American Economic Journal: Economic Policy 2022, 14(4): 343–370
https://doi.org/10.1257/pol.20200400

Reminders Work, but for Whom? Evidence from New York City Parking Ticket Recipients†

By Ori Heffetz, Ted O'Donoghue, and Henry S. Schneider*

We investigate heterogeneity in responsiveness to reminder letters among New York City parking ticket recipients. Using variation in the timing of letters, we find a strong aggregate response. But we find large differences across individuals: those with a low baseline propensity to respond to tickets—a natural nudge target—react least to letters. These low-response types, who incur significant late penalties, disproportionately come from already disadvantaged groups. They do react strongly to traditional, incentive-based interventions. We discuss how accounting for response heterogeneity might change one's approach to policy and how one might use our analysis to target interventions at low-response types. (*JEL* D04, D12, D91, H71)

There are many tasks that policymakers may want people to complete: paying taxes, bills, and loan installments; making child support payments; engaging in preventative health care; applying for government benefits; responding to public surveys (e.g., the census); and voting. Unfortunately, people do not always carry out these tasks, raising the question of what types of interventions might increase compliance. Traditionally, economists have focused on interventions that alter the real costs and benefits of compliance, such as a monetary penalty or a forgoing of potential benefits. Over the past decade, some economists have focused instead on "nudge" interventions (Thaler and Sunstein 2008) designed to influence behavior in a less intrusive manner. However, research is still in the early stages of understanding heterogeneity in the responses to nudges and how best to target them.

* Heffetz: Bogen Family Department of Economics and Federmann Center for Rationality, Hebrew University of Jerusalem and S. C. Johnson Graduate School of Management, Cornell University (email: oh33@cornell.edu); O'Donoghue: Department of Economics, Cornell University (email: edo1@cornell.edu); Schneider: Stephen J. R. Smith School of Business, Queen's University (email: henry.schneider@queensu.ca). Dan Silverman was coeditor for this article. This paper previously circulated under the title "Forgetting and Heterogeneity in Task Delay: Evidence from New York City Parking-Ticket Recipients." We thank Samuel Bufter, David Frankel, Christopher Quinn, and Andrew Salkin at the New York City Department of Finance and Janet Capuano at CGI Group for access to the data and many helpful conversations. We have benefitted from valuable comments from Amr Farahat and seminar participants at Ben Gurion, Cornell, Harvard, Hebrew U, IDC Herzliya, UCLA Anderson, UCSD Rady, USC, Wharton, Williams, Stanford, SITE 2014, and BEAM 2015. We also thank Maya Catabi, Ofer Glicksohn, Lawrence Jin, Lev Maresca, Yotam Peterfreund, and Lin Xu for excellent research assistance. The authors have no financial or other material interests related to this research to disclose. Online Appendices are available on the authors' websites. This research has been approved by the Institutional Review Boards of Cornell University (Protocol ID#: 1507005700) and Queen's University (GREB TRAQ #: 6020858).
† Go to https://doi.org/10.1257/pol.20200400 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

In this paper, we investigate for whom nudges work in one particular environment: people's responses to parking tickets. Using administrative data provided by the New York City Department of Finance (henceforth, DOF), we analyze response behavior associated with parking tickets issued in New York City between June 2011 and August 2013. Our core dataset consists of 6.6 million tickets issued to 2 million unique passenger vehicle license plates, totaling $424 million in fines and $85 million in late penalties. The nudges are notification letters, which appear to serve primarily as reminders to respond. We identify the impact of these nudges by exploiting exogenous variation in the timing of the letters, and we identify heterogeneity across people in how they respond by exploiting the fact that we observe the same individuals responding to multiple tickets over a two-year period.

Our paper makes four main contributions. First, we find that notification letters have a significant positive impact on aggregate response behavior—consistent with people forgetting about their tickets and letters serving as reminders. Second, we find clear evidence of persistent types that differ in their baseline propensity to respond to tickets and, moreover, that respond differently to notification letters. In particular, those with a low baseline propensity to respond to tickets—arguably the natural target for interventions—react the least to reminders. Third, we find that these low-response types, who incur significant late penalties, disproportionately come from already disadvantaged groups. Finally, we show that the low-response types do, in fact, react strongly to more traditional, incentive-based interventions. Based on these findings, we discuss how accounting for response heterogeneity might change one's approach to policy and how one might use our analysis to target interventions at low-response types.

In Section I, we describe our data and setting. After receiving a ticket on day 0, the recipient faces a series of three deadlines by which to respond (by either paying or contesting the ticket). These occur at day 30, the first Monday after day 61, and the first Friday after day 100, with escalating and additive late penalties of $10, $20, and $30. If the third deadline is missed, DOF additionally enters a default judgment in court against the plate owner, after which more serious actions might be taken, including towing or booting the vehicle.

The windshield ticket clearly states the first deadline and late penalty. At various later times, the ticket recipient receives notification letters from DOF to keep her informed of her current situation and to specify updated deadlines and penalties. The key policy variation in our data occurred on June 18, 2012, about one year into our two years of data. On that date, DOF changed the timing of the first notification letter from roughly day 40 to roughly day 20—which we label a shift from the OLD regime to the NEW regime. Our primary identification of the impact of letters exploits this variation.

In Section II, we analyze aggregate response behavior in the OLD and NEW regimes. Figure 1 depicts daily hazard rates of recipients' first responses as well as cumulative response rates. The horizontal axis indicates the number of days since the ticket was issued, and the three deadlines are highlighted by the vertical shaded bands. Figure 1 shows a striking impact of notification letters on aggregate responses. Relative to the OLD regime, where no letters are received prior to day 40, under the NEW regime, there is a dramatic increase in hazard rates following the

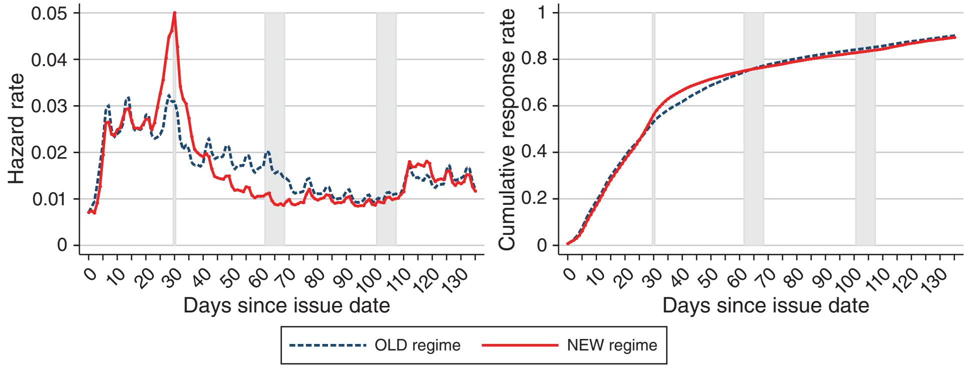

Figure 1. Response Rates in the OLD versus NEW Regimes

Notes: The figure depicts daily hazard rates (# first response on day d/# no response before day d) and cumulative response rates (# first response on or before day d/total # of tickets issued) in the OLD versus NEW regimes. All tickets have a first deadline at day 30, a second deadline at days 62–68, and a third deadline at days 101–107, indicated by the shaded areas. The latter two deadlines are a range because they depend on ticket-issuance day of the week. The first notification letter is received around day 40 in the OLD regime and around day 20 in the NEW regime. Based on 3,355,094 and 3,020,357 observations in the OLD and NEW regimes, respectively.

receipt of the letter at day 20. Quantitatively, the net hazard rate between day 20 and day 40 is 10 percentage points higher in the NEW regime relative to the OLD regime (46 percent versus 36 percent). Analogously, relative to the NEW regime, where no letters are received between day 20 and day 75, under the OLD regime, hazard rates increase following the receipt of the letter at day 40 by roughly the same magnitude.

To investigate the mechanism behind this strong aggregate response to notification letters, we worked with DOF on a field experiment that was implemented over five weeks for tickets issued July 13, 2013 through August 16, 2013. The recipient of each ticket was assigned to receive one of four versions of the first letter (at roughly day 20) that vary in their content and to receive or not receive an additional letter at roughly day 48 (a 4×2 experimental design). The findings—that the content of the first letter hardly matters and that the second letter, which contains no new information, generates an additional response—suggest that the letters served primarily as reminders.

Our demonstration of a significant aggregate response to a nudge is consistent with prior research (see below); however, the heart of our paper is an analysis of heterogeneity, to which we turn in Section III. For most ticket recipients in our data, we have little information on observable characteristics, and thus our primary analysis does not focus on heterogeneity based on observables. Instead, our analysis focuses on identifying unobserved persistent types by exploiting the fact that we observe the same individual responding to multiple tickets. As a proof of concept, we first demonstrate that even crude measures of response behavior on one's past tickets are highly predictive of response behavior on one's current ticket (Figure 4). We then pursue a more rigorous analysis in which we estimate a mixture model of unobserved types. We represent a type by a set of regime-specific hazard rates. We estimate each type's hazard rates jointly with the population distribution of types, allowing for up to four types.

Our heterogeneity analysis yields several findings. First, our estimated mixture model implies dramatic differences across types in their baseline propensity to respond to tickets: in our estimated three-type model, for instance, implied cumulative responses in the OLD regime by the time the first letter is sent (on roughly day 40) are 93 percent for high-response types, 60 percent for medium-response types, and 19 percent for low-response types. Moreover, the aggregate response to letters in Figure 1 masks enormous differences across types. The switch from the OLD to the NEW regime increases the net hazard rate between days 20 and 40 by 15 percentage points (from 73 to 88 percent) for the high-response types and 12 percentage points (from 47 to 59 percent) for the medium-response types, but only 1 percentage point (from 10 percent to 11 percent) for the low-response types. Hence, the economic impact of letters in our domain is far larger for the high- and medium-response types than it is for the low-response types.

Second, the low-response types, who incur significant late penalties, seem to disproportionately come from already disadvantaged groups. Specifically, for a subset of our data, we observe ticket recipients' addresses, which we match to census variables. Doing so, we find that the low-response types are more likely to reside in Census Block Groups that have lower income, less education, and higher proportions of "Black" or "other" racial groups.[1]

Third, the low-response types, who react little to reminders, in fact respond strongly to incentive-based interventions. In particular, they react to a combination of (i) a letter they receive shortly after the third deadline (at roughly day 110) informing them that their vehicle is now subject to the possibility of towing or booting and (ii) the actual towing or booting that occurs in the weeks that follow. The low-response types exhibit their largest hazard rates immediately after receiving this letter—even before any significant booting occurs.

