

An Experimental Evaluation of Notification
Strategies to Increase Property Tax Compliance:
Free-Riding in the City of Brotherly Love

Michael Chirico, Robert Inman, Charles Loeffler,
John MacDonald, and Holger Sieg*
University of Pennsylvania

Prepared for the 2015 NBER Conference on Tax Policy and the Economy

April 11, 2015

*We would like to thank Rob Dubow (Director of Finance), Clarena Tolson (Revenue Commissioner), and Marisa Waxman (Deputy Commissioner for Assessment of Properties) in the Department of Revenue of the City of Philadelphia for their help and support. We would also like to thank Reed Shuldiner and Chris Sanchirico as well as participants of numerous seminars for comments and suggestions. The views expressed here are those of the authors and do not necessarily represent or reflect the views of the City of Philadelphia.

Abstract

This study evaluates a set of notification strategies intended to increase property tax collection. We develop a field experiment in collaboration with the Philadelphia Department of Revenue. The different treatments emphasize aspects of behavior that play a large role in the tax compliance: deterrence, the need to finance the provision of public goods and services, and the desire for peer conformity. Our empirical findings provide evidence that both moral appeal to finance public goods and services and the desire for peer conformity modestly improve tax compliance, while deterrence notifications are no different from standard notifications.

KEYWORDS: Tax Compliance, Property Taxation, Field Experiment, Deterrence, Moral Appeal, Peer Conformity.

1 Introduction

The purpose of this study is to evaluate a set of notification strategies intended to increase property tax revenue collection. We design several notification-based treatments motivated by the theoretical literature on tax compliance. Extrinsic motivations are emphasized by Becker's deterrence model which explains tax compliance as a combination of penalties and costly monitoring; intrinsic motivations are captured by models that rely on social norms or social contracts. To weigh the empirical importance of these competing narratives, the basic strategy behind our controlled, randomized experiment is to vary the informational content of notification letters sent to citizens late or delinquent on their property tax payment. This field experiment was implemented in collaboration with the Philadelphia Department of Revenue (DoR).

While most of the previous tax compliance literature has focused on collection of income taxes, we focus on property tax collection.¹ Property taxes in the United States play an integral role in financing local municipalities and their school districts² In theory, levying a property tax is straightforward. It requires only valuation of the property and a registry of the identities of property holders. There is no need for tax payers to deal with complicated forms or to give up personal information as with an income tax. Since property cannot be hidden, scuppered away to a tax haven, or concealed in an electronic data system, the tax is difficult to evade. Moreover, there are some compelling theoretical arguments for taxing property instead of its

¹The empirical literature on tax compliance is discuss in detail in Hallsworth, List, Metcalfe and Vlaev (2014). Other notable papers that use field experiments to study tax compliance are Blumenthal et al (2001), Kleven et al. (2011), Ariel (2012), and Pomeranz (2013).

²It is worth noting here that the U.S. provides more tax authority to local municipalities than do most other developed countries.

alternatives (income, consumption, and wealth).³

The practice of property tax collection is a lot more ambiguous.⁴ In particular, property tax compliance is a significant problem in many large U.S. cities. Philadelphia is a prominent example. A recent study by the Pew Charitable Trust (2013) focused on a sample of 36 cities and found that Philadelphia had the fifth-highest delinquency rate in 2011, the last year for which common statistics were available. As of April 2014, the city and school district were owed \$595 million in delinquent taxes on 134,888 properties.

Enforcing property tax collection on delinquents⁵ is restricted by the limited set of sanctions that are available to local municipalities. In practice, a city typically has three options. First, a tax lien may be imposed on a property. Tax liens are an efficient tool to force delinquent tax payers to settle their debts for properties that are sold in official, arms-length market transactions. The outstanding debt can typically be recovered as a condition of sale when the title of the property needs to be transferred to a new owner. Unfortunately, this punishment hardly constrains any owner who has no intent of selling their property on the open market—for example, if they intend to unofficially bequest the title to their children or relatives—since the owner does not need to transfer the title of the property. Liens are also not effective for any owners that hold properties for purely speculative purposes, since such owners will only flip

³Two common arguments for the property tax are that a) the burden of the tax is fairly and equally distributed among the residents and firms located in a municipality; and b) that the excess burden of the tax is likely to be low.

⁴The property tax base, which is the assessed value of the house, is established in a separate assessment process in most cities in the U.S. While there are frequent disagreements about correct property tax assessments, tax enforcement is dictated by the assessed value of the property.

⁵Philadelphia operates with relatively wide latitude under the Pennsylvania Municipal Claims and Tax Lien Law, enacted in 1923. In Philadelphia, tax payers are considered delinquent if their debt is still outstanding at the close of the calendar year.

the property if they expect sufficiently large returns. Tax liens can also, in principle, be sold to collection agencies to generate immediate revenues for a city. Many cities, however, are reluctant to pursue this option, because there are some potentially large political costs associated with outsourcing tax collection enforcement.

Second, a city can start a foreclosure, a legal process in which the city as a lien holder attempts to recover the balance of the debt of the property owner by forcing the sale of the property on which a lien has been placed. In a tax foreclosure, the local jurisdiction petitions a court to award it the title based on an unpaid lien. Properties with court-ordered foreclosures can be sold in an auction or at a sheriff's sale.⁶ Obtaining a court-ordered foreclosure is a difficult and costly process. While it may be sensible for a bank to bear these costs when a lender defaults on a large mortgage, it seems as a property tax collection tool to be rather crude. Foreclosures can also be politically unpopular since they have the potential to generate the headlines which create frictions between voters and local politicians.

Lastly, a city can try to engage a tax delinquent in a bargaining process prior starting the foreclosure process. If a property is in danger of foreclosure due to a tax delinquency, DoR will, upon request of the debtor, attempt to work out a schedule for resolution of the balance due over a period of several months, usually a year⁷. While DoR has had some success in negotiating tax payment plans in Philadelphia, it should be emphasized that participation in these negotiations is purely voluntary. In general, the bargaining power in these negotiations is ultimately determined by

⁶The Office of the Philadelphia Sheriff completes about 60 Sheriff's Sales each month. A list of properties currently up for Sheriff's Sale by the City of Philadelphia can be found at <http://www.officeofphiladelphia sheriff.com/en/real-estate/foreclosure-listings>.

⁷In fact, as of October 2013, the structure of such an agreement is legally prescribed through Owner-Occupied Payment Agreement (OOPA) plans. OOPA plans basically determine a monthly payment as one of four income-dependent percentages of monthly income

the likelihood of enforcing liens or foreclosure.

Given the problems associated with the traditional methods of enforcing property tax compliance, there is some interest in exploring the efficacy of alternative, softer approaches. The main idea behind these softer approaches is to use the power of persuasion and design an effective notification strategy. The purpose of this paper is to provide new insights into the effectiveness of different notification strategies for increasing property tax collection. Our intervention focuses on three treatments that are motivated by the recent literature on tax compliance.

