Reminders Work, But for Whom? Evidence from New York City Parking-Ticket Recipients*

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

March 26, 2020

Abstract

We study response behavior of New York City parking-ticket recipients, analyzing administrative data on 6.6 million tickets issued to 2 million individuals over two years. Using variation in the timing of reminder letters, we find evidence consistent with significant forgetting. But we find large differences across individuals, and, importantly, those with a low baseline propensity to respond to tickets—a natural nudge target—react *least* to reminders. These low-response types, who incur significant late penalties, disproportionately come from already disadvantaged groups. They do react strongly to more incentive-based interventions. We discuss how accounting for effect heterogeneity might change one's approach to policy, and how one might use our analysis to target interventions at low-response types.

*Heffetz: S.C. Johnson Graduate School of Management, Cornell University, 324 Sage Hall, Ithaca, NY 14853, and Department of Economics, Hebrew University of Jerusalem (e-mail: oh33@cornell.edu); O'Donoghue: Department of Economics, Cornell University, 482 Uris Hall, Ithaca, NY 14853 (e-mail: edo1@cornell.edu); Schneider: Stephen J.R. Smith School of Business, Queen's University, 439 Goodes Hall, Kingston, ON, K7L 3N6 (e-mail: henry.schneider@queensu.ca). 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 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, SITE 2014, and BEAM 2015. We also thank Maya Catabi, Ofer Glicksohn, Lawrence Jin, 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.

1 Introduction

Life involves many tasks that must—or at least should—be completed. People must pay bills, file taxes, make doctors' appointments and fill prescriptions, apply for schools and jobs, and so forth. Despite the ubiquity of such tasks, research is still in the early stages of understanding who completes them and how to target interventions.

In this paper, we study a classic task-completion problem: when to respond to a parking ticket. Using an administrative dataset provided by the New York City Department of Finance (henceforth, DOF), we observe response behavior associated with the universe of parking tickets issued in New York City between June 2011 and August 2013. In our core dataset, we analyze response behavior for 6.6 million tickets issued to 2 million unique passenger vehicle license plates, totaling \$424 million in fines and \$85 million in late penalties.

Our paper makes four main contributions. First, by exploiting exogenous variation in the timing of notification letters, we find evidence consistent with significant forgetting. Second, we find large and persistent differences across individuals in their baseline propensity to respond to tickets and in their reaction to reminders. 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 incentive-based interventions. Based on these findings, we discuss how accounting for effect heterogeneity might change one's approach to policy, and how one might use our analysis to target interventions at low-response types.

In Section 2, we outline a model of task completion designed to roughly match our field context. The model assumes that, each day, people draw a cost of paying the ticket from a known distribution. Optimal behavior entails seeking a convenient (low-cost) time to pay one's ticket, while also taking into account the deadlines and penalties that one faces. The model further incorporates two mechanisms thought to be important for task completion: present bias and forgetting. Using this model, we illustrate the impact of each mechanism, and clarify our strategy for identifying the impact of forgetting. We also describe why we cannot identify present bias (Heidhues and Strack (2019) provide a formal proof of nonidentification of present bias in a context similar to ours).

In Section 3, 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 impacts of forgetting and reminders exploits this variation.

In Section 4, we analyze aggregate response behavior in the OLD and NEW regimes. Figure 1 depicts daily hazard rates of recipients' first responses (i.e., $\frac{\#\text{ first response on day } d}{\#\text{ no response before day } d}$) as well as cumulative response rates (i.e., $\frac{\#\text{ first response on or before day } d}{\text{total } \#\text{ of tickets issued}}$). 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 the first letter that is consistent with forgetting. 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 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% versus 36%). 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.

While the patterns in Figure 1 are consistent with forgetting, to rule out alternative explanations 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

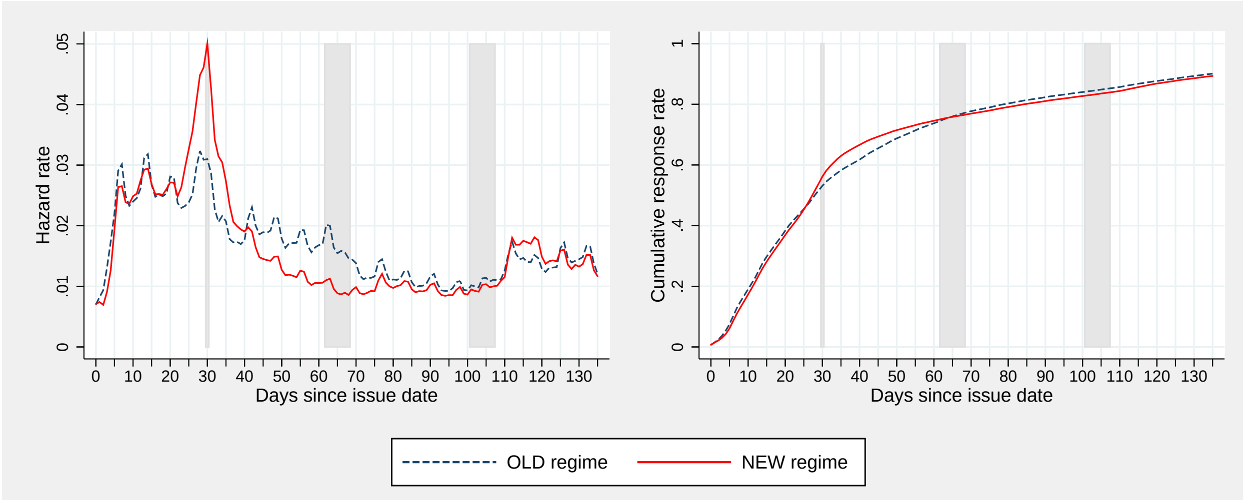

Figure 1: **Response Rates in OLD versus NEW Regimes.** Note: Daily hazard rates and cumulative response rates in OLD versus NEW regimes. All tickets have a first deadline at day 30, second deadline at days 62-68, and third deadline at days 101-107, indicated by the shaded areas (the latter two deadlines are a range because they depend on ticket-issuance day of the week). The first notification letter is received around day 40 (OLD) or day 20 (NEW). Based on 3,355,094 (OLD) and 3,020,357 (NEW) observations; see details in Sections 3 and 4.

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—especially that the content of the first letter hardly matters and that the second letter, which contains no new information, generates an additional response—seem to confirm that the letters served primarily as reminders.

Our demonstration of strong reminder effects is consistent with prior research (see below); however, the heart of our paper is an analysis of heterogeneity, to which we turn in Section 5. We first demonstrate the existence of large and persistent differences in behavior by showing that even crude measures of response behavior on one's past tickets are highly predictive of response behavior on one's current ticket (Figure 5). To investigate the nature of this heterogeneity, we estimate a mixture model in which we represent a type by a set of regime-specific hazard rates. We then 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 between 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% for high-response types, 60% for medium-response types, and 19% for low-response types. Moreover, our estimates of the aggregate impact of reminders mask 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%) for the high-response types, 12 percentage points (from 47% to 59%) for the medium-response types, but only 1 percentage point (from 10% to 11%) for the low-response types. Hence, the economic impact of reminders 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 are least likely to respond to deadlines and reminders and thus incur significant late penalties, disproportionately come from already disadvantaged groups. Specifically, for a subset of our data, we are able to match ticket-recipients' addresses 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 more 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) 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 a growing literature on task completion. The topic was not of central interest in economics until researchers started exploring sources of *suboptimal* delay. Initial theoretical work studied procrastination on tasks due to present bias (Akerlof (1991); O'Donoghue and Rabin (1999a,b, 2001)). Later theoretical work studied delay due to forgetting about (or inattention to) tasks (Holman and Zaidi (2010); Ericson (2011); Taubinsky (2014)). Ericson (2017) studies the interaction of present bias and forgetting.[2] Research on task completion can be divided into two categories. Most early work focused on

[1] Ghesla, Grieder, and Schubert (2019) 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] There is also work focused on the role of "mental goals" in inducing task completion (Koch and Nafziger (2011); Hsiaw (2013)), although this mechanism seems less relevant for our context.

one-time tasks with a deadline, such as filing one's taxes or writing a report. More recently, researchers have studied recurring tasks such as taking one's daily medicine or sticking to one's exercise regime. The two categories likely involve different (though overlapping) sets of mechanisms (Taubinsky (2014) explicitly compares forgetting in the two categories). Paying a parking ticket is an example of the former.

