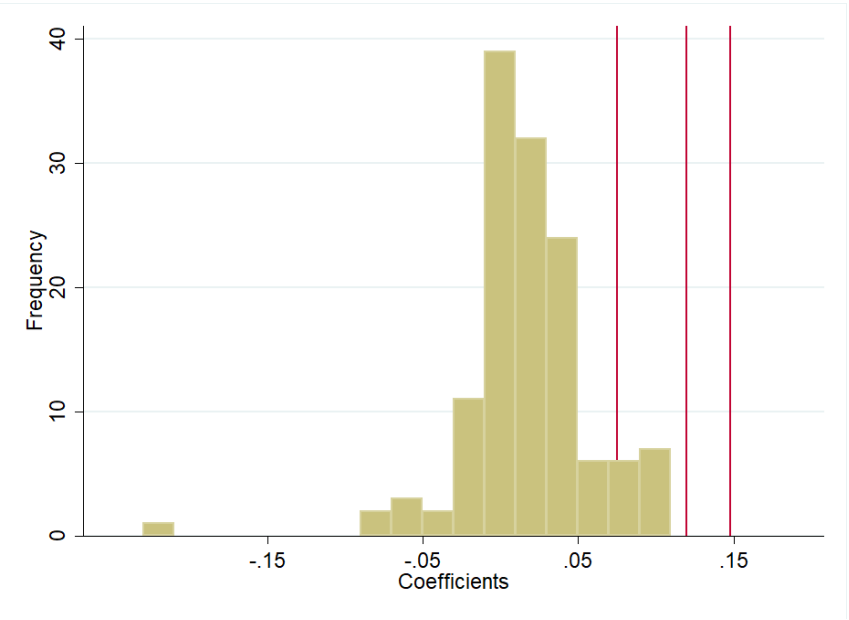
Note: The list of papers that studies the impact of notifications on property tax compliance.

The treatment in these papers is a notification sent to taxpayers on behalf of tax authorities through different channels (e.g., letter, tax bill, text message) and the dependent variable is the extensive margin compliance to property taxes. In all papers, the impact of only one interaction between tax authorities and taxpayers is formally evaluated. The average time between the intervention and the outcome measurement is around 11 weeks. In sum, the *LF* of the current paper is aligned with the prevailing setting in the literature.

Figure 2 plots the histogram of the treatment effect estimates of the mentioned papers (treatment effects are presented in bins of 0.02 or 2%). The vertical lines present the baseline estimates obtained in the current paper as shown in column 2 of Table 3. More specifically, the vertical lines going from the left to the right represent the *LF*, *MF*, and *HF* treatment effects as compared to the control group receiving no letter.

analysis. The treatment effect estimates reported for different sub-samples are not retrieved (i.e., the estimates reported in the heterogeneity analysis).

Figure 2: Histogram of treatment effect estimates



Note: Histogram of treatment effect estimates of papers studying the impact of notifications (nudges) on property tax payment. The vertical lines denote *LF*, *MF*, and *HF* estimates of the current paper as compared to the control group receiving no letter.