The first treatment is motivated by the deterrence model of tax compliance, which emphasizes external motivations of behavior (Becker, 1968, Allingham and Sandmo, 1972). Key elements of deterrence are the probability of detection (the monitoring probability) and the seriousness of the penalty or fine that is levied on tax evaders who are caught. Given the uncertainty of tax enforcement, tax compliance is primarily a function of subjective perceptions of the probability of detection and the seriousness of the penalties for being caught evading taxes. Our experiment does not explicitly vary the penalties associated with non-compliance, i.e. we do not randomize based on different fiscal incentives to pay taxes. Neither do we expose tax payers to different enforcement regimes. Instead, we inform the delinquent tax payers that the DoR will pursue their failure to pay property taxes, and that evading property taxes has serious repercussions.

Proceeding, the recent literature in public economics emphasizes that property taxes are used to provide local public goods and services. We can thus study the problem of tax compliance using insights of the literature on voluntary public good provision (Bergstrom, Blume & Varian, 1986). Paying property taxes may also be perceived by many households as the “right thing to do”, i.e. there may be some

warm glow associated with tax compliance (Andreoni, 1989).⁸ The second notification strategy focuses on this rationale for paying taxes. It reinforces the importance of public goods and services that are provided by the city and points out that these services cannot be provided without a fair distribution of the tax burden; we interpret this as a moral appeal to pay taxes. The treatment emphasizes the *quid pro quo* of tax compliance in the city—i.e., public goods and services (namely, education and safety provision) funded by compliant taxpayers.

The third notification treatment builds on the insights of the recent literature on social norms.⁹ In particular, tax payers may comply with tax laws due to peer pressure and the desire to conform to acceptable standards of behavior (Bernheim, 1994). In principle, one could vary the conformity costs within a randomized controlled experiment. For example, one could threaten delinquent tax payers with public exposure by printing names in newspapers or public web pages. In practice, such interventions faces large legal hurdles and are, therefore, onerous to implement; we instead focus on a simpler and softer strategy. The basic idea behind our peer conformity treatment is to notify non-compliant individuals that their peers and neighbors are paying taxes.

To implement our experiment we needed to integrate our three treatment conditions into the standard operating procedures of the billing system of the DoR. Fortunately, DoR assigns properties to billing cycles using a pseudo-randomized mechanism, in which assignments are primarily based on the last two digits of social security or employer identification numbers. We then assigned our three treatments (deter-

⁸ Parallels are drawn between the problem of raising taxes and the fund-raising problem faced by non-for-profit organizations (Andreoni, Erard and Feinstein, 1998). For example, publicly funded radio stations typically need to raise funds from their listeners and cannot rely on threats of enforcement.

⁹This literature also emphasizes concepts such as social fairness and reciprocity (Fehr and Gächter, 1998).

rence, social contract, peer conformity) and the control condition of basic payment notification to days in the billing cycles using randomly drawn four-day sequences.

We conducted our experiment in November 2014 over a period spanning 9 treatment days. We assessed the fidelity of our experiment by analyzing letters that were returned as undeliverable to DoR by the U.S. Postal Service. Based on our analysis of the returned letters we concluded that our treatment had been correctly assigned on these days; our final sample consists of 4,927 properties. Based on a variety of balance tests, we conclude that treatments were randomly assigned within our sample. The DoR started to share data on property tax payments immediately after the experiment. Our current sample is based on all payments through January 6th, 2015.

Our empirical analysis focuses on two sets of outcome measures. First, we follow the recent empirical literature on tax compliance and analyze discrete outcomes. Specifically, we would like to know whether each owner paid any or all property taxes during the follow-up observation period. Second, we examine the magnitude of payments received by DoR. We normalize the received property tax payments by the assessed value of the property. Our findings show some evidence that both moral appeal and peer conformity help to improve tax compliance. We find that the deterrence notification performs no different than the standard billing notification.¹⁰

The rest of the paper is organized as follows. Section 2 provides a detailed discussion of our three treatments and the control. Section 3 discusses the experimental design and the fidelity of its implementation. Section 4 presents a descriptive analysis summarizing the main effects of our experiment. Section 5 provides a formal analysis and discusses the estimates of the treatment effects. Section 6 offers some conclusions and discusses future research.

¹⁰Other empirical studies that have failed to provide support for the deterrence model are Alm (1999) and Togler (2002).

2 Treatments

To explore softer avenues for revenue augmentation, we determined that the most logistically feasible approach was to include a “message” in the bills regularly mailed to non-payers, the language of which was carefully chosen to target one of three aforementioned enforcement strategies: deterrence (the “Threat” treatment), social normality (the “Moral” treatment), and conformity (the “Peer” treatment). To properly isolate the treatment effects of receiving a specific message from the effects of receiving *any* message given the status quo of plain bills, we also randomly sent properties a nondescript message (the “Control” treatment).¹¹

The letters were designed carefully to differ only in the wording of their middle paragraphs; for clarity, the crux of the idiosyncratic wording is reiterated below. We also took care to minimize issues of communication for those with limited literacy by simplifying the language of the three different messages—shunning uncommon words and syntactic complexity by churning the text through the latest linguistic software for complexity analysis.

Further, in accordance with the DoR’s general desire to reach out to Philadelphia’s substantial immigrant and refugee populations, we agreed to include Spanish translations of each treatment’s text on the reverse of our stuffers.¹²

¹¹See Figure 1 for the original due letter and Figures 2-5 for the exact style and full wording of each treatment letter.

¹²As Puerto Ricans make up the plurality of Philadelphia’s non-English-speaking population, we targeted the vernacular of the translation towards Caribbean Spanish speakers.

2.1 Treatment 1: Deterrence

The goal of this treatment is to emphasize the repercussions of not complying with property tax requirements with the city, highlighting the policy tools available to the city. This message was intended to educate owners who may have poorly-formed notions of the extent of actions the city may be willing to take to recover taxes from each property. The message contained the following sentences:

Not paying your Real Estate Taxes is breaking the law. Failure to pay your Real Estate Taxes may result in seizure or sale of your property by the City. Do not make the mistake of assuming we are too busy to pursue your case.

2.2 Treatment 2: Moral Appeal & Public Services

The goal of this treatment is to emphasize the social contract between tax payers and the city. The city can only provide public goods if property taxes are paid punctually. In particular, we chose to highlight the correspondence between Real Estate Tax compliance and the City's provision of public education and safety services. The message contained the following sentences:

We understand that paying your taxes can feel like a burden. We want to remind you of all the great services that you pay for with your Real Estate Tax Dollars. Your tax dollars pay for schools to teach our children. They also pay for the police and firefighters who help keep our city safe. Please pay your taxes as soon as you can to help us pay for these essential services.

