

Breaking Habits

Daan van Soest[†] and Ben Vollaard[‡]

[†] CentER, TSC, Tilburg University. E-mail: d.p.vansoest@tilburguniversity.edu

[‡] CentER, TSC, Tilburg University. E-mail: vollaard@tilburguniversity.edu

January 5, 2019

Abstract

We provide rare evidence for how a temporary incentive can kick-start an improved habit. We conducted a natural field experiment involving 70,000 households in an urban area. We study the extent to which they separate their waste, a behavior that is prone to habit formation. The incentive consisted of a letter informing households that disposal of non-separated waste is illegal and punishable by a fine, followed by an intensive and highly salient crackdown of four weeks. The crackdown had a large and instantaneous effect on household behavior. Most of the effect was still apparent up to seven months later. The stable, higher rate of separating waste is likely to have been sustained by a new, improved habit: crowding out of intrinsic motivation by the transactional motive to avoid punishment is found to be only transitory. In contrast to the threat of punishment, two behavioral interventions, one conveying the injunctive norm, another one the descriptive norm, are not found to have any effect.

Key words: habit formation, intrinsic motivation, deterrence, crackdown, field experiment.

JEL Codes: D03, Q53.

* The authors would like to thank seminar audiences at London School of Economics, McGill University, Universiteit Leiden, SOFI Stockholm, Università degli Studi di Firenze, Erasmus University Rotterdam, and participants at the 2018 CEP/LSE-University of Glasgow Workshop on “The Economics of Crime and Policing” in London, the 2018 Maastricht Behavioral Economic Policy Symposium in Maastricht, the 2016 workshop on “Experiments for Environmental Policy Formation”, University of Stirling, the 2017 workshop on “Social preferences and environmental and resource economics”, University of Manchester, and at the 2018 CPB “Economics of the Circular Economy” workshop in The Hague for useful comments and suggestions. Daan van Soest gratefully acknowledges financial support by the European Commission via the Seventh Framework Program (FP7/2007-2013), grant agreement No. 613420 (Glamurs). Special thanks to Cees Paridaans, Piet van Oirschot and Edwin Haest from the Tilburg Department of Sanitation (BAT) for daily management of the field experiment, and to CentER Data for conducting the survey.

1. Introduction

Many decisions are made repeatedly. Think of how to travel to work or what to eat. Such decisions tend to become habitual: the influence of conscious thought processes diminishes and behavioral patterns become automatically activated as soon as a situational setting previously associated with the behavior is encountered. Habits economize on mental effort and as such contribute to efficient allocation of limited cognitive resources; they are a form of procedural rationality (Simon 1976). In the economics literature, habitual behavior has been defined as behavior that is affected by decisions made in the past (Becker and Murphy 1988, Hussam et al. 2016). More specifically, the net benefits from engaging in a certain behavior are assumed to increase with a greater stock of similar choices made in the past.¹

A large number of behaviors that are high on the policy agenda are habit-based, including dietary choices, energy consumption, and exercise. Policies targeted at these habitual behaviors are often found to have no or just a temporary effect. This is not surprising, since the weight of past behavior in current decision making dampens the impact of changes in costs and benefits in the present that are related to that decision. Consequently, habitual behavior is less sensitive to changes in incentives than non-habitual behavior. The habit-driven nature of many behaviors has been suggested to be a primary explanation for why numerous policy interventions fail to succeed. A recent review of the empirical literature of the effects of interventions on charitable giving, study performance, exercise, smoking cessation, weight loss and energy consumption only found a long-lasting effect on energy consumption – but the authors go on to show that the persistent effect is driven by purchase of technology rather than a change in habits (Brandon et al. 2017).²

Provided that it is not completely impossible to change habits – and there is some, albeit limited, evidence to that effect that we discuss later in this paper – the sticky nature of habit-driven behavior may also work to the policymaker’s advantage. If the policymaker manages to kick-start a new habit, then a long-lived stream of benefits will follow (Verplanken and Wood 2006). In this scenario, a one-off incentive targeted at maximizing initial take-up of the behavior is all that is needed, provided that the ‘training period’ is not too short. As discussed in Gneezy, Meier and Rey-Biel (2011), kick-starting a good habit in this manner may also avoid possible negative side-effects from extrinsic incentives on intrinsic motivation. An extrinsic incentive may transform behavior driven by intrinsic motivation into a transactional exchange. But if a habit sustains a new behavior, then removal of the incentive is unlikely to undo the initial behavioral response.

¹ See Gardner (2015) for a review of the psychological literature on habit formation.

² Calzolari and Nardotto (2017), not reviewed in Brandon et al. (2017), also find decay in the effect of a reminder by e-mail to attend the gym, even though the decay is slow, in particular for low-attendance users. Another study not included in the review, Gallani (2017), finds that a temporary incentive for hand hygiene in a hospital does not lead to persistent effects either.

In this paper, we study the effectiveness of a potent but temporary incentive aimed at improving households' habits to separate waste.³ We follow households for many months to see what happens after the incentive is removed. This context is ideally suited for our purposes for three reasons. First, our setting allows us to exclude an alternative channel by which a temporary incentive may have persistent effects: learning about the returns from the activity. As discussed in Gneezy et al. (2011: 194), once individuals experience the positive aspects of a healthy lifestyle for instance, perhaps their motivation will increase enough to help them continue their improved habits even without the extrinsic motivation. In contrast to exercise, separation of waste is not much fun in and of itself and in contrast to energy conservation measures, a financial benefit is absent in our study area because households pay a lump-sum fee for their waste disposal. In the absence of policy interventions, separating waste is typical prosocial behavior with no private benefits other than warm glow (Cecere et al. 2014).⁴ In addition, a setting with primarily social benefits lends itself well for a critical assessment of the possible crowding out of intrinsic motivation. Second, we can exclude purchase of technology as an alternative explanation for persistence. Unlike reducing energy consumption by purchasing more energy-efficient appliances, separation of waste cannot be mechanized.⁵ If we see a persistent change in separation of waste, we know that households have changed their behavior permanently. Finally, although successes have been reported for some interventions targeted at improved waste separation such as curbside fees for waste collection (Kinnaman 2006), it remains an area with a large potential for improvement, and thus a great need for habit-altering interventions. Increasingly, local authorities facilitate separate collection of various types of household waste, including organic waste, plastics, glass, etc. – but households are slow to follow (European Environment Agency 2014).

We conduct a natural field experiment involving 70,000 households in Tilburg, the Netherlands, the universe of households in the city serviced by curb-side collection. These households have two different 240-liter garbage containers, each consisting of two compartments. One container has compartments for residual waste and organic waste; the other container has compartments for plastics

³ The way households treat their waste is habitual in nature, given that past behavior has been found to be a strong predictor of current behavior (Cheung et al. 1999, Holland et al. 2006, Carrus et al. 2008). As worked out in Ouellette and Wood (1998), such a strong correlation signals the presence of a habit if the behavior was performed relatively frequently (daily or weekly), and under relatively stable conditions (in the same environment or circumstances, see Ge and Ho (2018) for a recent application in energy consumption). These conditions are likely to be met in our context. In addition, Prugsamatz et al. (2017) provide evidence for a high degree of automaticity when separating waste, a direct indicator of the presence of a habit (see Gardner, 2015).

⁴ The utility benefit to households from separating waste seems modest. Kinnaman (2006) reviews empirical work based on contingent valuation and reports that households are willing to pay on average €5.70 per month for recycling services (\$5.61 in 2006). It should be noted that in our context social approval for separating waste – another potential benefit – is absent because the activity occurs unobserved. Waste is collected in non-transparent containers featuring a lid that is generally closed.

⁵ See Brandon et al. (2017) for an overview of the literature on technology versus behavior within the context of household energy consumption.

and paper. The treatment consists of informing all households by way of a letter that not separating one's waste is illegal and punishable by a €90 fine, followed by a one-month intensive and highly salient enforcement campaign. During this month, enforcement officers open each and every garbage container put at the curbside for collection. If they detect a violation, they attach a brightly colored warning label to the container. After two warnings, households can be fined.⁶ We randomize the timing that Tilburg's 65 garbage collection routes, each servicing a little over 1,000 households, are to receive the treatment during an eight-month time window. We study both the immediate behavioral response to the treatment and what happens in the months after the incentive has been removed. To see how intrinsic motivation is affected by the treatment, we fielded a survey among 10,000 households with questions about the motivation to separate waste. The survey also included questions about means that facilitate separating waste such as receptacles for organic waste.

We find that receipt of the letter announcing the threat of being fined results in an immediate and sizable drop in residual waste collected and a concomitant increase in organic waste, paper and plastics collected. The response is enhanced by the one-month crackdown that starts three weeks later. A response *before* the crackdown is to be expected since households were not informed that actual enforcement would only start a few weeks later. The weight of residual waste collected is some 15 percent lower at the end of the crackdown compared to before receipt of the announcement letter; the combined weight of organic waste, paper and plastics is some nine percent higher. The behavioral response shows some decay after the end of the crackdown. The decay levels off within three months, however, leaving a stable nine percent drop in residual waste for another seven months.

Next to the threat of punishment, the treatment included two behavioral interventions that may have contributed to the behavioral effect. A persuasive appeal to separate waste in the letter sent to households before the start of enforcement conveyed the injunctive norm; the highly visible warning labels conveyed the descriptive norm, i.e. the degree of compliance among neighbors. To distinguish the effects of these two 'nudges' from the effect of enforcement, we administered two additional treatments. First, we devised another letter that *only* conveyed the injunctive norm – informing households what is considered acceptable social behavior, and how to implement it. This letter excludes any reference to the illegal nature of not separating waste and ensuing enforcement. We sent the letter to a random half of 21,000 households that had not been subjected to enforcement yet and find no effect on household waste separating behavior. Second, in 25 randomly selected routes, we strengthened the signal from the warning labels by attaching white labels to garbage containers when no wrongdoing was detected. This made the brightly colored warning labels stand out even more. We

⁶ Penalties for not separating waste exist and are enforced in several other European cities, such as [Lausanne](#), [Milan](#), [Rome](#) and [Varese](#). In addition, introducing a penalty for not separating waste is not as odd as it may seem given that penalties are a common way of altering habits in other contexts, think of fines for smoking outside designated areas and driving-under-influence.

do not find that increasing the visibility of ‘good’ behavior has any additional impact on behavior. Combined, the results of the two additional treatments suggest that households only responded to the threat of being warned or fined, not to the social norm nudges.

So why is it that a crackdown results in a *permanent* change in behavior? We hypothesize that households feel *forced* to change their behavior in the short run, causing them to gain experience in the required activity. With time, the fear of sanctions dissipates, but habits have changed, and the new behavior sticks. For more than 200 days, we observe that households’ daily decisions consistently result in a stable, higher rate of separation of waste than before the campaign. This holds for a behavior that is known to be susceptible to habit formation. If conscious decision making is behind the new behavior, then we should expect continuing decay after the end of the crackdown rather than a leveling off of the decay, just like any other study into the effects of crackdowns (e.g. Banerjee et al. 2017). Households discount past experiences with law enforcement (Lochner 2007, see also Sah 1991) after all, and pay less attention to the threat of punishment as enforcement becomes less salient over time (Dur and Vollaard 2018).

Our survey provides additional support for this hypothesis. We find a significant decrease in households’ intrinsic motivations to separate waste in response to the treatment, but this negative effect is short-lived. In other words, the transactional motive for doing the right thing – separating waste because it pays off in terms of avoiding punishment – is only transitory. Intrinsic motivations for separating waste quickly recover to their original levels after the crackdown. This suggests that the deterrent effect of the crackdown fully dissipates within a fairly short time frame, in line with the findings of earlier studies into crackdowns. With the deterrent effect gone, arguably it is the new habit that sustains the stable, higher rate of separating waste.

Technology does not stand in the way of interpreting the evidence as a change in habits. A greater reported use by treated households of in-house means to separate waste such as receptacles in the kitchen lowers the cost of separating waste, but still requires daily decision making, decisions that are likely to be habitual. Rather, the greater presence of these means indicate that treated households are in it for the long run. Households invest in this technology precisely because they wish to keep up a higher rate of separating waste in the foreseeable future.

The finding that a short intervention results in a long-lasting (or even permanent) improvement in environmental behavior plausibly has implications for policies outside the environmental domain as well. Introducing a penalty and strictly enforcing it around the time of the introduction with the aim of creating an improved habit may be applicable in contexts such as compliance with safety procedures, implementation of quality assurance procedures, etc. Related behaviors such as wearing protective gear and abiding with sanitary regulations are habitual in nature and enforcing a penalty for not abiding with the norm is feasible and likely to be acceptable.

