TI 2012-061/1
Tinbergen Institute Discussion Paper

The Power of a Bad Example – A Field Experiment in Household Garbage Disposal

Robert Dur[1]
Ben Vollaard[2]

[1] *Erasmus University Rotterdam, Tinbergen Institute, CESifo, and IZA;*
[2] *Tilburg University, CentER, and TILEC.*

The Power of a Bad Example – A Field Experiment in Household Garbage Disposal*

Robert Dur[†] and Ben Vollaard[‡]

July 3, 2012

Abstract

Field-experimental studies have shown that people litter more in more littered environments. Inspired by these findings, many cities around the world have adopted policies to quickly remove litter. While such policies may avoid that people follow the bad example of litterers, they may also invite free-riding on public cleaning services. This paper reports the results of a natural field experiment where, in a randomly assigned part of a residential area, the frequency of cleaning was reduced from daily to twice a week during a three-month period. Using high-frequency data on litter at treated and control locations before, during, and after the experiment, we find strong evidence that litter begets litter. However, we also find evidence that some people start to clean up after themselves when public cleaning services are diminished.

JEL-codes: C93, H40, K42

Keywords: littering, public services, free-riding, field experiment.

[*] We gratefully acknowledge comments and suggestions from Olivier Marie, Jan Stoop, Christian Traxler, participants of the Advances with Field Experiments 2011 Workshop at the University of Chicago, the New Directions in Welfare 2011 conference at the OECD in Paris, the 3rd Transatlantic Workshop on the Economics of Crime at Collegio Carlo Alberto in Moncalieri, the 10th Tiber Symposium on Psychology and Economics at Tilburg University, and the WZB Field Days in Berlin. We also thank seminar audiences at the University of Munich, Erasmus University Rottterdam, Stockholm University, and Tilburg University. Special thanks to the City of Rotterdam, the Charlois city district, and the community service providers Roteb and Stadstoezicht who made the field experiment possible. We are most grateful to Maya von Harras for providing daily project management, to Herman Coomans and Arthur Andjelic for collecting and processing the data, and to Peter van Dalen for his role in the early stages of the project.

[†]Erasmus University Rotterdam, Tinbergen Institute, CESifo, IZA. E-mail: dur@ese.eur.nl

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

1. Introduction

When a person breaks the law, he often inflicts more damage to society than just the direct negative consequences of his act. The transgression may be an invitation for others to also break the law. This knock-on effect constitutes an additional negative externality from illegal behavior.

As a case in point, consider littering. In most countries, littering is illegal and punishable by a fine. A litterer spoils a neighborhood's appearance and, in the longer run, harms the environment. In addition to these direct negative externalities, litterers set a bad example for others to follow. Mimicking may arise for several reasons. First, some people may simply imitate other people's behavior, thinking that "If everyone is doing it, it must be a sensible thing to do." (Cialdini et al. 1990: 1015). Second, people may infer from observing litter that littering is tolerated by the police and by the community at large, thus reducing the fear for formal and social sanctions (Kahan 1997). Third, people may behave in a conditionally cooperative manner (Gächter 2007), and so may be discouraged to contribute to a clean environment when they observe that others did not do so. Last, preferences for cleanliness may be nonconvex (Anderson and Francois 1997), meaning that the decline in well-being from an additional piece of litter is relatively small when the environment is already littered as compared to when the environment is clean.

These behavioral spillover effects can give rise to negative feedback loops, transforming a clean and orderly neighborhood into a dirty and messy place. A series of experimental studies have provided evidence supporting this idea. In these studies, researchers experimentally induced variation in the amount of litter present at places like grocery stores (Geller et al. 1977), picnic areas (Crump et al. 1977), a waiting room for participants to a lab experiment (Kraus et al. 1978), a parking garage (Reiter and Samuel 1980, Cialdini et al. 1990, and Reno et al. 1993), an amusement park (Cialdini et al. 1990), the lobby of a dormitory on a university campus (Cialdini et al. 1990), an academic department's common room (Ramos and Torgler 2010), and an alley in a large shopping area (Keizer et al. 2011).[1] With a few exceptions (Crump et al. 1977 and Reno et al. 1993), these experimental studies find that people litter significantly more often in littered environments as compared to clean

[1] Finnie (1973) and Schultz et al. (2011) also study littering in clean and dirty areas, but the cleanliness of the area is not experimentally varied, making it impossible to infer causal relationships.

environments.[2] In most studies, the effect is substantial. For instance, in Cialdini et al. (1990: 1017)'s study into littering of a handbill that was tucked under the windshield wiper of parked cars, 11 percent of subjects litter in a clean environment whereas 41 percent of subjects do so in a littered environment.

Based on the findings from these one-shot, small-scale field experiments, it has often been argued that frequent clean-ups of public places yield a double dividend (Reiter and Samuel 1980, Cialdini et al. 1990, Huffman et al. 1995, and Anderson and Francois 1997). First, the frequent clean-ups ensure that litter is quickly removed. Second, the resulting litter-free environment may keep people from further littering. As Anderson and Francois (1997) put it, government provision of cleaning efforts can *crowd in* rather than crowd out private contributions to a clean environment. Several local governments have followed this advice and have implemented intensive periodic cleaning programs in residential and nonresidential areas alike. Many practitioners believe that such programs have been highly effective in discouraging littering (Lewis et al. 2009, Schultz and Stein 2009, and Ouiller and Sauneron, 2011).

Yet, other studies suggest that a strategy of frequent public cleaning may backfire. In line with the standard theory of private provision of public goods (Bergstrom et al. 1986), some empirical evidence suggests that government clean-ups of public spaces invites free-riding by citizens. For example, surveys conducted in the US reviewed by Beck (2007) suggest that people litter more if they know that someone else will clean up after them. Similar findings are reported by Lewis et al. (2009) for the UK. They note that "many people think that if someone else is paid to clean up […] then littering can be justified. More specifically, those questioned often named this "someone else" as the local authority." (Lewis et al. 2009: 19). Clearly, if many people think along these lines, government provision of cleaning efforts will *increase* rather than decrease littering.