2.3 Treatment 3: Social Pressure & Conformity

The goal of this treatment is to socially shame delinquents by underlining their non-conformity compared to their neighbors. The message contained the following sentences:

You have not paid your Real Estate Taxes. Almost all of your neighbors pay their fair share– 9 out of 10 Philadelphians do so. Paying your taxes is your duty to the city you live in. By failing to pay, you are abusing the good will of your Philadelphia neighbors.

2.4 Control

The control group message was designed to be as plain and detached as possible, with the goal of the extra informational content being orthogonal to each of the three treatment messages. The control condition contained the dispassionate billing notification:

The enclosed bill details your outstanding Real Estate Taxes due to the City of Philadelphia.

3 Experimental Design

3.1 Timing of Property Tax Collection in Philadelphia

Real estate taxes in Philadelphia are levied annually on a property-level basis. The Office of Property Assessment evaluates the market value of each property, 1.34 % of which must be paid to the Philadelphia Department of Revenue. The city then splits

property tax revenue with the School District of Philadelphia, with the former getting approximately 45% revenues. Tax bills are mailed by DoR in batches throughout December and early January each year; owners have until March 31st to remit their balance to the City, after which time the owners are considered late and the bill begins to accrue penalties and interest.

The DoR actively begins pursuing non-paying properties in the September following nonpayment. First, the city delegates roughly $\frac{2}{3}$ of the debts to the authority of two designated external law firms ($\frac{1}{3}$ each), which have contracts with the city for delinquent property tax collection.¹³ These law firms are free to pursue the collection of the debt as they see fit, and are rewarded with a portion of any debts recovered for the City.

For those debts that remain targeted by DoR itself, the city traditionally leverages one of the several legal options described above as threat and punishment for non-payment. Beginning on March 31st, the city regularly sends plain bills to properties still in hock—roughly once every 10 weeks.¹⁴ More substantive enforcement strategies begin when the properties become officially delinquent on December 31st. Our experiment was conducted in November, a few weeks before the tax payer officially became delinquent.

3.2 Randomization

Our approach to randomization was constrained by the logistics of DoR’s enforcement faculties. We concluded after several discussions with our collaborators at DoR that it would be in practice impossible to assign individual properties at random to different

¹³Currently, these law firms are Linebarger, Goggan, Blair & Sampson and Goehring, Rutter & Boehm.

¹⁴See the section on implementation below for exact details.

treatments. Instead, we chose to exploit the pseudo-random assignment of properties to billing cycles and randomized treatments across them. To understand this decision it is useful to discuss the current practice of posting reminder letters by DoR.

Mailing of delinquent real estate tax bills works essentially as follows. Every property in the city is assigned to one of 50 mailing cycles. Since it is cheaper and simpler to send at once all bills to those owners owing taxes on multiple properties, assignment to cycles is done at the owner level, so that each mailing cycle has roughly the same number of owners. Every morning, a printer at DoR taps the in-house accounting system to find all properties that a) owe taxes to the City and b) are in the current day's mailing cycle, with the numbered cycles progressing in sequence day-by-day. After identifying the bills to be printed for the day, the printer merges into the bill several other pieces of information stored with the delinquent balance such as the mailing address and an in-house ID associated with the property. The 1200 or so bills that are printed each day are then brought to the City's mailing room, wherein they are stuffed into envelopes and delivered to the property owners.

Given the volume of bills printed each day and the existing infrastructure for processing them, especially the machine-automated process of envelope stuffing, we determined the most practical solution would be to randomize treatment at the mailing cycle level, so that every bill printed on the same day would be paired with the same message. Randomization of mailing days was handled by the authors. We elected to randomize 4-day cycles—for each 4-day period, we picked at random among the $4! = 24$ possible arrangements of treatments over the subsequent 4 days. Our experiment was conducted on 9 days in November 2014, between the 4th and the 25th.

While we are certain of the sanctity of our mailing cycle-level randomization process, one may be concerned about the assignment of properties to mailing cycles by

the city. Fortunately, however, the city uses a pseudo-random mechanism to assign owners to billing cycles, which means that we achieve proper full-scale two-stage randomization of the properties through our process of day-level randomization.

In particular, the city assigned properties to cycles based on the last two digits of an in-house ID number; those with final two digits 01 and 02 are mapped to cycle 1, those with final digits 03 and 04 are mapped to cycle 2, and so on. The in-house ID itself is motley in nature. For many properties, DoR has on file the owner’s Social Security Number (SSN); for many others, mainly commercial properties, the DoR stores their Employer Identification Number (EIN); and for the remainder of properties, DoR assigns its own in-house ID number. This last is a 9-digit code which is assigned sequentially to property owners who cannot be matched to either of the federal ID numbers. While this assignment based on SSN or EIN is not purely random, it is a pseudo-random assignment. It is hard to believe that there would be any significant sorting or self-selection based on the last two digits of SSN or EIN.

3.3 Implementation Fidelity

To assess the fidelity of the experimental design, we leveraged a unique feature of the system. The Department of Revenue regularly posts envelopes destined for addresses that are either unattended (vacant) or do not exist in the first place due to typos. Either before or after an attempted delivery to such an address, the postal service flags down these missives and returns the bills to DoR, which then processes them and attempts, if they can identify a suitable alternative address, to re-deliver the tax bill. We took advantage of the fact that a subset of bills made their way back to DoR to check first-hand the extent of treatment fidelity. Our final sample consists of the nine treatment days for which greater than 90% fidelity was achieved.

3.4 Sample Size

Our sample is derived from the database of 134,888 delinquent properties in the City of Philadelphia. Figure 6 plots the properties using different colors to indicate the total outstanding tax liabilities. We find that the delinquent properties are fairly evenly spaced throughout the city. As some confirmation of this, Figure 7 plots the Rittenhouse Square neighborhood in Center City, one of the most affluent neighborhoods in Philadelphia. Despite the uniformly high property values there, we find interspersed there a fair number of delinquent properties.

Figure 8 plots the properties with different colors indicating the duration of the tax liability. We find that most liabilities are less than 2 years old. Nevertheless, there are also a fair number of cases involving debt that has been outstanding for more than a decade. These cases seem to be more prevalent in South Philadelphia.

From this original sample, we obtained the final sample of 4,927 properties that was used in our experiment by using the following screening devices:

1. Payment agreement (23%=31456)
2. Any tax abatement (5% = 4706)
3. Not handled by DoR (62%=61170)
4. Sheriff's Sale (11%=4098)
5. Bankruptcy (3%=948)
6. Sequestration (3%=1130)
7. Returned mail flag (5%=1429)
8. Not mailed during treatment period (83%=24800)

9. Paid off all but \$0.61 of debt by mailing (4%=224)

Note in particular that this sample selection means that our sample consists only of properties that are not in the purview of the two law firms that DoR uses as collection agencies. It is therefore useful to compare briefly the properties that are kept in-house with those that are assigned to the law firms. We find that properties kept in-house have lower balances, with a median of \$1,000, as compared to \$1,700 overall. However, in-house properties have higher market values—the DOR median is \$91,000 vs. \$66,100 overall. Properties handled by DoR have younger debt—an average of 4 years vs. 7 and 11 for the two law firms. Even conditional on age of debt, in-house balances are low. DoR-managed accounts are more likely to be owner-occupied, less likely to be in payment agreements, and more likely to result in a sheriff’s sales. In summary, it appears that the outside firms are holding properties which, even given other characteristics, have the highest potential returns.