More importantly, our analysis contributes to the literature on reminders. We are not the first to demonstrate an impact of reminders in a field setting. Much of this research focuses on recurring tasks: 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 their scheduled payments on installment loans; and Karlan, McConnell, Mullainathan, and Zinman (2016) study the impact of monthly text messages reminding people to make their planned monthly deposits into commitment savings accounts. One study focuses on a one-time task: Chirico, Inman, Loeffler, MacDonald, and Sieg (2018) study the impact of sending reminder letters to property owners who were tardy in paying their property taxes.[3]

However, none of this research on reminders involves the type of heterogeneity analysis that we focus on—investigating heterogeneous types based on behavior as opposed to based on observable characteristics. Indeed, our heterogeneity analysis raises a key question for assessing the impact of reminders and other nudges: Does the nudge actually help the intended target? In our domain, the impact of reminders is an order of magnitude weaker for those who delay the most. Moreover, in our concluding Section 6, we calculate that, in the switch from the OLD to the NEW regime, 91% of the extra spending on earlier notification letters and 87% of the gains in terms of reduced penalties accrue to the higher-response types—i.e., not 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 we show that such targeting could be relatively easy to implement based on crude measures. This discussion further highlights the importance of analyzing heterogeneity in nudge effects prior to giving policy advice.[4]

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

[4]It is worth noting that our policy discussion does not assume that the low baseline response rates of the

2 An Organizing Model

To motivate our empirical analysis, we develop a model of task completion. In this section, we provide an overview of the model and its predictions; full details are included in Appendix 1. Our model can be thought of as a variant of the McCall (1970) job-search model. It is similar in structure to that used in Holman and Zaidi (2010), Taubinsky (2014), Ericson (2017), and Altmann, Traxler, and Weinschenk (2017), although our environment has multiple deadlines. Heidhues and Strack (2019) also use a similar structure to demonstrate the nonidentification of present bias in many domains—we return to their work at the end of this section.

Suppose that a person receives a parking ticket on day $d = 0$ with a fine amount f. On each day $d \in \{0, 1, ...\}$ she decides whether or not to pay the ticket. If she has not paid by an initial deadline d_1, a late penalty a_1 is imposed. Similarly, if she misses a second deadline $d_2 > d_1$, a second late penalty a_2 is imposed, and if she misses a third deadline $d_3 > d_2$, a third late penalty a_3 is imposed, after which more serious consequences occur. In our data, we observe hazard rates for a fixed set of deadlines and penalties that do not change over time.

After receiving a ticket, a person seeks a convenient (low-effort-cost) time to respond. Let c_d denote the realized effort cost on day d, drawn i.i.d. from some known (to the person) distribution F. Each day d she learns the realization c_d and then decides whether to pay the ticket on that day. The cost distribution F has a major impact on predicted behavior, and yet there is no obvious assumption to make about the nature of this distribution. Below, we highlight how the resulting flexibility gives rise to major identification issues, and we do so even while limiting attention to a simple two-parameter functional form $F(c) = v + c/w$, defined for $c \in [0, (1 - v)w]$.[5]

The person seeks to minimize her expected discounted total cost (effort cost plus monetary cost) with β, δ discounting as in Laibson (1997) and O'Donoghue and Rabin (1999a)

---

low-response types are suboptimal. Rather, it merely 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.

[5]This functional form is convenient because the two parameters capture two key aspects of the cost distribution: v captures the mass at (or, more generally, near) zero, which has a major impact on the level of hazard rates, and w captures the spread of possible costs, which plays a major role in determining the magnitude of spikes in hazard rates near deadlines.

with $\delta = 1$.[6] $\beta \leq 1$ represents a time-inconsistent present bias. We further make the natural assumptions that the effort cost c_d is experienced on day d (i.e., it is effort exerted now), while any monetary cost is experienced in the future (i.e., it requires forgone future consumption). We consider both sophistication (full awareness) and naivete (full unawareness) of future present bias.

We also permit that a person might forget about the need to make a payment. Each day, the ticket can either be on the mind—in which case the person actively decides whether to pay it—or off the mind—in which case the person necessarily does not pay the ticket. We let λ^Y be the probability that the ticket is on the mind tomorrow given that it is on the mind today, and λ^N be the probability that the ticket is on the mind tomorrow given that it is off the mind today. We also assume an exogenous probability Λ_0^Y that the ticket is on the mind on day 0 to reflect that people might not fully register receiving the ticket. Finally, we consider both full awareness and full unawareness of one's future propensity to forget.

Figure 2 illustrates hazard rates predicted by this model under different parameter values.[7] Consider first the solid orange line labelled "Baseline," which is the same in each of the four panels. This line illustrates the qualitative pattern of hazard rates for the case where there is no present bias ($\beta = 1$) and no forgetting ($\lambda^Y = \lambda^N = \Lambda_0^Y = 1$). Intuitively, the person faces a trade-off: she would like to pay the ticket before the next deadline (to avoid the penalty), but she would also like to find a convenient time. Well in advance of a deadline, it is safe to wait for a future low-cost day. As that deadline approaches, however, the incentive to pay rises. Once the deadline passes, the incentive to pay drops immediately, but then rises again as the next deadline approaches. Finally, the spike for the second deadline is larger because the penalty for missing it is larger.

The short-dashed pink line in each of the four panels reflects the direct impact of in-

[6] Specifically, if $\gamma_{d'}$ is the (monetary or effort) cost incurred on day d', the expected discounted total cost from the perspective of day d is $E\left[\gamma_d + \beta \sum_{d'=d+1}^{\infty} \delta^{d'-d} \gamma_{d'}\right]$. The assumption $\delta = 1$ seems reasonable when studying daily decisions, but it is not important.

[7] Figures 2 and 3 below assume $d_1 = 30$, $d_2 = 65$, $d_3 = 100$, $a_1 = \$10$, and $a_2 = \$20$, all of which (roughly) apply to all individuals in our data. It also assumes a fine amount of $f = \$65$, although the fine amount has little impact on qualitative behavioral patterns. Finally, we close the model by assuming that if a person delays beyond deadline $d_3 = 100$, there is an exogenous continuation cost $z = \$130$ (see Appendix 1 for details). Figures 2 and 3 present predicted hazard rates only through day 70 because later behavior becomes sensitive to assumptions about z.

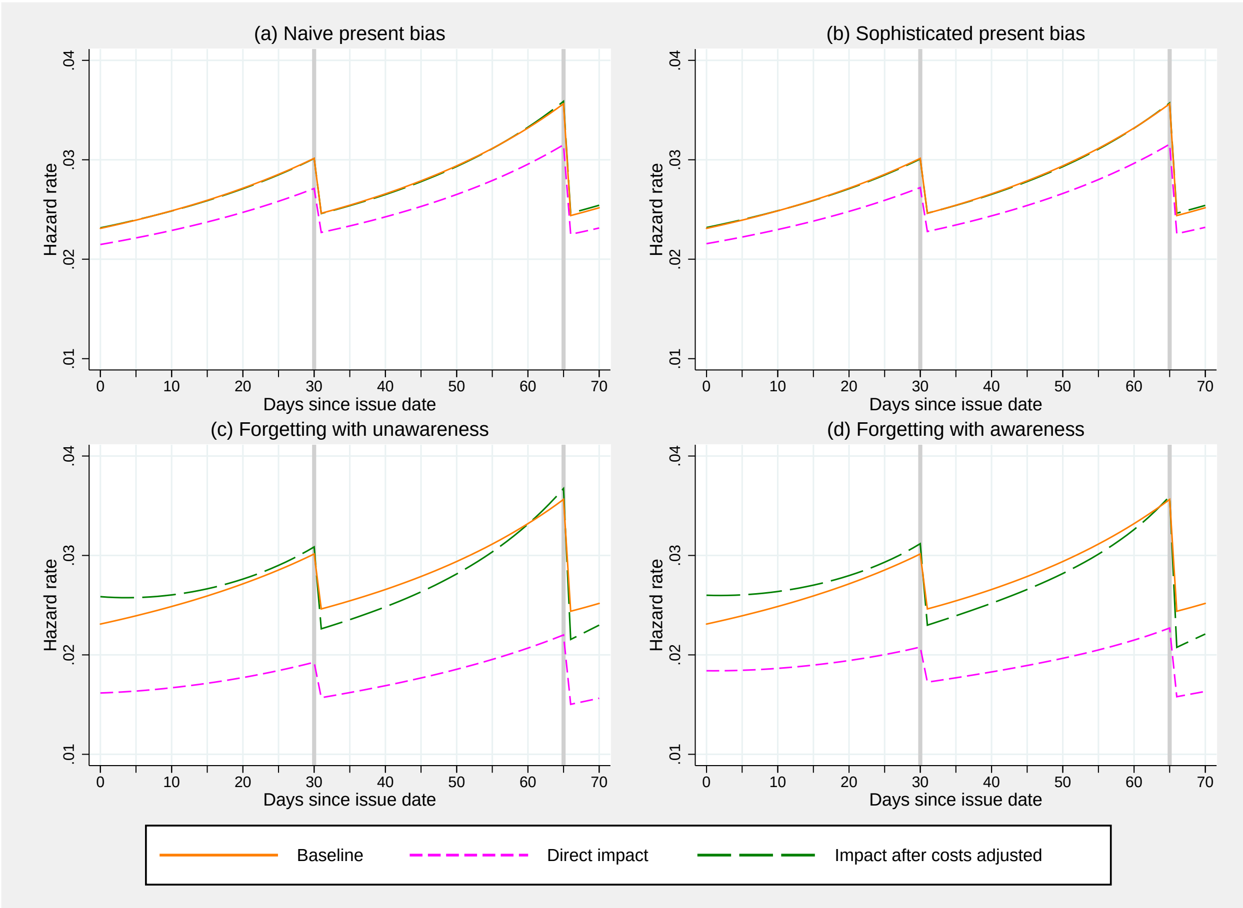

Figure 2: **Hazard Rates Predicted by Model under Different Parameters**. Note: Baseline, which is the same in all four panels, assumes $\beta = 1$ and $\lambda^Y = \lambda^N = \Lambda_0^Y = 1$, and uses cost parameters $v = 0.015$ and $w = 1750$. Relative to Baseline, Direct impact and Impact after costs adjusted assume $\beta = 0.8$ in panels (a) and (b), and $\lambda^Y = 0.98$, $\lambda^N = 0.05$, and $\Lambda_0^Y = 0.7$ in panels (c) and (d). In each case, Direct impact uses the same cost parameters as Baseline, while Impact after costs adjusted uses cost parameters that bring the curve as close as possible to Baseline.