Our paper contributes to four different strands of the literature. First, we contribute to the literature on habit formation. We provide rare evidence on persistence in the behavioral response to a temporary incentive that can be attributed to habit formation. As argued in Hussam et al. (2016) and Brandon et al. (2017), existing studies that show a persistent response to a temporary incentive often leave room for other explanations than habit formation, such as the purchase of a technology and the process of learning about the rewards from the activity. In our context, these two alternative mechanisms can be excluded. Within the specific context that we study, Holland et al. (2006) is closely related to our paper. They find that waste recycling habits at the workplace can be altered by first eliciting a personal plan how to deal with waste and then providing a clear situational cue that prompts the planned behavior. The cue came in the form of an eye-catching personal recycling box that was placed near the desk of employees. In response to the treatment, the amount of non-separated waste fell to almost zero for up to two months after the intervention. A related paper in a different context that also reports persistent effects is Yang and Long Lim (2017). They find that offering free pre-peak trips on the Singapore subway for 2.5 months led to a small but persistent increase in pre-peak subway use up to 7 months later.

Second, we add to the literature on crowding out of intrinsic motivation, in two ways. One, we show that a strong extrinsic incentive can indeed harm intrinsic motivation, but that the effect may fade quickly after the incentive has been removed. Uniquely, we directly measure indicators of intrinsic motivation rather than deduce changes in intrinsic motivation from the observed behavioral response. What our findings show is that any immediate negative effects on intrinsic motivation may not be indicative of what happens in the longer run. This is an important caveat to studies in this area that focus on the immediate effects after withdrawal of the incentive only (see Goswami and Urminsky 2017). Our findings are in line with other field studies such as those by Halpern et al. (2015) on smoking cessation and Jackson (2010) on academic performance. They find no evidence for negative effects on the desirable behavioral response after withdrawal of the incentive. Two, the results of our study underpin the argument that kickstarting a good habit stems concerns about crowding out of intrinsic motivation more generally. If habits sustain a new behavior, then any harm to intrinsic motivation may not lead to an undesirable response after the incentive has been removed (Gneezy et al. 2011).

Third, given our particular incentive to alter behavior, a short period of intensive law enforcement, we contribute to the literature on deterrence. As far as we are aware, we are the first to provide evidence that police crackdowns may have lasting behavioral effects when habits are altered in the process.⁷ The behavioral effects of law enforcement activity, in particular crackdowns, are generally believed to be short-lived. Based on a review of a great number of studies of police crackdowns, Sherman (1990)

⁷ Our treatment is related to an optimal policing strategy that takes the form of random, pre-announced crackdowns worked out in Eeckhout, Persico and Todd (2010). We do not provide a test of their idea, however, because we do not compare random crackdowns with other possible monitoring strategies.

finds evidence for short-term deterrent effects, but hardly any evidence for effects after the crackdown has ended. In line with his review, the time that the effects of speed limit enforcement on drivers' behavior continue after enforcement has ended, the so-called 'time halo', is found to be very limited (Elliott and Broughton 2005). Similarly, effects of efforts in various countries to increase seat belt use and limit driving-under-influence that focus on law enforcement activity mostly fade in one year (Ross 1984, Campbell 1988). Banerjee et al. 2017 evaluate an anti-drunk driving crackdown in India which is implemented in a randomized fashion. They also find decay in the behavioral effect after the end of the crackdown: at the estimated coefficient values it would take three months for the impact to fully dissipate. Existing studies regularly hint that potential offenders return to their old ways because the injunctive norm (e.g. do not drink and drive) was not internalized but do not point towards the possibility that deterrent effects do not stick because habits are difficult to change.

Fourth, our paper speaks to the literature on behavioral interventions or 'nudges' (Thaler and Sunstein 2008). The persuasive appeal by means of a letter and the use of white labels to increase the visibility of socially desirable behavior can be seen as social norm treatments. Our paper is fairly unique in testing the effectivity of social norm treatments alongside more traditional policy instruments. Both nudges were found to have no effect on household behavior (compare Schultz 1999), in contrast to the traditional policy of introducing a penalty.

The paper proceeds as follows. The next section provides an overview of the experiment, including the contextual background, the treatment and related behavioral mechanisms, and the data collection. In Section 3, we present the estimation results. Section 4 discusses what happens to intrinsic motivation habits and in-house technology. Section 5 concludes.

2. Experiment overview

A. Context

Our experiment targets 70,000 Tilburg households living in single-family homes, all of which are serviced by curbside collection.⁸ A household has two wheeled garbage containers of 240 liters each. Both containers have two separate compartments, each for a different type of waste. A container's lid is two-colored and indicates what should go where. The container with the grey and green colored lid has compartments for residual and organic waste; the container with the blue and orange colored lid has compartments for paper and plastic waste.⁹

⁸ 25,000 Tilburg households who live in multi-story buildings have collective waste disposal facilities that provide little opportunity for separating waste; they are not considered in this paper.

⁹ Separating other types of waste such as metal, glass, batteries, chemical waste, discarded appliances and textiles is facilitated by the municipality as well, but not by way of curbside collection.

On collection day, households wheel their container to a nearby pick-up point, typically within 20 meters from their home. A pick-up point is typically shared by about ten to twelve households (Figure 1).

The Tilburg Department of Sanitation divided the city into 65 collection routes, each servicing about 1,050 households. Thirteen garbage trucks cover one route on each of the five weekdays, hence the 65 routes. In our study period the container for organic and residual waste was serviced every week; the container for paper and plastic waste every other week. Households pay a lump-sum fee for garbage collection, and prior to the intervention there was no active enforcement of (proper) waste sorting.



Figure 1. Containers put at the curbside, ready for collection.

An analysis of residual waste collected by the Department of Sanitation shows that waste sorting prior to our experiment was still far from its potential. In terms of weight, the residual waste of a randomly selected sample of containers contained some 70 percent recyclable materials, including first and foremost organic waste such as discarded food, tea bags, and coffee grounds (about a third of all residual waste), and to a lesser extent paper and cardboard (7 percent) and glass (5 percent) (Eureco 2015). In terms of volume, recyclable plastics are an important fraction too. The goal of the Tilburg City Council is to reduce residual waste to 140 kilos per person annually by 2019. Given an average of 190 kilos of residual waste per person in 2014, this implies a 25 percent decrease within 5 years.

B. Treatment

The treatment consists of informing all households by way of a letter from the municipality of Tilburg that not separating waste is illegal and punishable by a fine of €90, followed by a one-month, intensive and highly salient enforcement campaign. Households receive the announcement letter three weeks prior to the crackdown (a copy of the letter is provided in the Appendix). The letter says that offending households risk a fine; it does not say anything about the start or end of the one-month crackdown.¹⁰ The letter is put in fairly harsh terms, but it also points out why separating waste is important ('it saves the environment and it keeps local taxes down') and includes a one-pager with the types of waste that are to be separated. The legal basis for the treatment is a municipality by-law stating that separating

¹⁰ The campaign lasted one month for 55 out of the 65 routes. For ten routes that were treated first, the campaign lasted an additional two weeks. For the first five routes that were treated, the additional two weeks were intended to leave extra room for learning about possible teething problems. For the second five routes, the two additional weeks were the result of political turmoil in the City Council about the treatment. Only once the turmoil subsided, the Department of Sanitation switched to the next series of routes, which happened to be after six rather than four weeks after the start of the campaign in these routes. In the sensitivity analysis, we show that excluding these ten routes from the sample does not affect our results.

waste is mandatory. Prior to the experiment, enforcement of this by-law was largely absent.¹¹ It seems safe to assume that most households were unaware that incorrect separation of waste was at all illegal.¹²

During the crackdown, enforcement officers employed by the Department of Sanitation opened each and every garbage container put at the curbside for collection, visually inspecting its contents (Figure 2). On occasion, when the officers suspected that incorrectly separated waste was hidden deeper down in the container, they would conduct a more thorough



Figure 2. Visual inspection of a waste container

inspection. In that case, the officers would search through the trash. If the officers detected a violation, then they attached a brightly colored warning label to the container. The labels were very visible, enhancing the salience of the campaign. The color of the label indicated how many times the household had received a warning (they did not reflect the severity of the violation). The first warning label was yellow, the second one orange. After two warnings, households received a red warning label and could be fined. Obviously, orange labels only appeared as of the second week of the campaign; red labels as of the third week. On the label, the officers indicated the nature of the violation, for instance the presence of recyclable materials in the residual waste compartment. The officers kept track of the number of violations per household on a handheld device by scanning the chip on the container that contains a unique identifier. As legal evidence, they also took a digital photograph of the contents of the container. Out of practical concerns, enforcement was limited to only one of the households' two containers – the one for residual and organic waste. The presence of recyclable materials (organic waste, plastics, glass, textiles, etc.) in the residual waste compartment and the presence of non-

¹¹ From conversations with management of the Department of Sanitation we learned that sanitation workers would sometimes point out to households that their waste was not separated correctly. In some cases, they attached labels with this message to waste containers; labels that resembled those designed for the experiment. Given that the job of sanitation workers is to collect waste rather than to enforce regulations, this was exceedingly rare, limited to only the most motivated of sanitation workers, limited to the most extraordinary cases of incorrect separation of waste, and very rarely led to fines.

¹² The surprise of warnings and fines for something like improper separation of waste may explain the public outcry that followed after the start of the treatment. A common complaint was that City Hall had to stay out of people's trash and that City Hall did not get their priorities right (see local newspaper *Brabants Dagblad*, '[Kritiek op 'klikopolitie' Tilburg](#)', January 21, 2015). See Benabou and Tirole (2011) for another reason why the introduction of material incentives to increase private contributions to the public good is often met with strong criticism.

biodegradable waste in the organic waste compartment were considered violations of the municipality bylaws.¹³

C. Behavioral mechanisms

Several mechanisms may explain the effect of the enforcement campaign. Deterrence is the first and most obvious mechanism, given the threat of punishment. The possibility of being fined increases the costs of not separating waste, making it more likely that households contribute to waste recycling. The fine of €90 (\$100) is fairly high; it is similar to a fine for driving 15 km/h too fast on a 50 km/h street. The probability of being caught is anything but zero. As we discuss below, on average some 10 percent of households were caught in the first week of the crackdown. If 80 percent of households do not correctly separate their waste, which is a guess, then the probability of getting caught is 12.5 percent.

Given that the norm is set very high – even the slightest mistake is considered a violation – the municipality preferred to have multiple warnings before imposing a fine. The warnings in and of themselves are likely to increase the costs of not separating as well, since they serve as a reminder of the threat and since every warning makes a fine more likely.¹⁴ The visibility of the warning labels may further increase the costs of not (properly) separating one’s waste because of reputational concerns about being exposed to neighbors as an offender (see Kahan 2005 on shaming). Even though all containers are identical, shaming is a possible channel because house numbers are clearly marked on containers to allow households to retrieve their own container from the pick-up point after it has been emptied. The visibility of the warning labels also serves to make law enforcement activity very salient, which helps households to quickly update their beliefs about the chance of getting caught compared to a situation of learning about the punishment by getting caught first (Dur and Vollaard 2018).

Whether a deterrent effect extends beyond the one-month crackdown is ambiguous. On the one hand, as we discussed in the introduction, the evidence tends to suggest that any effects of crackdowns are short-lived. Given the discounting of past experiences with law enforcement once the campaign is over, the expected penalty is likely to be adjusted downwards (Lochner 2007) and it may also become less salient, gradually lowering the incentive to comply over time. Moreover, intrinsic motivations to

¹³ Focusing enforcement on the container for residual and organic waste was also an efficient use of the Department of Sanitation’s budget. The general finding from regular checks of all waste categories at the waste treatment facility is that contamination of waste flows is not a concern for plastic, paper and organic waste. In other words, households fall short in terms of throwing too much waste in the residual waste compartment, not in terms of putting the wrong types of waste in the other three compartments.

¹⁴ Escalating punishment for repeat offenders can be seen as a separating contract that helps law enforcement to only fine those households that are least likely to increase efforts to separate waste in response to a simple warning. Households reveal themselves to be of the high offense propensity-type by not being deterred by the yellow and orange warning labels (cf. Polinsky and Rubinfeld 1991). Higher punishment for repeat offenders can also correct under-deterrence of households that are least likely to change their behavior in response to a warning label without the need of raising punishment across the board, which may be prohibitively costly and also not politically feasible (Polinsky and Shavell 1998).

separate waste may be lowered in response to the crackdown, which could also lead to a decay in the behavioral response once the incentive has been removed (Gneezy, Meier and Rey-Biel 2011). On the other hand, the treatment may have changed households' habits of separating waste. If the higher rate of separating waste has become an automatic response rather than a conscious decision, then this may halt or at least limit the decay in the behavioral effect – even in the absence of enforcement.