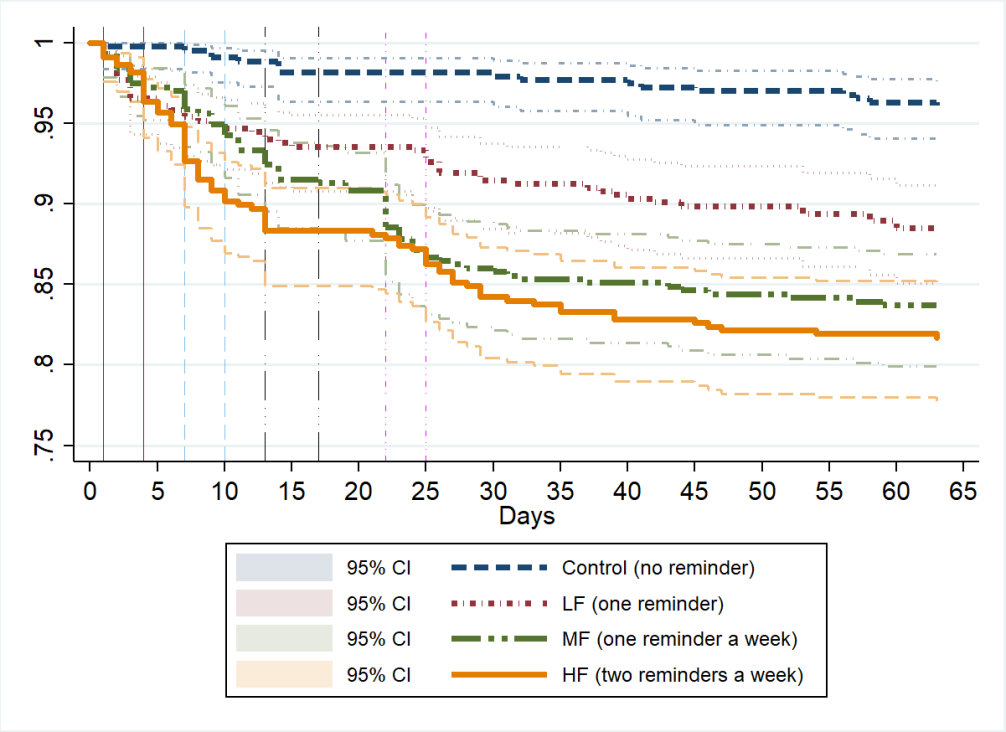
Figure 2 suggests that the treatment effect in *LF* is in the upper percentiles of the treatment effects distribution. Three other papers – John and Blume (2018), Chirico et al. (2019) and Brockmeyer et al. (2021) – find treatment effects larger than 0.07. The treatment effects in *MF* and *HF* outperform all the treatment effects depicted in the literature. We can only speculate why our *LF* estimate is larger than what most of the previous literature has found. This may for instance be attributed to the fact that we ran the experiment in China and the Chinese taxpayers are somewhat sensitive to reminders sent by authorities. Another candidate is the specifics of the framework we are working in, since the taxpayers in our sample are relatively wealthy compared to the overdue payments amount.

4.3. Survival Analysis

To uncover the dynamics of tax compliance over time, in Figure 3, we investigate the Kaplan-Meier survivor curves for all four treatments under consideration. The horizontal axis represents time in days ranging from September 16, 2019 (i.e., one day before the start of the experiment) until November 18, 2019 inclusive (i.e., the end of the trial), meanwhile the vertical axis illustrates the probability of not paying the property tax (alternatively, the share of taxpayers that remain non-

compliant).[20] The vertical lines on the graph depict the reminder days. The lines of the same color and shape represent the reminders disbursed within the same week. To recall, i) the subjects in *LF* receive only the first message of the first week; ii) the subjects *MF* receive only the first message of each week (4 messages in total); iii) the subject in *HF* receive all messages sent during the trial (8 messages in total).

Figure 3: Kaplan-Meier survival estimates



Note: The figure reports Kaplan-Meier survivor curves by treatment with associated 95% confidence intervals.

The inspection of Figure 3 suggests (relatively) big drops in the share of non-compliant taxpayers on most of the reminder days within each treatment compared to non-reminder days as well as on most of the reminder days in *MF* and *HF* compared to the same days in *LF* and *Control* (notice the relatively large steps on the reminder days).

4.4. Heterogeneity analysis

In this subsection, we assess the sensitivity of different subgroups to reminder frequency. More specifically, we distinguish between subgroups based on variables that proxy observable

[20] We set the start date to September 16, 2019 to be able to capture the survival on the first day of the intervention.

differences across individuals (mainly gender and previous payment history) and estimate equation (1) for each subgroup separately. Table 4 collects the estimates.

Table 4: Heterogeneity analysis

	(1)	(2)	(3)	(4)
LPM	Male	Female	Never Paid	Paid
LF	0.061**	0.097***	-0.001	0.108***
	(0.024)	(0.028)	(0.002)	(0.026)
MF	0.087***	0.175***	0.002	0.175***
	(0.024)	(0.033)	(0.002)	(0.028)
HF	0.126***	0.181***	0.006	0.214***
	(0.026)	(0.033)	(0.006)	(0.030)
Age	0.001	0.001	0.000	0.002*
	(0.001)	(0.001)	(0.000)	(0.001)
Tax Debt	0.061**	0.097***	-0.001	0.108***
	(0.024)	(0.028)	(0.002)	(0.026)
Gender			-0.006	-0.010
			(0.006)	(0.023)
Number of Months	-0.006***	-0.005***	0.003	-0.003***
Taxes Paid	(0.001)	(0.001)	(0.003)	(0.001)
Constant	0.237***	0.143**	-0.135	0.104*
	(0.055)	(0.072)	(0.125)	(0.056)
F stat.	10.184	9.883	0.174	12.081
P>F	0.000	0.000	0.991	0.000
Adj. R^2	0.068	0.069	0.092	0.056
N	1,081	661	561	1,181

Note: The remarks of Table 3 apply.