3.5 Sample Balance on Observables

To confirm whether or not we indeed achieved randomization, we performed a series of balance-on-observables tests. The null hypotheses of these tests are that a given observable data moment is identical across mailing cycles. We turn now to the results of those tests.

Analysis of balance on observables is complicated by the random assignment at the owner level. Because there are some large holders of property—thousands of properties owned by public entities like the City of Philadelphia, the Philadelphia Housing Authority, and the Redevelopment Authority of Philadelphia, hundreds owned by many others such as the University of Pennsylvania and Drexel University—a simple analysis of balance at the property level will likely be skewed by these outliers. In

addition, it is not clear how to aggregate many of the property-level characteristics to the owner level meaningfully, especially geographic variables, complicating the task of testing balance at the owner level. Our compromise was to examine sample balance on the subset of properties for which a) the owner is unique, and b) any tax exemption claimed by the property is related to abatements for new construction.¹⁵

Most of the observed characteristics are categorical variables, so we can test balance using standard χ^2 tests. We had two days of threat treatment, four days of moral treatment, two days of peer treatment and one day of controls.¹⁶ Hence the expected population frequencies should be approximately 0.22, 0.44, 0.22 and 0.11 respectively. Tables 1 and 2 report the observed frequencies for each variable as well as p -values for the categorical χ^2 test. In addition, we have three continuous variables: balance due, market value and land area. Since there are some outliers in the data, we convert each of these continuous variables into categorical variables using quartiles of the underlying empirical distribution to mitigate their influence. We then used the same categorical χ^2 test to determine whether our sample is balanced on these continuous measures.

As can be seen in Tables 1 and 2, randomization appears to have been successful. The properties are strongly randomly distributed by location (their political ward, of which there are 66 in Philadelphia), category (type of property usage), property size (as measured by the number of rooms or by the size of the tract), case assignment (this

¹⁵We ran several other similar specifications, with the qualitative results remaining unchanged. We also ran tests on the subsample of properties for which we could obtain the secure ID used by the City, for which the putative mapping was violated; again, the results are qualitatively identical. See the Appendix for details.

¹⁶Initially, we had hope to have four days for each treatment, but the experimental design was compromised on 6 of the days in which we ran the experiment. We, therefore, had to exclude these observations from our sample.

Table 1: Tests of Sample Balance on Observables

	Threat	Moral	Peer	Control	<i>p</i> -value
Balance Due Quartiles					
<\$300	0.22	0.4	0.28	0.1	0.2
[\$300,\$1300)	0.24	0.47	0.22	0.08	
[\$1300,\$3300)	0.23	0.45	0.2	0.11	
> \$3300	0.18	0.48	0.23	0.11	
Market Value Quartiles					
<\$46k	0.24	0.43	0.21	0.12	0.2
[\$46k,\$82k)	0.22	0.46	0.23	0.1	
[\$82k,\$152k)	0.21	0.45	0.25	0.09	
> \$152k	0.21	0.45	0.24	0.1	
Land Area Quartiles					
<800 sq. ft.	0.22	0.45	0.23	0.1	0.83
[800,1200) sq. ft.	0.23	0.43	0.24	0.1	
[1200,1800) sq. ft.	0.21	0.47	0.22	0.1	
>1800 sq. ft.	0.21	0.44	0.24	0.1	
Distribution of Properties	0.22	0.45	0.23	0.1	0.08
Expected Distribution	0.22	0.44	0.22	0.11	

Table 2: Tests of Sample Balance on Observables (cont.)

	Threat	Moral	Peer	Control	<i>p</i> -value
# Rooms					
0-5	0.22	0.44	0.23	0.11	0.32
6	0.21	0.46	0.23	0.09	
7+	0.22	0.44	0.24	0.1	
Years of Debt					
1 Year	0.23	0.43	0.24	0.09	0.32
2 Years	0.22	0.44	0.24	0.1	
3-5 Years	0.2	0.48	0.22	0.1	
6+ Years	0.2	0.47	0.2	0.13	
Category					
Residential	0.22	0.45	0.23	0.09	0.07
Hotels&Apts	0.2	0.45	0.23	0.12	
Store w. Dwell.	0.21	0.48	0.22	0.09	
Commercial	0.15	0.5	0.24	0.11	
Industrial	0.27	0.42	0.2	0.11	
Vacant Land	0.25	0.39	0.23	0.13	
Expected Distribution	0.22	0.44	0.22	0.11	

variable captures, if applicable, to which outside law firm a property is assigned), and whether the property is in sequestration or has entered a payment agreement with the city. The number of properties assigned to each treatment is further exactly as expected, given the unequal number of mailing days in our treatment.

4 Descriptive Statistics of Treatment Effects

We consider results for three different subsamples. The first sample (I) is the full sample and consists of all 4927 observations; The second sample (II) eliminates commercial property owners, which reduces the sample to 4749 observations; the third sample (III) eliminates owners of multiple properties, resulting in a sample size of 3888.

Table 3 summarizes the impact of our experimental intervention on revenue collection. The table reports the total balance owed, the amount generated, and the number of mailing days for the three treatments and the control groups. It also reports the percent of properties that paid the City anything and the percent that paid off their full debt in our sample period.

We also report the dollars in revenue raised per day, which ranges from \$60,292 in the control group to \$111,931 in the moral treatment group. Note that the average payments per day is higher in all three treatment group. A simple difference between the treatment and the control group provides an estimate of the overall effectiveness of the intervention. These estimates range from \$10,883 for the treat treatment to \$51,639 for the moral treatment. Summing over all treatment groups and days suggests that our experiment generated approximately \$250,000 for the DoR in just nine days.