Our analysis is relevant for the growing literature on nudges following Thaler and Sunstein (2008). Most closely related is the literature on reminders for task completion. Several papers demonstrate an impact of reminders in a field setting: for instance, Calzolari and Nardotto (2017) study the impact of weekly emails reminding gym members to use the gym; Cadena and Schoar (2011) and Karlan, Morten, and Zinman (2016) study the impact of regular text messages to remind people to make payments on their installment loans; Karlan et al. (2016) study the impact of monthly text messages reminding people to make their planned deposits into commitment savings accounts; and Chirico et al. (2019) study the impact of letters reminding property owners who were tardy in paying their property taxes to pay those taxes.[2,3]

[1] Ghesla, Grieder, and Schubert (2020) analyze the differential impact on different socioeconomic groups for a different type of nudge: the use of choice defaults to induce households to choose more "green" electricity contracts. In their case, the nudge seems to reduce welfare for poorer households.

[2] See also Taubinsky (2014) and Tasoff and Letzler (2014), who study the impact of reminders in getting people to complete experimental tasks.

[3] Our analysis is also related to the literature on delay in task completion due to present bias (Akerlof 1991; O'Donoghue and Rabin 1999a,b, 2001) and forgetting about (or inattention to) tasks (Holman and Zaidi 2010; Ericson 2011; Taubinsky 2014). Unfortunately, the nature of our data does not permit us to investigate the impact of each mechanism in this domain (Heffetz, O'Donoghue, and Schneider 2022b). See also recent work on the

However, none of this research on reminders involves the type of heterogeneity analysis that we focus on—typing individuals based on behavior rather than observable sociodemographic characteristics. Indeed, our heterogeneity analysis raises a key question for assessing the impact of reminders and other nudges: are they targeted effectively? In our domain, the impact of reminders is an order of magnitude weaker for those who delay the most. Moreover, in our concluding Section IV, we calculate that in the switch from the OLD to the NEW regime, 91 percent of the extra spending on earlier notification letters and 86 percent of the gains in terms of reduced penalties accrue to the higher-response types—instead of to the low-response types that arguably ought to be the target of the nudge. We then discuss an alternative regime with notification letters that would target the population of low-response types and show that such targeting could be relatively easy to implement based on crude measures of past behavior. This discussion further highlights the importance of analyzing heterogeneity in nudge effects prior to giving policy advice.[4]

For other types of nudges besides reminders, there is also relatively little focus on heterogeneous treatment effects. Indeed, three recent papers synthesize and analyze nudge interventions from a large number of past studies (Benartzi et al. 2017; Hummel and Maedche 2019; DellaVigna and Linos 2022), and all three focus exclusively on the aggregate response to each nudge without discussing the possibility of heterogeneous treatment effects. There are some recent papers (see our discussion in Section IV) that use large administrative datasets to study how the impact of nonreminder nudges depend on sociodemographic observables. Relative to these, our analysis highlights how people's own past behavior might be a particularly powerful predictor of the impact of nudges, as we discuss in Section IV.

I. New York City Parking Tickets

A. *Data Description*

The data (New York City Department of Finance 2014; Heffetz, O'Donoghue, and Schneider 2022a) come from the New York City DOF, which handles most city revenue. Online Appendix 1 contains a detailed description of our data; here we summarize the most important details. The full dataset contains information on 20,874,688 tickets, covering virtually all tickets issued between June 1, 2011, and August 31, 2013. The data include ticket issue date, violation type, fine amount, issuing agency, and other details. In addition, the data allow us to construct each ticket's history of "events" through late January 2014. Events are actions taken by either the ticket recipient (e.g., making a payment or contesting the ticket) or DOF (e.g., imposing a late penalty or sending a notification letter).

The *core dataset* that we analyze is comprised of the 6,646,540 tickets that satisfy various restrictions. The full set of restrictions is described in online Appendix

difficulty of identifying present bias from data on task completion (Martinez, Meier, and Sprenger 2017; Heidhues and Strack 2021).

[4] It is worth noting that our policy discussion does not assume that the low baseline response rates of the low-response types are suboptimal. Rather, it points out how an alternative policy suggested by our analysis might lead to more timely payments from low-response types, without imposing larger penalties on them.

1; the vast majority of excluded tickets are excluded due to one or more of three criteria: (i) they are issued to a commercial vehicle or a vehicle that is part of a fleet program (32 percent of the full dataset), (ii) they are not issued for parking violations (another 14 percent), and/or (iii) the plate owner does not have DOF's highest address verification level (another 21 percent). The core dataset only includes tickets issued to New York plates.[5]

The first column of Table 1 presents descriptive statistics for the core dataset. The most common violations are for expired parking meter (36 percent), no parking due to street cleaning (26 percent), and parking in a general no-parking zone (9 percent). The most common fine amounts are \$35 (30 percent), \$45 (24 percent), and \$115 (23 percent). The vast majority (97 percent) of tickets are issued by parking ticket agents. The bottom panel of Table 1 presents the distribution of payment types for the 80 percent of tickets in the core dataset that have payments made by day 135. Four payment methods are available: online (54 percent), by mail (32 percent), by phone (3 percent), and in person at 1 of 5 DOF business centers (11 percent).

B. *The OLD and NEW Regimes*

Figure 2 summarizes the timelines of key events under the OLD and NEW regimes (and also under the EXP regime, which we describe in Section IIB). These timelines are identical except for one thing: DOF changed the timing and content of the first notification letter. The rest of this section provides detail.[6]

Timeline in the OLD Regime.—Tickets are issued on (what we define as) day 0. The ticket and an envelope are placed on the windshield of the offending car and together indicate the violation type, fine amount, the (first) due date of day 30, the (first) late-payment penalty amount of \$10, and information on how to pay or contest. They also mention that failure to respond may result in additional penalties and a default judgment being entered, after which the vehicle may be towed. Online Appendix 12 contains sample tickets and the relevant part of the envelope, as well as samples of all notification letters described below.

If there is no response by the first deadline, DOF mails a notification letter to the plate owner (OLD letter 1) on the Tuesday that is day 35–41. This letter, titled "NOTICE OF OUTSTANDING VIOLATION," shows an updated balance due that includes the \$10 late penalty. It also provides a new due date, the Monday that is 27 days after that Tuesday (day 62–68), and states that failure to respond in time will

[5] Ex ante, we chose to include vehicles with only the highest verification level to maximize the probability that the address DOF has on file matches that of the driver of the violating vehicle. This criterion turns out to exclude essentially all summonses for vehicles with out-of-state plates while including 88 percent of summonses for passenger vehicles with New York State plates. Excluding out-of-state vehicles has the added benefit of guaranteeing that all included ticket recipients have a similar relationship with the governmental authority issuing the ticket. At the same time, because this criterion excludes only 12 percent of New York vehicles, there is limited scope for any sample selection effects.

[6] In this section, we describe deadlines and penalties as they are presented to plate owners, which seem a good proxy for people's perceptions of those deadlines and penalties. In practice, they were implemented in a slightly different way—see online Appendix 1 for details.

Table 1—Descriptive Statistics

	Core dataset	OLD regime	NEW regime	EXP regime
Total number of tickets	6,646,540	3,355,094	3,020,357	271,089
Violation type				
Expired meter ($35/$65)	36.23%	37.52%	34.92%	34.88%
Street cleaning ($65/$45)	26.18%	25.38%	27.01%	26.88%
General no-parking zone ($65/$60)	9.21%	9.27%	9.14%	9.28%
General no-standing zone ($115)	6.70%	6.58%	6.78%	7.24%
Fire hydrant ($115)	5.59%	5.24%	5.95%	5.78%
Double parking ($115)	4.75%	4.91%	4.63%	4.00%
Bus stop ($115)	2.40%	2.30%	2.50%	2.39%
Truck loading/unloading ($95)	2.17%	2.09%	2.24%	2.22%
Authorized vehicles only ($95/$65/$60)	1.94%	2.04%	1.85%	1.73%
In commercial zone ($115)	1.35%	1.25%	1.40%	2.09%
In crosswalk ($115)	1.02%	0.90%	1.15%	1.10%
On sidewalk ($115)	0.68%	0.70%	0.66%	0.64%
Parking longer than limit ($65/$60)	0.37%	0.43%	0.32%	0.23%
In a driveway ($95)	0.30%	0.30%	0.30%	0.30%
Not as marked ($65)	0.23%	0.23%	0.22%	0.30%
In pedestrian ramp ($165)	0.22%	0.20%	0.25%	0.27%
In a safety zone ($115)	0.22%	0.20%	0.24%	0.23%
In a bike lane ($115)	0.17%	0.16%	0.18%	0.17%
No standing/taxi stand ($115)	0.14%	0.13%	0.15%	0.16%
In handicapped zone ($180)	0.13%	0.16%	0.11%	0.11%
Ticket amount				
$35	30.11%	31.30%	28.91%	28.70%
$45	23.89%	23.12%	24.69%	24.47%
$60	8.20%	8.25%	8.15%	8.29%
$65	10.45%	10.64%	10.23%	10.49%
$95	3.97%	3.95%	4.00%	3.87%
$115	23.00%	22.36%	23.64%	23.80%
$165	0.22%	0.20%	0.25%	0.27%
$180	0.13%	0.16%	0.11%	0.11%
Other/missing	0.02%	0.02%	0.01%	0.01%
Ticket issuer				
Parking ticket agent	97.16%	97.28%	97.03%	96.98%
New York City police department	2.84%	2.72%	2.97%	3.02%
Payment type				
Payment made by day 135	5,333,147	2,721,947	2,397,666	213,534
Mail	32.34%	33.50%	31.23%	29.94%
Online	53.81%	51.11%	56.55%	57.48%
Phone	2.76%	2.10%	3.36%	4.33%
In person	11.09%	13.28%	8.85%	8.25%
Unknown	0.00%	0.00%	0.00%	0.00%

Notes: The primary analysis uses the core dataset of tickets issued to passenger vehicles for which the plate owner has DOF's highest address verification level. The OLD regime applies to tickets issued June 1, 2011 through June 17, 2012; the NEW regime applies to tickets issued June 18, 2012 through July 12, 2013, and August 17, 2013 through August 31, 2013; the EXP regime applies to tickets issued July 13, 2013, through August 16, 2013. For all but payment type, percentages are relative to the total number of tickets in that regime (listed in the first line). For payment type, percentages are relative to the number of tickets with payment received by day 135.

result in an additional late penalty of $20 and can lead to a default judgment entry, after which various actions may be taken, including towing the owner's vehicle(s).

If there is still no response by the second deadline, DOF mails a second notification letter (letter 2) on the subsequent Tuesday (day 70–76). Letter 2 shows an updated balance that includes the second late penalty, and it provides a new due date, which is the Friday that is 31 days after that Tuesday (day 101–107). The plate owner is again warned that failure to respond can lead to a default judgment entry.