troducing (a) naive present bias, (b) sophisticated present bias, (c) forgetting with full unawareness, and (d) forgetting with full awareness. Specifically, each mechanism is introduced while maintaining the same cost distribution as in the Baseline hazard rates. Clearly, each mechanism significantly affects behavior; however, the impact of each is qualitatively quite similar, for the most part leading to lower hazard rates but with the same qualitative pattern. Intuitively, present bias (panels (a) and (b)) implies a person overly weights immediate effort costs and is thus less willing to act now, but the person is still reacting to deadlines. Similarly, forgetting (panels (c) and (d)) means a person might not have the ticket on the mind and thus is less likely to act now, but anyone with the ticket on the mind is still reacting to deadlines. Finally, being aware of future delay—due to either sophisticated present bias or anticipated forgetting—makes a person more likely to pay now, but this also does not change qualitatively how a person reacts to deadlines (and, for present bias, the impact of sophistication is sufficiently small that it is not visible in Figure 2).

This similarity in responses suggests that it would be difficult to separately identify the mechanisms from one another. The long-dashed green line in each panel further highlights the identification challenge by presenting hazard rates under each mechanism after adjusting the parameters of the cost distribution to make the long-dashed line come as close as possible to the solid line.[8] In all four cases, we cannot perfectly replicate the solid line. It follows that, with enough data—and enough confidence that the model is correctly specified—one could separately identify the mechanism and the two-parameter cost distribution. However, for each mechanism, and even with the simple two-parameter cost distribution, the long-dashed line lies very close to the solid line and has much the same shape. Hence, in practice, one cannot identify the mechanism separately from the cost distribution if all we observe are reactions to a fixed set of deadlines and penalties.

Identifying a particular mechanism, then, requires some variation in the environment that primarily impacts behavior through that mechanism. In our context, the shift from the OLD regime to the NEW regime changed only the timing of the first letter sent to the ticket recipient. Under the model of forgetting, to the extent that this letter serves as a reminder that puts the ticket back on the mind (at least for some people), this regime shift should have a dramatic impact on hazard rates. To illustrate, Figure 3a presents predicted hazard rates under the model of forgetting with full unawareness for the case where a reminder letter is received on day 20 versus day 40. The figure assumes that the letter puts the ticket back on the mind for 50% of those for whom it had fallen off the mind.[9] The predicted contrast in hazard rates between the two regimes is qualitatively similar to what we observe in Figure

[8] As detailed in Appendix 1, the adjusted cost parameters are chosen to minimize the sum of the squared distances between the long-dashed and solid lines. For Figure 2, the adjusted cost parameters are: (a) $v = 0.016$ and $w = 1400$, (b) $v = 0.016$ and $w = 1420$; (c) $v = 0.029$ and $w = 700$, and (d) $v = 0.019$ and $w = 700$.

[9] The cost parameters in Figure 3 are the same as those used for the Impact-after-costs-adjusted line in Figure 2c (see Footnote 8).

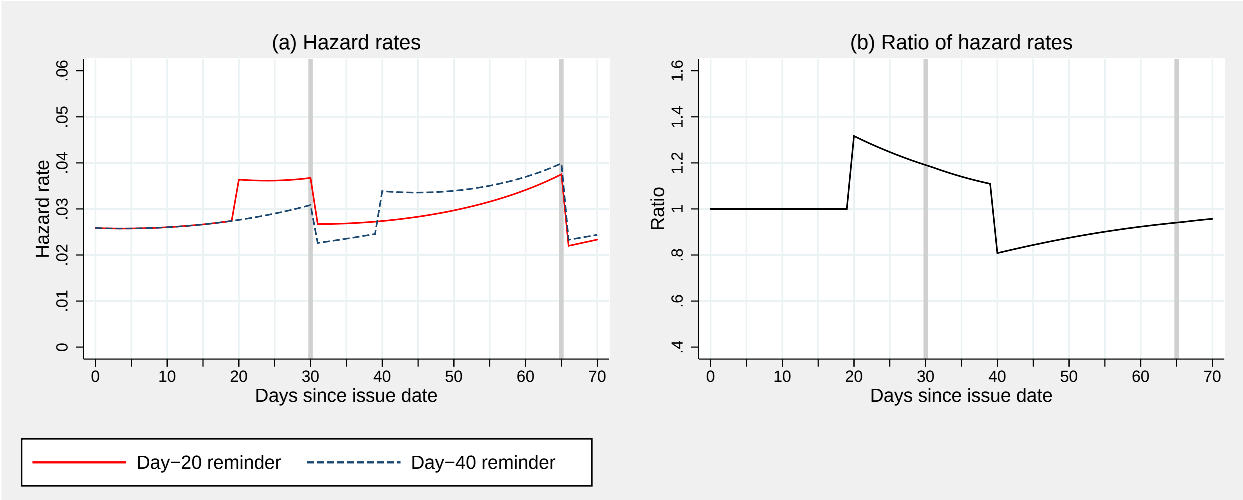

Figure 3: **Predicted Impact of Day-20 Versus Day-40 Reminder Letter**. Note: Panel (a) presents hazard rates assuming letter puts the ticket back on the mind for 50% of those for whom it had fallen off the mind. Both curves assume cost parameters $v = 0.029$ and $w = 700$, $\beta = 1$, $\lambda^Y = 0.98$, $\lambda^N = 0.05$, $\Lambda_0^Y = 0.7$ and full unawareness of forgetting. Panel (b) presents ratio of hazard rates in Panel (a).

1. (The predictions are much the same for the case with full awareness—see Appendix 1.)

Figure 3b depicts the predicted hazard rate given a day-20 letter divided by the predicted hazard rate given a day-40 letter. Under the model of forgetting with full unawareness, this ratio corresponds to the proportion of people with the ticket on the mind given a day-20 letter divided by the proportion of people with the ticket on the mind given a day-40 letter. As such, this ratio is a useful way to represent the impact of reminder letters. We return to this ratio in our empirical analysis to study the differential impact of letters on different types.

While seeing an impact of the timing of a reminder letter is evidence for the existence of forgetting, in practice we cannot identify specific parameters. Moreover, we also cannot identify whether people are aware of their future propensity to forget. Assessing such awareness requires observing some change in the environment that impacts perceptions of future forgetting or remembering—e.g., informing people that they will receive a future reminder. Unfortunately, our dataset does not contain this sort of variation.

Finally, while present bias is likely to be important in this domain, our dataset does not allow us to identify this mechanism. Indeed, the difficulty in separately identifying present bias from the distribution of effort costs in task-completion environments is only recently being appreciated. In parallel work, Heidhues and Strack (2019) independently provide a similar though more stylized example, and then formally prove in the context of a single deadline and a fully flexible cost distribution that present bias and the cost distribution cannot be separately identified. Identifying present bias might be possible if one's dataset contained independent variation in effort costs versus monetary costs, because effort costs are experienced immediately whereas monetary costs imply forgone future consumption, and thus present bias impacts reactions only to the former.[10] That said, our inability to identify present bias does not undermine our strategy for identifying forgetting.

3 New York City Parking Tickets

3.1 Data Description

The data come from the New York City Department of Finance (DOF), which handles most incoming revenue to the city. Appendix 2 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, location, and many other details. In addition, the data allow us to construct each ticket's history of "events" through late January 2014. Events are actions taken either by the ticket recipient (e.g., making a payment or contesting the ticket) or by 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 Appendix 2; 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% of the full dataset); (ii) they are not issued for parking violations (another 14%); and (iii) the plate owner does not

[10] Alternatively, one can estimate present bias in a task-completion environment by putting restrictions on the cost distribution, and then assessing robustness to varying those restrictions. Martinez, Meier, and Sprenger (2017) pursue this approach using data on tax-filing behavior.

have DOF's highest address verification level (another 21%). The core dataset only includes tickets issued to New York plates.

The first column of Table 1 presents descriptive statistics for the core dataset. The most common violations are for expired parking meter (36.2%), no parking due to street cleaning (26.2%), and parking in a general no parking zone (9.2%). The most common fine amounts are $35 (30.1%), $45 (23.9%), and $115 (23.0%). The vast majority (97.2%) of tickets are issued by parking-ticket agents.

The bottom panel of Table 1 presents the distribution of payment types for the 80.2% of tickets in the core dataset that have payments made by day 135. Four payment methods are available: online (53.8%), by mail (32.3%), by phone (2.8%), or in person at one of five DOF Business Centers (11.1%).