In this paper, we put the two conflicting theories of how people respond to publicly provided cleaning services to the test. We report the results of a natural field experiment where, in a

[2] Cialdini et al. (1990) provide some evidence that the effect is nonlinear: a single piece of litter seems to make an anti-litter norm more salient than a completely clean environment. They also find that the positive effect of a heavily littered environment on littering is significantly greater when people observe an experimental confederate littering (see also Reno et al. 1993). Keizer et al. (2008) show that also other signs of disorder, such as unreturned shopping carts standing around in disarray in a parking garage and illegally set off fireworks, cause littering.

randomly assigned part of a residential area, public cleaning services were drastically reduced during a period of three months. In contrast to the existing one-shot experiments, which focus on the spontaneous behavioral response, the repeated setting allows for learning effects on the side of residents. The experiment took place in Rotterdam, the second largest city in The Netherlands, from December 2010 to February 2011. Using high-frequency data on litter at treated and control locations before, during, and after the experiment, we examine whether people litter more when the environment is less often cleaned up, how this behavior develops over time, and whether it is persistent after the treatment has ended.[3]

The focus of our study is on littering around underground containers for collection of household garbage in a densely populated area. The containers feature a lid that is to be opened to dispose of garbage. Putting garbage on the street is illegal and subject to a 115 euro fine. Discarded household items that are too big to be put into the container should either be brought to a nearby garbage collection depot or be picked up by municipality workers after making an appointment by telephone, at no charge.

Despite the fine and the free legal alternatives, littering around underground containers occurs frequently. In surveys, littered streets feature highly on residents' lists of greatest annoyances, together with speeding, dog dirt, illegal parking, and nuisance from youth (IVM 2009). In response to this public concern and inspired by some of the above-mentioned experimental studies, the municipality decided a few years ago to provide cleaning services at all container locations in the area on a daily basis rather than the regular frequency of two or three times a week. Our experimental treatment is to abandon these extra services and to return to the regular cleaning frequency in a randomly assigned part of the area for a period of three months, while keeping the frequency of emptying the container constant. The natural setting of the experiment does not allow us to isolate the specific behavioral mechanisms driving the response to a change in public cleaning services. Our primary aim is to put previous work on behavioral mechanisms to the test within people's own habitat, with particular attention to crowding in versus crowding out of private contributions to a public good.

[3] We do not consider effects on other offenses or other outcomes because of data limitations.

The results of our field experiment are as follows. First, using high-frequency data on litter at treatment and control locations, we find strong evidence that litter begets litter. The effect is substantial: not cleaning a location in the morning increases the likelihood of encountering *additional* litter at the location the next morning by about fifty percent. While this result lends strong support to the hypothesis that public cleaning services crowd in private contributions, we also find evidence for crowding-out effects from data on the number of telephone appointments for pick-up of discarded household items, one of the legal alternatives to littering. We find that the reduction in cleaning activities increases the number of appointments by about a third. Hence, our findings suggest that the provision of public cleaning services crowds in private contributions to a clean environment of some people, but crowds out private contributions of others, with the former effect dominating the latter.

Our paper contributes to several strands in the literature. First, we make a contribution to the empirical literature on littering behavior, a literature that we briefly reviewed above. Existing studies are limited to short-term interventions (lasting one day or less) and focused on people's spontaneous response (to litter or not to litter). Our field experiment is the first to study the effects of an intervention that lasted several months, while also measuring behavior in the month after the treatment. This repeated setting allows us to study possible learning effects. Arguably, it takes a while before people realize that they can no longer free ride on frequent public cleaning services.

Second, our paper contributes to a recent literature in economics showing that people tend to follow examples set by others. This has for instance been shown in the context of donations to charities (Frey and Meier 2004, Martin and Randal 2008, Shang and Croson 2009), tax evasion (Fellner et al. 2011), welfare participation (Bertrand et al. 2000), and in public good games played in laboratory experiments (Burlando and Guala 2005, Gächter 2007, Beckenkamp et al. 2009, Engel et al. 2011).

Lastly, our paper is related to the empirical literature on private contributions to public goods (see Nyborg and Rege 2003 and Payne 2009 for surveys). Our findings point to a delicate balance between publicly provided services and voluntary contributions from citizens. Too low a level of public services may result in a bad equilibrium in which most citizens choose not to cooperate. Too high a level of public services may crowd out private contributions, and leave it up to the state to provide the public good.

The remainder of the paper is structured as follows. The next section describes the context and design of our experiment. Section 3 presents the econometric model and reports the estimation results. Section 4 concludes.

2. Experimental design

A. Littering and cleaning activities

The experiment was conducted in a residential area of some 4,000 households within the city of Rotterdam, from November 29, 2010 until March 3, 2011. The residential area is part of Charlois, a densely populated and relatively poor city district. The area counts 41 locations with underground garbage containers.[4] The containers are placed on the sidewalk. The part of the container that is visible at street level features a lid that is to be opened to dispose of garbage (see Figure 1). Each container serves some 60 households.

[FIGURE 1]

It is illegal to place garbage next to the container. Every container includes a large warning label stating the 115 euro fine for illegal disposal of garbage. Discarded household items that are too big to fit in the container should either be brought to a nearby garbage collection depot or be picked up by municipality workers after making an appointment by telephone. On average, 18 households make an appointment per week.

Many times people break the law by putting large household items next to the container. It is tempting to take the easy way out: it takes time and effort to bring garbage to a nearby collection depot or to set up an appointment for pick-up. Moreover, when making an appointment, it often takes a few days before garbage can be picked up, implying that people need to store garbage in their house or backyard for some time (as a matter of fact, appointments could be made for Tuesdays only). When garbage is illegally placed on the sidewalk, it is almost always put next to the container rather than somewhere else on the sidewalk. Supposedly, the container is a safe location: at this common collection point it is close to impossible to trace the offender, quick removal is secured since this is where service

[4] The total number of containers in the area equals 70: the 41 locations include 15 with one container, 23 with two containers, and three with three containers.

workers stop to clean up, and neighbors may be less annoyed by items placed next to the container rather than someplace else (e.g. in front of someone's door).

Why someone would place a garbage bag next to a container is less obvious. If one takes the effort of bringing a bag all the way from home to the container, then why not put the bag into the container? Based on observations of residents and non-structured interviews with residents and service workers, we deduced three reasons. First of all, it is a way of avoiding the effort of opening the lid and pulling the lever. It is a small effort, but it requires the use of some force and one may get his hands wet or dirty. Second, people sometimes put garbage bags next to the container because the container is stuck or full. Lastly, illegally disposed garbage may simply block access to the container. Leaving alone rare situations of piles of garbage next to the container, residents are always able to reach the container and open the lid.[5] Further note that, even if a container is stuck or full or cannot be accessed, it is still illegal to put garbage next to the container. In those cases, citizens are supposed to use a nearby container. Alternative container locations are rarely more than one block away.