In almost all subgroups, we identify a robust treatment effect in the sense that the probability of paying the overdue tax increases following the reminder interventions. In line with column 3, the subgroup in which taxpayers did not pay their 2016 and 2017 property taxes as of November 18, 2019 is an exception.[21,22] Two potential mechanisms can be accountable for such a result. First, these highly non-compliant taxpayers (or in more general terms individuals who are hard to be convinced) may not be susceptible to reminders, irrespective of reminder frequency. This may be especially relevant for those taxpayers who are in the first tertile of the tax debt distribution

[21] As a robustness check, we again run additional analysis to account for multiple hypotheses testing. More specifically, we distinguish between four sub-groups (males who paid before, males who never paid before, females who paid before, females who never paid before) and compare the tax compliance behavior across the treatment arms and the control within each subgroup applying the procedure developed by List et al. (2019). Our conclusions remain intact. The results of this analysis are reported in Table A2 in Appendix A.

[22] The average 2018 property tax for the taxpayers who never paid the property tax in the last two years and for the taxpayers who paid at least once is roughly equal (4509.279 RMB (631.299 USD) and 4373.088 RMB (612.232 USD), respectively). Thus, the stark behavioral differences across the two groups cannot be explained by the extraordinarily high tax debt of the former group. The two groups are also similar in terms of gender composition ($\chi^2(1) = 0.879$, p-value=0.349).

(N=156), have a low tax burden (approximately 1,000 RMB on average), and are less likely to face liquidity constraints unlike taxpayers with higher tax debts. Second, a fraction of taxpayers who never complied before may face liquidity constraints hindering them to fulfill their tax obligations. This argument may be relevant for taxpayers who are in the second (N=215) or the third (N=190) tertiles of the tax debt distribution and have relatively high tax obligations (on average 3,400 RMB and 8,800 RMB, respectively). Indeed, there will be taxpayers in the first tertile of the tax debt distribution with liquidity constraints, as well as taxpayers in the second and the third tertiles of the tax debt distribution with no liquidity constraints, albeit insensitive to reminders. We believe that this finding can be important for policy purposes, given the excitement around nudge interventions in order to enhance tax compliance (Antinyan and Asatryan, 2020; Mascagni, 2018).

Interestingly, the taxpayers' gender is sensitive to reminder frequency. Low and medium frequency reminders have a similar impact on males, given that the probability of tax payments in *MF* is not statistically different from that in *LF* (F=0.58, p-value=0.446). On the contrary, females are more responsive to medium frequency than to low frequency reminders (F=3.9, p-value=0.049). High frequency reminders have a similar effect on the representatives of both genders compared to the low frequency reminders (for males F=4.01, p-value=0.046; for females F=3.99, p-value=0.046). The comparison of the respective treatment dummy coefficients across subsamples (e.g., the coefficient of *HF* treatment in the female subsample with that of *HF* treatment in the male subsample) suggests that females are more sensitive to medium frequency reminders than males (χ^2=4.74, p-value=0.030), while high frequency reminders have similar impact on both genders (χ^2=1.71, p-value=0.191).

5. Discussion

Our paper leads to two novel results. First, we find that recurrent reminders are more potent instruments to enhance tax compliance compared to a one-off reminder. Second, we illustrate that beyond a certain frequency the effectiveness of additional reminders seems to decline. In this section, we uncover potential mechanisms that may drive these results. We start the discussion by detailing why a one-off reminder can be effective. Next, we build on these considerations to

explain the dominance of recurrent reminders. Lastly, we discuss why the effectiveness of high-frequency reminders may decline compared to medium-frequency reminders.