Table 3: Summary of Effectiveness of Treatment

Sample	Group	Treated Days	No. Treated	Total Debt Owed	Percent Ever Paid	Percent Paid in Full	Dollars Received	Dollars Per Day Treated	Dollars above Control Per Day	Total Generated over All Days
I	Threat	1	499	\$1,839,826	14	8	\$71,176	\$71,176	\$10,883	\$ 10,883
I	Moral	4	2,211	\$8,003,148	15	7	\$447,728	\$111,932	\$51,639	\$206,557
I	Peer	2	1,142	\$3,794,900	18	12	\$152,217	\$76,109	\$15,816	\$ 31,632
I	Control	2	1,075	\$3,294,516	16	10	\$120,585	\$60,293	\$ 0	\$ 0
II	Threat	1	480	\$1,657,379	15	8	\$71,176	\$71,176	\$11,142	\$11,142
II	Moral	4	2,122	\$7,024,458	15	7	\$288,758	\$72,189	\$12,155	\$48,621
II	Peer	2	1,099	\$3,350,147	19	12	\$146,227	\$73,114	\$13,079	\$26,158
II	Control	2	1,048	\$2,930,759	16	10	\$120,069	\$60,034	\$ 0	\$ 0
III	Threat	1	406	\$1,437,902	15	9	\$51,309	\$51,309	\$18,011	\$ 18,011
III	Moral	4	1,754	\$6,956,034	16	7	\$418,767	\$104,692	\$71,393	\$285,572
III	Peer	2	891	\$3,331,168	20	13	\$130,016	\$65,008	\$31,710	\$ 63,419
III	Control	2	837	\$3,007,232	16	9	\$66,597	\$33,299	\$ 0	\$ 0

The results are qualitatively similar for the two other samples. However, there are some important quantitative differences. If we restrict attention to the subsample of non-commercial properties, we find that the moral treatment raises a much smaller amount. That suggests that these differences are driven by a relatively small number of commercial property owners. If we restrict attention the properties of sole owners, all treatments appear in a more positive light as the intake of the control group drops precipitously.

5 Analysis of Treatment Effects

To analyze the causal effects of our treatments, we first consider the outcome that indicates whether the delinquent tax payer ever made a payment. We can measure this outcome over time starting with the date of our experiment. Figure 9 plots the time series for the three treatment groups and the control group using the full sample. To allow for postal delivery, we define time zero as five days after our experiment was implemented for each cycle. Note that the four time series look similar for the first 15 days of our sample. After that, we find that tax payers in the peer treatment group outpace those in the other three groups in participation rates. Moreover, tax payers in the moral group are less likely to pay some of their debt than those in the control.

To formalize the analysis, control for observables and test for statistical significance, we consider Logit regressions. Let $y_i = \mathbb{1}[x_i > 0]$, where $\mathbb{1}[\cdot]$ is an indicator taking the value one when its argument is true and 0 otherwise. Let x_i be the cumulative remittance to the city at the conclusion of the sample period by property i . Given the random assignment of treatments we can obtain a consistent estimator of the causal impact of treatment on y_i by using the following logistic regression model:

$$y_i = X_i^T \beta + D_{T,i} \gamma_T + D_{M,i} \gamma_M + D_{P,i} \gamma_P + \epsilon_i, \quad \epsilon \text{ logistic} \quad (1)$$

The $D_{k,i}$ are indicators for the three treatments, i.e. $D_{k,i} = \mathbb{1}[\text{treatment}_i = k]$, $k \in \{T, M, P\}$ for Threat, Moral, and Peer, respectively. The coefficients γ_k , then, measure the causal impacts of the treatments on the likelihood of some degree of remittance to the city, relative to the control treatment of a plain message.¹⁷ We report robust standard errors that are clustered to deal with multiple ownership.

Table 4 summarizes the estimates and the estimated standard errors for the three samples that we considered above. As can be seen from Table 4, the moral and the threat treatments had no significant effect at the conclusion of the current sample period. The Peer treatment is consistently positive and significant when we restrict the sample to sole owners.

Next we investigate whether there is heterogeneity in response to the treatment. It is plausible that tax payers who owe small amounts of money behaved differently than those who owe larger amounts. To gain some additional insights we define four quartiles of the distribution of outstanding balance of the tax debt within each sample. In the full sample, the first quartile consists of tax payers that owe less than \$300, the second quartile consists of tax payers that owe between \$300 and \$1400, the third quartile consists of tax payers that owe between \$1400 and \$3600, and the highest quartile consists of tax payers that owe more than \$3600. These cutoffs do not change substantially in the other subsamples. Figure 10 plots the time series by quartile for the three treatment groups and the control group for the full sample.

Figure 10 provides some important insights. First, note that response rates in

¹⁷To improve efficiency we also include some controls such as land area, maturity of debt (more or less than 5 years), geographic location (as approximated by City Council District), usage category, property exterior condition (whether or not the property was categorized as sealed/compromised by the city), whether the property took a homestead exemption (an indicator of owner occupancy), balance at mailing, and market value.

Table 4: Logistic Regressions – Ever Paid

	Full Sample	Non-Commercial	Sole Owner
Intercept	-1.69*** (0.08)	-1.67*** (0.08)	-1.68*** (0.10)
Moral	-0.07 (0.10)	-0.10 (0.10)	0.04 (0.12)
Peer	0.21 (0.11)	0.19 (0.11)	0.30* (0.13)
Threat	-0.09 (0.15)	-0.06 (0.15)	-0.03 (0.17)
Log Likelihood	-2136.16	-2068.89	-1758.95
Num. obs.	4927	4749	3888

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

all four subsamples decline rather dramatically across quartiles of the balance distribution. In the first quartile the response rate is almost 30 percent by the end of the sample. In the fourth quartile, the response rate is typically less than 10 percent. Moreover, there are important differences by treatment. We find that the peer treatment is more effective for delinquent tax payers in the low quartiles. In contrast, the moral appeal appears to be more powerful for tax payers in the high quartiles of the balance distribution.

To investigate these issues more formally we create dummy variables that indicate whether a tax payer is in a given quartile of the balance distribution or not. We then include these dummies into the Logit regression to capture the declining participation rate by balance. We also interact these dummies with the treatment indicators to capture heterogeneity in treatment. Table 5 summarizes the estimates and the estimated standard errors for our three subsamples.

Table 5 shows the quartile dummy variables are significantly negatively correlated with participation. The more a delinquent tax payer owes the city, the less likely he is to pay his tax bills. Moreover, we find that the effect of the peer treatment is large and significant for the second quartile. The effect of the moral treatment is significantly different from zero and positive for the fourth quartile in all three samples. We thus conclude that the peer treatment seems to be most effective for tax payers that owe small amounts, while the moral treatment is more effective for those who owe larger amounts. The Threat treatment appears to have a negative effect on tax payers that owe larger amounts.

