NBER WORKING PAPER SERIES

THE SHORT-RUN AND LONG-RUN EFFECTS OF BEHAVIORAL INTERVENTIONS: EXPERIMENTAL EVIDENCE FROM ENERGY CONSERVATION

Hunt Allcott
Todd Rogers

Working Paper 18492
http://www.nber.org/papers/w18492

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2012

We thank Ken Agnew, David Cesarini, Gary Charness, Xavier Gabaix, Francesca Gino, Uri Gneezy, Sendhil Mullainathan, and the senior evaluation analyst at the utility for feedback and helpful conversations. Thanks to Tyler Curtis, Lisa Danz, Rachel Gold, Arkadi Gerney, Marc Laitin, Laura Lewellyn, and many others at Opower for sharing data and insight with us. We are grateful to the Sloan Foundation for financial support of our research on the economics of energy efficiency. Stata code for replicating the analysis is available from Hunt Allcott''s website. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation
Hunt Allcott and Todd Rogers
NBER Working Paper No. 18492
October 2012
JEL No. D03,D11,L97,Q41

ABSTRACT

Interventions to affect repeated behaviors, such as smoking, exercise, or workplace effort, can often have large short-run impacts but uncertain or disappointing long-run effects. We study one part of a large program designed to induce energy conservation, in which home energy reports containing personalized feedback, social comparisons, and energy conservation information are being repeatedly mailed to more than five million households across the United States. We show that treatment group households reduce electricity use within days of receiving each of their initial few reports, but these immediate responses decay rapidly in the months between reports. As more reports are delivered, the average treatment effect grows but the high-frequency pattern of action and backsliding attenuates. When a randomly-selected group of households has reports discontinued after two years, the effects are much more persistent than they had been between the initial reports, implying that households have formed a new "capital stock" of physical capital or consumption habits. We show how assumptions about long-run persistence can be important enough to change program adoption decisions, and we illustrate how program design that accounts for the capital stock formation process can significantly improve cost effectiveness.

Hunt Allcott
Department of Economics
New York University
19 W. 4th Street, 6th Floor
New York, NY 10012
and NBER
hunt.allcott@nyu.edu

Todd Rogers
Harvard Kennedy School
todd_rogers@hks.harvard.edu

1 Introduction

While some of the decisions people make are occasional, such as buying a car or enrolling in a retirement savings plan, many of our choices are constantly repeated, such as whether or not to smoke, exercise, eat healthfully, work or study hard, and pay bills on time. Sometimes, our choices differ from those that would maximize social welfare, or perhaps even our own long-run welfare. Individuals, employers, and government agencies have therefore experimented with many different kinds of interventions to "improve" behaviors, including financial incentives, information provision, commitment contracts, appeals to the public good, and social comparisons. In an effort to produce useful and timely insights about these interventions, evaluations often only examine short-run effects. The studies that do examine long-run effects often find that it is difficult to sustainably change behaviors.[1]

Given this, several related questions are crucial to designing and evaluating programs aimed at changing repeated behaviors. First, is it helpful to repeat an intervention, or do responses eventually attenuate? Second, how persistent are effects after the intervention ends? Third, do longer interventions cause more persistent post-intervention effects? These questions often determine whether an intervention is cost-effective, and understanding the answers can help optimize program design.

In this paper, we study the short-run and long-run effects of an intervention that aims to induce energy conservation by sending "Home Energy Reports" that feature personalized feedback, social comparisons, and energy conservation information. The reports are mailed to households monthly or every few months for an indefinite period of time. The program, which is managed by a company called Opower, is typically implemented as a randomized control trial, giving particularly credible estimates of its effects on energy use. Opower's programs have been implemented at 70 utilities across the United States, and there now 8.4 million households in their experimental populations, including five million in treatment. Utilities hire Opower to implement the intervention primarily because the resulting energy savings help to comply with state regulations requiring energy conservation.

[1] For example, see Cahill and Perera (2009) for a review of the long-run effects of interventions to encourage smoking cessation, as well as a number of studies of exercise, weight loss, school performance, and other behaviors that we discuss later in the introduction.