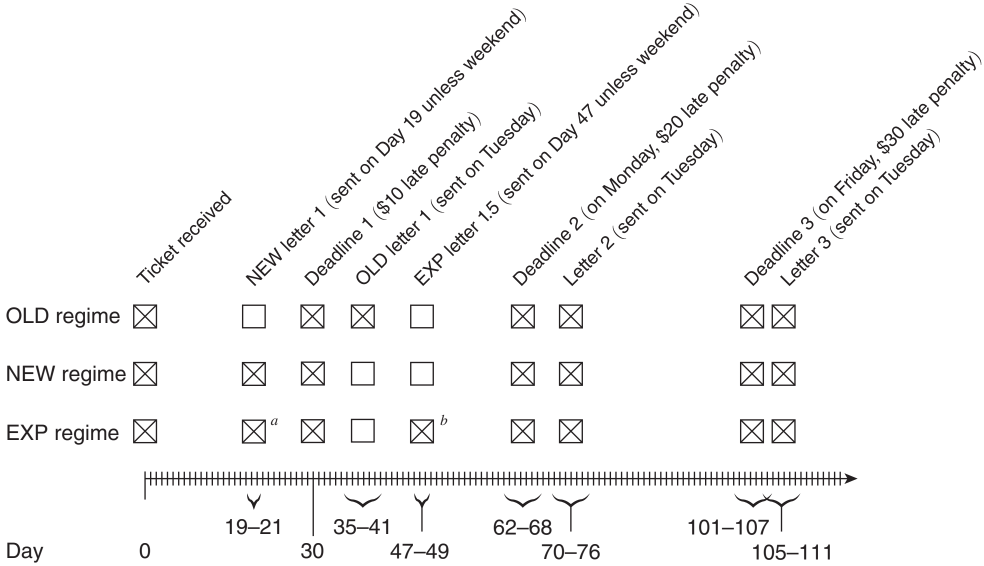

Figure 2. Timeline for Each Regime

Notes: The OLD regime applies to tickets issued June 1, 2011 through June 17, 2012; the NEW regime applies to tickets issued June 18, 2012 through July 12, 2013, and August 17, 2013 through August 31, 2013; the EXP regime applies to tickets issued July 13, 2013 through August 16, 2013. In the EXP regime, eligible tickets were randomized [a] to receive one of four versions of NEW letter 1 and [b] to receive or not receive EXP letter 1.5.

However, the letter does not explicitly mention the amount—$30—of the impending third late penalty.

If the plate owner misses the third deadline, DOF sends a third notification letter (letter 3) on the subsequent Tuesday (day 105–111). Letter 3 lists a deadline of "IMMEDIATELY." It further states that a default judgment has been entered and that the owner is now subject to immediate actions, including seizing any of the owner's vehicles.

Change to the NEW Regime.—Under the OLD regime, OLD letter 1 was sent shortly after the first deadline (on day 35–41). Beginning with tickets issued on June 18, 2012, DOF moved this letter to before the first deadline. Specifically, this letter (NEW letter 1) is generated on day 18 and sent on day 19, unless day 18 occurs on a weekend, in which case the letter is generated on the subsequent Monday (day 19 or 20) and sent on Tuesday (day 20 or 21).

In addition, while most of the content of NEW letter 1 is identical to that of OLD letter 1, DOF made three changes. First, the title is changed to "PRE-PENALTY NOTICE OF UNPAID VIOLATION." Second, instead of stating the second deadline (day 62–68) and the second penalty ($20), NEW letter 1 states the first deadline (day 30) and the first penalty ($10). Finally, unlike OLD letter 1, which mentions the possibility of a default judgment entry and uses a bold font to highlight various future penalties, NEW letter 1 does not mention default judgment and does not contain any bold font.

It is worth highlighting the difference in information between the OLD and NEW regimes. Specifically, under the OLD regime, the plate owner is informed about the second deadline and its \$20 penalty in OLD letter 1. Under the NEW regime, in contrast, the plate owner is not informed about the existence of the second deadline and its \$20 penalty until they receive letter 2 (sent on day 70–76)—i.e., not until after the second deadline has passed and the penalty has been imposed. Our field experiment was designed in part to test whether this difference in information might drive some of the observed behavioral differences across the two regimes. As discussed below, we conclude that it does not.[7]

Descriptive Statistics.—The OLD regime applies to tickets issued between June 1, 2011, and June 17, 2012. The NEW regime applies to tickets issued between June 18, 2012, and July 12, 2013, and between August 17, 2013, and August 31, 2013. Finally, tickets issued between July 13, 2013, and August 16, 2013, were part of a field experiment (the EXP regime, described in Section IIB). As Table 1 shows, the distributions of violation type, ticket amount, ticket issuer, and payment type are all similar across all three regimes. The main difference is that as time passes and we move from the OLD to NEW to EXP regime, there is a modest shift from in-person and mail payments to online and phone payments.

II. Aggregate Responses to Letters

A. *Aggregate Responses in the OLD versus NEW Regimes*

A person's first response to a ticket can be either a payment or a contest. Our analysis focuses on the timing of the first response, pooling the two response types together. See online Appendix 2 for the rationale behind this approach, along with descriptive statistics for type of first response.

We measure a person's first response to a ticket in number of days since issue date. We then analyze first responses using survival analysis. Each ticket is a single observation, and we estimate daily hazard rates by dividing, for each of days 0–135, the number of first responses on that day by the number of tickets with no first response before that day.

As described in the Introduction, Figure 1 depicts estimated hazard rates in the OLD and NEW regimes.[8] Prior to day 20, behavior is roughly the same under the two regimes, as expected given that there is not yet any differential treatment between regimes. Then, when NEW letter 1 hits in the NEW regime, hazard rates

[7] There are two additional idiosyncratic differences between the two regimes: (i) there was a settlement program in place for part of the OLD regime that was not in place during the NEW regime, and (ii) Hurricane Sandy occurred during the NEW regime. See online Appendix 1 for further details about each and online Appendix 3.2 for evidence that neither impacts our conclusions.

[8] Throughout, we provide graphical depictions of behavior without confidence bands (as in Figure 1) because those confidence bands are mostly indistinguishable from the depicted point estimates, and essentially any visible difference in our figures is statistically significant. Online Appendix 3.1 reproduces the major figures with 95 percent confidence bands.

Table 2—Responses Analyzed by Period

Definition of periods (same for both regimes)					
Period 1: from day 0 to the day NEW letter 1 is sent					
Period 2: from the day after NEW letter 1 is sent to deadline 1					
Period 3: from the day after deadline 1 to the day OLD letter 1 is sent					
Period 4: from the day after OLD letter 1 is sent to deadline 2					
Period 5: from the day after deadline 2 to the day letter 2 is sent					
Period 6: from the day after letter 2 is sent onward					
Start and end dates for each period (same for both regimes)					
Period	1	2	3	4	5
Start	day 0	day 20–22	31	day 36–42	day 63–69
End	day 19–21	30	day 35–41	day 62–68	day 70–76
Days in period	20–22	9–11	5–11	27	8
Average daily hazard rates					
Period	1	2	3	4	5
OLD	2.28%	2.69%	2.00%	1.86%	1.32%
NEW	2.17%	3.51%	2.88%	1.33%	0.90%
Cumulative response rates					
Period	1	2	3	4	5
OLD	37.63%	53.14%	60.17%	76.02%	78.44%
NEW	36.18%	56.18%	65.27%	75.79%	77.48%

Notes: The number of days in some periods is a range due to differences in the day of the week on which tickets were issued. Average daily hazard rates within each period and cumulative response rates after each period are calculated using weighted averages across the different days of the week. Online Appendix 4 contains these calculations.

are increasingly larger and more obviously spike at day 30 (deadline 1) relative to hazard rates in the OLD regime. Analogously, from roughly day 40, when OLD letter 1 hits in the OLD regime, hazard rates are larger than hazard rates in the NEW regime. After the second deadline, hazard rates converge again.[9]

The hazard rates in Figure 1 do not control for any fixed effects. A natural concern is a day-of-the-week effect, and indeed we show in online Appendix 3.3 that there is a weekly cycle in hazard rates, with lower hazard rates on weekends. Another concern is a day-of-the-month effect. To remove such effects, and as a means to quantify some of the differences seen in Figure 1, we next analyze behavior after partitioning days into six natural "periods" (we use these same periods when we estimate a mixture model in Section IIIB). Table 2 delineates the start and end dates for each period.[10]

[9] Online Appendix 3.2 demonstrates that these conclusions are robust to the settlement program and to Hurricane Sandy (see footnote 7) and further suggests that the small differences between the two regimes prior to NEW letter 1 and after the second letter are primarily an artifact of Hurricane Sandy.

[10] Online Appendix 3.3 provides an alternative approach to controlling for a day-of-the-week effect and shows that the main conclusions are robust to this alternative approach. Also, a third concern is a month-of-the-year effect. Given that we have only one year of data under each regime, combined with the fact that there are some idiosyncratic events that affect behavior in a few specific months (see online Appendix 3.2), we do not see a good way to control for such effects. Nonetheless, we think it unlikely that a month-of-the-year effect would alter the key patterns in Figure 1 and Table 2.

Table 2 presents estimated average daily hazard rates within each period and cumulative response rates through each period.[11] As in Figure 1, response rates prior to day 20 (in period 1) are roughly the same across the two regimes, but then NEW letter 1 leads to a dramatic increase in response rates relative to the OLD regime, both before the first deadline (in period 2) and for a while after the first deadline (in period 3). Using the cumulative response rates by period, the net hazard rate over periods 2 and 3 combined is 45.6 percent in the NEW regime, relative to 36.1 percent in the OLD regime.[12] Analogously, relative to the NEW regime, under the OLD regime, there is a dramatic increase in hazard rates following OLD letter 1—both before the second deadline (in period 4) and for a while after the second deadline (in period 5). Like Figure 1, Table 2 shows that the cumulative response rate by the time letter 2 is sent (through period 5) is roughly the same under the NEW and OLD regimes.[13]

Our analysis in this section demonstrates that notification letters generate a large increase in aggregate responses in the weeks shortly after letters are received. We can also investigate how this aggregate response depends on characteristics of the ticket (Section IIID investigates the impact of characteristics of the plate owner). Within each regime, we estimate daily hazard rates separately for (i) the six most common violation types, (ii) the six most common fine amounts, and (iii) the two issuing agencies (online Appendix 3.4 contains the figures). From this analysis, we draw two conclusions. First, while there are noticeable differences across subgroups of ticket types, there is nothing systematic that relates naturally to some underlying mechanism. Second, within each subgroup of ticket types, a qualitative comparison between the OLD and NEW regimes yields essentially the same conclusions.

B. *A Field Experiment* (*the EXP Regime*)

To investigate the mechanism behind the strong aggregate response, we worked with DOF to conduct a field experiment. The experiment included random variation along three dimensions: (i) NEW letter 1 might or might not include additional information (described below), (ii) NEW letter 1 might or might not include "forceful" language (also described below), and (iii) there might or might not be an additional notification letter between the first and second deadlines (which we label EXP letter 1.5). Hence, there are eight experimental cells, as described in Table 3.

This design addresses three issues in the comparison of the OLD versus NEW regimes. First, as discussed in Section IB, ticket recipients learn the schedule of deadlines and penalties in a piecewise fashion, and there are differences in this information across regimes. To explore whether these differences in information drive some of the differences in behavior between the OLD and NEW regimes, individuals in the *info* and *info forceful* treatments received a modified version of

[11] We focus on average daily hazard rates within a period rather than the aggregate hazard rate across the whole period because these periods have different lengths for different tickets depending on the day of the week on which the ticket is issued. See online Appendix 4 for details of how Table 2 is created.