The treatment contained two other elements or 'nudges' that may affect household behavior. First, the letter announcing enforcement of waste separation regulations included a persuasive appeal regarding the desirability of waste recycling and information on the types of waste that were to be separated (see Appendix 1). Conveying the *injunctive* norm – what households should be doing – by way of an appeal may further increase the 'moral cost' of not separating waste, with the cost being the disutility from perceived social disapproval that comes with not complying with the norm. Obviously, introducing punishment for incorrect separation of waste sends the same message about the social norm; the persuasive appeal should be seen as additive.¹⁵ Second, the highly visible warning labels conveyed the *descriptive* norm, what other households are doing. The labels signaled the share of neighboring households that were caught for incorrect separation of waste. This signal may increase waste separation if households update their perception of the share of other households doing to right thing upwards (Schultz et al. 2007). If more households come to believe that they are not meeting the norm, then the treatment may further increase the moral cost of not separating waste. Whether any of the behavioral effects of these two elements of the treatment are persistent is not clear. It is not uncommon to see that behavioral interventions such as the ones we study only have short-term effects on behavior (Allcott and Rogers 2014). As households discard or forget about the persuasive appeal and the signal from the warning labels, they may return to the behavior that they showed before the intervention.

D. Randomization

We randomize the timing that a garbage collection route is to receive the treatment over a period of eight months. Randomization involved 60 of the 65 routes; the first five were pre-selected by the municipality because of implementation considerations.¹⁶ This phase-in design came down to randomizing the order in which the twelve remaining collection routes for a specific day of the week were to be treated (e.g. the twelve Monday routes; one Monday route was left out of randomization). To increase the chance that on average similar routes were treated at different times, we used stratified

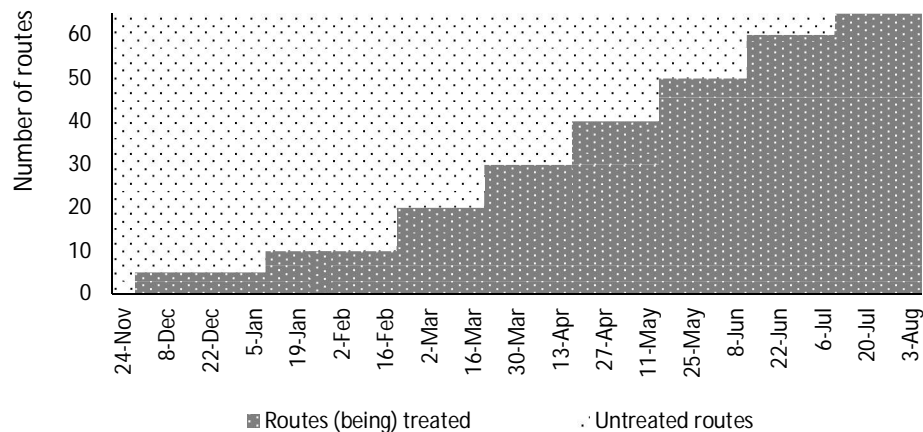
¹⁵ The very fact of saying that those who do not separate their waste are in violation of the law sends a message about society's values, and hence values, and hence about the norms according to which behavior is likely to be judged (cf. the literature on expressive law, see Benabou and Tirole 2011). The treatment turned something that was most likely perceived as voluntary into something that is mandatory after all.

¹⁶ Excluding the first five routes from randomization is another reason why routes that were treated first may be different from routes treated later. Recall that in the first ten routes the crackdown lasted six rather than four weeks (footnote 12). Excluding the first ten routes from the sample does not alter our findings (results available upon request).

random ordering of the timing of the treatment based on two variables: the share of residual waste in total waste pre-treatment (September–November 2014), and the percentage of terraced houses on a route in 2014. Both variables are indicators of the (lack of) potential for separation of waste. Once the ordering of the twelve routes for each day of the week was determined, we combined them into groups of five for a complete week of treatment in the Tilburg area. Most of the time, two teams of enforcement officers were available, which allowed for simultaneous treatment of two groups of five routes, otherwise only one group of five routes was treated.

The first enforcement campaign started on Monday, December 1, 2014; the last campaign started on Monday, July 13, 2015. By Friday, August 7, 2015, all 70,000 households in the 65 garbage collection routes had been treated. The phase-in of the treatment is shown in Figure 3.

Figure 3. Phase-in of the enforcement campaign



Note: The dates on the horizontal axis denote the start of enforcement campaign; the announcement letter was sent out three weeks earlier. As discussed in section 2 (D), the first five routes that were treated were outside randomization and the enforcement campaign in the first ten routes lasted six rather than four weeks.

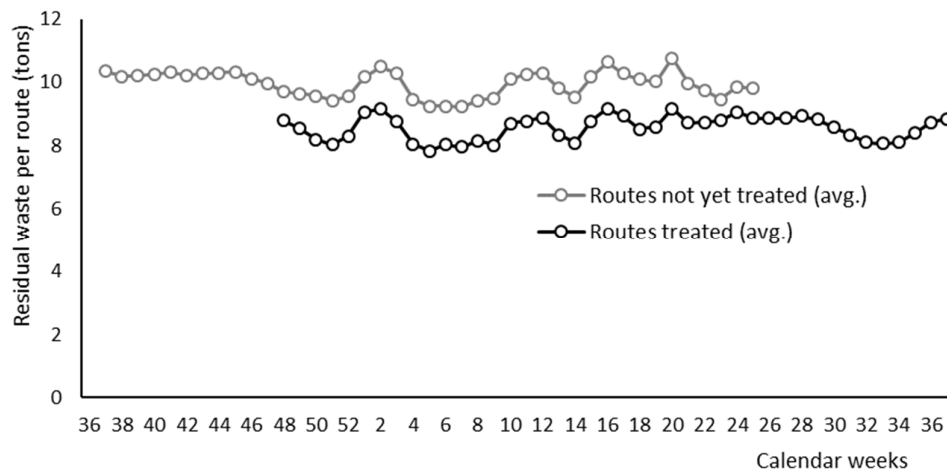
E. Data collection

The Department of Sanitation provided us with administrative data on the weight of the four fractions of waste collected by way of the compartmented containers for each week in which the fraction was collected. Waste is weighed when a garbage truck empties its contents at the waste treatment facility, which is why our unit of analysis is the garbage collection route rather than the individual household. The data are either weekly (residual and organic waste) or bi-weekly (plastic and paper waste). The data cover a period of 15 months (September 1, 2014 – December 1, 2015); three months prior to the start of the enforcement campaign in the first set of routes and four months after completion of the campaign in the last set of routes. Only for a few weeks, scattered over the 15-months period, data are missing (less than 1 percent).

When estimating the treatment effect, we assume that if households decide to throw more of their waste in the compartments for paper, plastic and organic waste rather than the compartment for residual waste, then they do so correctly. In other words, we interpret reductions in residual waste as evidence of improved waste sorting, not by more residual waste ending up in the compartments for the other three fractions. This is in line with the results of regular checks by the Department of Sanitation of all four waste categories at the waste treatment facility. Contamination of waste flows is not found to be a concern for plastic, paper and organic waste. In other words, households fall short in terms of throwing too much waste in the residual waste compartment, not in terms of putting the wrong types of waste in the other three compartments. Consequently, measurement error in our outcome variables is likely to be small, but perhaps not zero. For instance, toys made of hard plastic may end up in the plastic compartment even though they should not.

Figure 4 plots the weight of residual waste collected averaged over routes that have not been treated yet and over routes that are being treated or have been treated. Differences between the treatment and control group reflect a treatment effect, which may or may not vary with calendar time, time since start of the treatment, and route, but possibly also a composition effect given the phase-in design. Residual waste shows some peaks and troughs related to specific events such as holidays, but overall the variation is fairly limited.

Figure 4. Average weight of residual waste collected, by treatment status and by calendar week



Note: To smooth the weekly data, three-week moving averages are presented. The horizontal axis shows calendar weeks up to the start of the second enforcement campaign, which we discuss later.

Table 1 provides summary statistics for the period before any route was treated split by the 30 routes that were to be treated first and the remaining 35 routes. This split indicates whether randomization of the timing of the treatment was successful in obtaining balance. The table includes the five routes that were hand-picked to received treatment first; excluding these routes does not affect the results. The only variable for which we find a statistically significant difference between the two groups is residual

waste. The difference is small, however (244 kilos on an average of more than 10,000 kilos), and is statistically significant because of the large number of observations.¹⁷

Table 1. Descriptive statistics for baseline period

	30 routes			35 routes			p-value difference
	treated first			treated last			
	mean	st.dev.	N	mean	st.dev.	n	
Residual waste (tons)	10.13	1.40	357	10.37	1.33	417	0.01
Organic waste (tons)	3.50	1.77	357	3.44	1.64	417	0.82
Plastic waste (tons)	0.51	0.54	179	0.45	0.54	210	0.26
Paper waste (tons)	2.64	2.74	179	2.32	2.81	210	0.25
Value of home (euros)	205,511	55,214	30	211,301	63,644	35	0.70
Terraced houses (%)	74.4	20.1	30	74.7	17.8	35	0.94

Note. Observations by garbage collection route and by week (residual waste, organic waste), by two-weekly period (plastic, paper) or time invariant (value of home, terraced houses). Baseline period is September-November 2014.

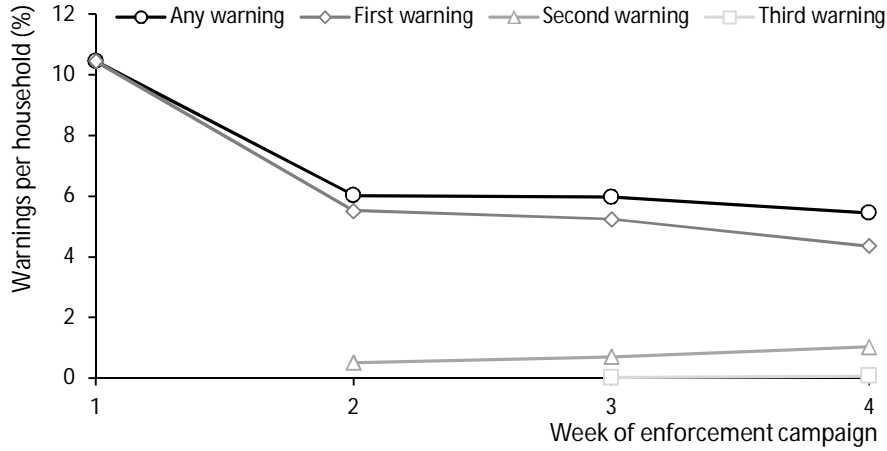
F. Treatment compliance

The phase-in of the treatment was implemented as planned. For all routes together, 19,049 warnings were issued during the four weeks of the enforcement campaign, of which 17,612 first warnings, 1,395 second warnings and 42 third warnings.¹⁸ Figure 5 shows the trend in the number of warnings per household over the four weeks of the campaign. In the first week, more than 10 percent of households received a yellow warning label, on average. The intensity of the enforcement activity remained constant during the campaign, and yet the number of warnings dropped sharply after the first week. This suggests that households changed their behavior in response to experiencing law enforcement activity (in line with Lochner 2007). Another indication for this process of learning through experience is the relatively small number of second and third warnings. Once a household received a warning, it is unlikely to be caught again. On average, the chance that a household received a second warning in the second week conditional on having received a first warning in the first week is less than five percent. If these households would not have changed their behavior in response to the first warning, then this chance should have been about twice as high.

¹⁷ Not only is the difference in the means small, it is also of no consequence since we use the difference-in-differences estimator in our analysis of treatment effects.

¹⁸ These totals exclude the warnings issued during the additional two weeks of the enforcement campaign in the first ten routes.