Table 6 shows the marginal predictions for the probability that properties in each treatment group enter the fold for each quartile. These values represent odds ratios predicted near the center of the control variables—specifically, the sample average levels of all dummy variables and the median of the continuous regressors (land area and

Table 5: Logistic Regressions – Ever Paid

	Full Sample	Non-Commercial	Sole Owner
Balance Q2	-0.46*	-0.52*	-0.33
	(0.21)	(0.22)	(0.24)
Balance Q3	-1.03***	-0.97***	-1.54***
	(0.24)	(0.24)	(0.30)
Balance Q4	-1.25***	-1.15***	-1.36***
	(0.30)	(0.30)	(0.33)
Moral	-0.30	-0.34	-0.34
	(0.18)	(0.19)	(0.20)
Moral*Balance Q2	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Moral*Balance Q3	0.06	0.08	-0.11
	(0.16)	(0.16)	(0.18)
Moral*Balance Q4	0.40*	0.42*	0.33
	(0.20)	(0.20)	(0.21)
Peer	0.16	0.13	0.21
	(0.19)	(0.19)	(0.21)
Peer*Balance Q2	0.54**	0.58***	0.58**
	(0.17)	(0.17)	(0.18)
Peer*Balance Q3	-0.08	-0.07	-0.19
	(0.17)	(0.17)	(0.18)
Peer*Balance Q4	-0.49**	-0.45*	-0.61**
	(0.17)	(0.17)	(0.19)
Threat	-0.05	-0.01	-0.13
	(0.26)	(0.26)	(0.29)
Threat*Balance Q2	0.07	0.10	0.11
	(0.16)	(0.16)	(0.17)
Threat*Balance Q3	-0.50**	-0.49**	-0.64***
	(0.17)	(0.18)	(0.19)
Threat*Balance Q4	-0.04	-0.03	-0.20
	(0.16)	(0.16)	(0.17)
Log Likelihood	-2010.55	24 -1948.32	-1639.28
Num. obs.	4927	4749	3888

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Control coefficients omitted for brevity; see Appendix.

Table 6: Marginal Predictions – Ever Paid

	Q1	Q2	Q3	Q4
Control	23.40	16.10	9.80	8.00
Moral	18.50	12.70	12.10	11.40
Peer	26.40	15.20	14.40	8.20
Threat	22.40	12.20	13.40	7.10

market value). Considering the first quartile, the table shows us that properties given the Peer treatment were about 3 percentage points more likely to have contributed something to DoR prior to the sample’s conclusion than those who received the control treatment.

Next, we examine a more restrictive yes-no participation outcome—namely, whether or not the property offered not just perhaps token repayment in our sample period, but whether its debts were paid back in full.¹⁸ The ever-paid outcome used above does not differentiate between tax payers that made full payments and tax-payers that made partial payments. Figure 11 shows the time series for this outcome for the four treatment groups. The overall patterns are similar as in the previous outcome. The peer conformity treatment outpaces all other treatments after two weeks.

Again we can formalize the analysis using Logit models. Table 7 summarizes the results from Logit regressions that use an indicator for whether the Department of Revenue received full payment by the end of the sample period as the outcome variable. We report the estimates from the three samples.

¹⁸Due to some measurement issues, it is not possible to track on a day-to-day basis exactly the balance due for each property—accrual of interest and other charges is hard to pinpoint exactly. In the main results below, we actually measure full repayment as submission of at least 95% of the balance due.

Table 7: Logistic Regressions – Paid in Full

	Full Sample	Non-Commercial	Sole Owner
Intercept	-2.23*** (0.10)	-2.22*** (0.10)	-2.29*** (0.12)
Moral	-0.42** (0.13)	-0.44** (0.14)	-0.29 (0.15)
Peer	0.24 (0.14)	0.24 (0.14)	0.41** (0.16)
Threat	-0.21 (0.19)	-0.18 (0.20)	-0.04 (0.21)
Log Likelihood	-1435.15	-1395.06	-1175.05
Num. obs.	4927	4749	3888

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

We find positive effects for the peer treatment in all three samples. The effects are significantly different from zero in the sole owner sample. Moreover, we find negative effects for the moral treatment.

Moving on, we investigate whether there is heterogeneity in response to the treatment. Figure 12 plots the time series for the three treatment groups and the control group by quartile of the outstanding balance. Figure 12 reinforces our finding that response rates in all four subsamples declines rather dramatically across quartiles of the balance distribution. In the first quartile the response rate is almost 30 percent by the end of the sample. In the highest quartile, the response rate is typically less than 2 percent. The plots are also consistent with our previous finding that the peer treatment is more effective for delinquent tax payers in the first and second quartile.

Again we formalize the analysis by estimating Logit models with interactions. Table 8 summarizes the estimates and their standard errors for the three samples that we considered above.

Table 8 shows a very steep decline in the willingness to pay the tax bills as the amount owed increases. These effects are even larger than in the previous model, which suggests that many delinquent tax payers in the higher quartiles only make partial payments when they respond to the letters. The estimates of the treatment effects are similar to the ever-paid outcome. Again we find some evidence that the peer treatment works for tax payers in the lower quartiles while the moral treatment appeals to those in the highest quartile. The treatment seems to be counterproductive for delinquent tax payers in the third quartile. The findings are reinforced by the marginal predictions reported in Table 9.

To evaluate the overall impact of the different notification strategies, we would also like to know how much revenue can be raised by the different approaches. Figure

Table 8: Logistic Regressions – Paid in Full

	Full Sample	Non-Commercial	Sole Owner
Balance Q2	-1.28*** (0.27)	-1.42*** (0.28)	-1.35*** (0.31)
Balance Q3	-2.32*** (0.39)	-2.18*** (0.37)	-3.18*** (0.61)
Balance Q4	-3.27*** (0.74)	-2.85*** (0.61)	-3.83*** (1.03)
Moral	-0.45* (0.19)	-0.49* (0.20)	-0.49* (0.22)
Moral*Balance Q2	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Moral*Balance Q3	-0.27 (0.23)	-0.23 (0.24)	-0.54* (0.26)
Moral*Balance Q4	0.61* (0.26)	0.65* (0.26)	0.49 (0.27)
Peer	0.25 (0.19)	0.22 (0.19)	0.29 (0.22)
Peer*Balance Q2	1.01* (0.42)	0.99* (0.42)	1.06* (0.46)
Peer*Balance Q3	-0.18 (0.25)	-0.13 (0.25)	-0.33 (0.26)
Peer*Balance Q4	-0.39 (0.23)	-0.32 (0.23)	-0.54* (0.25)
Threat	-0.07 (0.27)	-0.03 (0.27)	-0.05 (0.29)
Threat*Balance Q2	0.21 (0.22)	0.25 (0.22)	0.13 (0.24)
Threat*Balance Q3	-0.64** (0.25)	-0.65** (0.25)	-0.80** (0.27)
Threat*Balance Q4	0.20 (0.22)	0.22 (0.22)	-0.03 (0.24)
Log Likelihood	-1150.17	28 -1120.45	-919.68
Num. obs.	4927	4749	3888

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Control coefficients omitted for brevity; see Appendix.

Table 9: Marginal Predictions of Logistic Regressions – Paid in Full

	Q1	Q2	Q3	Q4
Control	19.90	6.40	2.40	0.90
Moral	13.60	4.00	2.90	1.20
Peer	24.10	6.90	3.10	1.80
Threat	18.80	3.50	4.20	0.90

13 depicts the trajectories of repayments by each group, normalized by the number of properties in the group. Here we find that all three treatments increased the cumulative average payment, as initially noted in Table 3. Moreover, the moral appeal treatment caused the largest increase of payment. This result is confirmed in Figure 13, which depicts the trajectories of the percentage of mailing day debt owed to the city by each group. As discussed in detail above, the key finding that the moral treatment raised more revenue than the other treatments or the control is largely due to a small number of delinquent tax payers that owed large amounts of money. The sample size is not large enough to find significant effects. This can be shown by estimating a variety of regressions and Tobit models intended to predict the magnitude of the received payment. We do not report these results here since we did not find any significant treatment effects.