[12] These numbers are derived from Table 2—e.g., $45.6\% = (65.27\% - 36.18\%)/(100\% - 36.18\%)$.

[13] Some plates have multiple overlapping tickets, which could cause interactions in response behavior. However, the key patterns in Figure 1 and Table 2 are unchanged if we restrict attention to plates with only one ticket.

Table 3—Letters Sent in the Eight Experimental Cells

		EXP-letter-1.5 treatment	
		not sent (50%)	sent (50%)
NEW-letter-1 treatment	baseline (20%)	NEW letter 1	NEW letter 1, EXP letter 1.5
	info (40%)	NEW letter 1*i*	NEW letter 1*i*, EXP letter 1.5
	forceful (20%)	NEW letter 1*f*	NEW letter 1*f*, EXP letter 1.5
	info forceful (20%)	NEW letter 1*if*	NEW letter 1*if*, EXP letter 1.5

Notes: NEW letter 1 is the standard letter received in the NEW regime around day 20. NEW letter 1*i* and NEW letter 1*if* include full information about deadlines and late fees. NEW letter 1*f* and NEW letter 1*if* include forceful language. EXP letter 1.5 is an additional letter sent to some ticket recipients in the EXP regime around day 48. EXP letter 1.5 contains full information about the remaining deadlines and late fees. Randomization probabilities are reported in parentheses.

NEW letter 1 that lists the full set of (individualized) deadlines and penalties. For instance, for a ticket issued on July 15, 2013, with a fine amount of $65, this would read as follows:

AMOUNT DUE IF PAID BY 8/14/13:	$65
AMOUNT DUE IF PAID BY 9/16/13:	$75 (INCLUDES $10 PENALTY FOR LATE PAYMENT)
AMOUNT DUE IF PAID BY 10/25/13:	$95 (INCLUDES $30 PENALTY FOR LATE PAYMENT)
AMOUNT DUE IF PAID AFTER 10/25/13:	$125 (INCLUDES $60 PENALTY FOR LATE PAYMENT)

If no payment is received by 11/1/13, Finance may boot or tow your vehicle.

Second, the language used in NEW letter 1 is different from that used in OLD letter 1. In particular, OLD letter 1 mentions the possibility of a default judgment entry and the associated actions and, moreover, uses a bold font to highlight the various future penalties, while NEW letter 1 does not. To investigate the impact of such language differences, individuals in the *forceful* and *info forceful* treatments received a modified version of NEW letter 1 that contains more forceful language. Specifically, the letter had the following header in large bold-faced letters:

WARNING: PENALTY APPROACHING
DON'T MISS THE DEADLINE

In addition, NEW letters 1*i*, 1*f*, and 1*if* all mention that failure to respond might result in one's vehicle being booted or towed, and in NEW letters 1*f* and 1*if*, this is mentioned in a larger font size.

Third, the comparison of the OLD versus NEW regimes reveals the impact of changing the timing of a notification letter. To test the impact of an additional notification letter, some individuals received an additional letter (EXP letter 1.5) between the two deadlines. Specifically, if there is no response by day 45, then a letter is generated on day 46, mailed on day 47, and (most likely) received on day 48, except for tickets issued on Tuesday or Wednesday, for which day 46 occurs on a weekend and the letter is generated on the subsequent Monday. The content of this letter is identical to that in NEW letter 1*i*, except that (i) the first amount due in the information box is omitted (since it is no longer relevant) and (ii) the letter is titled "NOTICE OF OUTSTANDING VIOLATION."

The experimental (EXP) regime applied to all tickets issued July 13, 2013, through August 16, 2013. For tickets issued during these five weeks, if a NEW letter 1 was triggered, it was randomly assigned to one of the four NEW letter 1's according to the probabilities in Table 3. The EXP-letter-1.5 treatment applied for tickets issued July 22, 2013, through August 10, 2013. For tickets issued during this period, if an EXP letter 1.5 was triggered, with 50 percent chance an EXP letter 1.5 was sent (independent of which NEW letter 1 was sent).[14]

C. *Aggregate Responses in the EXP Regime*

We analyze daily hazard rates in each of the eight experimental cells. The numbers of observations in the 4 cells without EXP letter 1.5 are 38,009 (1), 76,602 (1*i*), 38,199 (1*f*), and 38,156 (1*if*), and the number of observations in the 4 cells with EXP letter 1.5 are 16,060 (1), 32,041 (1*i*), 15,976 (1*f*), and 16,046 (1*if*).[15]

Figure 3, panel A depicts hazard rates for the four experimental cells assigned not to receive an EXP letter 1.5. It reveals that the four versions of NEW letter 1 lead to almost identical hazard rates. This result suggests that the large differences in behavior between the OLD and NEW regimes are not driven by differences in information or language.[16] Figure 3, panel B, in which the four NEW-letter-1 treatments are pooled, reveals that EXP letter 1.5 has a noticeable impact. In other words, even after getting a letter shortly after day 18, getting a second letter shortly after day 46 increases response rates. The combined results—that the content of the first letter hardly matters and that the second letter, which contains no new information, generates an additional response—suggest that letters serve primarily as reminders.

Quantitatively, however, the impact of EXP letter 1.5 is smaller than the impact of NEW letter 1. Recall that the net hazard rate between day 20 and day 40 is 9.5 percentage points higher in the NEW regime relative to the OLD regime (45.6 versus 36.1 percent). Here, the net hazard rate between day 48 and day 76 is only 4.7 percentage points higher for tickets assigned to receive EXP letter 1.5 relative to tickets not assigned to receive that letter (30.4 versus 25.7 percent). This difference provides an initial hint of heterogeneous responses, because it suggests that the population still present when EXP letter 1.5 arrives is less responsive to letters than the population that was present when NEW letter 1 arrived. The next section focuses more directly on heterogeneous responses.

[14] Each randomization was done by ordering plates alphanumerically and then assigning plates to treatments via a preset pattern. For plates that received multiple tickets in the EXP regime, it was possible to receive different treatments for the different tickets. Our results in Section IIC are unchanged if we consider only plates that received exactly one ticket in the EXP regime (see online Appendix 5.2).

[15] Because randomization occurred only when letters were generated and not when tickets were issued, we create the eight experimental cells by performing an ex post random assignment that assigns each ticket to one of the eight experimental cells. See online Appendix 5.1 for details.

[16] Online Appendix 5.2 contains the analogue for Figure 3, panel A for the four experimental cells assigned to receive an EXP letter 1.5. Again, the four versions of NEW letter 1 lead to almost identical hazard rates.

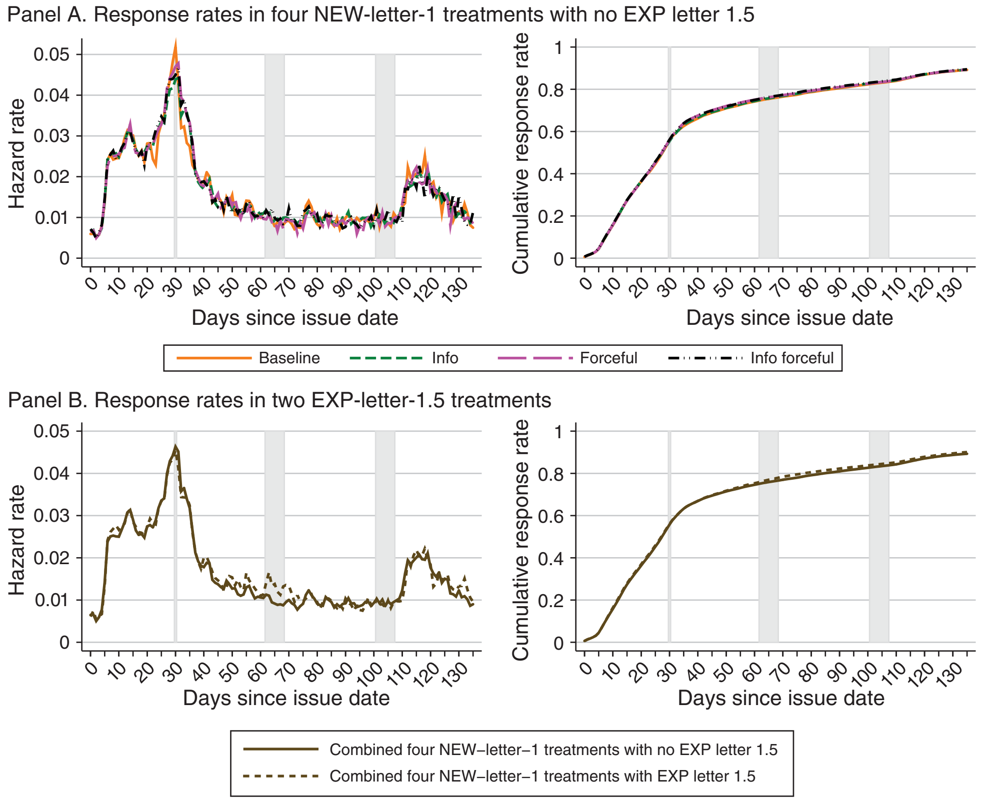

FIGURE 3. RESPONSE RATES IN THE EXP REGIME

Notes: The figure depicts daily hazard rates and cumulative response rates in the EXP regime. Panel A shows only the four experimental cells in which EXP letter 1.5 is not sent. In panel B, for each EXP-letter-1.5 treatment, the four NEW-letter-1 treatments are pooled. EXP letter 1.5 is received around day 48.

III. Heterogeneous Responses to Letters

A. *Simple Evidence of Persistent Types*

As a proof of concept, we begin with a crude approach that clearly establishes the existence of persistent types. Moreover, this approach highlights how even crude statistics about recent behavior can provide strong signals about types and, thus, how it might be easy for a policymaker such as DOF to implement a policy targeting specific types. We return to this latter point in Section IV.

Specifically, we first identify all license plates that received exactly three tickets under the OLD regime and divide them into four groups based on responses to the first two tickets: (i) both tickets have a response by day 30, (ii) the first but not the second ticket has a response by day 30, (iii) the second but not the first ticket has a response by day 30, and (iv) neither ticket has a response by day 30. Then, for each of these four groups, we estimate daily hazard rates for each plate's third ticket. We carry out the same exercise for all plates that received exactly three tickets under the NEW regime.

Figure 4 depicts these hazard rates. In each regime, third-ticket hazard rates for plates in group (i) are roughly twice the hazard rates in Figure 1, while those in group (iv) are less than half the rates in Figure 1. Those in groups (ii) and (iii) are in between. Clearly, response behavior on one's past tickets is highly predictive of response behavior on one's current ticket, indicating the existence of persistent types.

B. *Estimating a Mixture Model of Types*

We now pursue a more rigorous approach. Conceptually, we consider an underlying model in which each type is characterized by a survival function that depends on the regime, and we estimate a mixture model of these unobserved types. In the end, we are interested in the estimated survival function for each type so that we are able to compare differences across types in terms of baseline response rates and responsiveness to letters.