Figure 5. Warnings issued per household (%), average over 65 collection routes



3. Results

A. Event time model

We specify an event time model that non-parametrically estimates the effect of the treatment on the various waste fractions – residual, organic, plastic and paper waste. The event time model allows us to graphically evaluate changes in the treatment response. We estimate the following equation:

$$(1) \ln W_{i,t} = \sum_{\tau=-T}^T \alpha_{\tau} E_{i\tau} + Z_{i,t} \beta + \sigma_i + \lambda_t + \varepsilon_{i,t}.$$

The dependent variable $W_{i,t}$ is the weight in tons of waste collected in route i in week t . As explanatory variables, we include a vector E_{τ} of event time indicator variables. Event time is measured relative to the week of introduction of the treatment in a route, with event time being equal to 1 in the week that the announcement letter was sent out. The indicator variable E_{i1} equals 1 for the week that the treatment is introduced in route i and is zero otherwise; E_{i2} equals 1 for the first week after the introduction in route i and is zero otherwise, and so on. We leave out the indicator variable for event time 0, i.e. the last week before the start of the treatment. Event time spans a period of 42 weeks before and 55 weeks since the announcement letter was sent out. Given the small numbers of observations at the tails of the event time distribution, we bin the $E_{i\tau}$ by creating a single indicator variable for the end periods for every route i .¹⁹ Hence, α_{τ} indicates the percentage change in waste collected in week τ relative to $\tau=0$, the last week before the start of the treatment. Given large differences in baseline levels of waste collected between routes, and the resulting positive skewness in the data, we estimate the effect in percentage terms by taking the natural logarithm of the dependent variable. For the types of waste that

¹⁹ For the weekly data, we bin observations for 15 weeks at the left tail of the distribution and for 15 weeks at the right tail of the distribution. For the biweekly data, we bin observations for an equivalent number of two-weekly periods at the left tail and the right tail of the distribution.

are collected biweekly, plastic and paper, subscripts t and τ in Equation (1) stand for two-week periods rather than one-week periods.

$Z_{i,t}$ is a vector that denotes two other treatments randomized at the level of garbage collection routes that may affect household behavior. We discuss these other treatments in section 4. Equation (1) includes route-fixed effects σ_i to account for time-invariant, route-specific factors that affect household waste production. Equation (1) also includes calendar-week fixed effects λ_t to account for common shocks in household waste production across the Tilburg area, such as seasonal patterns, holidays, etc. The treatment rolled out in waves over time, and hence the common shocks in the outcomes can be isolated from effects of the treatment on the outcomes. $\varepsilon_{i,t}$ is the idiosyncratic error. To allow for serial correlation in observations by route, we cluster the standard errors at the level of routes using the Huber/White/sandwich-estimator.

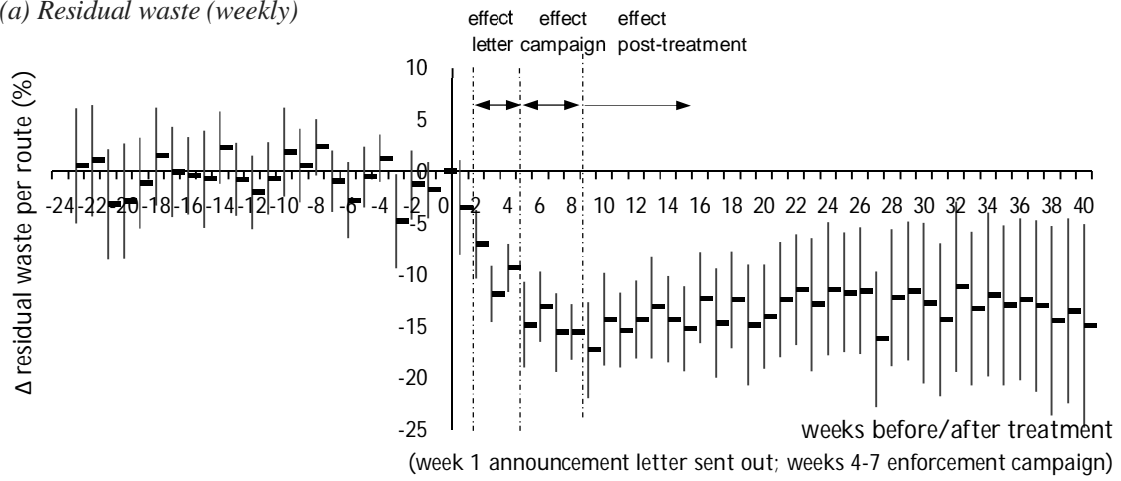
Figure 6 plots the weekly or two-weekly coefficients, α_τ , from estimating Equation (1) for the four fractions of household waste. Event time is plotted on the horizontal axis. The vertical axis shows the average percentage change in tons of waste collected relative to the last week or two-weekly period before the announcement letter was sent out.

The topmost graph shows that the weight of residual waste collected by the Department of Sanitation drops in the week after receipt of the announcement letter at event time 1. Given the weekly routine of waste collection, the treatment response is always lagged one week, which is why the effect only materializes in event time week 2. The enforcement campaign, which takes place in weeks 4-7 and hence directly affects recycling behavior in weeks 5-8, leads to a further drop in residual waste collection. Residual waste is down by some 14 percent on average. After the end of the crackdown, the behavioral response shows some decay (from week 9 onward). The decay seems to level off after about three months, however, leaving a persistent 9 percent drop in residual waste. Due to a declining number of observations at the tail of the event time distribution, the confidence bounds increase at high values of event time, however. Consequently, simply eyeballing the graph is not going to tell us whether the decay indeed levels off. In part (B) of this section, we provide statistical tests and a further discussion of the persistence in the behavioral response.

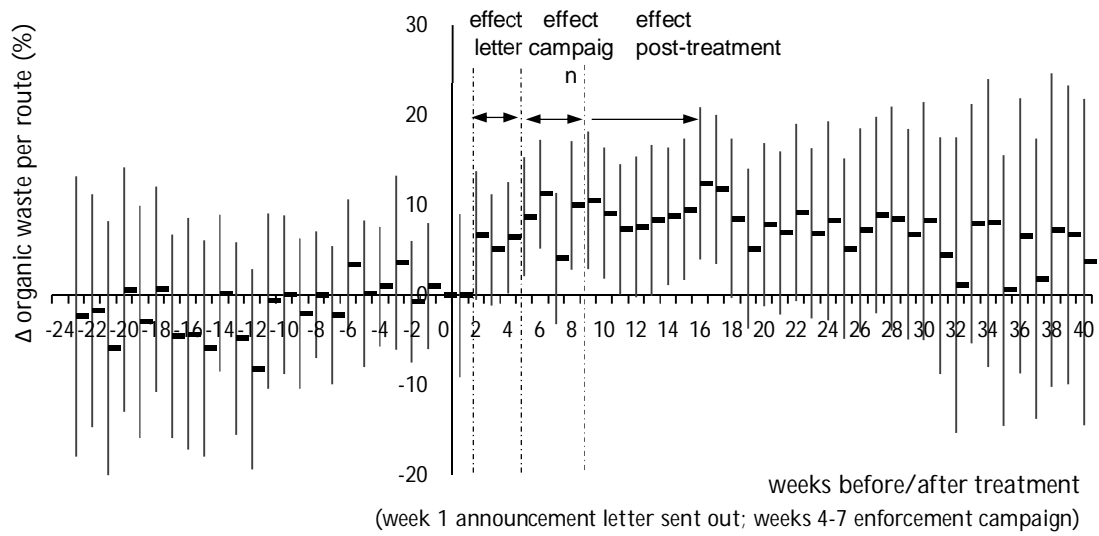
For organic waste (shown in the second panel of Figure 6) we find an increase of some 10 percent during the enforcement campaign relative to the week before the start of the treatment, on average. This amounts to some 0.4 tons of organic waste extra per route. In other words, out of a total drop of 0.9 tons in residual waste on average per route (9 percent of 10 tons), 0.4 tons or 40 percent is accounted for by an increase in organic waste. Thus the weight of separated fractions other than organic waste must also have increased. Effects beyond the garbage container that was inspected – the container for residual and organic waste – are to be expected given that plastics, paper and many other types of waste have to be kept out of the residual waste compartment as well.

Figure 6. Percentage change in waste collected relative to the last week before treatment

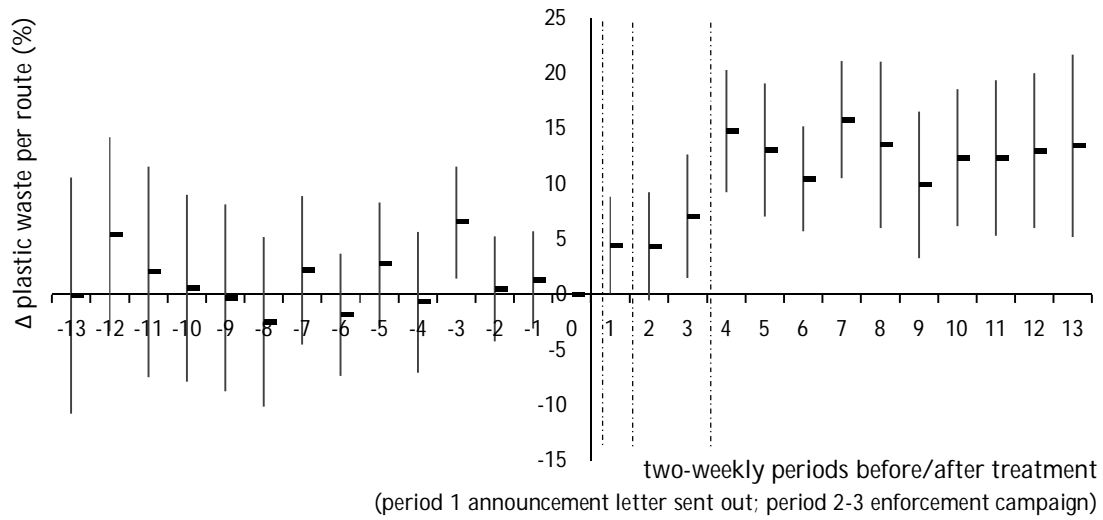
(a) Residual waste (weekly)



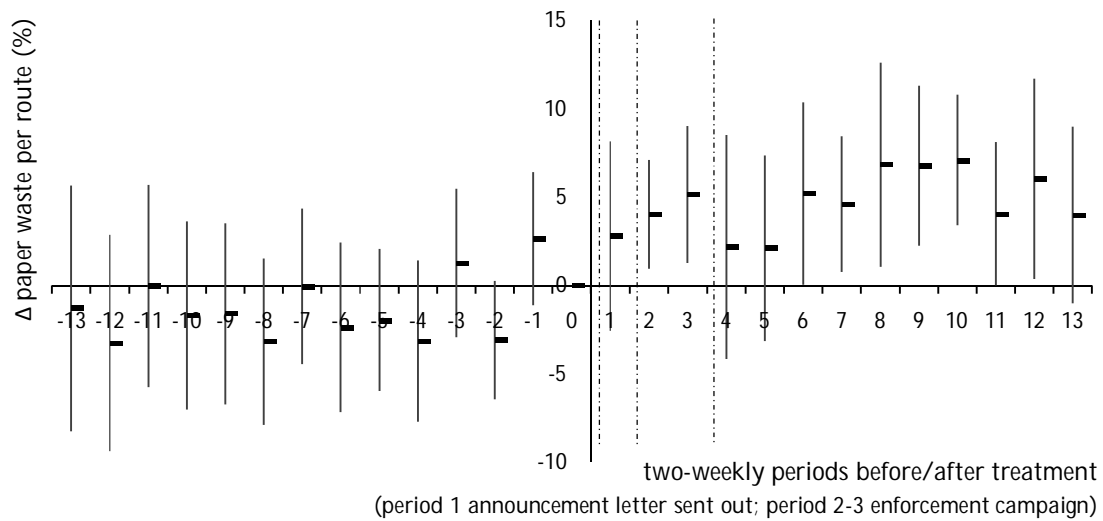
(b) Organic waste (weekly)



(c) Plastic waste (biweekly)



(d) Paper waste (biweekly)



Note. The figures plot coefficients α_t from estimation of Equation (1). The bars reflect the 90 percent confidence intervals, with robust standard errors clustered by routes. In panels (a) and (b) the observations are weekly, in panels (c) and (d) biweekly. Number of observations: 4,144 (panels a and b); 2,052 (panels c and d).

The two bottom panels in Figure 6 show the behavioral response to the treatment for plastic and paper waste, respectively. Because these fractions are collected biweekly, the horizontal axes reflect two-weekly periods. For plastic waste, we find a strong and seemingly persistent increase of about 13 percent or 0.03 ton in weekly terms (13 percent of 0.5 ton, divided by two). For paper waste, we find a relatively imprecisely estimated increase of some 6 percent or 0.08 ton in weekly terms.