3.2 The OLD and NEW Regimes

Table 2 summarizes the timelines of key events under the OLD and NEW regimes (and also under the EXP regime, which we describe in Section 4.2). These timelines are identical except for one thing: DOF changed the timing and the content of the first notification letter. The rest of this section provides detail.[11]

3.2.1 Timeline in the OLD Regime

Tickets are issued on (what we define as) day 0 as a paper ticket that is placed, along with an orange envelope, on the windshield of the offending car. The ticket and envelope together provide information on 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. Appendix 13 contains sample tickets and the relevant part of the envelope, as well as samples of all notification letters described below.

[11] 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 Appendix 2 for details.

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 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 vehicles.

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 provides yet 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. 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 towing the owner's vehicles.

3.2.2 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. For these tickets, 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 from "NOTICE OF OUTSTANDING VIOLATION" 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 that 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 we discuss below, we conclude that it does not.[12]

3.2.3 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 4.2). 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 OLD to NEW to EXP regime, there is a modest shift from in-person and mail payments to online and phone payments.

4 Evidence of Forgetting

4.1 Behavior under 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

[12] 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 Appendix 2 for further details about each, and Appendix 4.2 for evidence that neither impacts our main conclusions.

Appendix 3 for the rationale behind this approach, along with descriptive statistics for type of first response.[13]

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 Section 1, Figure 1 depicts estimated hazard rates in the OLD and NEW regimes.[14] Prior to day 20, behavior is roughly the same under the two regimes, as expected given there is not yet any differential treatment between regimes. Then, when NEW letter 1 hits in the NEW regime, hazard rates 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 quickly converge again.[15]

Because the timing of deadlines and letters depends on the day of the week a ticket is issued, and as a means to quantify some of the differences seen in Figure 1, we also analyze behavior across six natural "periods" (we use these same periods when we estimate a mixture model in Section 5.3). Table 3 provides the definitions of each period along with the start and end dates.[16]

Table 3 presents estimated average daily hazard rates within each period and cumulative response rates through each period.[17] 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

[13] The main concern about our approach is that perhaps regime changes have more important effects on the type of first responses rather than on their timing. Our analysis in Appendix 3 suggests this is not the case.

[14] 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. Appendix 4.1 reproduces the major figures with 95% confidence bands.

[15] Appendix 4.2 demonstrates that these conclusions are robust to the settlement program and to Hurricane Sandy (see footnote 12), 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.

[16] Appendix 4.3 provides an alternative approach to controlling for the day of the week on which a ticket is issued, and shows that the main conclusions are robust to this alternative approach.

[17] 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 Appendix 5 for details of how Table 3 is created.

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% in the NEW regime relative to 36.1% in the OLD regime.[18] 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 3 shows that the cumulative response rate by the time letter 2 is sent (through period 5) is roughly the same under the NEW versus OLD regimes.

Our analysis in this section demonstrates that ticket recipients react strongly to the timing of the first notification letter, thus providing evidence of forgetting. We conclude this section by investigating the impact of characteristics of the ticket (Section 5.4 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 (Appendix 4.4 contains the figures). From this analysis, we draw two conclusions. First, while there are noticeable differences across sub-groups, there is nothing systematic that relates naturally to some underlying mechanism (e.g., while hazard rates are different for different fine amounts, higher fines are associated with neither higher nor lower hazard rates). Second, within each sub-group, the qualitative comparison between the OLD versus NEW regimes is essentially the same—that is, our results appear robust to observed heterogeneity in violation type, fine amount, and ticketing agency.

4.2 A Field Experiment (the EXP Regime)

After some initial comparisons of behavior under the OLD versus NEW regimes, DOF agreed 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 some "scary" 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

[18] These numbers are derived from Table 3—e.g., 45.6% = (65.27% − 36.18%)/(100% − 36.18%).

as described in Table 4.[19]

This design addresses three issues of interpretation in the comparison of the OLD versus NEW regimes. First, as discussed in Section 3.2.2, 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 versus NEW regimes, individuals in the *info* and *info scary* treatments received a modified version of 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 *scary* and *info scary* treatments received a modified version of NEW letter 1 that contains arguably scarier (or more forceful) language. Specifically, the letter had the following header in large bold-faced letters:

WARNING: PENALTY APPROACHING

DON'T MISS THE DEADLINE

[19] It was not feasible to vary penalty amounts or the timing of deadlines. Moreover, while we were able to vary the language and content of letters to some extent, we had to do so within the constraints DOF faced under current state and local legislation.

In addition, NEW letters $1i$, $1s$, and $1is$ all mention that failure to respond might result in one's vehicle being booted or towed, and in NEW letters $1s$ and $1is$ 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 $1i$, 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 from 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 4. 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% chance an EXP letter 1.5 was sent (independent of which NEW letter 1 was sent).[20]

4.3 Behavior under the EXP Regime

We analyze daily hazard rates in each of the eight experimental cells. The number of observations in the four cells without EXP letter 1.5 are 38,009 (1), 76,602 ($1i$), 38,199 ($1s$), and 38,156 ($1is$), and the number of observations in the four cells with EXP letter 1.5 are 16,060 (1), 32,041 ($1i$), 15,976 ($1s$), and 16,046 ($1is$).[21]

Figure 4a depicts hazard rates for the four experimental cells assigned not to receive an

[20] Each randomization was done by ordering plates alphanumerically and then assigning plates to treatments via a pre-set 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 4.3 are unchanged if we consider only plates that received exactly one ticket in the EXP regime (see Appendix 6.2).

[21] 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 Appendix 6.1 for details.

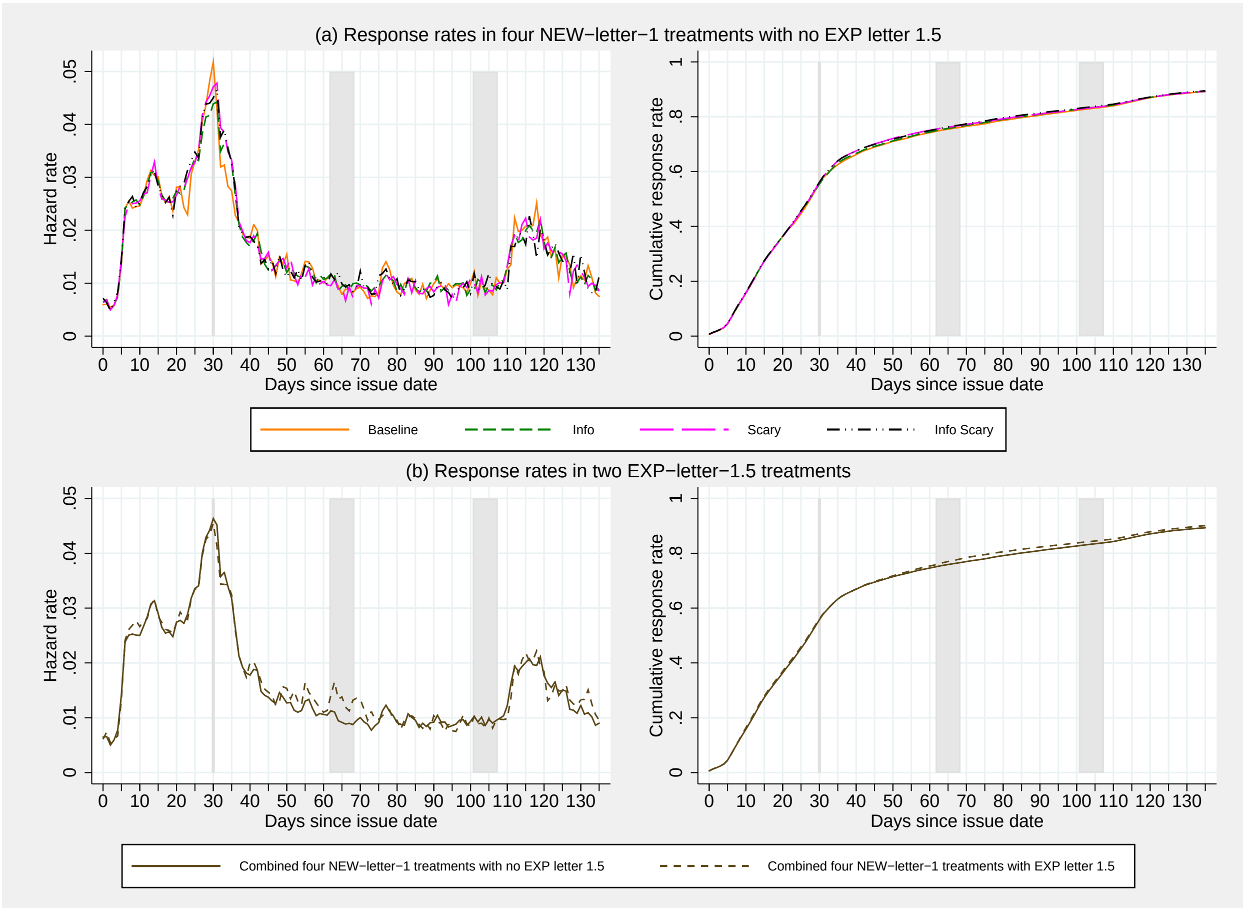

Figure 4: **Response Rates in EXP Regime.** Note: Daily hazard rates and cumulative response rates in the EXP regime. Figure 4a includes only the four experimental cells in which EXP letter 1.5 is not sent. Figure 4b pools the four experimental cells assigned not to receive, and to receive, an EXP letter 1.5.