The garbage containers are emptied two or three times a week, either in the early morning or the late afternoon, according to a fixed schedule. Most of the containers in the area are emptied on Monday and Thursday afternoon. Before emptying, the area around the containers is cleaned of all litter. That is necessary to prevent litter from falling into the exposed underground space when the container is being emptied. In addition to the clean-up before emptying, the area around the containers is cleaned every morning by a crew of two, both of them driving a small truck. Taken together, it was standard practice to clean the area around the containers at least once a day, and sometimes even twice a day.

B. Data collection

Littering around the garbage containers was monitored twice daily on weekdays from Monday, September 20, 2010 until Friday, April 1, 2011. During these six months, the same two non-uniformed observers monitored the area, firstly from 8.00 to 9.30 am, which is before the daily cleaning round in the morning, and secondly from 1.30 to 3.00 pm. For each of the container locations, the observers noted the number of garbage bags next to the container as well as the number of discarded items such as household appliances and

[5] Based on the data collection that we discuss in the next paragraph, we conclude that in 90 percent of the cases that litter is present, the number of bags and discarded household items is equal to 3 or less.

furniture. They also noted whether the container was stuck or full. Whether a container is stuck rather than full is not always easy to determine, this is why we created one indicator for the two conditions.

In the two months before the start of the intervention, on average 24 percent of locations were littered with one or more bags or garbage items at the time of the morning round and 9 percent of locations at the time of the afternoon round. The containers were stuck or full 13 percent of the time in the morning and 6 percent in the afternoon on average during the baseline period. The distribution of littering over locations was roughly stable, but heavily skewed. The reasons for why some locations tend to be more littered than others within the same neighborhood and sometimes even the same street are not clearly understood. Correlational evidence suggests that the number of homes with a view on the container and distance to shops are positively related to the cleanliness of a location (details are available upon request).

C. Treatment and randomization procedure

To examine how littering behavior changes in response to public cleaning services, the frequency of cleaning around the garbage containers was reduced for a period of three months in a randomly assigned part of the area. The daily cleaning by the crew of two in the morning was cancelled in the treatment area. As a consequence, cleaning was reduced from one or two times a day to two or three times a week. The treatment was not communicated to residents so as to avoid confounding effects arising from communication. Hence, the main channel by which residents learn about cleaning activities is the cleanliness of the area around the garbage container. Since most residents dispose of their garbage multiple times per week, and also live in close proximity to the container, they have frequent exposure to the condition of this area.[6] Some residents may also have noticed that the two service workers no longer clean up litter around the garbage container in the morning.

The lower frequency of cleaning in the morning increased the presence of litter in the early afternoon. During the experiment, with the morning cleaning round cancelled, presence of litter in the early afternoon strongly increased in the treatment group (from 8 to 21 percent of locations) and remained stable in the control group (from 10 to 11 percent of locations). We

[6] Clearly, residents in both the treatment and control area are also exposed to conditions at container locations that they do not use, which may lead to treatment dilution.

focus the empirical analysis on the accumulation of garbage next to the container from the early afternoon until the next morning. During these hours build-up of litter is not distorted by cleaning activities, except when the container is emptied late afternoon. We drop all observations of build-up from afternoon to next morning that are distorted by cleaning activities. This way the number of hours during which garbage is built up is similar across observations.

Garbage can be either bags or household items. We are mainly interested in the overall response in littering behavior to the treatment, but we will report responses for bags and household items separately as robustness checks. Given the heterogeneity in illegally disposed items, with some items tiny and others large, we create an indicator variable that is one in case of build-up of any type of garbage and zero otherwise. Another advantage of using this indicator variable rather than the number of accumulated items is that it is not distorted by the disposal of several items by one individual. If litter begets litter, then we expect build-up of garbage at more locations when the cleaning frequency is reduced compared to the control area. If people learn to clean up after themselves, then we expect to see build-up of garbage at fewer locations compared to the control area in the long run.

For every week, we have four observations of the build-up of garbage per container location: Monday to Tuesday, Tuesday to Wednesday, Wednesday to Thursday, and Thursday to Friday.[7] As mentioned above, we exclude observations for locations that were cleaned in the afternoon due to scheduled emptying of the container. These regular cleaning activities distort the build-up of garbage from the early afternoon until the next morning. The 2 or 3 days per week that the containers are emptied differ from location to location. Since a great majority of the containers are emptied on Monday and Thursday afternoon, we exclude the build-up from Monday to Tuesday and from Thursday to Friday completely. Another reason to exclude these observations is enforcement activities in the afternoon that were scheduled for Mondays and Thursdays in both the control and treatment area. These activities, albeit at a low level and conducted haphazardly, includes removal of garbage bags to search for address labels, distorting the build-up of garbage. In total, we have observations for 50 days

[7] We exclude data on build-up from Friday to Monday because cleaning activities take place throughout the neighborhood on Sundays.

for the period September 20, 2010 until April 1, 2011. The total number of observations including the month after the treatment is equal to 1,730.[8]

Randomization happened at the group level rather than the individual level. Logistical constraints made co-location of treatment locations necessary. That meant that the residential area had to be cut in two parts. The cut of choice was the one for which the two parts were most similar in terms of presence of illegally disposed garbage in the early afternoon in the months before the start of the intervention. A coin toss determined which of the two halves was chosen as treatment group. The treatment group includes 21 container locations; the other 20 locations serve as control group. Figure 2 presents a map of the area. In the sensitivity analysis, we show spatial correlation between co-located container locations not to be an issue: our results are robust to collapsing the data to two groups – treatment and control.

[FIGURE 2]

Table 1 summarizes the baseline characteristics of container locations in our sample. As discussed above, the randomization was targeted at balance in the presence of garbage in the early afternoon. The percentage of locations that were littered in the afternoon was similar between the two areas before the start of the intervention (September 20-November 26, 2010). In the treatment group, litter was observed at 8 percent of locations in the early afternoon; in the control group at 10 percent of locations. An F-test shows the two averages not to be statistically significantly different. The empirical analysis is based on a related outcome variable; the build-up of garbage from early afternoon to the early morning the next day. Build-up of garbage from early afternoon to early morning the next day was somewhat larger in the control area than in the treatment area. The difference is not statistically significant, however. Naturally, in the empirical analysis, we control for pre-existing differences in the build-up of garbage.[9]

---

[8] Over the course of four months, we miss one day due to sickness of the observers and one day due to a late start of their morning round. We miss one day in the week before the start of the treatment, because of miscommunication to the service workers about the nature of the treatment. Due to adverse weather conditions in December, cleaning activities were altered during two days, resulting in another two missing days (December 1 and 8, 2010). Finally, we miss one day due to closure of garbage containers around New Year's Eve.

[9] Occasionally, the registered amount of garbage declined. This happened 4% of the time during the baseline period. In the empirical analysis, these observations are classified as "no build-up of garbage". Our results do not change when we code these observations as missing.