Taken together, plastic and paper waste increase by some 0.11 ton in weekly terms, which is equal to 12 percent of the drop in residual waste. Adding up the increase in organic waste implies that almost 60 percent of the decrease in residual waste is accounted for by the increase in organic waste, paper

and plastic waste. More than 40 percent of the additional waste separated, or 0.4 ton per week, must have been collected in another manner, such as glass, textile, and discarded appliances. Household are expected to dispose these types of waste at disposal containers scattered throughout the city or at the central collection depot. Alternatively, some of the decline in residual waste may be due to an overall reduction in waste generated by households. For instance, households may switch to products with less packaging material or households may use more of their organic waste as garden compost.²⁰

B. Parametric model

To provide a statistical test of the average treatment effect, we estimate the following model:

$$(2) \ln W_{i,t} = \alpha \text{Letter}_{i,t} + \beta \text{Campaign}_{i,t} + \delta \text{PostCampaign}_{i,t} + Z_{i,t} \beta + \sigma_i + \lambda_t + \varepsilon_{i,t}.$$

$\text{Letter}_{i,t}$, is an indicator variable aimed at capturing the effect of the announcement letter. It is 1 as of the first week after the announcement letter has been sent out to households in route i until three weeks later and is zero otherwise. This term is also one in the first week of the enforcement campaign, because in that week households still respond to the letter and not to the campaign: the contents of the container at time t reflects separation of waste from $t-1$ to t after all. $\text{Campaign}_{i,t}$ stands for the effect of the enforcement campaign in route i at time t . It is 1 as of the second week of the enforcement campaign in route i and for the following three weeks and is zero otherwise.²¹ The period after the end of the enforcement campaign is denoted by $\text{PostCampaign}_{i,t}$. Similar to Equation (1), we include a vector of control variables $Z_{i,t}$, route-fixed effects σ_i and calendar-week fixed effects λ_t .

Table 2 provides the estimation results for Equation (2), with point estimates being similar to what we found in the event time analysis. We find treatment effects that are statistically significantly different from zero for each of the four types of waste, with the average treatment effect during the post-campaign period of 10 months being fairly similar to the effect during the enforcement campaign.

Next, we test for decay in the behavioral effect. We allow the effect post treatment $\text{PostCamp}_{i,\tau}$ to vary per six-week periods τ . We estimate the following equation:

$$(3) \ln W_{i,t} = \alpha \text{Letter}_{i,t} + \beta \text{Campaign}_{i,t} + \sum_{\tau=1}^5 \gamma_{\tau} \text{PostCampaign}_{i\tau} + Z_{i,t} \beta + \sigma_i + \lambda_t + \varepsilon_{i,t}$$

The estimated coefficients γ_{τ} denote the average treatment effect conditional on post-treatment period τ . Other than that, Equation (3) is exactly equal to Equation (2).

²⁰ Dumping of waste is unlikely since households can dispose of any type of waste at no charge.

²¹ As noted in Section 2, the enforcement campaign lasted six rather than four weeks in the ten routes that were treated first. The variable is coded accordingly.

Table 2. Effect of treatment on waste collected for each of the four fractions

Dependent variable:	ln(Residual waste)	ln(Organic waste)	ln(Plastic waste)	ln(Paper waste)
	(1)	(2)	(3)	(4)
Announcement letter	-0.086 (0.009)***	0.068 (0.020)***	0.044 (0.020)**	0.049 (0.016)***
Enforcement campaign	-0.138 (0.011)***	0.096 (0.020)***	0.109 (0.022)***	0.041 (0.019)**
Post-Campaign	-0.123 (0.012)***	0.093 (0.023)***	0.131 (0.016)***	0.054 (0.010)***

Note. This table presents estimates of Equation (2). Number of observations 4,185 (weekly data in columns 1, 2); 2,056 (bi-weekly data in columns 3, 4). Not shown are estimated coefficients for route-fixed effects, time-fixed effects and two covariates denoted by $Z_{i,t}$. Standard errors between parentheses are robust, clustered by route. * Statistically significant at 90 percent confidence level; ** at 95 percent level; *** at 99 percent level.

We also estimate Equation (3) separately for the autumn/winter-season and the spring/summer-season, given that the coefficients γ_τ may reflect seasonal variation in the treatment response. The observations post treatment are anything but equally distributed by season. For instance, during the first two months of the treatment, half of all observations occurred in spring, and this share drops to zero several months later. To have sufficient statistical power for estimating γ_τ by season, we focus our parametric test on estimating the extent to which separation of residual waste bounced back after the enforcement spell ended. If we find evidence for persistent treatment effects in residual waste, then it follows that there are persistent effects in separated types of waste as well – whether collected by way of the compartment containers or collected otherwise – or that households generate less waste overall.

Table 3 provides the estimation results for Equation (3). The results in the first column suggest the presence of some decay in the behavioral response during the first two six-week periods (i.e. first three months), with a statistically significant drop in the treatment effect between the first and second and second and third six-week post-campaign period. During the next seven months, the treatment effect is stable, suggesting a persistent decrease in residual waste of 9 percent. Compared to the treatment effect during the enforcement campaign, the long-term effect is some 4 percentage points smaller. When we limit our observations to the autumn/winter-season (column 2) or the spring/summer-season (column 3), we again find that the decay levels off within three months. Only the *level* of the treatment effect is somewhat higher in the autumn/winter-season than in the spring/summer-season. To conclude, our findings suggest that the change in behavior is persistent – at least for a period of seven months after end of the treatment.

Table 3. Persistence in the treatment effect for collection of residual household waste

Dependent variable:	all seasons	autumn/winter only	spring/summer only
ln(Residual waste)	(1)	(2)	(3)
Announcement letter	-0.081 (0.009)***	-0.075 (0.008)***	-0.086 (0.017)***
Enforcement campaign	-0.131 (0.011)***	-0.154 (0.018)***	-0.120 (0.016)***
Post-Campaign			
first 6 wks	-0.124 (0.013)***	-0.170 (0.020)***	-0.111 (0.020)***
second 6 wks	-0.110 (0.015)***	-0.158 (0.027)***	-0.096 (0.023)***
third 6 wks	-0.091 (0.020)***	-0.134 (0.037)***	-0.077 (0.027)***
fourth 6 wks	-0.091 (0.023)***	-0.133 (0.040)***	-0.070 (0.033)**
later wks	-0.096 (0.027)***	-0.143 (0.043)***	-0.075 (0.037)**
F-tests			
1 st vs. 2 nd period	3.45*	0.53	2.68
2 nd vs. 3 rd period	4.39**	1.12	3.55*
3 rd vs. 4 th period	0.00	0.01	0.45
4 th vs. 5 th period	0.34	0.67	0.21

Note. This table presents estimates of Equation (3). Number of observations 4,185 (column 1); 2,326 (column 2); 1,859 (column 3). Not shown are estimated coefficients for route-fixed effects, time-fixed effects and two covariates denoted by $Z_{i,t}$. Standard errors between parentheses are robust, clustered by route. * Statistically significant at 90 percent confidence level; ** at 95 percent level; *** at 99 percent level.

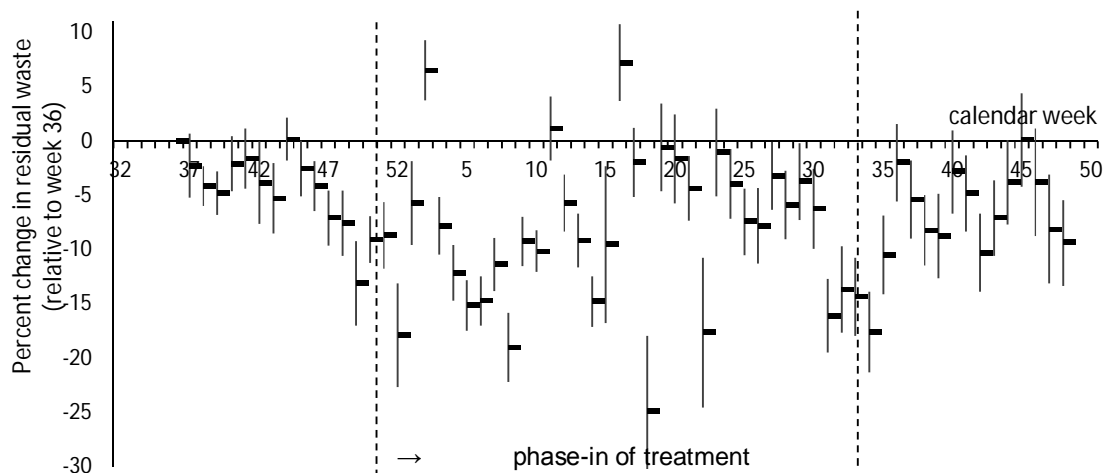
C. Sensitivity analysis

So far, we assumed that routes were either treated or not treated. It is not unlikely, however, that the treatment spilled over to other, yet to be treated areas. The local newspaper carried some fairly negative stories about the treatment while it was being rolled out. A possible behavioral response to the media coverage by households living in areas that had not been treated yet may have been reinforced by observing warning labels elsewhere in the city, namely in those areas that were undergoing treatment at that time. As a consequence, households in yet untreated routes may have improved their recycling behavior *before* the start of the treatment. If so, then this drives down the marginal treatment effect, given that our estimates are based on within-variation: we estimate how waste collected in a route develops relative to the last period before the start of the treatment.²²

²² The positive treatment spillovers may be attenuated by households' casual inferences about the effectiveness of enforcement. Households in routes that have not been treated yet may be surprised to see warning labels elsewhere but not on their own garbage container or any other container in their street. They may infer that enforcement is weak, possibly lowering their incentive to separate their waste.

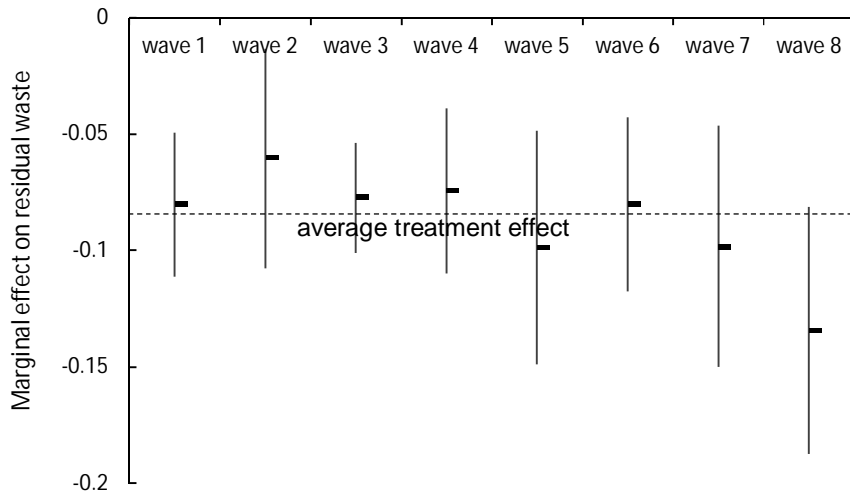
We follow two strategies to evaluate the extent to which treatment spillovers affected the estimated treatment effect. First, we look at the time-fixed effects in Equation (2), our default estimation equation. If the estimated effect of the treatment on residual waste collected is affected by treatment spillovers, then we expect a downward trend in the time-fixed effects. Households in the remaining control routes become also more likely to separate more of their waste after all. As Figure 7 shows, this is not what we find. The estimated week-fixed effects show no clear downward trend during the phase-in period. If anything, the fixed effects vary along with the seasonal variation in the weight of organic waste: during winter, the weight of organic waste is considerably lower than in other seasons.

Figure 7. Week-fixed effects in the regression model for residual waste, based on Equation (2)



Second, we allow the effect of the treatment to vary with calendar time. More specifically, in Equation (2), we interact the initial effect of the treatment – the behavioral response to the letter announcing the treatment – with indicator variables for each of the eight consecutive waves by which the treatment was phased in. In the presence of treatment spillovers, we would expect an upward trend, reflecting a diminishing impact on residual waste disposal. Figure 8 shows the estimation results. A joint test shows that we cannot reject the null hypothesis that all eight estimates are similar.

Figure 8. Estimated effect of the announcement letter on residual waste, conditional on treatment wave



Note: Estimates based on Equation (2), allowing the effect of the announcement letter on residual waste to vary by treatment wave. The vertical lines show the 95 percent confidence intervals.

D. Social norm treatments

As discussed in section 2, next to the threat of punishment, we subjected households to two behavioral interventions: conveying the injunctive norm by way of a persuasive appeal in the letter announcing the treatment and conveying the descriptive norm by way of the highly visible warning labels. To identify the effects of these interventions, we administered two additional treatments. These treatments also help to rule out alternative ways by which the enforcement campaign changed behavior

To see whether the persuasive appeal had an effect in and of itself, we devised a new letter.²³ The new letter closely resembles the original letter but excludes any reference to the illegal nature of not separating waste and the threat of being warned or fined. The letter only includes an appeal to separate waste – to save the environment and to keep local taxes down – and some additional information on what types of waste should be separated. We sent this letter to 10,000 households in 10 out of 20 randomly selected routes that were yet to be treated. To identify the treatment effect, we included an indicator variable in Equation (2) which is one for routes that received the persuasive appeal and zero otherwise. The variable is part of the vector of covariates $Z_{i,t}$.