It is well known that identification of unobserved heterogeneity in single-spell hazard models is challenging (see, for instance, Heckman and Singer 1984). In the spirit of Honoré (1993), our identification is based on observing multiple spells (tickets) for the same individuals (plates). However, much of the literature uses a proportional-hazard-rates structure—assuming that all types share an underlying qualitative pattern and differ only in the (proportional) level of their hazard rates. Such a structure would force all types to have the same (proportional) response to letters. Because our goal is to investigate whether different types respond differently to letters, we do not use a proportional-hazard-rates structure.

Suppose there is a discrete set K of types in the population, where π_k denotes the proportion of the population that is type $k \in K$ and $\sum_{k \in K} \pi_k = 1$. Although we conceptualize each type to have a set of daily hazard rates, to reduce the dimensionality of the estimation, we conduct this analysis in terms of the six periods introduced in Table 2. Hence, each type k is characterized by hazard rates $\left(p_1^k, p_2^k, p_3^k, p_4^k, p_5^k\right)$. The hazard rate p_t^k is the probability that the person responds to a ticket in period t conditional on not having responded prior to period t. The hazard rate in the last (open-ended) period 6 is, by definition, equal to 1. Since hazard rates depend on the regime γ, we write $p_t^k(\gamma)$ for each $t \in \{1,2,3,4,5\}$ and $k \in K$.

For plate i, we can write observed behavior as a vector

$$\theta^i \equiv \left(J_i, m_1^i, \gamma_1^i, m_2^i, \gamma_2^i, \ldots, m_{J_i}^i, \gamma_{J_i}^i\right),$$

where J_i is the total number of tickets received by plate i, $m_j^i \in \{1,2,3,4,5,6\}$ is the period in which plate i's owner responded to ticket j, and γ_j^i is the regime that applies to ticket j for plate i. Then, conditional on receiving J_i tickets, the likelihood that type k would generate observed behavior θ^i is

$$\ell_k(\theta^i) = \prod_{j=1}^{J_i} \left(\left[p_1^k(\gamma_j^i)\right]^{I\{m_j^i = 1\}} \prod_{t=2}^{6} \left(\left[\prod_{t'=1}^{t-1} \left(1 - p_{t'}^k(\gamma_j^i)\right) p_t^k(\gamma_j^i) \right]^{I\{m_j^i = t\}} \right) \right),$$

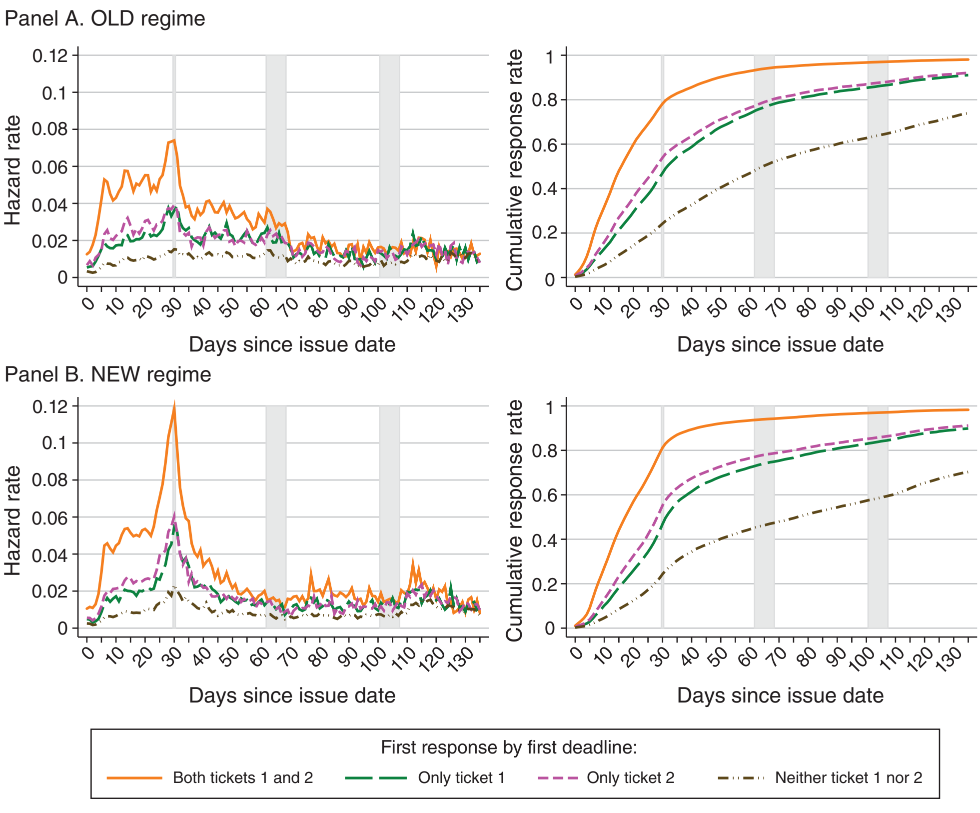

Figure 4. Past Response Behavior Predicts Future Response Behavior

Notes: The figure depicts daily hazard rates and cumulative response rates for third tickets for plates that received exactly three tickets in the OLD regime (panel A) and NEW regime (panel B). In each regime, plates are split into four groups according to whether each of the first two tickets had a response by the first deadline (day 30). The number of plates in each group is 56,035, 19,872, 20,429, and 41,559 in the OLD regime and 55,783, 17,510, 17,166, and 35,111 in the NEW regime.

where I is the identity function. Since type is unobserved, the likelihood that plate i generates observation θ^i is

$$\ell\left(\theta^i\right) = \sum_{k=1}^{K} \pi_k \ell_k\left(\theta^i\right).$$

Finally, assuming that the number of tickets received J_i is independent of one's type k, the sample log-likelihood can be written as

$$\log \mathcal{L} = \sum_i \log \ell\left(\theta^i\right).$$

This model makes several simplifying assumptions: (i) the population distribution of types π_k is the same for each regime γ, (ii) the number of tickets received J_i is independent of one's type k,[17] and (iii) within a type, the $p_t^k(\gamma)$'s are the same

[17] Under the assumption that the number of tickets received J_i is independent of one's type k, the actual sample log-likelihood is $\sum_i \log\left(\Pr(J_i)\ell\left(\theta^i\right)\right)$. Since $\Pr(J_i)$ is assumed to be independent of the model's parameters, it does

for all tickets received under regime γ (this assumption rules out "learning" in the sense that one's experience on prior tickets does not change one's response behavior on the current ticket as well as any other form of interaction across tickets). These assumptions are primarily made for reducing dimensionality. In online Appendix 6.1, we provide evidence that, while not fully consistent with the data, these assumptions seem reasonable for our purposes.

To estimate this model, we need a sample of plates that have multiple tickets. However, we are also concerned that if a plate has too many tickets within our roughly two-year window, some of those tickets could be outstanding at the same time, and this might change the nature of the decision-making problem. Keeping this trade-off in mind, in our main estimation, we use all plates that received $J \in \{3,4, \ldots, 12\}$ tickets across the OLD and NEW regimes combined—657,890 plates that received 3,366,145 tickets.[18] Before estimating the model, for each plate, we put one randomly chosen ticket into a *holdout sample*. Using the remaining 2,708,255 tickets for the 657,890 plates—the *estimation sample*—we estimate the mixture model above for $K \in \{1,2,3,4\}$.

Table 4 reports, for each $K \in \{1,2,3,4\}$, the estimated average daily hazard rates for each type in each period, along with the estimated proportion of each type.[19] Our analysis in the next three subsections focuses on the $K = 3$ model, but similar messages emerge from the $K = 2$ and $K = 4$ models. In the $K = 3$ model, we refer to the type with the highest hazard rate in all periods as the high-response type (*HR*s), the type with the lowest hazard rate in all periods as the low-response type (*LR*s), and the other type as the medium-response type (*MR*s).[20]

Before we discuss the results in Table 4, we briefly discuss an alternative to our structural approach. In particular, one might worry that some of our conclusions below are driven by the particular structure that we assume. In online Appendix 9, we provide an alternative approach in which we type plates in a reduced-form manner using median days to first response and then reproduce our analyses in this and the subsequent three subsections. We demonstrate in online Appendix 9 that our conclusions would be much the same while also highlighting the limitations of that reduced-form approach.

not impact the estimation, and thus we suppress it from the sample log-likelihood.

[18] Having 12 tickets across the 2 regimes would imply, on average, receiving a ticket roughly once every 65 days. We do not use data from the EXP regime in estimating this model because regime-specific hazard rates are identified from plates that have multiple tickets within a regime, and few plates received multiple tickets within any one cell in the EXP regime.

[19] The estimation technique described in the text yields *per-period* hazard rates (reported in online Appendix Table A8). For interpretation, we convert each per-period hazard rate into an average *daily* hazard rate using the average number of days in each period and use the delta method to convert the standard errors. Details of this transformation are available in online Appendix 6.2.

[20] While in principle we could have used a statistical criterion (such as BIC) to select the number of types, we chose not to for two reasons. First, our goal is not to obtain an accurate estimate of the number of types but rather to understand the qualitative nature of the heterogeneity, and the $K = 4$ model and the $K = 3$ model already yield much the same conclusions. Second, given the size of our sample, we suspect that such an approach would select a large number of types, and finding the optimal number of types would be computationally burdensome (the $K = 4$ model already takes quite a while to compute, and the BIC strongly selects it over the $K = 3$ model).