[TABLE 1]

Further evidence for balance between treatment and control is provided by a survey on crime and disorder conducted in 2009 (IVM, 2009). 131 randomly selected households from the area were interviewed for the survey, 70 in the treatment area and 61 in the control area. Table 2 presents the items from the survey that are related to our study. Perceived disorder from littering, worry about crime, age of residents, and employment status are not statistically significantly different between the two areas. Given the similarity in the actual and perceived presence of litter before the experiment, we conclude that the randomization was successful in achieving balance between the treatment and control group.

[TABLE 2]

The estimated treatment effect of less frequent public cleaning may be confounded in two ways, both dealing with activities other than removal of garbage by services workers who do the cleaning round in the morning. First, it is possible that containers are stuck more often in the treatment area during the treatment period. Occasionally, the service workers pull the lever of the container when it seems stuck, fix a container that has become stuck, or take out garbage that is sitting in the opening of a stuck container. With the morning round cancelled during the treatment, greater accumulation of garbage next to the container from the early afternoon to the next morning may simply be the result of a greater frequency of stuck containers in the early afternoon in the treatment area relative to the control area. If so, our results are mainly driven by the specific design of the container, limiting the external validity of our results. To exclude this channel, we include an indicator of a container being stuck or full in the early afternoon as an additional covariate in the estimation equation in the sensitivity analysis. We find this to be of no consequence to our results.[10]

Second, there was a change in the level of law enforcement. The reduced frequency of cleaning lowered the level of law enforcement in the treatment area because the two service workers regularly open garbage bags to search for the identity of the offender. If they find sufficient evidence that links a bag to a household, they impose a 115 euro fine. The fine

[10] We cannot rule out, however, that residents see the presence of garbage next to the container as a signal of the container being stuck or full. This could be an additional channel explaining our results.

arrives in the mail two weeks later. For lack of similar identifiers, people who illegally dispose household items can only be caught *in flagrante*, which almost never happens. Community service provider Stadstoezicht Rotterdam provided us with container-level data on fines for illegal disposal of garbage. The data show that the number of fines per container location per month was more than halved in the treatment area during the experiment, whereas it remained more or less stable in the control area. The effect on littering is likely to be small, however, because of the very low number of fines. Before the experiment, per month some 25 out of the 4,000 households received a fine for illegal disposal of garbage, which comes down to about one per location every two months on average. Given the number of bags observed, the implied chance of getting caught is smaller than 1 percent. Data for previous years show similar low number of fines. Still, to prevent a possible estimation bias, we include the number of fines per location as a covariate in the estimation equation in the sensitivity analysis, and find it to be inconsequential for our results.

In theory, the brief presence in the morning of two service workers may have a deterrent effect in and of itself. If so, there may be an upward bias in the estimated effect of reduced cleaning on littering behavior. This potential bias is limited because of the low level of law enforcement, the short time of the visit of the two workers, the fact that they also had to drive through a couple of treatment streets on their daily route during the experiment, and the focus of the empirical analysis on what happens from the afternoon until the next morning, i.e. in the absence of the service workers.

3. Econometric model and estimation results

A. Non-parametric evidence

To analyze how littering behavior changed as a result of the reduced frequency of cleaning, we non-parametrically estimate the difference in build-up of garbage between the control and treatment area for each day before, during and after the experiment. This flexible estimation framework allows us to analyze trends before the start of the intervention, changes in the behavioral response during the months of the intervention, and persistency in the behavioral response after cleaning frequency went back to its original level. We estimate the following equation:

$$I(\Delta L_{i,t}{=}1) = \sum_{t=1}^{T} D_{i,t}\, \alpha_t + \gamma_i + \delta_t + \varepsilon_{i,t} \qquad (1)$$

$I(\Delta L_{i,t}{=}1)$ is an indicator function which is one in case of build-up of garbage from the early afternoon until the next morning of day *t* at container location *i* and zero otherwise. $D_{i,t}$ are daily dummy variables which are one for container locations in the treatment area and zero for container locations in the control area. We include container location-fixed effects γ_i to control for location-specific factors driving accumulation of litter that are constant over time. We include day-fixed effects δ_t to account for common shocks in littering, such as weather conditions and official holidays. $\varepsilon_{i,t}$ is the idiosyncratic error.

[FIGURE 3]

In Figure 3 the daily coefficients, α_t, from estimation equation (1) are plotted in light gray – as open dots during the treatment period and closed dots outside the treatment period. On the horizontal axis, time 0 corresponds with the first day of the experiment (November 29, 2010). The vertical axis shows the percentage point difference in build-up of garbage between the treatment and control area. The coefficients are estimated relative to the first day for which we have data (September 22, 2010).[11,12] The separately estimated monthly coefficients, plotted in black and superimposed on the daily coefficients in Figure 3, illustrate the major trend in littering behavior before, during, and after the intervention.

First of all, Figure 3 shows the absence of a difference in trend between treatment and control before the start of the treatment, which is in line with the common trend assumption underlying causal interpretation of the parameter of interest. Second, we find littering to go up sharply as soon as cleaning services are reduced. Build-up of litter at container locations goes up by some 10 percentage-points on average. The figure suggests that the frequency of positive build-up increased by about 50 percent (the baseline frequency is 18.5%, see Table 1). Below, we conduct a test whether the average rate of littering during the experiment in the treatment area relative to the control area is statistically significantly different from zero. Finally, Figure 3 suggests some persistency in behavior: littering behavior does not return to its original level once the cleaning round in the morning is re-established – at least not in the

[11] As discussed in Section 2, monitoring started two days earlier. We exclude Mondays and Tuesdays, which is why the first observation is for Wednesday, September 22.

[12] We scale the coefficients to the mean difference in build-up during the baseline period.

first month after the intervention for which we have data. Residents show an immediate response to less frequent cleaning of litter, but seem slow in adjusting to more frequent cleaning.

B. Average treatment effect

Next, we test whether the treatment had on average a statistically significant effect on littering behavior. We estimate the following equation:

$$I(\Delta L_{i,t}=1) = \alpha T_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t} \qquad (2)$$

where T is the treatment dummy, which is one in the experimental area during the intervention period and zero otherwise. Other than replacing the daily dummies by T, estimation equation (2) is identical to equation (1). Parameter of interest α is the estimated change in incidence of accumulation of garbage from the afternoon to the morning the next day.