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 versus NEW regimes are not driven by differences in information or language.[22]

Figure 4b, in which the four NEW letter 1 treatments are pooled together, reveals that EXP letter 1.5 has a noticeable impact, providing further evidence of forgetting. Even after getting a letter shortly after day 18, getting a second letter shortly after day 46 increases response rates. Quantitatively, the net hazard rate over the duration from day 48 through day 76 is 30.4% for tickets in cells assigned to receive EXP letter 1.5 relative to 25.7% in cells not assigned to receive that letter.

[22] Appendix 6.2 contains the analogue for Figure 4a 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.

To summarize, the findings from this field experiment—especially that the content of the first letter hardly matters and that the second letter, which contains no new information, generates an additional response—seem to confirm that notification letters serve primarily as reminders.[23]

5 Persistent Types and Heterogeneous Responses

5.1 Simple Evidence of Persistent Types

We begin with a crude approach that clearly suggests the existence of persistent types. First, we 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 an identical exercise for all plates that received exactly three tickets under the NEW regime.[24]

Figure 5 depicts these hazard rates. 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. Hence, response behavior on one's past tickets is highly predictive of response behavior on one's current ticket, indicating the existence of persistent types. Moreover, the behavioral differences between types are quite large.

In the next subsection, we pursue a more rigorous approach to identifying persistent types. Here, we started with this simple approach because it highlights how even crude statistics about recent behavior can provide strong signals about underlying types. Hence,

[23] Because EXP letter 1.5 contains information about deadlines 2 and 3 that NEW letters 1 and 1s do not, it is possible that the EXP letter 1.5 effect could partially be due to providing new information in those cells. We show in Appendix 6.2 that this seems not to be the case, as already suggested by the fact that the four versions of NEW letter 1 lead to almost identical hazard rates.

[24] In the OLD regime, the number of plates in each group are 56,035, 19,872, 20,429, and 41,559. In the NEW regime, the number of plates in each group are 55,783, 17,510, 17,166, and 35,111.

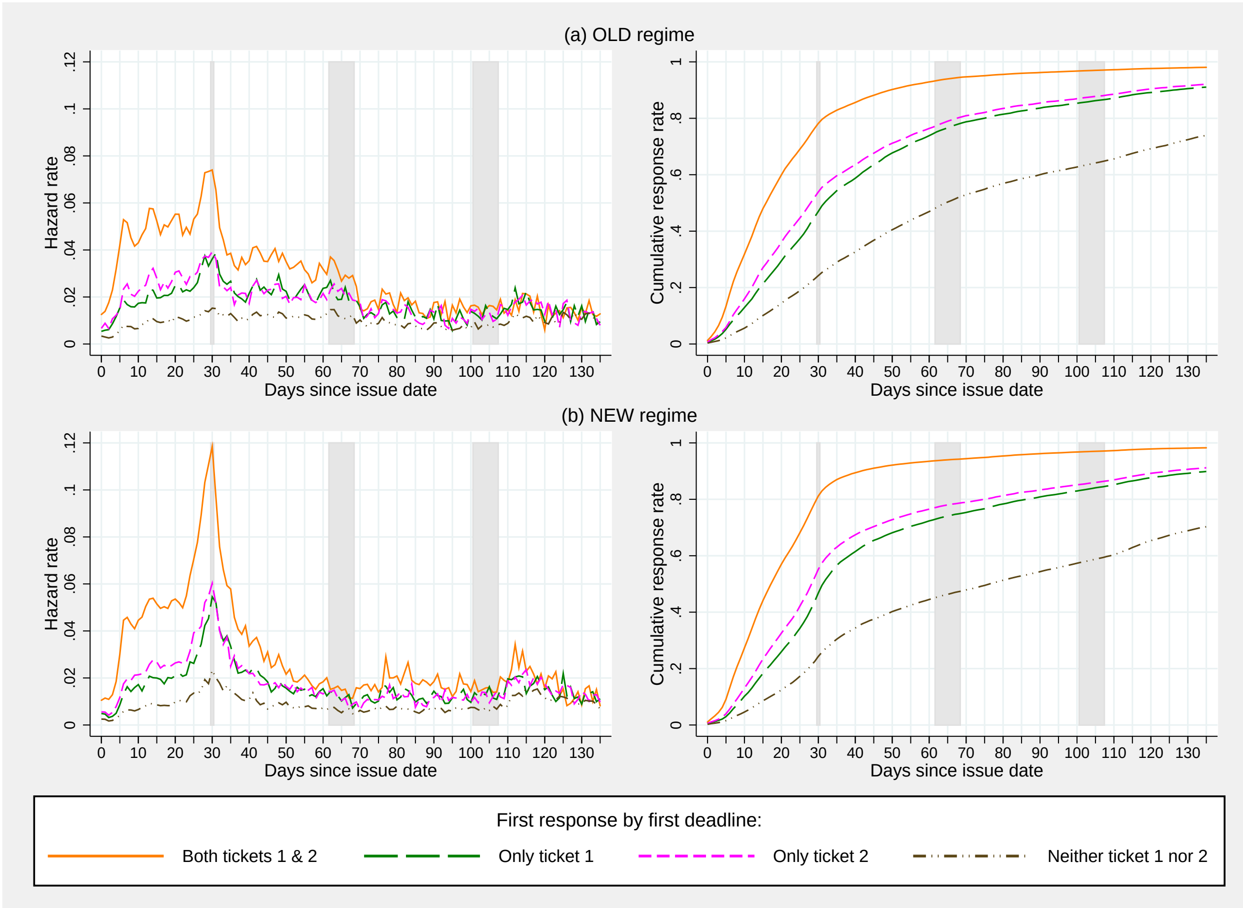

Figure 5: **Past Response Behavior Predicts Current Response Behavior.** Note: Daily hazard rates and cumulative response rates for third ticket, split by the ticket recipient's response behavior on two prior tickets. Based on recipients of exactly three tickets within a regime.

it may be easy for DOF (or another entity in some other context) to implement a policy targeting specific types. We return to this point in Section 6.

5.2 Estimating a Mixture Model

We now investigate more rigorously the nature of this heterogeneity. Conceptually, we imagine 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 unobserved types using this structure. In principle, if our conceptualization is reasonable, then this structural mixture model is the right approach. However, one might worry that some of our conclusions are driven by the particular structure that we assume. In Appendix 10, we provide an alternative approach. Specifically, 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 Appendix 10 that our conclusions are much the same, while also highlighting the limitations of that reduced-form approach.

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 Honore (1993), the key to our identification is that our dataset contains 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. We avoid that structure because one of our goals is to investigate whether different types exhibit different qualitative patterns in their hazard rates. Rather, we model an unobserved type as a survival function, and permit a finite number of types.

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 3. Hence, each type k is characterized by hazard rates $(p_1^k, p_2^k, p_3^k, p_4^k, p_5^k)$. 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

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

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 $\boldsymbol{\theta}^i$ is

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

where I is the identity function. Since type is unobserved, the likelihood that plate i generates observation $\boldsymbol{\theta}^i$ is

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

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

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

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, and (iii) within a type, the $p_t^k(\gamma)$'s are the same 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).[25] These assumptions are primarily made for reducing dimensionality. In Appendix 7.1, we provide evidence that, while not fully consistent with the data, these assumptions seem reasonable for our purposes.

In our main estimation, we use all plates that received $J \in \{3, 4, ..., 12\}$ tickets across the OLD and NEW regimes combined—657,890 plates that received 3,366,145 tickets.[26] 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 5 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.[27] Our analysis in

[25] Under the assumption that the number of tickets received J_i is independent of one's type k, the actual sample log-likelihood is $\sum_i \log\left(\Pr(J_i)\ell(\boldsymbol{\theta}^i)\right)$. Since $\Pr(J_i)$ is assumed to be independent of the model's parameters, it does not impact the estimation, and thus we suppress it from the sample log-likelihood.

[26] 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 receive multiple tickets within any one cell in the EXP regime.

[27] As described above, our estimation technique yields per-period hazard rates (reported in Appendix Table

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).[28]

5.3 Heterogeneous Behavior and Responses to Reminders

The estimates in Table 5 confirm the large and persistent differences across individuals suggested by Figure 5. In the estimated three-type model, 34% of the population is estimated to be *HR*s, 41% to be *MR*s, and 25% 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 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% for the *HR*s, 60% for the *MR*s, and 19% for the *LR*s. By the time the second letter is sent (on day 70-76), implied cumulative responses are 99%, 90%, and 35%.

These differences in cumulative responses 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%, respectively), and by the time Letter 2 is sent, the vast majority of the remaining population consists of *LR*s (78%). These strong selection effects help to explain why the aggregate hazard rates in Figure 1 show little reaction to the second and third deadlines.[29]

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 Appendix 7.2.

[28] 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 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 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 very strongly selects it over the $K = 3$ model).