Next, we devised a way to see whether the warning labels changed behavior because they conveyed information about how uncommon improper separation of waste is. The most straightforward way to test this would be to deliver warnings by letter rather than by way of the conspicuous warning labels on some routes. Unfortunately, the municipality did not allow us to do so. Instead, we were allowed to use an additional, white label saying that no evidence of wrongdoing was detected. In 25 randomly

²³ We did not have the discretion to devise an alternative announcement letter that only included the threat, at the expense of the persuasive appeal and the information what to separate. As a consequence, we cannot rule out the possibility that the persuasive appeal had an effect conditional on the threat.

selected routes, during the first week of the crackdown, every household that did not receive a warning label received a white label. Given the detection rate discussed previously, this held for almost 90 percent of households on average. The resulting sea of white labels attached to garbage containers is likely to have strengthened the signal from the warning labels, since they made the warning labels stand out even more.²⁴ If we do not find any evidence for a behavioral effect from the use of the white labels, then this suggests that the warning labels in and of themselves did not have much of an additional behavioral effect. This statement also holds if the white labels affected household behavior in other ways, for instance by increasing salience of the enforcement campaign. To identify this particular treatment effect, we included an indicator variable to Equation (2) which is one for routes that received the white labels in the first week of the crackdown and zero otherwise. The variable is also part of the vector of covariates $Z_{i,t}$.

Table 4. Effects of additional treatments on collected household waste

Dependent variable:	ln(Residual waste)	ln(Organic waste)	ln(Plastic waste)	ln(Paper waste)
	(1)	(2)	(3)	(4)
Persuasive appeal	0.009 (0.014)	0.009 (0.041)	-0.032 (0.037)	0.005 (0.019)
Enforcement campaign				
without white labels	-0.135 (0.012) ^{***}	0.102 (0.024) ^{***}	0.105 (0.031) ^{***}	0.060 (0.018) ^{***}
with white labels	-0.144 (0.015) ^{***}	0.087 (0.025) ^{***}	0.069 (0.036) [*]	-0.028 (0.054)
Post-Campaign				
without white labels	-0.123 (0.013) ^{***}	0.096 (0.025) ^{***}	0.097 (0.026) ^{***}	0.039 (0.020) [*]
with white labels	-0.125 (0.012) ^{***}	0.088 (0.027) ^{***}	0.094 (0.029) ^{***}	0.037 (0.021) [*]

Note. This table presents estimates of Equation (2), in particular the treatment variables $Z_{i,t}$ that were not reported in Table 2. Number of observations 4,185 (weekly data in columns 1, 2); 2,056 (bi-weekly data in columns 3, 4). Not shown are estimated coefficients for the announcement letter, route-fixed effects, and time-fixed effects. Standard errors between parentheses are robust, clustered by route. * Statistically significant at 90 percent confidence level; ** at 95 percent level; *** at 99 percent level.

The estimation results are shown in Table 4, for each of the four types of waste. The results in the first row show that we do not find evidence for an effect of the persuasive appeal. The average treatment effect is not significantly different from zero, and we do not even detect a short-run effect (not shown). The results for the persuasive appeal stand in sharp contrast to the impact of having received the

²⁴ Enforcement started only three weeks after receipt of the announcement letter. From the perspective of households, not seeing any warning labels during those three weeks could be due to both low levels of law enforcement and an inability to detect offenders. Given this uncertainty, we consider it to be unlikely that households inferred a high rate of compliance from not seeing any warning labels.

announcement letter that we reported previously. The results in the following rows show that we do not find evidence for any effect of the use of white labels alongside the colored warning labels either. In all cases, the estimated effects for routes with and without the white labels are not statistically significantly different. To conclude, these results suggests that the two nudges that were part of the intervention were not effective in changing behavior, while the temporary and salient enforcement spell was (cf. Ferraro et al. 2011 and Allcott and Rogers 2014).

4. Intrinsic motivation, habits and technology

We find the intervention to have a substantial effect on the behavior of households, both in the short term and in the longer term, and that the effects are driven by the threat of punishment. Next, we discuss what happened to the intrinsic motivation of households to separate waste, whether a new habit is likely to have formed and what the role of in-house technology that facilitates separating waste is in formation of a new habit.

A. Intrinsic motivation

We analyze the possible crowding out of the intrinsic motivation to separate waste by the extrinsic incentive based on survey. Midway the phase-in of the intervention, April 2015, we mailed a letter with an invitation for an online survey to 10,000 households. The invitees included 5,000 randomly selected households from the 35 routes that had already been subjected to enforcement at that time, and 5,000 randomly selected households from the 30 routes were yet to be treated. We sent a reminder to 3,000 households two weeks later.

In the survey, next to some background characteristics, households were asked to say whether they agree with statements like ‘separating waste is a societal duty’ and ‘separating waste is important for the environment’ and to indicate their use of in-house means that facilitate separating waste. All questions were phrased neutrally, i.e. without any reference to the treatment. We thus did not ask for *changes* in attitudes or *changes* in waste management practices over, say, the past 6 months, because we feared that such questions may provoke strong emotions for the respondents on the treated routes.

The Department of Sanitation allowed us to merge our survey data with their household-level administrative records on warnings and fines issued between the start and the end of the enforcement activities (December 1, 2014 – August 7, 2015). This allowed us to see whether selection into the survey and also the survey responses themselves were affected by having been issued a warning. To mitigate concerns that households that were issued a warning by the city of Tilburg were more or less likely to respond to the survey than other households, we used university letter head for the survey invitation.

We received 2,398 responses, 1,211 from households in treated routes and 1,187 from households in routes that were yet to be treated. The response rate was just below 25 percent, and about half of the respondents had received the treatment. Table 5 presents the descriptive statistics, bifurcated into the two groups of households. We find very similar household characteristics between treated and yet-to-be-treated households, suggesting that randomization was successful in achieving balance.

Table 5. Summary statistics for survey; households grouped by treatment status at time of the survey

	Yet-to-be-treated	Treated	Difference	p-value
<i>Household characteristics</i>				
Terraced home	0.581 (0.014)	0.554 (0.014)	0.026 (0.020)	0.191
Value of the home	255,840 (3,945)	250,309 (4,185)	5,531 (5,753)	0.336
Number of household members	2.472 (0.037)	2.536 (0.038)	-0.064 (0.053)	0.230
Male	0.526 (0.014)	0.495 (0.014)	0.030 (0.020)	0.139
Aged 60 or over	0.376 (0.014)	0.353 (0.014)	0.022 (0.020)	0.257
Does most of housekeeping chores	0.650 (0.014)	0.644 (0.014)	0.005 (0.020)	0.781
<i>Warnings (yet to be) issued</i>				
One warning	0.189 (0.011)	0.168 (0.011)	0.020 (0.016)	0.195
Two warnings	0.014 (0.003)	0.012 (0.003)	0.003 (0.005)	0.550
Three warnings	0.000 (0.000)	0.001 (0.001)	-0.001 (0.001)	0.322
Number of observations	1,187	1,211	2,398	

The similar shares of households that were issued one, two or three warnings either before or after the date of the survey (December 1, 2014 – August 7, 2015) suggests that responding to the survey is unrelated to being caught. In other words, the similar shares indicate that households that were issued a warning before the survey date were not more or less likely to respond to the survey than households

who were issued a warning after the survey date.²⁵ Otherwise, the share of households responding to the survey would have been higher for the treated group than the yet-to-be treated group. Appendix 2 provides two further tests of non-random selection into the survey, which confirm that treatment status did not have an economically significant effect on the likelihood to respond to the survey.

Now that we have addressed non-random non-response, we identify the effect of the treatment on intrinsic motivation and other survey outcomes. We estimate the following equation:

$$(4) Y_i = \alpha \text{Campaign}_i + \beta \text{Campaign}_i \text{Yellow}_i + \beta' \text{Yellow}_i + \gamma \text{Campaign}_i \text{OrangeRed}_i + \gamma' \text{OrangeRed}_i + \delta \text{WhiteLabel}_i + \eta \text{Appeal}_i + X_i \kappa + H_i \rho + \varepsilon_i$$

Y_i is one of the dichotomous outcome variables for household i , including whether a respondent agrees or not with a particular statement or whether her household makes use of an in-house technology. Campaign_i is one for household i that was subjected to enforcement prior to the survey date and is zero otherwise. α is the parameter of interest; the change in the outcome variable as a result of the treatment. We allow both the level and the treatment effect to differ between households that were issued a warning during the enforcement campaign and households that were not issued a warning. As discussed above, it is conceivable that actual experience with the long arm of the law affects the type of responses on a survey about separating waste. Yellow_i denotes having been issued one warning label at the survey date only; OrangeRed_i denotes having been issued two or more warning labels at the survey date (three warnings, the red label, is too rare to include, see section 2). X_i is a vector of responder-specific characteristics, including age, gender, and indicators for having received a reminder for completion of the survey, doing most household chores, living in a terraced house and for being a home owner rather than a tenant. H_i is a vector with neighborhood characteristics, including the share of residual waste in all collected waste, the share of terraced homes, and average house value.

Table 6 presents the results based on estimation of Equation (4), for the perception of separating waste as a social duty (columns 1 and 2) and for concern for the environment as motivation to separate waste (columns 3 and 4). We find that being subjected to enforcement is on average negatively related to agreeing with either of the two statements, but the effect is not statistically significantly different from zero (row 1 of columns 1 and 3). The effect is also small: the 2.9 percentage point change should be compared to a mean of 79 percent; the 1.3 percentage point change to a mean of 93 percent.

²⁵ We selected the period Dec 2015 – Feb 2016 to make sure that all routes have been equally exposed to the intervention. Alternatively, we might have chosen to compare the share of households having picked up labels for Dec 2014 – April 2015 for the treated routes, and May 2015–Sep 2015 for the yet untreated routes. Controlling for seasonal effects is easy in the panel analysis for the waste flows, but the surveys just offer us cross-section data. Using the full period for the balance test controls for seasonal (or any other time) differences.

Table 6. Effect of treatments on intrinsic motivation

Dependent variable:	Share of households agreeing with the following statement:			
	'Separating waste is a societal duty'		'Separating waste is important for the environment'	
	(1)	(2)	(3)	(4)
Subjected to enforcement campaign (y/n)	-0.029 (0.020)	-0.001 (0.028)	-0.013 (0.013)	0.012 (0.018)
... * recently (≤ 1.5 mnths ago)		-0.048 (0.030)		-0.043 (0.020)**
... * received one warning	-0.012 (0.044)	-0.020 (0.046)	0.027 (0.020)	0.022 (0.022)
... * received \geq two warnings	-0.022 (0.165)	-0.044 (0.174)	-0.206 (0.185)	-0.249 (0.197)
Received one warning only, before/after survey date	-0.063 (0.031)**	-0.062 (0.031)**	-0.025 (0.021)	-0.023 (0.021)
Received at least two warnings, before/after survey	-0.216 (0.108)**	-0.212 (0.109)*	-0.041 (0.075)	-0.038 (0.073)
Subjected to white labels (y/n)	0.030 (0.024)	0.048 (0.025)**	0.014 (0.013)	0.027 (0.012)**
Received persuasive appeal letter (y/n)	0.003 (0.022)	0.003 (0.022)	0.018 (0.014)	0.018 (0.014)

Note. This table presents estimates of Equation (3). Number of observations: 2,390. Not shown are coefficients for: respondent characteristics (type of home, home owner, male, senior, does most household chores, responded to reminder) and for neighborhood characteristics (share residual waste in total waste, share of terraced homes, average house value). Marginal effects; robust standard errors in parentheses, clustered at the route level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In columns (2) and (4) of Table 6 we allow the effect of the treatment to differ with the time since the end of the enforcement campaign. We find that any negative effects on intrinsic motivation are limited to recent experience with the crackdown (row 2). Recent experience led to an estimated 4 to 5 percentage point drop in the share of respondents agreeing with either of the two statements; the effect is statistically significant at the 3% level for the “importance question” (column 4), and only just fails to be statistically significant at the 10% level for the “societal duty question” (column 2, $p = 0.11$).