Table 3 presents the estimation results. We find that the incidence of littering is on average 9 percentage points higher when the cleaning frequency is reduced. The effect is statistically significant at the 95 percent confidence level. In the second column of Table 3, we allow the size of the treatment effect to vary during the intervention period. As can be expected from the pattern in littering behavior observed in Figure 3, the estimated coefficient for the second half of the period is similar to the coefficient for the first half of the period. A Wald-test shows that the difference between the two parameter estimates is not statistically significantly different from zero.

We test for post-intervention persistency in the behavioral effect of the treatment in the third column of Table 3. Some of the theories discussed in the introduction naturally predict that treatment effects do not disappear immediately when the frequency of cleaning returns to its regular level. Particularly, it may take a while before citizens' beliefs about the risk of formal and social sanctions and about other people's contributions to a clean environment adjust to the new context. Further, it may take some time before people realize that the local government is again cleaning up on a daily basis. We find weak evidence for persistency. During the first month after the treatment has ended, the treatment effect is of similar size to the treatment effect during the intervention. The coefficient is statistically significantly at the 90 percent confidence level.

[TABLE 3]

C. Evidence for crowding-out

The results so far lend strong support to the hypothesis that public cleaning services crowd in private contributions. The data on litter at treatment and control locations clearly show that when public cleaning services are diminished, private contributions to a clean environment go down on average. The average treatment effect may hide contrasting behavioral responses, however. To further analyze crowding effects, we focus on the use of a legal alternative for disposal of garbage: making a telephone appointment for pick-up of discarded oversized household items. If crowding in holds, then we expect to see a fall in the number of telephone appointments in response to the treatment. In the case of crowding out, we expect to see the reverse. Community service provider Roteb provided data on these telephone appointments. The data are weekly as appointments could be made for Tuesdays only. Moreover, the data relate to all of the treatment area and all of the control area rather than to specific container locations.

Similar to the analysis of build-up of illegally disposed garbage, first we non-parametrically estimate the effect of the treatment on the number of appointments. We estimate the following equation:

$$A_{i,t} = \sum_{t=1}^{T} D_{i,t}\, \alpha_t + \gamma_i + \delta_m + \varepsilon_{i,t} \qquad (3)$$

where $A_{i,t}$ is the number of appointments for pick-up of discarded household items made in area *i* and week *t*. $D_{i,t}$ are weekly dummy variables that are one for the treatment area and zero for the control area. We include area-fixed effects γ_i and month-fixed effects δ_m.[13]

Figure 4 plots the weekly coefficients, α_t, from estimation equation (3) in light gray. On the horizontal axis, time 0 corresponds with the first week of the experiment (first week of

[13] Given the number of parameters to be estimated in equation (3), we opt for month-fixed effects rather than week-fixed effects. Later, when we estimate the average treatment effect, we use week-fixed effects (equation 4).

December, 2010). For purposes of illustration, we extend the time window that we have used so far with an additional 6 weeks before and 6 weeks after the period of the daily monitoring of illegal disposal of garbage (August, 2010-May, 2011). The vertical axis shows the difference in the number of telephone appointments for pick-up of discarded household items between the treatment and control area. The coefficients are estimated relative to the first week for which we have data on appointments.[14] To illustrate the major trend in the number of appointments, we separately estimate monthly coefficients, which are plotted in black in Figure 4 and superimposed on the weekly coefficients.

[FIGURE 4]

In line with the common trend assumption and similar to our analysis of build-up of illegally disposed garbage in Figure 3, we do not observe a trend before the start of the treatment in Figure 4. As soon as cleaning services are reduced, we find the number of telephone appointments to *go up* in the treatment area relative to the control area. An increase in the number of appointments is in line with the crowding-out hypothesis. At least some residents seem to have learned that discarded household items are no longer quickly removed when they are put next to the container. Own initiative is necessary to make sure that those discarded items disappear. During the treatment period, discarded items such as couches and appliances could be seen in the street for days. Apparently, these items served as visible pointers to the residents, which resulted in the observed change in behavior.

This finding stands in contrast to the average treatment effect that we found before. The effect is also small relative to our previous results for illegal disposal of garbage. In the treatment area, the number of appointments for pick-up of discarded household items went up by 2 to 3 *per week*, while the number of container locations with build-up of garbage went up by about 2 *per day*.[15] The behavioral effect is not insignificant within the context of this legal alternative, however. Given an average number of appointments in the treatment area before the start of the intervention of 8.6 per week, this amounts to about a one-third increase.

[14] We scale the coefficients to the mean difference in the number of appointments during the baseline period.
[15] As discussed in Section 3, the frequency of build-up went up by 9%-points, from about 18.5% in the baseline period. With 21 container locations in the treatment area, this comes down to about 2 additional sites with build-up of garbage.

Again, in line with the findings for build-up of illegally disposed garbage, the behavioral effect of the treatment seems to be persistent. The number of telephone appointments in the treatment area relative to the control area remains higher even when the morning cleaning rounds are re-established.

To test whether the effect of the treatment on this type of legal disposal of household garbage is statistically significantly different from zero, we estimate the following equation:

$$A_{i,t} = \alpha T_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t} \qquad (4)$$

where T is the treatment dummy, which is one in the experimental area during the treatment period and zero otherwise. Parameter of interest α is the estimated change in the number of telephone appointments during the period of the treatment. We include area-fixed effects γ_i and week-fixed effects δ_t.

Table 4 presents the estimation results. In the first column, we estimate the average treatment effect using data for the same period as in Table 3 (September, 2010-March, 2011). We find the number of appointments for pick-up of discarded household items to go up by about 3 in response to the reduced frequency of cleaning. The effect is statistically significant at the 95% confidence level. In the second column of Table 4, we allow the size of the effect to vary over the treatment period. The point estimates are not statistically significantly different. In the third column, we also consider the month after the treatment. We find some evidence for persistency in the effect of the treatment on appointments. The point estimates for the treatment effect and the effect during the period after the intervention are similar in size; the latter is, however, not statistically significantly different from zero. In the last column, similar to Figure 4, we extend the time window of the analysis with 6 weeks before and 6 weeks after the period of the daily monitoring. The estimation results are similar to those based on the shorter period reported in column (3). The post-treatment effect on appointments becomes statistically significant when using this extended time window, providing further evidence for persistency in the behavioral effect.

[TABLE 3]

D. Sensitivity analysis

So far, we assumed variation in littering to be independent across the 41 container locations. Given the randomization at the area level rather than the container level, for reasons discussed in Section 2, the variation in littering among co-located container locations within the treatment area and within the control area may not be fully independent, rendering our statistical tests of the treatment effect invalid. To see how robust our results are to the possibility of spatial correlation among container locations within the treatment and control area, we collapse the data by area. In the second column of Table 5 we present the results for the difference-in-difference estimator. Even though we greatly reduce the number of observations that we use to identify the effect of the treatment, we find the results to be similar. This finding provides support for our assumption that spatial correlation due to co-location of treatment and control locations does not bias our results.