Table 4—Estimated Mixture Models with Average Daily Hazard Rates

	Type	π_k	Regime	p_1	p_2	p_3	p_4	p_5
$K = 1$			OLD	2.15%	2.66%	1.99%	1.84%	1.30%
		1.000		(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
		—	NEW	2.05%	3.48%	2.84%	1.31%	0.87%
				(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
$K = 2$			OLD	3.51%	5.25%	4.41%	5.26%	4.30%
	HR	0.641		(0.01%)	(0.01%)	(0.02%)	(0.02%)	(0.05%)
		(0.001)	NEW	3.29%	7.04%	7.28%	3.87%	2.62%
				(0.01%)	(0.02%)	(0.03%)	(0.02%)	(0.04%)
			OLD	0.58%	0.90%	0.87%	1.06%	0.98%
	LR	0.359		(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
		(0.001)	NEW	0.51%	1.05%	1.15%	0.82%	0.67%
				(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
$K = 3$			OLD	6.24%	8.14%	5.00%	5.65%	2.35%
	HR	0.340		(0.02%)	(0.05%)	(0.06%)	(0.06%)	(0.11%)
		(0.001)	NEW	5.78%	11.34%	9.83%	3.52%	1.73%
				(0.02%)	(0.06%)	(0.11%)	(0.06%)	(0.10%)
			OLD	1.42%	3.39%	3.28%	3.91%	3.59%
	MR	0.411		(0.01%)	(0.01%)	(0.02%)	(0.02%)	(0.03%)
		(0.001)	NEW	1.29%	4.37%	4.97%	2.94%	2.16%
				(0.01%)	(0.02%)	(0.02%)	(0.01%)	(0.02%)
			OLD	0.50%	0.58%	0.51%	0.63%	0.65%
	LR	0.249		(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
		(0.001)	NEW	0.45%	0.61%	0.60%	0.49%	0.45%
				(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
$K = 4$			OLD	7.60%	7.00%	3.34%	4.93%	2.31%
	HR	0.261		(0.03%)	(0.06%)	(0.06%)	(0.06%)	(0.12%)
		(0.001)	NEW	7.09%	10.64%	7.65%	2.94%	1.68%
				(0.03%)	(0.08%)	(0.12%)	(0.06%)	(0.10%)
			OLD	2.08%	6.24%	6.17%	6.39%	4.56%
	MHR	0.277		(0.01%)	(0.04%)	(0.05%)	(0.06%)	(0.13%)
		(0.002)	NEW	1.84%	7.97%	9.65%	4.77%	2.65%
				(0.01%)	(0.04%)	(0.07%)	(0.05%)	(0.10%)
			OLD	1.11%	1.64%	1.68%	2.52%	2.51%
	MLR	0.295		(0.01%)	(0.01%)	(0.01%)	(0.02%)	(0.02%)
		(0.001)	NEW	1.02%	2.07%	2.52%	1.91%	1.57%
				(0.01%)	(0.02%)	(0.02%)	(0.01%)	(0.02%)
			OLD	0.36%	0.44%	0.35%	0.35%	0.37%
	LR	0.167		(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
		(0.001)	NEW	0.31%	0.44%	0.35%	0.26%	0.26%
				(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)

Notes: Mixture models are estimated for $K = 1$, 2, 3, and 4 types. The p_t's are the estimated average daily hazard rates by period for each type, and the π_k's are the estimated proportions of each type. Standard errors are in parentheses. Online Appendix 6.2 provides additional details.

C. *Type-Specific Response Behavior*

The estimates in Table 4 confirm the large and persistent differences across individuals suggested by Figure 4. In the estimated three-type model, 34 percent of the population is estimated to be *HR*s, 41 percent to be *MR*s, and 25 percent to be *LR*s. Average daily hazard rates for the *HR*s are roughly twice those for the *MR*s and ten times those for the *LR*s.

These large differences in hazard rates imply substantial differences in cumulative response rates (see online Appendix Table A10). For instance, implied cumulative responses in the OLD regime by the time the first letter is sent (on day 35–41) are 93 percent for the *HR*s, 60 percent for the *MR*s, and 19 percent for the *LR*s. By the time the second letter is sent (on day 70–76), implied cumulative responses are 99, 90, and 35 percent.

Table 4 highlights how the aggregate response to letters from Section IIA masks significant differential response rates across higher and lower types. The *HR*s have a strong response to letters: adding NEW letter 1 (relative to the OLD regime) increases average daily hazard rates from 8.14 to 11.34 percent in period 2 and from 5.00 to 9.83 percent in period 3. In contrast, the *LR*s have a weak response: adding NEW letter 1 increases average daily hazard rates from 0.58 to 0.61 percent in period 2 and from 0.51 to 0.60 percent in period 3. The *MR*s have an intermediate response.

To better appreciate the economic impact of these differences, consider the net hazard rate over periods 2 and 3 combined. Recall from Section IIA that in aggregate, the switch from the OLD to the NEW regime increases the net hazard rate over periods 2 and 3 by 9.5 percentage points (from 36.1 to 45.6 percent). In our estimated three-type model, the corresponding numbers are 14.9 percentage points for the *HR*s (from 72.9 to 87.8 percent), 11.8 percentage points for *MR*s (from 46.8 to 58.6 percent), and only 1.0 percentage point for *LR*s (from 9.7 to 10.7 percent). Hence, our estimates imply that the economic impact of reminders is an order of magnitude larger for the *HR*s and *MR*s than it is for the *LR*s.

The differences in cumulative response rates imply very strong selection effects. By the time OLD letter 1 is sent, the remaining population consists primarily of *MR*s and *LR*s (42 and 52 percent, respectively), and by the time letter 2 is sent, the vast majority of the remaining population consists of *LR*s (78 percent). These strong selection effects, combined with the differential responses to letters discussed above, helps to explain why the aggregate response to NEW letter 1 is significantly larger than the aggregate response to EXP letter 1.5 (as discussed in Section IIC).[21]

Recall that the estimates in Table 4 are based on the estimation sample. We next look at behavior of "typed" plates in the holdout sample. Doing so provides an out-of-sample validation of the estimates in Table 4 while also providing a way to analyze daily hazard rates by type. Specifically, given the estimated parameters for the π_k's and the $p_t^k(\gamma)$'s, the predicted probability that plate i with observed behavior θ^i is type k is

$$\hat{\pi}\left(k \mid \theta^i\right) = \frac{\pi_k \ell_k\left(\theta^i\right)}{\sum_{k'} \pi_{k'} \ell_{k'}\left(\theta^i\right)}.$$

In principle, we could just assign plate i to the type k that maximizes $\hat{\pi}\left(k \mid \theta^i\right)$. However, one might worry about plates that are barely assigned to one type relative to another. Instead, we assign plate i to the type k that maximizes $\hat{\pi}\left(k \mid \theta^i\right)$ only if that k yields $\hat{\pi}\left(k \mid \theta^i\right) > 0.60$. With this approach, we type 583,749 of the 657,890

[21] Much as discussed in job search research, these strong selection effects can also cause aggregate hazard rates to decline over time even as type-specific hazard rates increase over time. In Table 4, for instance, under the OLD regime, aggregate hazard rates (i.e., the $K = 1$ estimates) decline from period 3 to period 4, even though for every type, the type-specific hazard rates increase from period 3 to period 4 (for the $K = 2$, $K = 3$, and $K = 4$ estimates).

plates (88.7 percent). Of these, 36.4 percent are assigned as *HR*s, 39.3 percent as *MR*s, and 24.3 percent as *LR*s.[22]

Using the holdout sample, Figure 5 depicts the type-specific response behavior in the OLD versus NEW regimes. Figure 5 yields much the same message as Table 4.[23] The *HR*s and *MR*s behave qualitatively the same, with the *HR*s acting sooner and both types reacting strongly to notification letters. The *LR*s, in contrast, have low and relatively flat response rates from day 0 through the third deadline, and they exhibit barely noticeable reactions to NEW letter 1 in the NEW regime and to OLD letter 1 in the OLD regime.[24]

Our results in this subsection help to rule out two alternative interpretations of the impact of the first letter. First, while we interpret the first letter as a reminder, perhaps its main role is to inform the owner about a ticket or to make it easier to pay. However, given that *HR*s react most strongly to the first letter, if these mechanisms were the primary drivers of behavior, then it would need to be that *HR*s are also the types most prone to not know about the ticket or most prone to benefit from assistance in responding. But both of these seem inconsistent with *HR*s' high response rate even prior to receiving the first letter. Second, there might be something special about the first letter that one receives—perhaps it reveals that DOF knows where one lives. However, such effects would primarily apply for first offenders, and online Appendix 6 Tables A5 and A6 show that response patterns change very little across tickets for repeat offenders.

D. *Who Are the Low-Response Types?*

Given our ability to assign each plate an ex post probability of being an *HR*, *MR*, or *LR*, it is natural to ask whether any plate-owner characteristics are correlated with these probabilities. Again, for the vast majority of our data, we have little information on the observable characteristics of the ticket recipient. However, during the EXP regime, our data contain an address for every ticket for which a NEW letter 1 was sent, and thus we can match the associated plates to census demographics (using United States Census Bureau 2002). Specifically, of the 657,890 plates in our estimation sample, 60,529 received a ticket under the EXP regime for which (i) they were sent a NEW letter 1, (ii) we were able to match their address to a Census Block Group, and (iii) there were no missing values for the demographic variables.

For these 60,529 plates, we further update their predicted type probabilities (i.e., their $\hat{\pi}(k\,|\,\theta^i)$'s) based on their response behavior under the EXP regime—this

[22] The criterion $\hat{\pi}(k\,|\,\theta^i) > 0.60$ is chosen to balance sufficient confidence in the typing against typing sufficiently many plates. If we instead require $\hat{\pi}(k\,|\,\theta^i) > 0.50$, we type 99.1 percent of plates, whereas if we require $\hat{\pi}(k\,|\,\theta^i) > 0.75$, we type 68.8 percent of plates. See online Appendix 6.3 for details. In online Appendix 6.3, Figure A2 illustrates that Figure 5 would look much the same for other cutoffs besides $\hat{\pi}(k\,|\,\theta^i) > 0.60$, and Figure A3 presents Figure 5 with a separate panel for each type.

[23] Online Appendix 6.4 investigates behavior of the three types under the EXP regime. Each type exhibits the main aggregate findings from Section IIC—specifically, the content of the first letter hardly matters, and the second letter generates a noticeable additional response.

[24] Although Figure 5 also seems to suggest that the shift from the OLD to the NEW regime leads to slightly worse cumulative outcomes for the *LR*s (and possibly also the *MR*s), we show in online Appendix 6.3 that this feature is most likely an artifact of Hurricane Sandy.

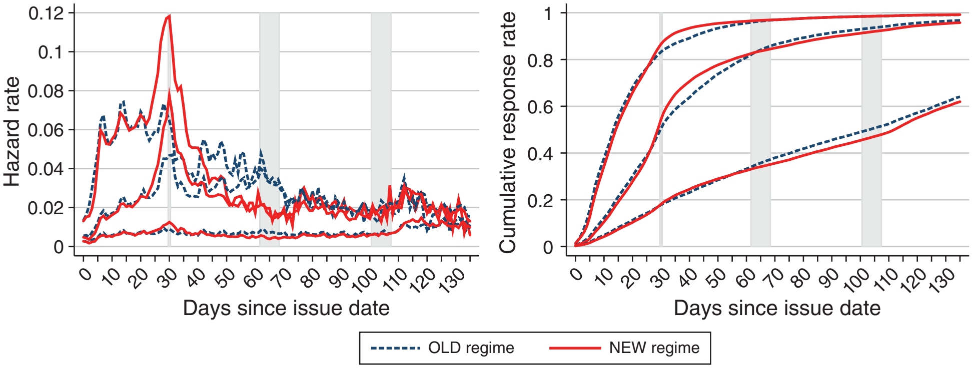

Figure 5. Response Rates by Predicted Type

Notes: The figure depicts daily hazard rates and cumulative response rates in the OLD versus NEW regimes for the 583,749 plates assigned to be 1 of the 3 predicted types (HRs, MRs, LRs). Types are assigned using the estimation sample, and the figure shows behavior from the holdout sample.

updating is necessary to correct for selection due to the fact that we observe an address only if a ticket is not paid by the end of period 1. We then regress the predicted likelihood of being an *LR* on census income, race, education, ability to speak English, and how one travels to work. Table 5 presents descriptive statistics for the census variables along with the regression results.[25]

Regression (1) in Table 5 presents OLS estimates including all of the census variables in a single specification. We find that the likelihood of a plate being an *LR* is larger when the owner lives in a Census Block Group that has lower income, less education, and higher proportions of "Black" or "other" racial groups. In other words, the *LR*s, who accumulate significant late penalties, seem more likely to come from already disadvantaged groups.