Why is a one-off reminder effective? Several factors can explain why individuals may have overdue property tax obligations. First, individuals may suffer from limited attention and memory (DellaVigna, 2009), which can make them forget either the payment or the deadline. Second, individuals may perceive the task as unpleasant because they have to invest time, mental energy, and financial resources to be compliant (Cranor et al., 2020). Given these circumstances, they may postpone the payment to the future, or in other words procrastinate (Akerlof, 1991; Rabin, 1998). Third, the financial (e.g., the penalty) and non-financial (e.g., trading restriction) consequences of having overdue tax obligations may not be salient enough for individuals (Slemrod, 2019; Cranor et al., 2020). Forth, individuals may face liquidity constraints, which may especially apply to those with high tax arrears. Most likely, the one-off reminder we sent effectively brought the pending payment to taxpayers' mind (Sunstein, 2014) and increased the salience of financial and non-financial consequences in case of non-compliance. This can explain the considerable increase in the probability of tax compliance in the low-frequency reminder treatment (*LF*) compared to the control treatment.

Why are frequent reminders more effective than one-off reminders? Upon receiving the one-off reminder there may still be individuals who will not immediately act upon this reminder. For instance, a few individuals may decide to make the overdue payment in the nearest future and again forget about it because of limited memory or attention.[23] Unlike a one-off reminder, recurrent reminders always keep the pending action fresh in these individuals' minds. Taxpayers' immediate reaction on reminder days captured by Kaplan-Meier survivor curves seems to support the conjecture that the frequent reminders constantly bring the pending payments to the taxpayers' mind inducing them to act. It can also be the case that after receiving a reminder a few individuals may still procrastinate and postpone the unpleasant task because of time, mental energy, and financial resources necessary to make the overdue tax payment.[24] By constantly reminding about the pending action, recurrent reminders may create pressure for these taxpayers and induce them to act. The identified gender differences provide additional evidence in favor of this claim. In

[23] For instance, the one-off reminder may arrive in the middle of the working day and an individual may decide to pay in the evening (or in several days) completely forgetting about the payment.

[24] For instance, an individual may face other competing and more attractive spending or investment activities and may be tempted to spend the money elsewhere (Calzolari and Nardotto, 2016).

general, females are less resistant to pressure than males (Gneezy et al., 2003; Shurchkov, 2012; De Paola and Gioia, 2016). This can explain why females react both to medium and high frequency reminders, while males react to high frequency reminders only that introduce the highest level of potential pressure. Lastly, recurrent reminders continually notify the taxpayers about the financial and non-financial sanctions, which can increase the salience of these sanctions.

Why does the effectiveness of additional reminders decline beyond a certain frequency? First, the additional 4 reminders sent in *HF* most likely do not provide tangible gains in the attention paid to the payment compared to that in *MF* where only 4 reminders were sent. Second, the additional 4 reminders sent in *HF* may not result in considerable gains in the salience of the financial and non-financial sanctions compared to that in *MF*. Third, the pressure triggered by two reminders a week may not be substantially different from the pressure triggered by one reminder per week.

6. Conclusion

In this paper, we add to the literature by studying the impact of low-cost reminder frequency on overdue tax compliance in Baoshan district in Shanghai. We show that frequent reminders considerably enhance the probability of paying overdue taxes compared to one-off reminders. We also illustrate that beyond a certain frequency the effect of additional reminders on the probability of tax compliance seems to decline, though the effect is still positive. Our interventions are highly cost-effective and result in considerable fiscal gains for the tax administration. More specifically, by spending around 173 RMB (24.22 USD) in *LF*, 692 RMB (96.88 USD) in *MF*, and 1,398 RMB (195.72 USD) in *HF* we were able to generate around 218,000 RMB (30,520 USD) tax revenue in *LF*, 306,000 RMB (42,840 USD) tax revenue in *MF*, and 361,000 RMB (50,540 USD) tax revenue in *HF*. In comparison, the tax revenue equaled to roughly 75,000 RMB (10,500 USD) in the control treatment.[25]

[25] To compute the tax revenue in each treatment we first calculate the estimated average benefits from a single taxpayer. To do so, we multiply the average tax debt of the entire sample by the probability of tax compliance in each treatment. The point estimates for the probability of tax compliance are borrowed from column 2 of Table 3. For instance, to calculate the estimated average benefit from a single taxpayer in *HF*, we multiply the average tax debt of the entire sample by 0.187 (0.039+0.148). To calculate the tax revenue, we multiply the estimated average benefits from a single taxpayer in each treatment by the number of taxpayers in the treatment.