6 Conclusions

This field experiment evaluated a set of different notification strategies intended to increase property tax compliance. We tested three of the most commonly suggested models of tax compliance: deterrence, moral appeal, and peer conformity. We have implemented the experiment in collaboration with the Philadelphia Department of

Revenue (DoR). Our findings provide some moderate evidence that both moral appeal and peer conformity may improve tax compliance. We find little evidence that supports the standard deterrence model compared to a traditional simple bill notification.

Our study provides ample scope for future research. Unlike several recent papers (Kleven et al. 2011; Slemrod, Blumenthal, and Christian 2001), which have found large increases in compliance after providing information about the threat of auditing, we find no evidence of a deterrence effect. It is probably not surprising that our deterrence treatment was not effective. Philadelphia is city with a history of high property tax delinquency. It is difficult to alter perceptions of beliefs of punishment by sending one letter in a city with already-high property tax delinquency. Any threats may be considered to be empty given previous attempts at collection and the lack of enforcement penalties. It would be more interesting to design an intervention that is based on a more credible threat. The peer conformity treatment is also subject to the same potential criticism. It may be more effective to randomly print a number of delinquent tax payers in the local news paper or to send mail informing delinquents' neighbors of their malfeasance. Of course, these types of intervention face much larger legal hurdles and are therefore much harder to implement. They may also backfire and increase tax noncompliance by encouraging defiance of the law (Sherman 1993).

Consistent with other recent tax compliance experiments (Fellner, Sausgruber, and Traxler 2013), we also find that the largest effect is observed for the subset of taxpayers at the margin of compliance. Beyond this traditional perspective on achieving higher rates of legal compliance through modifications to legal threats, others have examined how non-external factors contribute to tax compliance. Similar to other studies we find that providing social information about tax compliance provides some marginal increase in collection (Wenzel and Taylor 2004; Wenzel 2005; Hallsworth et al. 2014).

While this finding is by no means universal (Fellner, Sausgruber, and Traxler 2013), our results suggest that providing social norm information is noticeably more effective than either deterrent or moral persuasion messages.

References

Ariel, Barak (2012), "Deterrence and moral persuasion effects on corporate tax compliance: Findings from a randomized controlled trial." *Criminology*, 50 (1), 27-69.

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

Alm, James, Gary H. McClelland, and William D. Schulze (1992), "Why Do People Pay Taxes?" *Journal of Public Economics* 48: 21-38.

Alm, James (1999), "Tax compliance and administration." In: Hildreth, W. Bartley and James A. Richardson (eds.) *Handbook on Taxation*. New York, USA, Marcel Dekker, Inc., pp. 741-768.

Andreoni, James, Erard, Brian and Jonathan Feinstein (1998), "Tax compliance." *Journal of Economic Literature*, 36, 818-860.

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

Bernheim, B. Douglas (1994), "A Theory of Conformity." *Journal of Political Economy*, 102, 5, 841-877.

Blumenthal, Marsha, Christian, Charles and Joel Slemrod (2001), "Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota." *National Tax Journal*, 54 (1), 125 - 138.

Cowell, Frank A. and James P. F. Gordon (1988), “ Unwillingness to pay tax: tax evasion and public provision.” *Journal of Public Economics*, 36, 305-321.

Fehr, Ernst and Simon Gächter (1998), “Reciprocity and economics: The economic implications of homo reciprocans.” *European Economic Review* 42 (3-5), 845-59.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler (2013), “Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information.” *Journal of the European Economic Association* 11, 3, 634-60.

Frey, Bruno S., and Lars P. Feld (2002), “Deterrence and Morale in Taxation: An Empirical Analysis.” CESifo Working Paper no. 760, August 2002.

Hallsworth, Michael., John List, Robert Metcalfe and Ivo Vlaev (2014), “The Behavioralist as Tax Collector,” Using Natural Field Experiments to Enhance Tax Compliance.” NBER Working Paper 20007.

Harrison, Glenn W. and John A. List (2004), “Field Experiments.” *Journal of Economic Literature*, 42 (4), 1009-1055.

Kleven, Henrik J., Knudsen, Martin B., Kreiner, Claus T., Pedersen, Soren and Emmanuel Saez (2011), “Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark.” *Econometrica*, 79 (3), 651-692.

Pew Charitable Trust (2013), “Delinquent Property Tax in Philadelphia.” Technical Report.

Pomeranz, Dina (2013), "No taxation without information: Deterrence and self-enforcement in the Value Added Tax." Harvard Business School Working Paper.

Reckers, Philip M. J., Sanders, Debra L. and Stephen J. Roark (1994), "The influence of ethical attitudes on taxpayer compliance." *National Tax Journal*, 47 (4), 825-836.

Sherman, Lawrence (1993), "Defiance, deterrence, and irrelevance: A theory of the criminal sanction." *Journal of Research in Crime and Delinquency*, 30, 445-473.

Slemrod, Joel (2007), "Cheating ourselves: The economics of tax evasion." *Journal of Economic Perspectives*, 21 (1), 25-48.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian (2001), "Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota." *Journal of Public Economics* 79, 3, 455-83.

Torgler, Benno (2002), "Moral-suasion: An alternative tax policy strategy? Evidence from a controlled field experiment in Switzerland." *Economics of Governance* 5 (3), 235-253.

Torgler, Benno (2012), "A field experiment on moral-suasion and tax compliance focusing on under-declaration and over-deduction." QUT School of Economics and Finance Working Paper no. 285.

Wenzel, Michael (2005), "Misperceptions of social Norms about Tax Compliance:

From Theory to Intervention.” *Journal of Economic Psychology* 26, 6, 862-83

Wenzel, Michael and Natalie Taylor (2004), “An experimental evaluation of tax-reporting schedules: a case of evidence-based tax administration.” *Journal of Public Economics*, 88 (12), 2785-2799.