A particular feature of our experimental design is the great number of time periods that we use to estimate the treatment effect. The period before and during the treatment encompasses observations for 42 days over a period of four months. If we would ignore the potential serial correlation in observations by container location, then we would be likely to underestimate the standard errors (Bertrand, Duflo, and Mullainathan 2004). To address this issue, we have clustered the standard errors by container location in the default specification. Not clustering the standard errors would not lead to different conclusions, but it does lower the standard errors by some 10 percent. An alternative solution to the potential problem of serial correlation is to collapse the time series information into a pre- and post-period. The results are presented in the third column of Table 5. Even though the number of observations is greatly reduced, we still find the effect to be statistically significantly different from zero at the 90 percent confidence level. The standard errors are similar to those of the default specification, suggesting that we successfully correct for potential serial correlation within clusters.

In the fourth and fifth column of Table 5, we estimate the treatment effect separately for garbage bags and for discarded household items. We find both estimated coefficients to be positive but statistically insignificant. When studying the data at this disaggregated level, the incidence of illegal behavior becomes too small to identify the treatment effect in a very precise way. However, the estimated effects relative to the mean build-up for the two types of garbage suggest that residents respond similarly to the treatment, justifying our outcome measure that combines the two.

As discussed in section 2, the treatment effect of less frequent public cleaning services may be confounded in two ways, by a change in law enforcement and by a change in the frequency of stuck or full containers. The sixth and seventh column of Table 5 report estimates of the treatment effect when controlling for these factors. First, in the sixth column, we include the number of fines per location as a covariate in the estimation equation in the sensitivity analysis. We lag the effect by 14 days because of the time it takes to process the fine. We assume the fine to have an effect on littering for 14 days. In other words, we count fines written at most four weeks ago and at least two weeks ago.[16] As can be seen from the sixth column of Table 5, controlling for fines does not affect the estimated treatment effect. Likewise, in the seventh column of Table 5, we control for whether a container is stuck or full early afternoon. We find the estimated effect not to be changed, indicating that this specific design feature of the garbage containers is not driving our results.

Lastly, from November to January, two large cargo containers were placed during the weekend as an extra means to dispose of discarded household items. So as not to distort the experiment, one container was placed in the center of the treatment area and one container in the center of the control area. It could be that for some unknown reason residents in one area made greater use of the cargo container than residents in the other area. In that case, estimation of the treatment effect could be biased. Community service provider Roteb provided us with the exact weights of the garbage collected at the two cargo containers. When we include these weights as covariate in the empirical analysis, we find the same treatment effect (see the last column of Table 5), suggesting that the two temporarily placed cargo containers do not affect our results.

[TABLE 5]

[16] Our results are similar when we assume the fines to have an effect on littering behavior for more than two weeks.

4. Conclusions

Litter is a serious problem in many communities around the world (Lewis et al. 2009; Schultz et al. 2009). Quick removal of litter is seen by many as an effective policy response, not only because it ensures a clean environment, but also because the presence of litter may encourage further littering. In this view, government provision of cleaning services crowds in private contributions to a clean environment, and so yields a 'double dividend'. Others have argued that government provision may crowd out private contributions. Why clean up after yourself if government service workers stop by frequently to do the job? Existing empirical work on this issue provides mixed evidence. On the one hand, a number of small-scale, one-shot field experiments show that litter begets litter. On the other hand, survey evidence suggests that many people think that littering can be justified if the local authority cleans up.

We have reported the results from a large-scale and long-lasting field experiment, in which we experimentally reduced the frequency of public cleaning services for a period of 3 months in people's own habitat. Using high-frequency data on litter at treatment and control locations before, during, and after the treatment, we find clear support for the idea that government provision of cleaning services crowds in private contributions. In areas where the frequency of cleaning was reduced, the tendency to litter went up by almost 50 percent. At the same time, residents in the treatment area started to make significantly more appointments for pick-up of discarded household items, a legal alternative to disposal of garbage. This indicates that at least some people learned to clean up after themselves. Overall, our findings suggest that the provision of public cleaning services crowds in private contributions to a clean environment of some people, even in the longer term, but crowds out private contributions of others, with the former effect dominating the latter.

References

Anderson, Siwan, and Patrick Francois (1997), Environmental Cleanliness as a Public Good: Welfare and Policy Implications of Nonconvex Preferences, *Journal of Environmental Economics and Management*, 34(3): 256-274.

Beck, R.W. (2007), Literature Review – Litter; A Review of Litter Studies, Attitude Surveys and Other Litter-Related Literature, Report for Keep America Beautiful.

Beckenkamp, Martin, Christoph Engel, Andreas Glöckner, Bernd Irlenbusch, Heike Hennig-Schmidt, Sebastian Kube, Michael Kurschilgen, Alexander Morell, Andreas Nicklisch, Hans-Theo Normann, and Emanuel Towfigh (2009), Beware of Broken Windows! First Impressions in Public-good Experiments, Preprints of the Max Planck Institute for Research on Collective Goods Bonn 2009/21.

Bergstrom, Theodore, Lawrence Blume, and Hal Varian (1986), On the Private Provision of Public Goods, *Journal of Public Economics*, 29(1): 25-49.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), How Much Should We Trust Differences-in-Differences Estimates?, *Quarterly Journal of Economics*, 119(1): 249-275.

Bertrand, Marianne, Erzo F.P. Luttmer, Sendhil Mullainathan (2000), Network Effects and Welfare Cultures, *Quarterly Journal of Economics*, 115(3): 1019-1055.

Burlando, Roberto M., and Francesco Guala (2005), Heterogeneous Agents in Public Goods Experiments, *Experimental Economics*, 8(1): 35-54.

Cialdini, Robert B., Raymond R. Reno, and Carl 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(6): 1015-1026.

Crump, S. Larry, Dennis L. Nunes, and E.K. Crossman (1977), The Effects of Litter on Littering Behavior in a forest Environment, *Environment and Behavior*, 9(1): 137-146.

Engel, Christoph, Sebastian Kube, and Michael Kurschilgen (2011), Can We Manage First Impressions in Cooperation Problems? An Experimental Study on “Broken (and Fixed) Windows”, Preprints of the Max Planck Institute for Research on Collective Goods, Bonn 2011/5.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler (2011), Testing Enforcement Strategies in the Field: Threat, Moral Appeal, and Social Information, *Journal of the European Economic Association*, forthcoming.