From a policy perspective, our results suggest that policy outcomes can be improved through frequent reminders. Nonetheless, beyond a certain frequency of reminders, sending additional reminders may not result in a considerable change in human behavior. Our results also suggest that when opting for letters and reminders to enhance rule or tax compliance, policymakers may want to consider the past behaviors of individuals as a potential proxy for their sensitivity to such interventions. More specifically, individuals who were non-compliant in the past may not respond both to frequent and non-frequent reminders.

Acknowledgments

We thank Baoshan Tax Administration for allowing us to run the experiment. We also thank Marie Claire Villeval, Alexis Belianin, Albrecht Bohne and the audiences at EBE (New York), SABE (Moscow), HSE (Moscow), ZEW (Mannheim), and TARC (Exeter) webinars for helpful suggestions. Chenyixin provided excellent research assistance. The ethical committee of the Renmin University of China approved the experimental protocol of the study (RUCecon-201909-1).

References

Akerlof, G. A. (1991). Procrastination and obedience. *The American Economic Review, 81*(2), 1-19.

Alm, J., Lai, W., & Li, X. (2021). Housing market regulations and strategic divorce propensity in China. *Journal of Population Economics*, 1-29.

Antinyan, A., & Asatryan, Z. (2020). Nudging for tax compliance: A meta-analysis. *CESifo Working Paper No. 8500*.

Antinyan, A., Bertoni, M., & Corazzini, L. (2021). Cervical cancer screening invitations in low and middle income countries: Evidence from Armenia. *Social Science & Medicine, 273*, 113739.

Altmann, S., & Traxler, C. (2014). Nudges at the dentist. *European Economic Review, 72*, 19-38.

Apesteguia, J., Funk, P., & Iriberri, N. (2013). Promoting rule compliance in daily-life: Evidence from a randomized field experiment in the public libraries of Barcelona. *European Economic Review, 64*, 266-284.

Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., & Tungodden, B. (2019). You've got mail: A randomized field experiment on tax evasion. *Management Science*.

Brockmeyer, A., Estefan, A., Arras, K. R., & Serrato, J. C. S. (2021). *Taxing property in developing countries: Theory and evidence from mexico* (No. w28637). National Bureau of Economic Research.

Cadena, X., & Schoar, A. (2011). *Remembering to pay? Reminders vs. financial incentives for loan payments* (No. w17020). National Bureau of Economic Research.

Calzolari, G., & Nardotto, M. (2017). Effective reminders. *Management Science*, *63*(9), 2915-2932.

Carril, A. (2017). RANDTREAT: Stata module to randomly assign treatments uneven treatments and deal with misfits.

Cho, S., & Choi, P. P. (2014). Introducing property tax in China as an alternative financing source. *Land use policy*, *38*, 580-586.

Cleves, M., Gould, W., Gould, W. W., Gutierrez, R., & Marchenko, Y. (2016). *An introduction to survival analysis using Stata*. Stata press.

Cranor, T., Goldin, J., Homonoff, T., & Moore, L. (2020). Communicating tax penalties to delinquent taxpayers: Evidence from a field experiment. *National Tax Journal*, *73*(2), 331-285.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic literature*, *47*(2), 315-72.

De Neve, J. E., Imbert, C., Spinnewijn, J., Tsankova, T., & Luts, M. (2019). How to improve tax compliance? Evidence from population-wide experiments in Belgium. *Evidence from Population-Wide Experiments in Belgium (May 05, 2019). Saïd Business School WP*, *7*.

De Paola, M., & Gioia, F. (2016). Who performs better under time pressure? Results from a field experiment. *Journal of Economic Psychology*, *53*, 37-53.

Doerrenberg, P., & Schmitz, J. (2017). Tax compliance and information provision–A field experiment with small firms. *Journal of Behavioral Economics for Policy*, *1(1)*, 47-54.

Du, Z., & Zhang, L. (2015). Home-purchase restriction, property tax and housing price in China: A counterfactual analysis. *Journal of Econometrics*, *188*(2), 558-568.

Ericson, K. M. (2017). On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation. *Journal of the European Economic Association*, *15*(3), 692-719.