[29] 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 are increasing over time. In Table 5, 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 $K = 2$, $K = 3$, or $K = 4$ estimates).

Table 5 also suggests that the aggregate results in Section 4.1 on the impact of reminders mask significant differential responses between higher and lower types. The *HR*s have a strong response: adding NEW letter 1 (relative to the OLD regime) increases average daily hazard rates from 7.47% to 10.43% in period 2 and from 5.00% to 9.85% in period 3. Using the ratio of the daily hazard rates as a proxy for the relative likelihood of the ticket being on the mind (as in Figure 3b), these estimates imply the ticket is 40% more likely to be on the mind under the NEW regime relative to the OLD regime in period 2, and 97% more likely in period 3. In contrast, the *LR*s have a weak response: adding NEW letter 1 increases average daily hazard rates from 0.52% to 0.56% in period 2 and from 0.51% to 0.60% in period 3. These estimates imply the ticket is only 8% more likely to be on the mind under the NEW regime relative to the OLD regime in period 2, and only 18% more likely 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 4.1 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%). In our estimated three-type model, the corresponding numbers are 14.8 percentage points for the *HR*s (from 73.0% to 87.8%), 11.8 percentage points for *MR*s (from 46.9% to 58.7%), and only 1.1 percentage point for *LR*s (from 9.6% to 10.7%). Hence, our estimates imply that, while reminders affect all three types, the economic impact is an order of magnitude larger for the *HR*s and *MR*s than it is for the *LR*s.

The estimates in Table 5 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 5, 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 $\boldsymbol{\theta}^i$ is type k is

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

In principle, we could then merely assign plate i to the type k that maximizes $\hat{\pi}(k|\boldsymbol{\theta}^i)$.

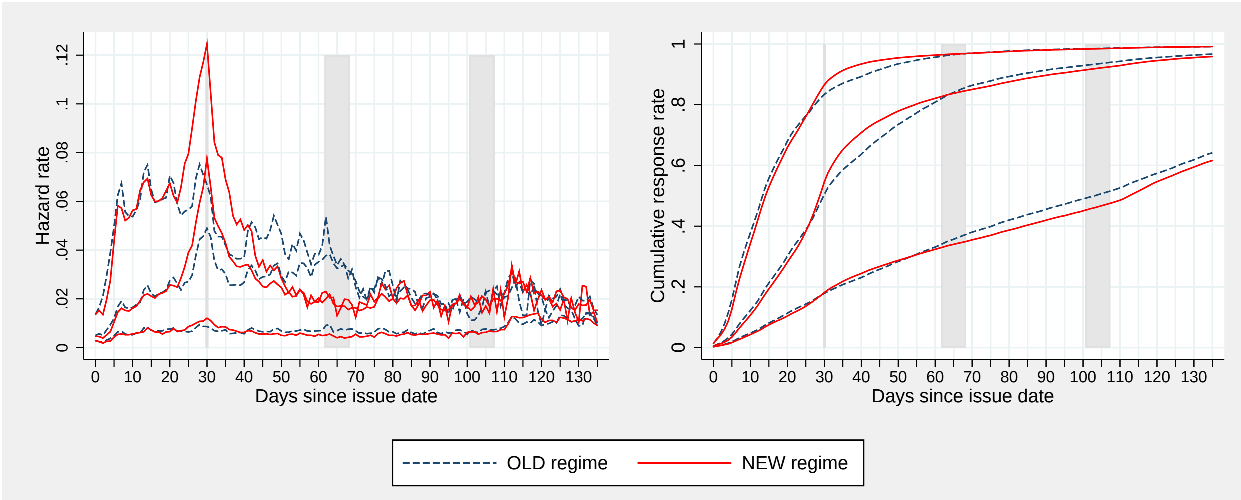

Figure 6: **Response Rates by Predicted Type.** Note: Daily hazard rates and cumulative response rates in OLD versus NEW regimes for each of the three predicted types (*HR*s, *MR*s, *LR*s) from the estimated three-type mixture model. Based on 582,065 summonses in the holdout sample.

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 $\hat{\pi}(k|\boldsymbol{\theta}^i)$ only if that k yields $\hat{\pi}(k|\boldsymbol{\theta}^i) > 0.60$. With this approach, we type 582,065 of the 657,890 plates (88.5%). Of these, 36.1% are assigned as *HR*s, 39.5% as *MR*s, and 24.4% as *LR*s.[30]

Using the holdout sample, Figure 6 depicts the type-specific response behavior in the OLD versus NEW regimes.[31] Figure 6 yields much the same message as Table 5. 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.[32]

To further highlight the heterogeneous responses to notification letters, we again use the ratio of the daily hazard rates under the two regimes as a proxy for the relative likelihood of the ticket being on the mind under the two regimes (as in Figure 3b). Figure 7 presents this

[30] The criterion $\hat{\pi}(k|\boldsymbol{\theta}^i) > 0.60$ is chosen to balance sufficient confidence in the typing against typing sufficiently many plates. If we instead require $\hat{\pi}(k|\boldsymbol{\theta}^i) > 0.50$, we type 99.3% of plates, whereas if we require $\hat{\pi}(k|\boldsymbol{\theta}^i) > 0.75$, we type 69.0% of plates. See Appendix 7.3 for details.

[31] In Appendix 7.3, Figure A5 illustrates that Figure 6 would look much the same for other cutoffs besides $\hat{\pi}(k|\boldsymbol{\theta}^i) > 0.60$; and Figure A6 presents Figure 6 with a separate panel for each type.

[32] Although Figure 6 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 Appendix 7.3 that this feature is most likely an artifact of Hurricane Sandy.

Figure 7: **Ratio of Hazard Rates in NEW versus OLD Regimes**. Note: For each type, the figure presents the ratio of the hazard rates depicted in Figure 6.

ratio for the hazard rates depicted in Figure 6. This ratio is roughly one until NEW letter 1 is received under the NEW regime. At this point, the ticket is more on the mind under the NEW regime for all three types, but more so for the *HR*s relative to the *LR*s. However, once OLD letter 1 is received in the OLD regime, the ticket is now more on the mind under the OLD regime for all three types, again more so for the *HR*s relative to the *LR*s.

The difference between types is much more pronounced in Figure 6 than in Figure 7. To understand this difference, recall that a type's response to reminders depends on a combination of (i) the extent to which reminders put the ticket back on the mind for members of that type and (ii) the type's propensity to respond conditional on the ticket being on the mind. Figure 6 shows that the hazard rate for *LR*s changes very little with the switch from the OLD to the NEW regime. This result could derive from two sources. First, it could be that (ii) is very small for *LR*s, and thus, even if reminders have a large impact in getting the letter back on the mind, they would have a limited impact on hazard rates. Second, even if (ii) is not small, it could be that the baseline propensity to remember is very low for *LR*s. If so, then even though reminders might get the ticket back on the mind for 30-40% more people (as suggested by Figure 7), it is still on the mind for very few people. Unfortunately, our data does not permit us to distinguish the magnitudes of (i) versus (ii).[33]

[33] Appendix 7.4 investigates behavior of the three types under the EXP regime. Each type exhibits the main aggregate findings from Section 4.3—specifically, the content of the first letter hardly matters, and the

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—e.g., perhaps the windshield ticket was never noticed—or to make it easier to pay—e.g., perhaps the original ticket/envelope with the instructions on how to pay was misplaced. 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 not to know about the ticket or most prone to have misplaced the original ticket/envelope. 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, or that tickets really do need to be paid. However, such effects would primarily apply for first offenders, and the results in this subsection derive from data on repeat offenders (Appendix Tables A5 and A6 also show that response patterns change very little across tickets for repeat offenders).

5.4 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. While for most of our data we have limited information about plate owners, 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. 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 $\hat{\pi}(k|\boldsymbol{\theta}^i)$'s) based on their response behavior under the EXP regime—this 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

second letter generates a noticeable additional response.

6 presents descriptive statistics for the Census variables along with the regression results.[34]

Regression (1) in Table 6 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 10th percentile ($18,973) rather than at the 90th percentile ($72,105). Regression (2) implies that the likelihood of that plate being an *LR* is 37% higher (36.4% versus 26.6%). Analogously, suppose all we know is that a plate comes from a Census block group with proportion black at the 90th percentile (0.81) rather than at the 10th percentile (0.00), with the remainder assumed to be white. Regression (4) implies that the likelihood of that plate being an *LR* is 63% higher (40.1% versus 24.6%).

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 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%).[35] These results are consistent with the income results in Table 6.

5.5 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 6, hazard rates for *LR*s

[34] See Appendix 8 for the details behind the analysis of this section. Appendix Table A12 presents these regressions when the dependent variable is the likelihood that a plate is an *HR*, and Appendix Table A13 presents logistic regressions. Both yield the same conclusions.

[35] We classify a car as "new" if it is 0-3 years old. We classify a make as "luxury" if the majority of its models appear in the *Consumer Reports* "Luxury Car" category. For more details on the vintage and make variables, see Appendix 8.

jump to their highest level at roughly day 110—to a daily hazard rate of roughly 1.1-1.2%—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 (this can be seen even more clearly in Appendix Figure A6d).