We analyze the one Opower site that uniquely combines three features. First, the program has been running continuously since October 2008, allowing us to assess the durability of effects over a relatively long period. Second, a randomly-selected subset of treatment group households was dropped from treatment after two years, allowing us to measure the persistence of effects for two years after the intervention stops. Third, while most utilities manually record household electricity use on a monthly basis, this utility uses advanced meters that record consumption each day. Although in recent years, millions of households have been outfitted with similar "smart meters" (Joskow 2012, Joskow and Wolfram 2012), the granularity of these data has generated privacy concerns that make them especially difficult to acquire for research. In total, we have just over 200 million observations of daily electricity use over six years at 122,000 de-identified households.

The data reveal a remarkable pattern of "action and backsliding": consumers reduce electricity use markedly within days of receiving each of their initial reports, but these responses decay relatively quickly. The decreases in average daily electricity use within ten days of receiving each of the initial four reports add up to well over 100 percent of the average daily savings over the first year. This is mathematically possible only because consumers backslide: conservation efforts decay at a rate that might cause the treatment effects to disappear in well under a year in the absence of subsequent reports. Interestingly, the consumption decreases immediately after report arrival dates are significantly more positively correlated in treatment than in control. This means that the repeated immediate reductions in average treatment group consumption are not simply due to new households opening reports for the first time: instead, some of the same households are repeatedly having their attention re-directed to energy conservation.

This cyclical pattern of responses attenuates after the first few reports: the immediate consumption decreases after report arrivals become much smaller, and the decay rate between reports becomes statistically indistinguishable from zero. What remains is a durable treatment effect: for the group that continues to receive reports throughout our four-year sample, the effects continue to grow. For the group whose reports are discontinued after two years of treatment, the post-intervention effects decay six to twelve times more slowly than they had between the initial reports. This implies that as the intervention is repeated, people gradually develop a new "capital stock" that makes the effects persistent. This capital stock might be physical capital, such as new energy efficient lightbulbs, or "consumption capital" - a stock of energy use habits in the sense of Becker and Murphy (1988).

Tangibly, what are consumers doing in response to the intervention? The intervention does not induce many large-scale changes to physical capital stock: the utility subsidizes and tracks major household energy efficiency investments such as purchases of energy efficient washing machines and refrigerators, and the differences between treatment and control are not statistically or economically significant. Interestingly, however, treatment and control households also have the same probability of reporting that they have engaged in a broad swathe of energy conservation behaviors over the past year. Although these self-reports should be interpreted cautiously, one interpretation is that some of the behavior changes are on the "intensive margin," by which we mean that the program motivates households to do more of the same things that they already were doing.

After presenting the empirical results, we turn to the economic implications. We show how long-run persistence can play an important role in whether or not it is cost-effective to adopt an intervention. In this particular experiment, different assumption about persistence would have changed the ex-ante predicted cost effectiveness by more than a factor of two. One broader implication is that when deciding whether or not to scale up other interventions when long-run results such as these are not available, it will in some cases be optimal to first measure whether short-run effects are persistent - even if this measurement delays the decision-making process.

We also show how the dynamics of persistence can play an important role in designing behavioral interventions. In doing this, we conceptually distinguish two separate channels through which repeated intervention changes cumulative outcomes. The *durability effect* reflects the extent to which repeated intervention increases effects *during* the treatment period. The *persistence effect* captures the extent to which repeated intervention induces households to change physical capital or consumption habits, which in turn causes the changes in outcomes to last longer *after* the intervention ends. We quantify these effects in the context of the Opower program, showing that the persistence effect is the primary reason why repeated intervention improves cost effectiveness. For Opower and for other interventions in other contexts, this suggests that it is important to repeat an intervention until participants have developed a new "capital stock" of habits or other technologies. After that point, it may be optimal to reduce the frequency of intervention, unless the durability effect is very powerful.