Figure 11:

Percent of Group Having Paid in Full By Empirical Treatment Group

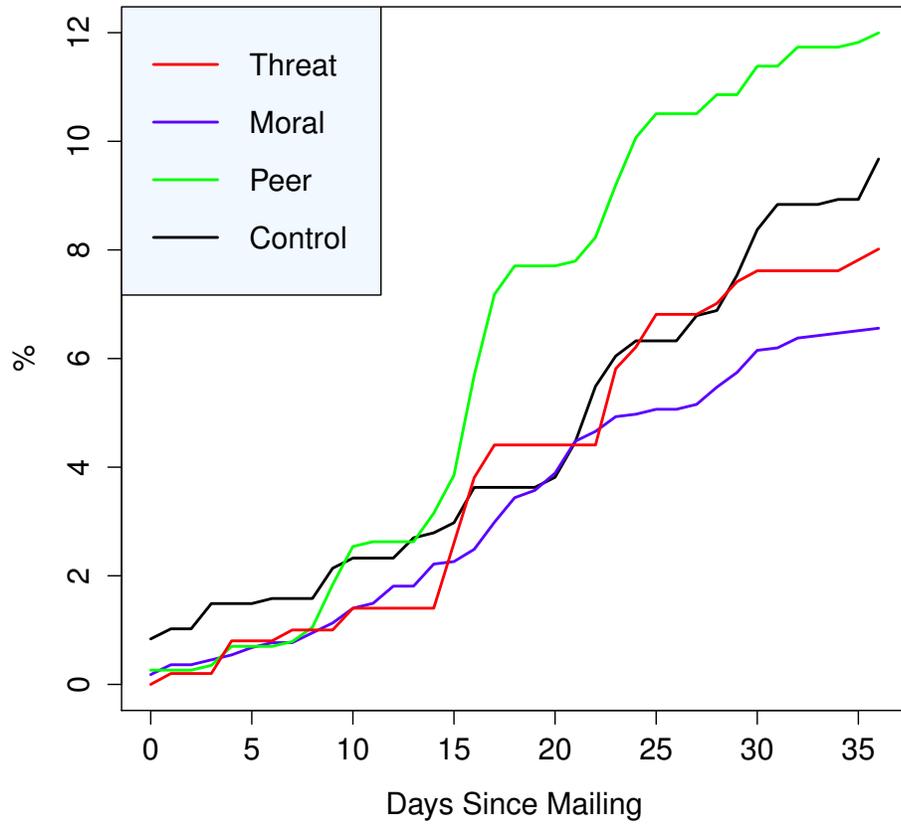


Figure 12:

Percent of Group Ever Having Paid
By Balance Quartile, Empirical Treatment

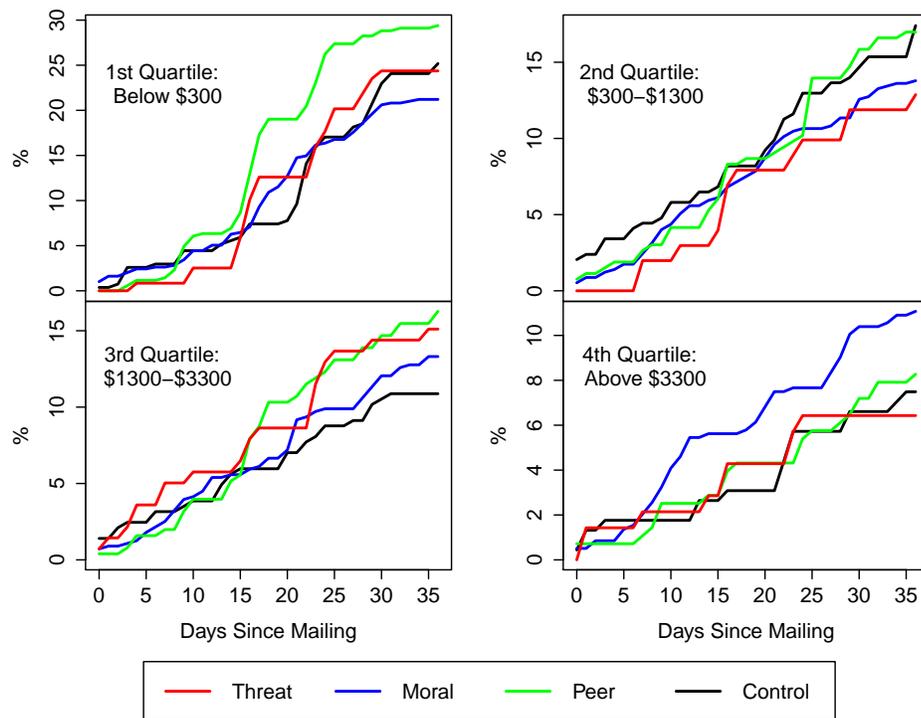


Figure 13:

Cumulative Average Payments Since Mail Date By Empirical Treatment Group

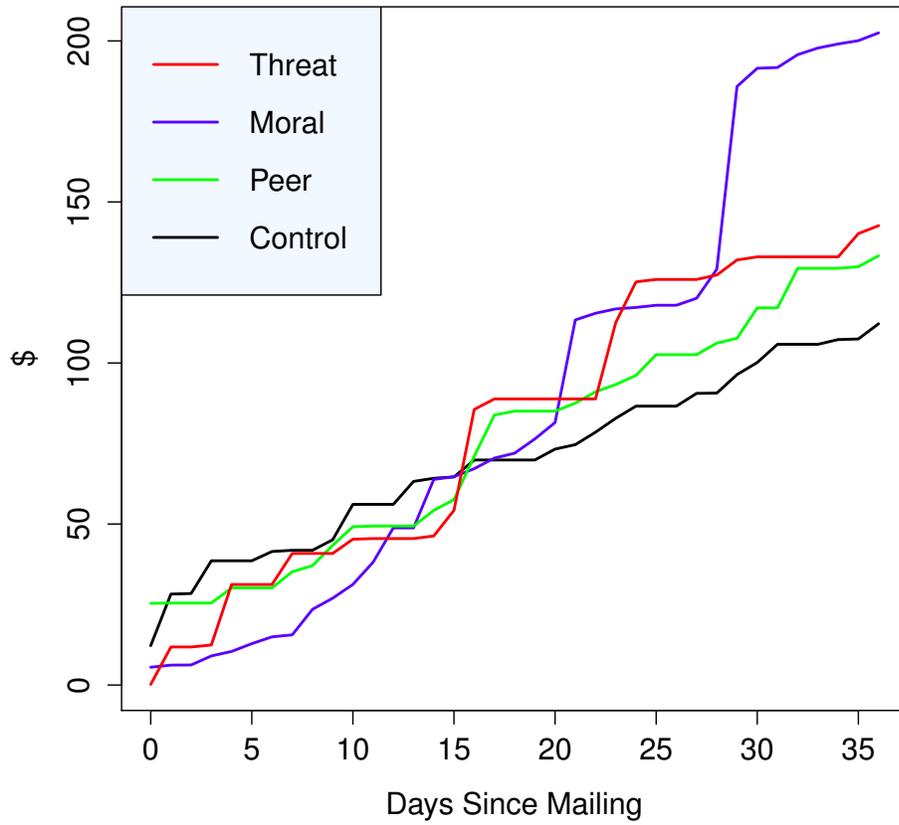


Figure 14:

Percentage of Mailing Day Debt Owed By Empirical Treatment Group