We do not find evidence for an additional effect on intrinsic motivation of having been issued either a first warning or a second warning before the survey date (rows 3 and 4). As much as households may have disliked a warning, apparently this did not have a clear negative effect on their attitudes. Next to the interaction between the treatment and the warnings, all regressions also include having been issued a warning before *or after* the survey date as a covariate. This variable shows whether intrinsic

motivation varies between the type of households that were issued a warning or not. The estimated coefficients are negative for both outcomes, but only statistically significantly different from zero for the social duty question (rows 5 and 6). In other words, households that do not see separating waste as a social duty are more likely to be caught for not separating their waste correctly. Given the likelihood that households' beliefs and actions are related, this is an indication that enforcement was not haphazard.

Finally, the results in row (7) show that the white labels, which conveyed the descriptive social norm, increased support for the two statements, but only if we allow for treatment heterogeneity by time since the crackdown (columns 1 and 3 vs. 2 and 4). The effect is small, however, which explains why we did not find any effect of the use of white labels on household behavior (see Table 4). The persuasive appeal had no discernable effect on households' opinions (row 8), which is also in line with the null findings of the appeal on household behavior that we reported previously.

To conclude, we find crowding out of intrinsic motivation by the transactional motive to avoid punishment to be only transitory. This suggests that the stable, higher rate of separation of waste is *not* sustained by a continuous fear for punishment. That is a very important finding as some 'residual' deterrent effect after the end of the crackdown is to be expected because households were never explicitly informed that enforcement would stop after one month. They may not perceive the enforcement spell as limited to four weeks. Moreover, households may be exposed to enforcement activities in other parts of the city, which adds to their uncertainty about the chance of getting caught.²⁶ Based on the results reported above, deterrence is unlikely to explain the persistent response in the longer term.

That raises the question how a similar level of intrinsic motivation and a withdrawn extrinsic incentive that is no longer felt can sustain a much higher rate of separating waste for an extended period. Below, we argue that a new, improved habit is a likely explanation.

B. Habit formation

Given the extended period for which we track household behavior, it is hard to believe that a new habit has *not* kicked in. For a repeated decision that is known to be prone to habit formation, we find a substantially higher rate of engaging in the behavior for about 200 consecutive days (see footnote 3 on the habitual nature of waste separation). If a new habit has *not* formed, then during all of these days, households must have consciously decided to separate more of their waste than they would have done in the absence of the fear of being warned and fined. This alternative scenario is not likely given the absence of any enforcement activity after the one-month crackdown.

²⁶ This spillover effect from routes that are being treated to routes that have been treated is likely to be small, given the finding that perception of the penalty is mostly affected by own experiences with law enforcement (Lochner 2007).

If conscious decisions are behind what we observe, then these decisions are likely to be affected by the discounting of past experiences with law enforcement, and a downwards adjustment of the chance of getting caught because households repeatedly get away with incorrect separation of waste (Sah 1991; see Lochner 2007 for empirical evidence). The incentive is weakened further as the punishment for not correctly separating waste becomes less salient in the months after the crackdown (cf. Dur and Vollaard 2018). But a steady decay is not what we find. Rather than a short-lived uptick in compliance, which is typical for the behavioral response to a crackdown (see the discussion in the introduction), we find some decay in the behavioral response in the first weeks after the crackdown (see Figure 6 and Table 3), but a stable, higher rate of separating waste for all of the following months. This suggests that the observed higher rate of waste separation has become an automatic response rather than a conscious decision out of fear of punishment. The insensitivity of household behavior to the discounting of past experiences and the decrease in salience can then be seen as a sign of automatic behavior.

The survey responses suggest that the persistent effect is not driven by a change in norms rather than a change in habits either. The survey questions that indicate intrinsic motivations to separate waste also reflect the prevailing social norm. These indicators are at a similar level before and after the treatment, suggesting that the norms are unchanged.

C. Technology

As we discussed in the introduction, the context of study is ideally suited for our purposes because purchase of technology is *not* an alternative explanation for persistence in the behavioral response. Having a bin, bag or receptacle lowers the cost of separating waste; it does not automate decision making. Daily decision making remains necessary, and that implies that separating waste remains prone to habit formation. Instead, one could argue that the purchase of these means makes it more likely that households stick to a new, improved habit. Purchase of these in-house means can be seen as an indication that households are in it for the long run. Why would someone purchase a receptacle for organic waste in the kitchen if he or she does not believe to keep separating organic waste for the foreseeable future?

Table 7 presents the results of estimating equation (4), with in-house means as dependent variable. We find that being subjected to enforcement resulted in a modest increase in the use of these means (row 1). The effect on receptacles for organic waste and a bin or bag for plastic is marginally statistically significant; the effect on a box for paper is outside conventional levels of statistical significance. The results in row (2) indicate that it takes some time before households adopt the new technology; we do not find a positive effect of the treatment on the use of these means in the short-term (adding the coefficients of row 2 and row 1 gives something close to zero). This suggests that, initially, households changed their behavior without these additional means, and only after some time re-organized their household in line with the new behavior. Given these results, technology can be seen as a cost-saving device that made it more likely for the new habit to stick.

Table 7. Effect of treatments on the usage of in-house means to facilitate separating waste

Dependent variable:	(1) 'Have receptacle for organic waste'	(2) 'Have bin/bag for plastic in the home'	(3) 'Have box for paper in the home'
Subjected to enforcement campaign	0.0554* (0.0305)	0.0557* (0.0316)	0.0266 (0.0319)
... * recently (≤ 1.5 months ago)	-0.0691** (0.0325)	-0.0828** (0.0344)	-0.0370 (0.0346)
Issued one warning only before/after survey date	-0.0292 (0.0262)	-0.0707*** (0.0273)	-0.0701** (0.0274)
Issued two or more warnings before/after survey date	-0.119 (0.0822)	-0.227*** (0.0865)	-0.123 (0.0917)
Subjected to white labels	0.00302 (0.0359)	0.0535 (0.0357)	0.0325 (0.0364)
Persuasive appeal	0.0266 (0.0291)	0.0101 (0.0294)	0.0265 (0.0295)

Note. See Table 6.

In line with our results for intrinsic motivation reported in Table 6, we find that households that are caught once or twice are different from households that were not caught. In this case, we find them to have a lower use of means that facilitate waste (rows 3 and 4).²⁷ Again, this suggests that enforcement was not haphazard. Finally, not surprising given previously reported results, we find a small and statistically insignificant effect on in-house means for the two behavioral intervention (rows 5 and 6).

5. Conclusion

In this paper we report the outcomes of a large-scale natural field experiment, involving 70,000 households living in single household dwellings in the city of Tilburg. The treatment involved informing households by way of a letter that not separating one's waste is illegal and punishable by a €0 fine, followed by an intensive and highly salient enforcement campaign. Service workers inspected each and every garbage container with separate compartments for organic and residual waste that was put at the curbside for collection. They attached brightly colored warning labels to containers whenever they found evidence of wrongdoing.

²⁷ Consistent with intuition, the effects are systematically more negative for those households that were issued/were yet to be issued two or more warnings, than for those that were issued/were yet to be issued just one warning. However, the coefficients of the former are measured less precisely, presumably because of the smaller sample size.

We find that the crackdown is highly effective in changing household behavior. The announcement letter alone resulted in an eight percent decrease in the weight of residual waste collected. During the crackdown that started three weeks later, residual waste was down by 13 percent. A slight decay after the end of enforcement levels off within three months, leaving a stable nine percent drop in residual waste for another seven months.

Arguably, the stable, higher rate of separating waste is sustained by a new habit. The primary alternative explanation is a continuing deterrent effect from the crackdown. Deterrence is likely to have only affected conscious decision making for a fairly short period, however, because we find crowding out of intrinsic motivation by the transactional motive – doing the right thing to avoid punishment – to be only transitory. A quickly disappearing deterrent effect once intensive enforcement is discontinued is also in line with the existing evidence on the effects of crackdowns and makes intuitive sense. Households discount past experiences with enforcement and adjust the rate of apprehension downwards every time they get away with incorrect separation of waste after all. The penalty also becomes less salient. With intrinsic motivation back at its original levels, the threat of punishment removed and no longer felt, it is likely to be a new habit that sustains the persistent response. In fact, it is hard to believe that a new habit has *not* kicked in during the 200 days that we observe a higher rate of separation of waste. In addition, the survey data show that households make greater use of bins, bags and other means to facilitate waste recycling, suggesting that they are in it for the long run. In contrast to other contexts such as energy consumption, technology only lowers the cost of separating waste and does not obviate the need to make daily decisions. Finally, the similar attitudes to separating waste before and after the treatment that we elicited in the survey indicate that the persistent effect is not driven by a change in norms rather than a change in habits.

To the best of our knowledge there is no consensus in the literature on how to include a change in habits in a welfare analysis, but the net benefits of the campaign are substantial. The municipality spent €250,000 euros on the enforcement campaign; the financial benefits from higher recycling revenues and lower residual waste treatment costs realized during the campaign amounted to €750,000. Taking into account (a) the long-lived stream of financial benefits that followed after the campaign; (b) the benefits to the environment; (c) the evidence from the survey indicating that households quickly shrug off the negative feelings of being monitored, the welfare benefits are indeed substantial.

This leaves the question why our treatment was so successful in changing habitual behavior. In line with John et al. (2018), we think that the high salience of the intervention, helped by the great intensity of enforcement, contributed to the effectivity of the treatment. Households could also not avoid enforcement during the crackdown, unlike drivers in traffic enforcement campaigns (Banerjee et al. 2018). Our findings suggest that police crackdowns can have persistent effects if habits are altered in the process, a hitherto unexplored channel in the deterrence literature.

Finally, we show that the effectiveness of introducing and temporarily enforcing regulations stands in contrast to the ineffectiveness of two ‘nudges’ that came with the intervention. A persuasive appeal in the letter announcing the treatment plus information on what waste should be separated did not have any identifiable effect on household behavior, not even in the short run. We do not find evidence that conveying the descriptive norm through the use of the highly conspicuous warning labels had any behavioral effect either. It was deterrence and deterrence only that initiated the strong behavioral response from households.

References

- Allcott, Hunt, 2011, Social norms and energy conservation, *Journal of Public Economics*, 95 (9-10), 1082–1095.
- Allcott, Hunt and Todd Rogers, 2014, The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation, *American Economic Review*, 104 (10), 3003-37.
- Banerjee, Abhijit, Raghavendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh, 2017, The Efficient Deployment of Police Resources: Theory and New Evidence from a Randomized Drunk Driving Crackdown in India, *mimeo*, Massachusetts Institute of Technology.
- Becker, Gary S., Kevin M. Murphy, 1988, A Theory of Rational Addiction, *Journal of Political Economy*, 96 (4), 675-700.
- Benabou, Roland and Jean Tirole, 2011, Laws and Norms, NBER working paper 17579.
- Brandon, Alec, Paul J. Ferraro, John A. List, Robert D. Metcalfe, Michael K. Price, Florian Rundhammer, 2017, Do the effects of social nudges persist? Theory and evidence from 38 natural field experiments, NBER working paper 23277, Cambridge (MA).
- Carrus, G., P. Passafaro, and M. Bonnes, 2008, Emotions, habits and rational choices in ecological behaviors: The case of recycling and use of public transportation, *Journal of Environmental Psychology*, 28, 51-62.
- Calzolari, Giacomo and Mattia Nardotto, 2017, Effective Reminders, *Management Science*, 63 (9), 2915-2932.
- Cecere, Grazia, Susanna Mancinell, and Massimiliano Mazzanti, 2014, Waste prevention and social preferences: the role of intrinsic and extrinsic motivations, *Ecological Economics*, 107, 163-176.

- Cialdini, R. B., R.R. Reno and C.A. Kallgren, 1990, A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places, *Journal of Personality and Social Psychology*, 58, 1015-1026.
- Cheung, Shu Fai, Darius Chan, and Zoe Wong, 1999, Re-examining the theory of planned behavior in understanding wastepaper recycling, *Environment and Behavior*, 31 (5), 587-612.
- Dur, Robert and Ben Vollaard, 2018, Salience of law enforcement: A field experiment, *Journal of Environmental Economics and Management*, forthcoming.
- Eeckhout, Jan, Nicola Persico, Petra Todd, 2010, A theory of optimal random crackdowns, *American Economic Review*, 100 (3), 1104-1135.
- Elliott, M.A., and J. Broughton, 2005, How methods and levels of policing affect road casualty rates, TRL Report 637, Wokingham: Transport Research Laboratory.
- Eureco, 2015, Sorteerrapportage huishoudelijk afval [in Dutch], Report commissioned by the City of Tilburg.
- European Environment Agency, 2014, Waste: a problem or a resource?, Copenhagen.
- Ferraro, Paul J., Juan Jose Miranda, and Michael K. Price, 2011, The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment, *American Economic Review*, 101(3), 318-22.
- Gallani, Susanna, 2017, Incentives, Peer Pressure, and Behavior Persistence, Harvard Business School Working Paper, No. 17-070.
- Gardner, Benjamin, 2015, A review and analysis of the use of ‘habit’ in understanding, predicting and influencing health-related behavior, *Health Psychology Review*, 9 (3), 277-295.
- Ge, Qi and Benjamin Ho, 2018, Energy Use and Temperature Habituation: Evidence from High Frequency Thermostat Usage Data, *Economic Inquiry*, forthcoming.
- Gneezy, U., Meier, S., and Rey-Biel, P., 2011, When and why incentives (don't) work to modify behavior, *The Journal of Economic Perspectives*, 25(4), 191-209.
- Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius, 2008, A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels, *Journal of Consumer Research*, 35(2), 472-482.
- Gosnell, G.K., J.A. List and R. Metcalfe, 2016, A New Approach to an Age-Old Problem: Solving Externalities by Incenting Workers Directly, NBER Working Paper 22316.