Our results are related to several different literatures. The action and backsliding in response to home energy reports is reminiscent of evidence that consumers "learn" about late fees and other charges as we incur them, but we act as if we forget that knowledge over time (Agarwal *et al.* 2011, Haselhuhn *et al.* 2012). For some consumers, the home energy report acts simply as a reminder to conserve energy, making this related to studies of reminders to save money (Karlan, McConnell, Mullainathan, and Zinman 2010) or take medicine (Macharia *et al.* 1992). Ebbinghaus (1885), Rubin and Wenzel (1996), and others have quantified the decay of memory and the functional form of "forgetting curves." Our results are novel in that they provide a clear illustration of how consumers' attention *repeatedly* waxes and wanes in response to repeated stimuli, but this cyclical action and backsliding eventually attenuates. There are also studies of the medium- and long-run effects of interventions to affect exercise (Charness and Gneezy 2009), smoking (Gine, Karlan, and Zinman 2010, Volpp *et al.* 2009), weight loss (Anderson *et al.* 2010, Burke *et al.* 2012, John *et al.* 2011), water conservation (Ferraro, Miranda, and Price 2011), academic performance (Jackson 2010, Levitt, List, and Sadoff 2010), voting (Gerber, Green, and Shachar 2003), charitable donations (Landry *et al.* 2010), labor effort (Gneezy and List 2006), and other repeated choices. Compared to these studies, we document relatively persistent changes in outcomes over a relatively long time horizon.

Aside from being of scientific interest, these results have very concrete practical importance. Each year, electric and natural gas utilities spend several billion dollars on energy conservation programs in an effort to both reduce energy use externalities and ameliorate other market failures that affect investment in energy-using durable goods (Allcott and Greenstone 2012). Traditionally, one significant disposition of these funds has been to subsidize energy efficient investments, such as Energy Star appliances or home energy weatherization. Recently, there has been significant interest in "behavioral" energy conservation programs, by which is meant information, persuasion, and other non-price interventions.[2] The Opower program is perhaps the most salient example of this approach. One of the foremost questions on practitioners' minds has been the extent to

[2] Abrahamse *et al.* (2005) is a useful literature review of behavioral interventions centered around energy conservation, and Allcott and Mullainathan (2010b) cite some of the more recent work.

which behavioral interventions have persistent long-run effects: while capital stock changes like new insulation are believed to reduce energy use for many years, it was not obvious what would happen after several years of home energy reports. Our results give some initial evidence on this issue.

The paper proceeds as follows. Section 2 gives additional background on the program and describes the data. Sections 3 presents short-run analysis using high-frequency data, while Section 4 presents the long-run analysis. Section 5 discusses some additional evidence on the channels through which the intervention acts. Section 6 carries out the cost effectiveness analysis, and Section 7 concludes.

2 Experiment Overview

2.1 Background

Aside from selling energy, most large electric and natural gas utilities in the U.S. also run energy conservation programs, such as home energy audit and weatherization programs and rebates for energy efficient light bulbs and appliances. While energy conservation can reduce revenues for private investor-owned utilities, many states require utilities to fund conservation programs out of small surcharges called System Benefits Charges, and in recent years many states have passed Energy Efficiency Resource Standards that require utilities to cause consumers to reduce energy use by some amount relative to counterfactual, often 0.5 to 1 percent per year.

Opower is a firm that contracts with utilities to help meet these energy conservation requirements. Their "technology" is unusual: instead of renovating houses or subsidizing energy efficiency, they send two-page Home Energy Report letters to residential consumers every month or every several months. Figure 1 reproduces a home energy report for an example utility. The first page features a "Social Comparison Module," which compares the household's energy use to that of 100 neighbors with similar house characteristics. The second page includes more personalized energy consumption feedback and an "Action Steps Module," which provides energy conservation tips. The exact content of the reports varies over time.

The initial proof of concept that social comparisons could affect energy use was developed in pair of papers by Nolan *et al.* (2008) and Schultz *et al.* (2007). There is also a body of evidence that social comparisons affect choices in a variety of domains, such as voting (Gerber and Rogers 2009), retirement savings (Beshears *et al.* 2012), and charitable giving (Frey and Meier 2004), as well as a broader literature in psychology on social norms, including Cialdini, Reno, and Kallgren (1990) and Cialdini *et al.* (2006).