Finnie, William C. (1973), Field Experiments in Litter Control, *Environment and Behavior*, 5(2): 123-144.

Frey, Bruno S., and Stephan Meier (2004), Social Comparisons and Pro-social Behavior: Testing "Conditional Cooperation" in a Field Experiment, *American Economic Review*, 94(5): 1717-1722.

Gächter, Simon (2007), Conditional Cooperation: Behavioral Regularities from the Lab and the Field and Their Policy Implications, In: Bruno S. Frey and Alois Stutzer (Eds.), *Economics and Psychology; A Promising New Cross-Disciplinary Field*, Cambridge, MA: MIT Press.

Geller, E. Scott, Jill F. Witmer, and Margaret A. Tuso (1977), Environmental Interventions for Litter Control, *Journal of Applied Psychology*, 62(3): 344-351.

Huffman, Kim T., William F. Grossnickle, John G. Cope, and Kenneth P. Huffman (1995), Litter Reduction: A Review and Integration of the Literature, *Environment and Behavior*, 27(2): 153-183.

IVM [Integrale Veiligheidsmonitor] (2009), Buurtproblemen in Rotterdam-Rijnmond.

Kahan, Dan M. (1997), Social Influence, Social Meaning, and Deterrence, *Virginia Law Review*, 83: 349-395.

Keizer, Kees, Siegwart Lindenberg, and Linda Steg (2008), The Spreading of Disorder, *Science*, 322: 1681-1685.

Keizer, Kees, Siegwart Lindenberg, and Linda Steg (2011), The Reversal Effect of Prohibition Signs, *Group Processes and Intergroup Relations*, 14(5): 681-688.

Krauss, Robert M., Jonathan L. Freedman, Morris Whitcup (1978), Field and Laboratory Studies of Littering, *Journal of Experimental Social Psychology*, 14(1): 109-122.

Lewis, Alan, Polly Turton, and Thomas Sweetman (2009), Litterbugs; How to Deal with the Problem of Littering, London: Policy Exchange.

Martin, Richard, and John Randal (2008), How is Donation Behaviour Affected by the Donations of Others?, *Journal of Economic Behavior and Organization*, 67(1): 228-238.

Nyborg, Karine, and Mari Rege (2003), Does Public Policy Crowd Out Private Contributions to Public Goods?, *Public Choice*, 115(3-4): 397-418.

Ouiller, Olivier, and Sarah Sauneron (2011), 'Green Nudges': New Incentives for Ecological Behavior, La Note d'Analyse No. 216, Centre d'Analyse Stratégique/Office of the Prime Minister, Paris.

Payne, A. Abigail (2009), Does Government Funding Change Behavior? An Empirical Analysis of Crowd Out, In: Jeffrey R. Brown and James M. Poterba (Eds.), *Tax Policy and the Economy*, Volume 23, Chicago: University of Chicago Press.

Ramos, João, and Benno Torgler (2010), Are Academics Messy? Testing the Broken Windows Theory with a Field Experiment in the Work Environment, FEEM Working Paper No. 104.

Reiter, Susan M., and William Samuel (1980), Littering as a Function of Prior Litter and the Presence or Absence of Prohibitive Signs, *Journal of Applied Social Psychology*, 10(1): 45-55.

Reno, Raymond R., Robert B. Cialdini, and Carl A. Kallgren (1993), The Transsituational Influence of Social Norms, *Journal of Personality and Social Psychology*, 64(1): 104-112.

Schultz, P. Wesley, and Steven R. Stein (2009), Litter in America: National Findings and Recommendations, Washington, DC: Keep America Beautiful.

Schultz, P. Wesley, Renée J. Bator, Lori Brown Large, Coral M. Bruni, and Jennifer J. Tabanico (2011), Littering in Context: Personal and Environmental Predictors of Littering Behavior, *Environment and Behavior*, forthcoming.

Shang, Jen, and Rachel Croson (2009), A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods, *Economic Journal*, 119: 1422-1439.

Table 1. *Baseline characteristics and randomization check*

	N	Treatment area	Control area	Difference
Illegally disposed garbage present in the early afternoon (%)	987 (T), 940 (C)	8.3 (27.6)	9.6 (29.4)	-1.3
Build-up of garbage from early afternoon to early morning next day (%)	308 (T), 326 (C)	14.9 (35.7)	21.8 (41.3)	-6.9

Note. Data relate to baseline period (September-November, 2010) and exclude days at which a container location was emptied. First row includes all business days; second row relates to build-up from Tuesday to Wednesday and Wednesday to Thursday. Standard deviations between parentheses. * Statistically significant at the 10 percent level; ** Statistically significant at the 5 percent level; *** Statistically significant at the 1 percent level.

Table 2. *Perceived crime and disorder and background characteristics of residents in treatment and control area*

	Treatment area	*Control area*	*Difference*
'In this neighborhood, roads, paths and squares are well-kept' (% agree)	46.4 (6.0)	48.3 (6.5)	-2.0
'In this neighborhood, streets are frequently littered' (% agree)	58.6 (5.9)	55.7 (6.4)	2.8
'In this neighborhood, the area around garbage containers is frequently littered' (% agree)	67.1 (5.7)	60.1 (6.3)	6.5
'I am frequently worried about my safety in this neighborhood' (% agree)	14.3 (4.4)	12.1 (4.3)	2.2
Age of residents†	40.6 (1.7)	41.4 (1.8)	-0.8
Paid work for more than 12 hours per week	61.4 (5.9)	57.4 (6.4)	4.1
Number of observations	70	61	

Source: IVM (2009).

Note. † Only residents aged 12 or over were interviewed. Standard deviations between parentheses. * Statistically significant at the 10 percent level; ** Statistically significant at the 5 percent level; *** Statistically significant at the 1 percent level.

Table 3. *The effect of reduced cleaning frequency on accumulation of garbage*

Dependent variable: build-up of garbage next to container from early afternoon until early morning next day (indicator)	(1)	(2)	(3)
Treatment	0.09 (0.04)**		0.09 (0.04)**
Treatment – first half		0.09 (0.05)*	
Treatment – second half		0.08 (0.05)*	
Treatment – post intervention			0.10 (0.06)*
Number of observations	1,474	1,474	1,730

Note. Observations by container location and by day. Standard errors between parentheses corrected for clustering at the container level. Other covariates are container location-fixed effects and day-fixed effects. * Statistically significant at the 10 percent level; ** Statistically significant at the 5 percent level; *** Statistically significant at the 1 percent level.