Regressions (2) through (4) present OLS estimates with only income, only education, or only race included. Given that the demographic variables are correlated, these regressions identify what one could predict about a plate if we only knew one dimension of its demographics. For instance, suppose all we know is that a plate comes from a Census Block Group with median income at the tenth percentile ($18,973) rather than at the ninetieth percentile ($72,105). Regression (2) implies that the likelihood of that plate being an *LR* is 36 percent higher (36.5 versus 26.7 percent). Analogously, suppose all we know is that a plate comes from a Census Block Group with proportion Black at the ninetieth percentile (0.81) rather than at the tenth percentile (0.00), with the remainder assumed to be White. Regression (4) implies that the likelihood of that plate being an *LR* is 61 percent higher (40.0 versus 24.8 percent).

Finally, we briefly mention two plate-owner characteristics that we have for the majority of plates in the data: car make and vintage. While these are crude measures

[25] See online Appendix 7 for the details behind the analysis of this section. Online Appendix Table A12 presents these regressions when the dependent variable is the likelihood that a plate is an *HR*, and online Appendix Table A13 presents logistic regressions. Both yield the same conclusions.

Table 5—Descriptive Statistics of Census Variables, and Regressions of Likelihood *LR* Type

	Descriptive statistics			Regressions			
				Dependent variable: likelihood *LR* type			
	Mean	Tenth percentile	Nintieth percentile	(1)	(2)	(3)	(4)
Median household income	44,403	18,973	72,105				
ln(Median household income)				−0.030	−0.073		
				(0.006)	(0.003)		
Education							
Less than high school	0.27	0.08	0.50				
High school	0.26	0.14	0.37	0.035		−0.084	
				(0.035)		(0.026)	
Some college	0.22	0.13	0.31	0.022		0.227	
				(0.033)		(0.028)	
College or more	0.25	0.06	0.51	−0.095		−0.316	
				(0.026)		(0.012)	
Race							
White	0.50	0.06	0.93				
Black	0.25	0.00	0.81	0.159			0.188
				(0.009)			(0.006)
Asian	0.08	0.00	0.24	−0.039			−0.161
				(0.017)			(0.014)
Other	0.17	0.01	0.43	0.246			0.177
				(0.017)			(0.011)
Language							
English only	0.54	0.19	0.86				
English very well	0.23	0.09	0.37	−0.066			
				(0.021)			
English well	0.12	0.02	0.23	−0.141			
				(0.031)			
English not well	0.09	0.00	0.21	−0.102			
				(0.039)			
English not at all	0.03	0.00	0.09	−0.107			
				(0.070)			
Transportation to work							
Public transportation	0.46	0.12	0.72				
Drive	0.43	0.13	0.80	0.119			
				(0.011)			
Other	0.11	0.02	0.22	0.035			
				(0.022)			
Constant				0.548	1.084	0.362	0.248
				(0.059)	(0.037)	(0.011)	(0.003)
Observations		60,529		60,529	60,529	60,529	60,529
R^2				0.04	0.01	0.02	0.03

Notes: The regressions are estimated using OLS, with standard errors clustered at the block group level. For each group of census variables, descriptive statistics are reported as proportions, and the omitted category in the regression is the one without a reported coefficient. Based on 60,529 plates in 9,481 Census Block Groups.

of socioeconomic status, we do find that *LR*s drive older cars than *HR*s (mean of 9.2 years old versus 7.7 years old) and are less likely to drive new luxury makes (3 versus 9 percent).[26] See online Appendix 7 for details. These results are consistent with the income results in Table 5.

E. *Low-Response Types Respond to Firmer Interventions*

While the *LR*s respond only weakly to deadlines and reminders, they do respond significantly to more consequential incentives. In particular, note that in Figure 5, hazard rates for *LR*s jump to their highest level at roughly day 110—to a daily hazard rate of roughly 1.1–1.2 percent—and remain at that level through day 135. In other words, hazard rates for *LR*s from day 110 to 135 are higher than they are for any earlier set of dates (online Appendix Figure A3d shows this even more clearly).

As described in Section I, DOF sends a third notification letter (letter 3) on the Tuesday that is day 105–111. Unlike the due dates on prior letters, which were always 10–31 days in the future, the due date on letter 3 is listed as "IMMEDIATELY," and the letter further states that the owner is now subject to immediate judgment enforcement actions. The rise in hazard rates for the *LR*s corresponds with receipt of this letter.

Moreover, the city indeed carries out the enforcement threat. Specifically, once an owner has more than $350 in outstanding judgment debt (across all her plates), that owner's cars are eligible to be towed or booted if they are identified on New York City streets. Prior to July 11, 2013, this meant a car was towed and then, if there was still no response within a few days, sold at auction. Starting on July 11, 2013, this instead meant that a car was initially booted, and if there were no response to the boot within a few days, then it would be towed and sold at auction a few days after that. In our data, we cannot identify the day on which a car is towed or booted; however, we can identify responses that occur after a car has been towed or booted.[27]

Figure 6 presents hazard rates and cumulative response rates for *LR*s while disaggregating responses into those that occur after towing/booting and those that do not. We note two key findings. First, towing/booting indeed occurs: shortly after day 110, we start to see responses that follow a tow or boot, and by day 135, nearly 50 percent of responses from *LR*s follow a tow or boot. Second, the increase in hazard rates for the *LR*s predates actual towing/booting: non–tow/boot hazard rates jump when letter 3 is received, before there is any significant towing/booting. While there is no treatment-control comparison here, there is a natural interpretation of these findings: *LR*s are reacting to the combination of (i) letter 3—with its "IMMEDIATELY" deadline and threat of more severe actions—and (ii) actual towing/booting taking place in the weeks that follow, making that threat credible.

These findings suggest three messages. First, it is not the case that the *LR*s are not responding because they are merely "disappearing" (e.g., moved away or were otherwise unavailable to respond). Second, it is not the case that the *LR*s are not

[26] We classify a make as "luxury" if the majority of its models appear in the "luxury car" category in *Consumer Reports* (2019).

[27] Specifically, we can identify responses linked to the units that release cars that have been towed or booted—see online Appendix 10 for details. Also, prior to full adoption on July 11, 2013, booting occurred at low levels under a pilot booting program that began on June 25, 2012.

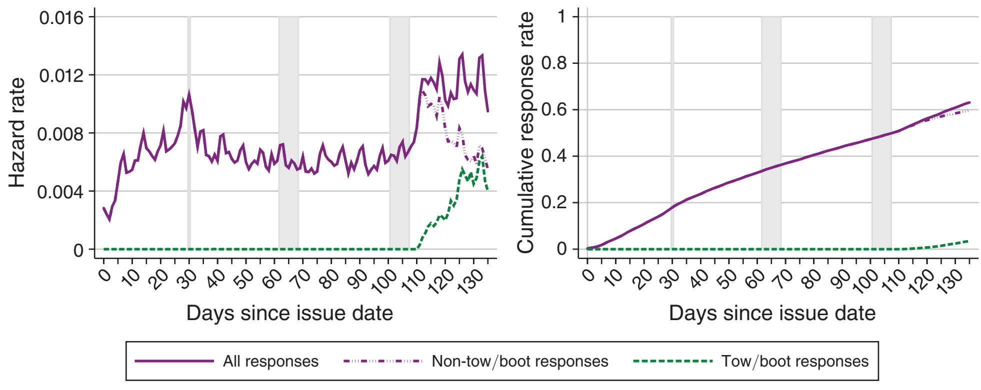

Figure 6. Response Rates for the *LR*s

Notes: The figure depicts daily hazard rates and cumulative response rates pooled across regimes for the 141,907 plates assigned to be *LR*s. (Online Appendix 8 contains regime-specific figures.) Types are assigned using the estimation sample, and the figure shows behavior from the holdout sample. Responses are decomposed into those that followed towing/booting and those that did not (using the same denominator for day *d*, which is the number of tickets without a response prior to day *d*).

responding because they do not react to letters—here they seem to exhibit a strong response to letter 3. Rather, a strong response seems to require incentives that are more consequential than a modest financial penalty to be applied only if a future deadline is missed. Third, while the *LR*s do respond to letter 3 and the eventual towing/booting, it is still at a relatively slow pace. Hence, there is an open question whether there are ways to get them to respond more quickly.

F. *Interpretation of Differences across Types*

Our analysis above reveals that types differ along two dimensions: baseline response rates and responsiveness to letters. We now discuss how one might interpret these differences.

Baseline response rates depend on a variety of factors. In traditional models of task completion, they depend on the variance in day-to-day effort and opportunity costs, which determines the value of waiting for a lower-cost future day. They can also depend on liquidity constraints, because one must be able to pay. In "behavioral" models of task completion, present bias might make one repeatedly prefer responding in the near future rather than now, and forgetting might lead to extended nonresponse due to not thinking about the ticket. Heterogeneity in baseline response rates could be due to differences in any of these factors.

If the primary role of letters is to remind people of a forgotten ticket, then the responsiveness to letters depends on a combination of (i) the likelihood that letters get noticed and thereby put tickets back on one's mind and (ii) the likelihood of responding when a ticket is on one's mind. Heterogeneity in responsiveness can be driven by heterogeneity in either. We suspect that the latter might be especially important in understanding the low responsiveness to letters among *LR*s. For instance, for people with liquidity constraints, it might not matter much whether the ticket is on the mind. Alternatively, people who mispredict future present bias or future forgetting may be unlikely to react much to letters that are well in advance of deadlines.

While understanding the relative impact of these different mechanisms is important, the variation in our data does not allow us to separately identify them.

IV. Discussion

In this paper, we investigate for whom reminders work among New York City parking ticket recipients. We demonstrate the existence of large and persistent differences in behavior across individuals, and importantly we find heterogeneous responses to notification letters: those with low baseline propensities to respond to tickets—arguably the natural target population for intervention—react least to letters. Moreover, these low-response types, who incur significant late penalties, disproportionately come from already disadvantaged groups and do respond to more traditional interventions. We conclude with some broader takeaways.

A key implication of our analysis is the value of analyzing heterogeneity in nudge effects prior to giving policy advice. There is, of course, a long tradition of incidence analysis for more incentive-based economic policies. Most of the existing literature on nudges focuses on aggregate impacts or, at best, studies the impact of observable demographics. In contrast, our large and longitudinal dataset allows us to investigate unobserved heterogeneity based on behavior, including in the impact of the nudge. We indeed find that aggregate analysis may yield misleading conclusions.