Gillitzer, C., & Sinning, M. (2020). Nudging businesses to pay their taxes: Does timing matter? *Journal of Economic Behavior & Organization*, *169*, 284-300.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The quarterly journal of economics*, *118*(3), 1049-1074.

Groves, P. M., & Thompson, R. F. (1970). Habituation: a dual-process theory. *Psychological review*, *77*(5), 419.

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

Hernandez, M. A., Jamison, J. C., Korczyc, E. J., Mazar, N., & Sormani, R. C. (2017). *Applying behavioral insights to improve tax collection: experimental evidence from Poland* (No. 116046, pp. 1-68). The World Bank.

Ito, K., Ida, T., & Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, *10*(1), 240-67.

Kalbfleisch, J. D., & Prentice, R. L. (2011). *The statistical analysis of failure time data* (Vol. 360). John Wiley & Sons.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, *53*(282), 457-481.

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, *62*(12), 3393-3411.

Mascagni, G. (2018). From the lab to the field: A review of tax experiments. *Journal of Economic Surveys*, *32*(2), 273-301.

Moulton, S., Collins, J. M., Loibl, C., Haurin, D. R., & Brown, J. (2019). Reminders to pay property tax payments: A field experiment of older adults with reverse mortgages. *Available at SSRN 3445419*.

Mogollón, M., Ortega, D., & Scartascini, C. (2021). Who's calling? the effect of phone calls and personal interaction on tax compliance. *International Tax and Public Finance*, 1-27.

Ortega, D., & Scartascini, C. (2020). Don't blame the messenger. The Delivery method of a message matters. *Journal of Economic Behavior & Organization*, *170*, 286-300.

Rabin, M. (1998). Psychology and economics. *Journal of economic literature*, *36*(1), 11-46.

Rapoza, K. (2017). Shanghai Housing Prices Completely Unsustainable. Retrieved from https://www.forbes.com/sites/kenrapoza/2017/03/19/shanghai-housing-prices-completely-unsustainable/. Retrieved on 27.07.2020.

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., ... & McSweeney, F. K. (2009). Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. *Neurobiology of learning and memory*, *92*(2), 135-138.

Shurchkov, O. (2012). Under pressure: gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, *10*(5), 1189-1213.

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature*, *57*(4), 904-54.

Sunstein, C. R. (2014). Nudging: a very short guide. *Journal of Consumer Policy*, *37*(4), 583-588.

Thompson, R. F., & Spencer, W. A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review*, *73*(1), 16.

Appendices

Appendix A: Multiplicity adjusted p-values based on List et al. (2019)

Table A1: Multiple treatments

Treatment Comparison	Differences in Means	Unadjusted p-values	Adjusted p-values (Theorem 3.1)	Improved p-values (Remark 3.7)
Control *vs.* LF	0.076	0.000	0.000	0.000
Control vs. MF	0.124	0.000	0.000	0.000
Control vs. HF	0.144	0.000	0.000	0.000
LF vs. MF	0.048	0.042	0.075	0.042
LF vs. HF	0.068	0.006	0.014	0.014
MF vs. HF	0.020	0.426	0.426	0.426

Note: p-values based on List et al. (2019) procedure.

Table A2: Multiple subgroups

Subgroup	Treatment Comparison	Differences in Means	Unadjusted p-values	Adjusted p-values (Theorem 3.1)	Improved p-values (Remark 3.7)
Male and Paid Before					
	Control vs. LF	0.088	0.011	0.011	0.011
	Control vs. MF	0.132	0.000	0.000	0.000
	Control vs. HF	0.191	0.000	0.000	0.000
Male and Never Paid Before					
	Control vs. LF	0.000	0.000	1.000	1.000
	Control vs. MF	0.000	0.000	1.000	1.000
	Control vs. HF	0.000	0.000	1.000	1.000
Female and Paid Before					
	Control vs. LF	0.135	0.002	0.003	0.003
	Control vs. MF	0.244	0.000	0.000	0.000
	Control vs. HF	0.232	0.000	0.000	0.000
Female and Never Paid Before					
	Control vs. LF	0.000	0.000	1.000	1.000
	Control vs. MF	0.000	0.000	1.000	1.000
	Control vs. HF	0.020	0.448	0.448	0.448

Note: p-values based on List et al. (2019) procedure.