As discussed in Section 3, 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 indicates 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 is still no response within a few days, sold at auction. Starting on July 11, 2013, this instead meant 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.[36]

Figure 8 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.[37] 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% of responses from *LR*s follow a tow or boot. Second, the increase in hazard rates for the *LR*s pre-dates actual towing/booting: non-tow/boot hazard rates spike right at the time letter 3 is received, and before there is any significant towing/booting. While, unlike for our reminder results, there is no treatment-control comparison here, there is a natural interpretation of these findings:

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

[37] Figure 8 uses tickets for *LR*s in the holdout sample, and pools them across regimes (see Appendix 9 for regime-specific figures). All three sets of hazard rates use the same denominator for day d, which is the number of tickets without a response prior to day d. Hence, for each day, the aggregate hazard rate is equal to the sum of the tow/boot and non-tow/boot hazard rates.

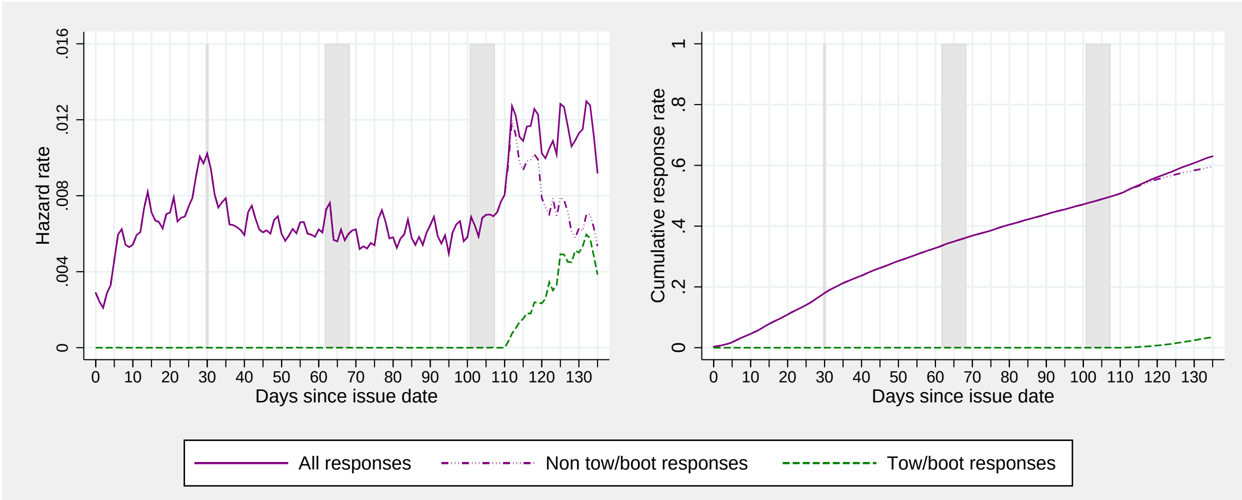

Figure 8: **Response Rates for the LRs.** Note: Daily hazard rates and cumulative response rates for the 141,959 plates predicted to be *LR*s, based on all 141,959 tickets received by those plates in the holdout sample. All responses are decomposed into those that followed towing or booting and those that did not.

*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 responding because they do not react to letters—here they seem to exhibit a strong response to letter 3. Rather, it seems that the letters just need to include 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.

6 Discussion

In this paper, we analyze the response behavior of New York City parking-ticket recipients. We identify forgetting as an important mechanism; demonstrate the existence of large and persistent differences in behavior across individuals; and find that those with a low baseline propensity to respond to tickets—arguably the natural target population for intervention—react least to reminders. Moreover, these low-response types, who incur significant late penalties, disproportionately come from already disadvantaged groups. We conclude with some broader takeaways.

A key implication of our analysis is the importance 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. Because many analyses of nudges are based on small samples, they focus on aggregate impacts, or at best study the impact of observable demographics. In contrast, our large and longitudinal dataset allows us to investigate unobserved heterogeneity based on *past 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% of tickets are sent first letters in the NEW regime versus 45% in the OLD regime—versus fewer first ($10) penalties are incurred—39% of tickets incur the first penalty in the NEW regime versus 45% in the OLD regime. While the NEW-versus-OLD trade-off involves a variety of monetary and non-monetary costs and benefits, we can quantify its direct aggregate monetary impact: In our core dataset, the NEW regime involves roughly $370,000 per year in extra notification-letter costs (the DOF cost per letter is approximately 50 cents; private communication) for a reduction of roughly $1.78 million per year in first penalties.[38]

Ultimately, it is for DOF (and the various constituencies of New York City) to decide whether this trade-off is worth it (of course, accounting for the indirect and non-monetary 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

[38] For details behind these calculations, see Appendix 11 Table A15.

for a reduction in late penalties of roughly \$1.45 million. In contrast, for the *LR*s, the NEW regime involves roughly \$35,000 in extra notification-letter costs for a reduction in late penalties of about \$210,000.[39] In other words, while the reduction in late penalties per extra notification letter is highest for the *LR*s, about 91% of the extra spending on notification letters and 87% 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 non-compliance.

Our analysis further suggests how one might design a policy of targeted reminders based on past behavior.[40] To illustrate, suppose 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 5 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 not be based 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 that, if our goal is to induce more timely payments from the *LR*s without imposing larger penalties on them, our analysis of their behavior suggests this alternative policy as one that might better achieve this goal.

A broader issue raised by our analysis is the relative value of socio-demographic 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 socio-demographic observables (for instance Beshears, Choi, Laibson, and Madrian (forthcoming) and De Neve, Imbert, Spinnewijn, Tsankova, and Luts (2019)).

[39] The type-specific numbers do not sum to the aggregate numbers because different samples are used in calculating them. Again, see Appendix 11 Table A15 for details.

[40] 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.

While we also find that socio-demographic 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 5 for which we also have Census variables.[41] A simple linear regression of the probability that a plate's third ticket is paid by day 30 on demographic variables (those in Table 6) 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 four indicator variables for response on the first two tickets by day 30 (corresponding to the four groups in Figure 5) 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 (personal income tax filing, bill payment, and so forth), not only to predict behavior within the same task, but also across tasks.[42]

Another important question is whether and how to reduce the regressivity of the current system. One area to focus on is the second and third deadlines. Despite having larger late penalties than the first deadline ($20 and $30 versus $10), they seem to have a smaller impact on behavior. While this smaller impact could in part be due to selection, it could also be that people are insensitive to the magnitudes of penalty amounts within this range.[43] Given their limited impact on behavior, one might consider reducing or even eliminating the second and third penalties as they are primarily incurred by LRs.[44]

Our analysis clearly identifies significant heterogeneity, and our theoretical model in Section 2 suggests possible sources of this heterogeneity: different distributions of effort costs, different levels of present bias, and different propensities to forget. One could imagine other

[41] Of the 263,465 plate-regime observations used in Figure 5, we have Census variables for 22,873 (the large reduction comes primarily from the fact that a small percentage of these plates also get a ticket in the EXP regime).

[42] 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 (pre-treatment consumption) is one of the two strongest predictors of the treatment effect.

[43] Some evidence on the impact of larger penalties comes from plates that received two tickets on the same day or two tickets spaced such that the first and second deadlines coincide, as in either case the effective penalty is larger. Such plates have much the same response patterns as other plates, suggesting that individuals are relatively insensitive to the magnitude of the penalty.

[44] Among typed plates in the holdout sample, the LRs incur 65% of the $20 penalties and 75% of the $30 penalties.

relevant forces not included in our model that also create heterogeneity. To optimize policy—especially targeted policy—one would want to understand the sources of heterogeneity. As highlighted in Section 2, however, identifying those sources would require variations that are not in our data. That said, important policy lessons, such as those discussed above, can be applied even without this knowledge.

Additional studies like ours would help to assess the generalizability of our findings. However, we see no a priori 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 and incidence 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, G. A. (1991): "Procrastination and Obediance," *American Economic Review*, 81(2), 1–19.

Altmann, S., C. Traxler, and P. Weinschenk (2017): "Deadlines and Cognitive Limitations," CESifo Working Paper No. 6761.

Beshears, J., J. Choi, D. Laibson, and B. C. Madrian (forthcoming): "Active Choice, Implicit Defaults, and the Incentive to Choose," *Organizational Behavior and Human Decision Processes*.

Cadena, X., and A. Schoar (2011): "Remembering to Pay? Reminders vs. Financial Incentives for Loan Payments," *Working Paper*.

Calzolari, G., and M. Nardotto (2017): "Effective Reminders," *Management Science*, 63(9), 2915–2932.

Chirico, M., R. Inman, C. Loeffler, J. MacDonald, and H. Sieg (2018): "Deterring Tax Delinquency in Philadelphia," NBER Working Paper No. 23243.

De Neve, J.-E., C. Imbert, J. Spinnewijn, T. Tsankova, and M. Luts (2019): "How to Improve Tax Compliance? Evidence from Population-wide Experiments in Belgium," Saïd Business School WP 2019-07.

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

——— (2017): "On the Interaction of Memory and Procrastination: Implications for Reminders, Deadlines, and Empirical Estimation," *Journal of the European Economic Association*, 15(3), 692–719.