Building on these initial studies, nearly all of Opower's programs have been implemented as randomized control trials (RCTs), with report recipients randomly selected from a population of residential consumers. This means that it is straightforward to evaluate the effects on energy use. Allcott and Mullainathan (2012) show that the average treatment effects across the first 14 Opower sites range from 1.4 to 2.8 percent of electricity use. Opower programs have also been studied by Allcott (2011), Ayres, Raseman, and Shih (2009), Costa and Kahn (2010), Davis (2011), and a number of consulting reports such as Violette, Provencher, and Klos (2009), KEMA (2012), and Opinion Dynamics (2012). Allcott and Rogers (2012) study the long-run effects at two other Opower sites, finding somewhat less persistence that at the site analyzed here but drawing qualitatively similar conclusions. This paper is significantly different than Allcott and Rogers (2012), as it includes analysis of high-frequency outcome data, more evidence on the mechanisms that underlie the treatment effects, and detailed discussion of cost effectiveness.

2.2 Experimental Design

The experiment we study takes place in the service territory of a large West coast utility which we have been asked not to identify directly. The experimental population comprises 78,887 households that use both natural gas and electricity, are relatively heavy energy users (more than 80 million British thermal units per year), live in single-family homes, have daily energy use data since the beginning of 2007, have at least 100 neighbors in similar-sized houses within a two-mile radius, have valid addresses, and do not have a solar photovoltaic system. The population was randomly assigned to treatment (34,942 households) and control (43,945 households). Two thirds of treatment group households were randomly assigned to receive monthly reports, while one third receive reports each quarter.

The first home energy reports were mailed in October 2008. Approximately 11,600 households were randomly selected to stop receiving reports after September 2010. We call this group the "dropped group." The remainder of the treatment group, which we call the "continued group," is still receiving reports at their original frequency. In February 2011, a "second wave" of 44,000 households from a nearby suburb was added to the program, with half assigned to bimonthly treatment and half to control.

We carry out two analyses in this paper. In the "short-run analysis," we analyze daily energy use data, testing for high frequency variation in the average treatment effects (ATEs). In the "long-run analysis," we collapse the data to the monthly level and measure the treatment effects over the past four years.

2.3 Data for Long-Run Analysis

Table 1 presents descriptive statistics. Baseline electricity usage is the household's average daily consumption during calendar year 2007. The average in the experimental population is 30.3 kilowatt-hours (kWh) per day, or 11,059 kWh per year. This is close to the national average of 11,280 (U.S. Energy Information Administration 2011). For context, one kilowatt-hour is enough energy to run either a typical new refrigerator or one standard 60-watt lightbulb for about 17 hours. This utility has an increasing block price schedule, with marginal prices of 8 to 11 cents/kWh.

While there appear to be very few errors in the dataset, there are a small number of very high meter reads that may be inaccurate. We exclude the 0.00035 percent of observations with more than 1500 kilowatt-hours per day. Baseline energy usage is balanced between treatment and control groups, as well as between the dropped and continued groups within the treatment group. Natural gas usage follows very different patterns than electricity, so for simplicity, we analyze only the latter.

We also observe temperature data from the National Climatic Data Center, which are used to construct heating degree-days (HDDs) and cooling degree-days (CDDs). The heating degrees for a particular day is the difference between 65 degrees and the mean temperature, or zero, whichever is greater. Similarly, the cooling degree days (CDDs) for a particular day is the difference between the mean temperature and 65 degrees, or zero, whichever is greater. For example, a day with average temperature 95 has 30 CDDs and zero HDDs, and a day with average temperature 60 has zero

CDDs and 5 HDDs. HDDs and CDDs vary at the household level, as households are mapped to different nearby weather stations.

In the average American home, heating and cooling are the two largest uses of electricity (U.S. Energy Information Administration 2005), and heating and cooling degree-days are thus important correlates of electricity demand. The higher electricity demand magnifies the level of potential energy conservation, and many Opower programs have highly seasonal effects. This experiment takes place in a moderate climate, with 25 heating degrees on an average day in January and 2.2 cooling degrees on an average day in July. Because the summers are so mild, only a small fraction of households in this site have air conditioners.