To illustrate, imagine a comparison of the OLD versus NEW regimes based on our aggregate results. As Figure 1 shows, the change in timing of the first letter had virtually no impact on the cumulative response rate by the second deadline. Hence, the main aggregate trade-off of sending notification letters at day 20 instead of day 40 is that more letters are sent (70 percent of tickets are sent first letters in the NEW regime versus 45 percent in the OLD regime) versus fewer first ($10) penalties are incurred (39 percent of tickets incur the first penalty in the NEW regime versus 45 percent in the OLD regime). While the NEW-versus-OLD trade-off involves a variety of monetary and nonmonetary costs and benefits, we can quantify its direct aggregate monetary impact: in our core dataset, the NEW regime involves roughly $369,000 per year in extra notification-letter costs (the DOF cost per letter is approximately $0.50; private communication) for a reduction of roughly $1.78 million per year in first penalties.[28]

Ultimately, it is for the various constituencies of New York City to decide whether this trade-off is worth it (of course, accounting for the indirect and nonmonetary impacts). However, one's assessment of the trade-off may change after incorporating our heterogeneity results. Quantifying the direct monetary impact as above, for the *HR*s and *MR*s combined, the NEW regime involves roughly $350,000 in extra notification-letter costs for a reduction in late penalties of roughly $1.42

[28] For details behind these calculations, see online Appendix 10 Table A15.

million. In contrast, for the *LR*s, the NEW regime involves roughly \$36,000 in extra notification-letter costs for a reduction in late penalties of about \$224,000.[29] In other words, while the reduction in late penalties per extra notification letter is highest for the *LR*s, about 91 percent of the extra spending on notification letters and 86 percent of their gains go to the *MR*s and *HR*s. Only a tiny fraction of this program helps the *LR*s, who represent nearly all of the serious noncompliance.

Our analysis further suggests how one might design a policy of targeted reminders based on past behavior.[30] To illustrate, suppose that DOF felt that the shift from the OLD to the NEW regime was too expensive, but it could return to the OLD regime and allocate *some* additional budget to send an extra reminder to some ticket recipients at day 20. Suppose further that DOF wanted to send those letters to *LR*s so as to help more of them to pay before the second deadline. As Figure 4 shows, even crude information on past behavior can identify the *LR*s—e.g., DOF could send the extra day 20 letter to any ticket recipient who had missed the first deadline on each of her two most recent tickets.

It is worth reiterating that this alternative policy of targeted reminders would be based not on individual characteristics (e.g., income, race, neighborhood) but only on past behavior—while statistically helping traditionally underserved populations to avoid penalties with a nonintrusive nudge. We further note that in proposing this policy, we are not assuming that the low baseline response rates of the *LR*s are suboptimal. Rather, we are pointing out a lower-cost policy that could induce more timely payments from the *LR*s without imposing larger penalties on them.

A broader issue raised by our analysis is the relative value of sociodemographic information versus simple measures of past behavior in predicting future behavior. Some prior analyses of nudges are based on large administrative datasets and study how the impact of nudges depends on sociodemographic observables (for instance, Beshears et al. 2021 and De Neve et al. 2021). While we also find that sociodemographic observables are a meaningful predictor of response behavior, our analysis suggests that simple measures of past behavior can be a more powerful predictor. For instance, we can compare the explanatory power of the two types of variables using the sample of plates from Figure 4 for which we also have census variables.[31] A simple linear regression of the probability that a plate's third ticket is paid by day 30 on demographic variables (those in Table 5) has $R^2 = 0.02$. In contrast, for the same sample, a linear regression of the probability that a plate's third ticket is paid by day 30 on the 4 indicator variables for response on the first 2 tickets by day 30 (corresponding to the four groups in Figure 4) has $R^2 = 0.17$, an order of magnitude larger. A potentially fruitful direction for future work is to investigate the predictive power of simple measures of past behavior on other tasks

[29] The type-specific numbers do not sum to the aggregate numbers because different samples are used in calculating them. Again, see online Appendix 10 Table A15 for details.

[30] While our discussion here speculates about alternative policies, and DOF has some latitude to set noticing and penalty policy, certain types of changes may in fact require state and local legislative action.

[31] Of the 263,465 plate-regime observations used in Figure 4, we have census variables for 22,873 (the large reduction is primarily due to only a small percentage of these plates also getting a ticket in the EXP regime).

(personal income tax filing, bill payment, and so forth), to predict behavior not only within the same task, but also across tasks.[32]

Another important policy question is whether and how to reduce the regressivity of the current system. In our domain, deadlines are associated with increasingly large late penalties—missing the first costs $10, missing the second costs an additional $20, and missing the third costs an additional $30 (now a total of $60). At the same time, later deadlines seem to have a smaller impact on behavior. Given their limited impact on behavior, one might consider reducing or even eliminating the second and third penalties, as they are primarily incurred by *LR*s.[33] Indeed, our impression is that similar schemes of increasingly large late penalties are common, and it is worth investigating whether other instances of such schemes lead to similar outcomes.

Additional studies like ours would help to assess the generalizability of our findings. However, we see no reason why our policy takeaways in the parking ticket setting—simple measures of past behavior can be used to identify types, these types may respond differently to interventions, and the characteristics of these types may vary in important ways—would not apply to other domains. Indeed, this kind of heterogeneity analysis has a long history in studies of traditional interventions. As we discuss above, however, this work is only beginning in the nudge literature.

REFERENCES

Akerlof, George A. 1991. "Procrastination and Obedience." *American Economic Review* 81 (2): 1–19.

Benartzi, Shlomo, John Beshears, Katherine L. Milkman, Cass R. Sunstein, Richard H. Thaler, Maya Shankar, Will Tucker-Ray, William J. Congdon, and Steven Galing. 2017. "Should Governments Invest More in Nudging?" *Psychological Science* 28 (8): 1041–55.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2021. "Active Choice, Implicit Defaults, and the Incentive to Choose." *Organizational Behavior and Human Decision Processes* 163: 6–16.

Cadena, Ximena, and Antoinette Schoar. 2011. "Remembering to Pay? Reminders vs. Financial Incentives for Loan Payments." NBER Working Paper 17020.

Calzolari, Giacomo, and Mattia Nardotto. 2017. "Effective Reminders." *Management Science* 63 (9): 2915–32.

Chirico, Michael, Robert Inman, Charles Loeffler, John MacDonald, and Holger Sieg. 2019. "Deterring Tax Delinquency in Philadelphia: An Experimental Evaluation of Nudge Strategies." *National Tax Journal* 72 (3): 479–506.

Consumer Reports. 2019. "Luxury Cars." Consumer Reports. https://www.consumerreports.org/cro/cars/luxury-cars.htm (accessed March 23, 2021).

De Neve, Jan-Emmanuel, Clément Imbert, Johannes Spinnewijn, Teodora Tsankova, and Maarten Luts. 2021. "How to Improve Tax Compliance? Evidence from Population-Wide Experiments in Belgium." *Journal of Political Economy* 129 (5): 1425–63.

DellaVigna, Stefano, and Elizabeth Linos. 2022. "RCTs to Scale: Comprehensive Evidence from Two Nudge Units." *Econometrica* 90 (1): 81-116.

Ericson, Keith M. Marzilli. 2011. "Forgetting We Forget: Overconfidence and Memory." *Journal of the European Economic Association* 9 (1): 43–60.

Ghesla, Claus, Manuel Grieder, and Renate Schubert. 2020. "Nudging the Poor and the Rich—A Field Study on the Distributional Effects of Green Electricity Defaults." *Energy Economics* 86: 104616.

[32] For a recent example, see Knittel and Stolper (2019), who study heterogeneous treatment effects for informational nudges to reduce household energy consumption. Using machine learning techniques, they find that a simple measure of past behavior (pretreatment consumption) is one of the two strongest predictors of the treatment effect.

[33] Among typed plates in the holdout sample, the *LR*s incur 47 percent of the $20 penalties and 75 percent of the $30 penalties.

Heckman, J., and B. Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52 (2): 271–320.

Heffetz, Ori, Ted O'Donoghue, and Henry S. Schneider. 2022a. "Replication data for: Reminders Work, but for Whom? Evidence from New York City Parking Ticket Recipients." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.38886/E135881V1.

Heffetz, Ori, Ted O'Donoghue, and Henry Schneider. 2022b. "A Note on the Identification of Present Bias and Forgetting from Task-Completion Data." Unpublished.

Heidhues, Paul, and Philipp Strack. 2021. "Identifying Present Bias from the Timing of Choices." *American Economic Review* 111 (8): 2594–622.

Holman, Jeff, and Farhan Zaidi. 2010. "The Economics of Prospective Memory." Unpublished.

Honoré, Bo E. 1993. "Identification Results for Duration Models with Multiple Spells." *Review of Economic Studies* 60 (1): 241–46.

Hummel, Dennis, and Alexander Maedche. 2019. "How Effective Is Nudging? A Quantitative Review on the Effect Sizes and Limits of Empirical Nudging Studies." *Journal of Behavioral and Experimental Economics* 80: 47–58.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. "Getting to the Top of Mind: How Reminders Increase Saving." *Management Science* 62 (12): 3393–411.

Karlan, Dean, Melanie Morten, and Jonathan Zinman. 2016. "A Personal Touch in Text Messaging Can Improve Microloan Repayment." *Behavioral Science and Policy* 1 (2): 25–31.

Knittel, Christopher R., and Samuel Stolper. 2019. "Using Machine Learning to Target Treatment: The Case of Household Energy Use." NBER Working Paper 26531.

Martinez, Seung-Keun, Stephan Meier, and Charles Sprenger. 2017. "Procrastination in the Field: Evidence from Tax Filing." Unpublished.

New York City Department of Finance. 2014. Summonses, Events, and Vendor Files, and Lookup Tables. City of New York. Multiple electronic files provided to the authors under the Research and Nondisclosure Statement of December 23, 2011.

O'Donoghue, Ted, and Matthew Rabin. 1999a. "Doing It Now or Later." *American Economic Review* 89 (1): 103–24.

O'Donoghue, Ted, and Matthew Rabin. 1999b. "Incentives for Procrastinators." *Quarterly Journal of Economics* 114 (3): 769–816.

O'Donoghue, Ted, and Matthew Rabin. 2001. "Choice and Procrastination." *Quarterly Journal of Economics* 116 (1): 121–60.

Tasoff, Joshua, and Robert Letzler. 2014. "Everyone Believes in Redemption: Nudges and Overoptimism in Costly Task Completion." *Journal of Economic Behavior and Organization* 107 (A): 107–22.

Taubinsky, Dmitry. 2014. "From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice." Unpublished.

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

United States Census Bureau. 2002. Summary File 3 Dataset: All 50 States and Puerto Rico. Electronic files XX00001.txt, XX00002.txt, XX00003.txt, XX00004.txt, XX00005.txt, XX00006.txt, XX00007.txt, and XXgeo.txt, Where the XX Correspond to Individual U.S. State Abbreviations." United States Census Bureau. https://www.census.gov/data/datasets/2000/dec/summary-file-3.html (accessed March 28, 2016).