Ghesla, C., M. Grieder, and R. Schubert (2019): "Nudging the Poor and the Rich - A Field Study on the Distributional Effects of Green Electricity Defaults," *Working Paper*.

Heckman, J. 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.

Heidhues, P., and P. Strack (2019): "Identifying Present-Bias from the Timing of Choices," *Working Paper*.

Holman, J., and F. Zaidi (2010): "The Economics of Prospective Memory," *Working Paper*.

Honore, B. E. (1993): "Identification Results for Duration Models with Multiple Spells," *Review of Economic Studies*, 60(1), 241–246.

Hsiaw, A. (2013): "Goal-Setting and Self-Control," *Journal of Economic Theory*, 148, 601–626.

Karlan, D., M. McConnell, S. Mullainathan, and J. Zinman (2016): "Getting to the Top of Mind: How Reminders Increase Saving," *Management Science*, 62(12), 3393–3411.

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

Knittel, C. R., and S. Stolper (2019): "Using Machine Learning to Target Treatment: The Case of Household Energy Use," NBER Working Paper No. 26531.

Koch, A. K., and J. Nafziger (2011): "Self-regulation through Goal Setting," *Management Science*, 113(1), 212–227.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112(2), 443–477.

Martinez, S.-K., S. Meier, and C. Sprenger (2017): "Procrastination in the Field: Evidence from Tax Filing," *Working Paper*.

McCall, J. J. (1970): "Economics of Information and Job Search," *Quarterly Journal of Economics*, 84(1), 113–126.

O'Donoghue, T., and M. Rabin (1999a): "Doing It Now or Later," *American Economic Review*, 89(1), 103–124.

——— (1999b): "Incentives for Procrastinators," *Quarterly Journal of Economics*, 114(3), 769–816.

——— (2001): "Choice and Procrastination," *Quarterly Journal of Economics*, 116(1), 121–160.

Tasoff, J., and R. Letzler (2014): "Everyone Believes in Redemption: Overoptimism and Nudges," *Journal of Economic Behavior and Organization*, 107A, 107–122.

Taubinsky, D. (2014): "From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice," *Working Paper*.

Table 1: Descriptive Statistics

	Core Dataset	OLD Regime	NEW Regime	EXP Regime
Total # 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%

Note: For all but payment type, percentages in each column are relative to the total number of tickets for that regime (listed on line 1). For payment type, percentages in each column are relative to number of tickets with payment made by Day 135.

Table 2: Timeline For Each Regime

Day	Event	OLD Regime	NEW Regime	EXP Regime
0	Ticket received	✓	✓	✓
19-21	NEW letter 1 (sent on Day 19 unless weekend)	--	✓	✓ (4 versions)
30	Deadline 1 ($10 late penalty)	✓	✓	✓
35-41	OLD letter 1 (sent on Tuesday)	✓	--	--
47-49	EXP letter 1.5 (sent on Day 47 unless weekend)	--	--	✓ (to 50%)
62-68	Deadline 2 (on Monday, $20 late penalty)	✓	✓	✓
70-76	Letter 2 (sent on Tuesday)	✓	✓	✓
101-107	Deadline 3 (on Friday, $30 late penalty)	✓	✓	✓
105-111	Letter 3 (sent on Tuesday)	✓	✓	✓

Note: The shaded areas indicate the communications corresponding to the regime-shift and experimental variations. See details in text.

Table 3: 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%

Note: See Appendix 5 for details of how average daily hazard rates and cumulative response rates are calculated.

Table 4: 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
	scary (20%)	NEW letter 1*s*	NEW letter 1*s* , EXP letter 1.5
	info scary (20%)	NEW letter 1*is*	NEW letter 1*is* , EXP letter 1.5

Table 5: Estimated Mixture Model with Average Daily Hazard Rates

	Type	π_k	Regime	p_1	p_2	p_3	p_4	p_5
K=1		1.000	OLD	2.26%	2.43%	1.99%	1.84%	1.30%
		----		(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
			NEW	2.15%	3.18%	2.84%	1.31%	0.87%
				(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
K=2	HR	0.640	OLD	3.70%	4.80%	4.42%	5.25%	4.32%
		(0.001)		(0.01%)	(0.01%)	(0.02%)	(0.02%)	(0.05%)
			NEW	3.45%	6.46%	7.30%	3.85%	2.60%
				(0.01%)	(0.02%)	(0.03%)	(0.02%)	(0.04%)
	LR	0.360	OLD	0.60%	0.83%	0.87%	1.06%	0.98%
		(0.001)		(0.00%)	(0.00%)	(0.01%)	(0.00%)	(0.01%)
			NEW	0.54%	0.96%	1.16%	0.83%	0.67%
				(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
K=3	HR	0.338	OLD	6.58%	7.47%	5.00%	5.69%	2.45%
		(0.001)		(0.02%)	(0.04%)	(0.06%)	(0.06%)	(0.11%)
			NEW	6.09%	10.43%	9.85%	3.52%	1.65%
				(0.02%)	(0.06%)	(0.11%)	(0.06%)	(0.10%)
	MR	0.413	OLD	1.50%	3.10%	3.29%	3.90%	3.57%
		(0.001)		(0.01%)	(0.01%)	(0.02%)	(0.02%)	(0.03%)
			NEW	1.35%	4.01%	4.98%	2.93%	2.18%
				(0.01%)	(0.02%)	(0.02%)	(0.01%)	(0.02%)
	LR	0.249	OLD	0.53%	0.52%	0.51%	0.63%	0.65%
		(0.001)		(0.00%)	(0.00%)	(0.01%)	(0.00%)	(0.01%)
			NEW	0.47%	0.56%	0.60%	0.49%	0.45%
				(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
K=4	HR	0.261	OLD	7.99%	6.46%	3.33%	4.99%	2.40%
		(0.001)		(0.03%)	(0.06%)	(0.06%)	(0.07%)	(0.12%)
			NEW	7.46%	9.74%	7.68%	2.98%	1.64%
				(0.03%)	(0.07%)	(0.12%)	(0.06%)	(0.10%)
	MHR	0.275	OLD	2.19%	5.73%	6.21%	6.37%	4.62%
		(0.002)		(0.01%)	(0.03%)	(0.05%)	(0.06%)	(0.13%)
			NEW	1.92%	7.36%	9.72%	4.71%	2.62%
				(0.01%)	(0.04%)	(0.08%)	(0.06%)	(0.10%)
	MLR	0.296	OLD	1.17%	1.51%	1.70%	2.54%	2.51%
		(0.001)		(0.01%)	(0.01%)	(0.01%)	(0.02%)	(0.02%)
			NEW	1.08%	1.91%	2.55%	1.94%	1.59%
				(0.01%)	(0.01%)	(0.02%)	(0.01%)	(0.02%)
	LR	0.169	OLD	0.38%	0.40%	0.35%	0.36%	0.39%
		(0.001)		(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)
			NEW	0.33%	0.40%	0.36%	0.27%	0.27%
				(0.00%)	(0.01%)	(0.01%)	(0.00%)	(0.01%)

Note: Estimated average daily hazard rates by period (p_t's) for each type, as well as estimated proportions of each type (π_k's). Standard errors in parenthesis. See Appendix 7.2 for details of the transformation from estimated per-period hazard rates to estimated average daily hazard rates.

Table 6: Descriptive Statistics and OLS Regressions for Census Variables

	Descriptive Statistics			Regressions			
				Dependent Variable: Likelihood Low-Response Type			
	Mean	10th pctl.	90th pctl.	(1)	(2)	(3)	(4)
Median household income	44,403	18,973	72,105				
ln(median household income)				-0.029	-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	0.032		-0.095	
				(0.034)		(0.026)	
Some College	0.22	0.13	0.31	0.013		0.219	
				(0.033)		(0.029)	
College or More	0.25	0.06	0.51	-0.085		-0.316	
				(0.026)		(0.012)	
Race							
White	0.50	0.06	0.93				
Black	0.25	0.00	0.81	0.167			0.191
				(0.009)			(0.006)
Asian	0.08	0.00	0.24	-0.047			-0.161
				(0.017)			(0.014)
Other	0.17	0.01	0.43	0.243			0.178
				(0.017)			(0.011)
Language							
English Only	0.54	0.19	0.86				
English Very Well	0.23	0.09	0.37	-0.055			
				(0.021)			
English Well	0.12	0.02	0.23	-0.129			
				(0.031)			
English Not Well	0.09	0.00	0.21	-0.094			
				(0.039)			
English Not At All	0.03	0.00	0.09	-0.085			
				(0.071)			
Transportation to Work							
Public Transportation	0.43	0.13	0.80				
Drive	0.46	0.12	0.72	0.122			
				(0.011)			
Other	0.11	0.02	0.22	0.035			
				(0.022)			
Constant				0.533	1.083	0.366	0.246
				(0.057)	(0.036)	(0.011)	(0.003)
Number of Observations		60,529		60,529	60,529	60,529	60,529
R^2				0.04	0.01	0.02	0.03

Note: Table presents descriptive statistics and regressions for 60,529 Unique Plates in 9,481 Census Block Groups. OLS regressions with standard errors clustered at the block group level.