Table 4. *The effect of reduced cleaning frequency on appointments for pick-up of discarded household items*

Dependent variable: number of telephone appointments for pick-up of garbage	(1)	(2)	(3)	(4)
Treatment	3.18 (1.23)**		3.18 (1.32)**	2.71 (1.22)**
Treatment – first half		3.90 (1.59)**		
Treatment – second half		2.67 (1.43)*		
Treatment – post intervention			2.70 (1.69)	2.53 (1.25)**
Number of observations	44	44	54	78

Note. Observations by area and by week. Standard errors between parentheses. Other covariates are area-fixed effects and week-fixed effects. * Statistically significant at the 10 percent level; ** Statistically significant at the 5 percent level; *** Statistically significant at the 1 percent level.

Table 5. *Sensitivity analysis*

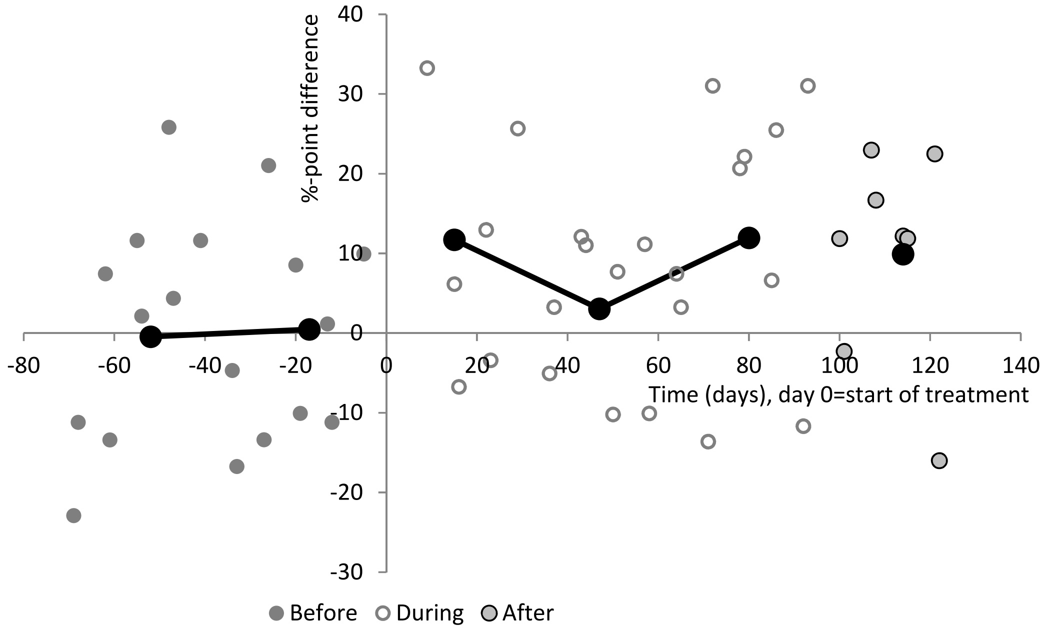
Dependent variable: build-up of garbage (indicator)	(1) Default	(2) Collapse to two areas	(3) Collapse to two periods	(4) Garbage bags only	(5) Household items only	(6) Control for fines	(7) Control for stuck or full container	(8) Control for use of cargo containers
Treatment	0.09 (0.04)**	0.09 (0.04)*	0.08 (0.04)*	0.04 (0.04)	0.04 (0.04)	0.09 (0.04)*	0.09 (0.04)**	0.09 (0.04)**
Number of observations	1,474	84	82	1,474	1,474	1,474	1,474	1,474

Note. Observations by container location and by day and standard errors between parentheses corrected for clustering at the container level, except for (2) and (3). Column (6) includes as a covariate the number of fines for illegal disposal of garbage written at most four weeks ago and at least two weeks ago. Column (7) includes an indicator of a stuck or full container in the early afternoon as covariate, column (8) the weights of garbage disposed at temporarily placed cargo containers. Other covariates are area-fixed effects and day-fixed effects (column 2), container location-fixed effects and period-fixed effects (column 3), and container location-fixed effects and day-fixed effects (all other columns). * Statistically significant at the 10 percent level; ** Statistically significant at the 5 percent level; *** Statistically significant at the 1 percent level.

Figure 1. *Underground garbage container*

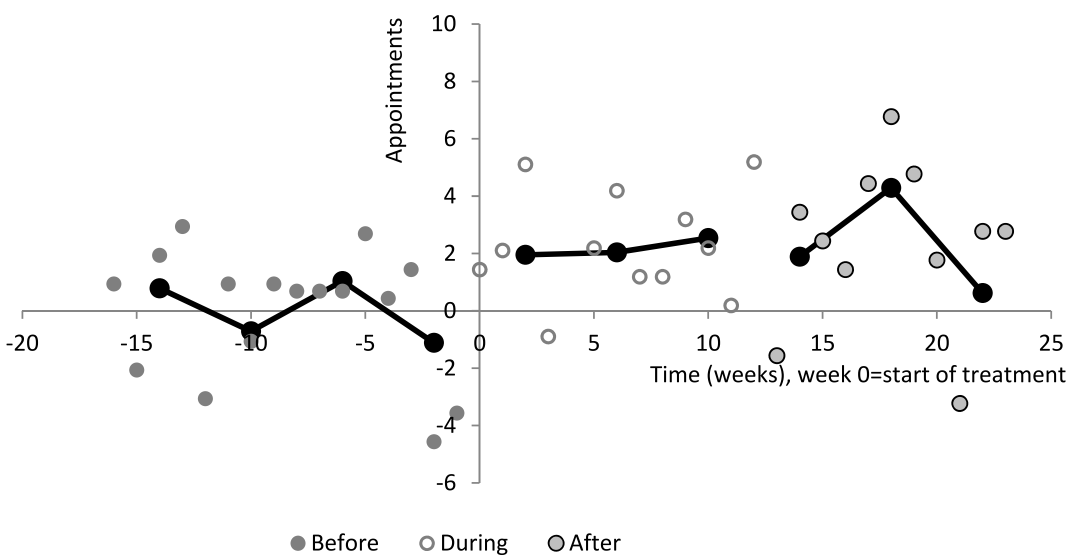
Figure 2. *Map of the area*

Figure 3. *Difference in accumulation of garbage between treatment and control area, by day and by month*



Note. Graph plots coefficients α_t from estimation equation (1). Daily coefficients are in light gray; separately estimated monthly coefficients are in black.

Figure 4. *Difference in number of appointments for pick-up of discarded household items between treatment and control area, by week and by month*



Note. Graph plots coefficients α_t from estimation equation (3). Weekly coefficients are in light gray; separately estimated monthly coefficients are in black.