Goswami, Indranil and Oleg Urminsky, 2017, The dynamic effect of incentives on postreward task engagement, *Journal of Experimental Psychology*, 146 (1), 1-19.

Hallsworth, M., J.A. List, R. Metcalfe and I. Vlaev, 2014. The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance. NBER Working Paper 2007.

Halpern, S. D., French, B., Small, D. S., Saulsgiver, K., Harhay, M. O., Audrain-McGovern, J., . . . Volpp, K. G., 2015, Randomized trial of four financial-incentive programs for smoking cessation, *The New England Journal of Medicine*, 372, 2108–2117.

Holland, Rob, Henk Aarts and Daan Langedam, 2006, Breaking and creating habits on the working floor: a field-experiment on the power of implementation intentions, *Journal of Experimental Social Psychology*, 42, 776-783.

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, Natalia Rigol, 2016, Habit Formation and Rational Addiction: A Field Experiment in Handwashing, working paper, MIT.

Jackson, C. K., 2010, A little now for a lot later: A look at a Texas advanced placement incentive program, *Journal of Human Resources*, 45, 591–639.

John, Leslie K., Francesca Gino, Luca Foschini, Katherine L. Milkman, and Bradford Tuckfield, 2018, The Ineffectiveness of Inconspicuous Incentives: A Field Experiment on Inattention, *Management Science*, forthcoming.

Kahan, Dan M., 2005, What’s really wrong with shaming sanctions, *Texas Law Review*, 84.

Kinnaman, Thomas C., 2006, Policy Watch: Examining the Justification for Residential Recycling, *Journal of Economic Perspectives*, 20 (4), 219–32.

Lochner, Lance, 2007, Individual perceptions of the criminal justice system, *American Economic Review*, 97 (1), 444-460.

Mullainathan, S. and R. Thaler, 2000, Behavioral Economics, NBER Working Paper 7948.

Ouellette, J. A. and W. Wood, 1998, Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior, *Psychological Bulletin*, 124, 54-74.

Prugsamatz, Sunita, Monika Tobolova, Alim Nayum and Christian A. Klöckner, 2017, Understanding the Mechanisms behind Changing People’s Recycling Behavior at Work by Applying a Comprehensive Action Determination Model, *Sustainability*, 9, 204.

Sah, Raaj K., 1991, Social Osmosis and Patterns of Crime, *Journal of Political Economy*, 99(6), 1272-1295.

Schultz, Wesley, 1999, Changing Behavior with Normative Feedback Interventions: A Field Experiment on Curbside Recycling, *Basic and Applied Social Psychology*, 21 (1), 25-36,

Schultz, Wesley, Jessica Nolan, Robert Cialdini, Noah Goldstein and Vidas Griskevicius, 2007, The Constructive, Destructive, and Reconstructive Power of Social Norms, *Psychological Science*, 18, 429–434.

Sherman, Lawrence W.. 1990, Police crackdowns: initial and residual deterrence, in: *Crime and justice. A review of research*, 12, edited by Michael Tonry and Norval Morris, University of Chicago Press, Chicago.

Simon, H.A., 1976, *Administrative behavior: a study of decision making processes in administrative organizations*, Harper, New York.

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

Verplanken, Bas and Wendy Wood, 2006, Interventions to break and create consumer habits, *Journal of Public Policy & Marketing*, 25 (1), 90-103.

Yang, Nan and Yong Long Lim, 2017, Temporary Incentives Change Daily Routines: Evidence from a Field Experiment on Singapore's Subways, *Management Science*, forthcoming.

Appendix 1. Letter announcing enforcement of regulations regarding separation of household waste



GEMEENTE TILBURG

Datum november 2014	Ons kenmerk
Postadres Postbus 90155 5000 LH Tilburg	Bezoekadres Schouwburgring 103

Controle afvalscheiding in uw wijk

Start controle op juiste sortering huishoudelijk afval

Via de nieuwe oranje/blauwe duocontainer hebt u de afgelopen tijd flink meer plastic verpakkingen en papier gesorteerd. Onze dank daarvoor. Toch blijkt uit steekproeven dat er nog veel goede grondstoffen in het restafval terecht komen.

Als we samen zorgen voor een betere sortering is dit goed voor het milieu en kan de gemeentelijke belasting (afvalstoffenheffing) omlaag. Daarom gaat de gemeente Tilburg binnenkort controleren of afval goed gesorteerd wordt.

Wist u dat afval sorteren ook verplicht is?

U bent verplicht om afval goed te sorteren. Hiervoor hebt u twee duocontainers. Is uw afval niet goed gesorteerd, dan krijgt u een gele kaart. De gele kaart is een eerste waarschuwing. Als u uw afval een tweede keer verkeerd aanbiedt dan krijgt u een oranje kaart. Uw container wordt dan niet of voor een deel niet leeggemaakt. Een derde keer is een rode kaart en betekent dat uw container niet of voor een deel niet leeggemaakt wordt, bovendien kunt dan u dan ook nog een boete kunt krijgen.

Meer informatie

Op de achterkant van deze brief vindt u meer informatie over de controle op afvalscheiding. Op www.tilburg.nl/afval leest u meer over hoe u het beste afval kunt sorteren.



BEWONERSBRIEF

Pagina 1 van 2

Vragen aan de gemeente? Kijk op www.tilburg.nl



Waarom gaan wij controleren?

Het gemeentebestuur wil in Tilburg de hoeveelheid restafval sterk terugdringen. Uit steekproeven blijkt dat er flinke hoeveelheden goede grondstoffen in het restafval zitten die nog geschikt zijn voor hergebruik. Denk aan groente- fruit- en tuinafval (GFT), papier, plastic verpakkingen, textiel, glas, elektrische apparaten en zelfs klein gevaarlijk afval (batterijen, verf e.d.). Wij willen graag dat deze grondstoffen beter gesorteerd worden zodat er veel minder goede grondstoffen verbrand worden.

Hoe gaan wij controleren?

Wij controleren de duocontainers op restafval en GFT. Speciale controleurs kijken in de duocontainers of er niet teveel goede grondstoffen zoals papier, plastic verpakkingen, GFT, textiel, elektrische apparaten en klein gevaarlijk afval in het restafval gedeelte zitten. Ook kijken ze of in het GFT-gedeelte niets anders zit dan GFT.

De controleurs hangen een gele, oranje of rode kaart aan de duocontainer. Via speciale apparatuur geven ze aan wat er fout is met de afvalscheiding. Deze gegevens worden digitaal opgeslagen via de chip in uw container en kunnen geraadpleegd worden door de controleurs bij een volgende aanbidding.

Gele kaart

Is het afval in uw duocontainer voor een deel niet goed gesorteerd dan krijgt u via een gele kaart een waarschuwing. Deze kaart hangt dan aan uw container. Op de kaart staat waarom u de gele kaart gekregen hebt en wat u kunt doen om deze niet meer te krijgen.

Oranje kaart

Is het afval in uw duocontainer voor de 2^e keer niet goed gesorteerd, dan krijgt u een oranje kaart aan uw duocontainer. Hierop is aangegeven waarom u deze oranje kaart krijgt en wat de gevolgen zijn. Als een deel van de duocontainer niet goed gesorteerd is, wordt dit gedeelte van de duocontainer deze week niet leeggemaakt. Zijn beide delen van de duocontainer niet goed gesorteerd dan wordt deze duocontainer deze week helemaal niet leeggemaakt.

Rode kaart

Is het afval in uw duocontainer voor de 3^e keer niet goed gesorteerd, dan krijgt u een rode kaart aan uw duocontainer. Hierop is aangegeven waarom u deze rode kaart krijgt en wat de gevolgen zijn. Als een deel van de duocontainer niet goed gesorteerd is, wordt dit gedeelte van de duocontainer deze week niet leeggemaakt. Zijn beide delen van de duocontainer niet goed gesorteerd dan wordt de duocontainer deze week helemaal niet leeggemaakt. Bovendien geven wij uw adres meteen door aan de gemeentelijk opsporingsambtenaar die u een boete kan opleggen vanwege overtreding van de Afvalstoffenverordening van de gemeente Tilburg. Deze boete is in 2014 € 90,00 per overtreding. Dit bedrag kan in 2015 misschien nog veranderen.

Wat moet ik dan met een volle duocontainer die niet is leeggemaakt?

Als u een oranje of rode kaart hebt ontvangen, dan moet u zelf zorgen dat uw afval goed gesorteerd wordt. Als uw afval weer goed gesorteerd is, dan kunt u de volgende week weer uw container aanbieden. Kunt u niet zo lang wachten dan kunt u zelf uw afval wegbrengen naar de milieustraat. Voor restafval betaalt u € 5,00 per aanbidding.

Hoe moet ik mijn afval nu precies sorteren?

Kijk op www.tilburg.nl/afval over hoe u afval moet sorteren.

BEWONERSBRIEF

Appendix 2. Testing for non-random selection into the survey

To conduct a further test that treatment status does not affect households' propensity to participate in the survey, we present two regression results. The first regression measures the impact of having received the treatment, controlling for a set of other characteristics like the tax assessment value of the property (possibly a proxy for income), number of people living in the home, whether it is a terraced or a (semi-) detached home, and also whether the invited household has received or is yet to receive one or more warnings. Those who are yet to receive a warning (a fact they are not aware of at the time they were considering filling out the survey) are arguably less likely to properly sort their waste, and hence they may also be less interested in filling out the survey.

Table A1 presents the estimation results. The results in the first column suggest that having received the treatment does not affect the propensity to respond to the survey invitation. We do, however, find that richer households (as proxied by the tax value of their home), smaller households and households living in terraced houses are more likely to participate, while those that have received or are yet to receive one or more warning are less likely to respond.

The coefficient on having received the treatment in the first column of Table A1 captures the effect of having received the treatment averaged over all households that were invited to participate in the survey – independent of whether they received a warning during the actual enforcement campaign, or not. In the second column of Table A1 we explore whether actually having received a warning affects the propensity to participate – because the household felt treated unjustly, and either wanted to air their dismay, or threw the invitation directly into the waste bin (and presumably more likely so in the residual waste bin than in the bin for paper). The results in the second column show that having received the treatment does not affect the propensity to participate in the survey of those households that did not receive any warnings (as shown by the coefficient in the first row), and those that actually received two or more warnings are not more or less likely to complete the survey than those that are yet to receive two or more warnings (as the coefficient in row 8 turns out to be insignificant). The coefficient in row 7 shows that households that actually received a warning are still less prone to fill out the survey than those households that did not receive a warning, but also that they are slightly more prone to do respond than their peers that are yet to receive a warning. Overall, the impact of having received the treatment seems to be small, or even negligible, implying that the difference in answers to the survey between those households that received the treatment and those that are yet to be treated can be interpreted causally.

Table A1. Factors affecting invited households to participate in the survey, treated versus yet-untreated households

	(1)	(2)
Treated	0.00674 (0.0131)	-0.00289 (0.0129)
Value of the home (in logs)	0.126*** (0.0155)	0.126*** (0.0154)
Number of household members	-0.0120*** (0.00277)	-0.0121*** (0.00278)
Terraced house	0.0258** (0.0124)	0.0255** (0.0124)
Has received/is yet to receive one warning by survey date	-0.0846*** (0.00934)	-0.107*** (0.0123)
Has received/is yet to receive two warnings or more by survey date	-0.127*** (0.0398)	-0.132** (0.0560)
Received a warning by survey date		0.0524*** (0.0172)
Received two or more warnings by survey date		0.00552 (0.0800)
<i>Wald Chi2</i>	155.03***	186.84***
<i>n</i>	9490	9490

Note. Marginal effects; robust standard errors, clustered at the route level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.