There is one source of attrition from the data: households that become "inactive," typically when they move houses. In some cases we observe an account-holder's electricity use at a different location after he or she moves, but we drop these observations, and these people no longer receive reports from the program. As Table 1 shows, 20 percent of households move in the four years after treatment begins, or about four to five percent per year. This is balanced between treatment and control groups, as well as between dropped and continued groups.

There is also a source of attrition from the program: people in the treatment group can contact the utility and opt out of treatment. In this site, 1.8 percent of the treatment group has opted out since the beginning of the program. The majority of this happens within the first two years of treatment: only 0.55 percent of the continued group opts out after October 2010, the beginning of the period when the dropped group has reports discontinued. We continue to observe electricity bills for households that opt out, and we of course cannot drop them from our analysis because this would generate imbalance between treatment and control. We estimate an average treatment effect (ATE) of the program, where by "treatment" we more precisely mean "receiving reports or opting out." Our treatment effects could also be viewed as an intent-to-treat estimate, where by the end of the sample, the Local Average Treatment Effect on the compliers who do not opt out is about 1/0.982 larger than our reported ATE. Because the opt-out rate is so low, we do not make any more of this distinction in our analysis. However, when calculating cost effectiveness, we make sure to include costs only for letters actually sent, not letters that would have been sent to households that opted out or moved.

2.4 Data for Short-Run Analysis

Table 2 presents the daily electricity use data for the short-run analysis. We separate the households into three different groups: the monthly and quarterly groups that begin treatment in October 2008, and the bimonthly group from the second wave that begins in February 2011. Within each group, each household was scheduled to receive reports on the same set of days. For this part of the analysis, we exclude the dropped group households after their reports are discontinued. This reduces the sample size somewhat after September 2010 but does not generate imbalance because the households were randomly selected.

While pre-treatment usage is balanced between treatment and control for the monthly and quarterly groups, this is not the case for the bimonthly group that begins in February 2011: the treatment group's average pre-treatment usage is lower than its control group by 0.69 kWh/day, with a robust standard error of 0.20 kWh/day. The reason is that the partner utility asked Opower to allocate these households to treatment and control based on odd vs. even street address numbers. Thus, it is especially important that we d control for baseline usage when analyzing this group of households. After controlling appropriately, the imbalance does not appear to significantly bias the results, although readers may feel free to focus on the results from the monthly and quarterly groups in the first wave.

In order to analyze how daily average treatment effects respond to the receipt of home energy reports, we must predict when the reports actually arrive. In this experiment, all of the reports to be delivered in a given month for any of the three frequency groups are generated and mailed at the same time. Opower's computer systems generate the reports between Tuesday and Thursday of the first or second week of the month. The computer file of reports for all households in each utility is sent to a printing company in Ohio, which prints and mails them on the Tuesday or Wednesday of the following week. According to the U.S. Postal Service "Modern Service Standards," the monthly and quarterly groups are in a location where expected transit time is eight USPS "business days," which include Saturdays but not Sundays or holidays. The bimonthly group is in a nearby suburb where the expected transit time is nine business days. Of course, reports may arrive before or after the predicted day, and people may not open the letters immediately.

3 Short-Run Analysis

3.1 Graphical

We begin by plotting the average treatment effects for each day of the first year of the experiment for the monthly and quarterly groups, using a seven-day moving window to smooth over idiosyncratic variation. These ATEs are calculated simply by regressing Y_{it}, household i's electricity use on day t, on treatment indicator T_i, for all days t within a seven-day window around day d. We control for a vector of three baseline usage variables $\mathbf{Y}_i^b$: average baseline usage (January-December 2007), average summer baseline usage (June-September 2007), and average winter baseline usage (January-March and December 2007). We also include a set of day-specific constants π_t. For each day d, the regression is:

$$Y_{it} = \tau^d T_i + \boldsymbol{\theta}\mathbf{Y}_i^b + \pi_t + \varepsilon_{it}, \qquad \forall t \in [d-3, d+3] \tag{1}$$

Figure 2 plots the ATEs τ^d with 90 percent confidence intervals. In this regression and all others in the paper, standard errors are robust and clustered at the household level to control for arbitrary serial correlation in ε_{it}, per Bertrand, Duflo, and Mullainathan (2004). Here and everywhere else in the paper, superscripts always index time periods; we never use exponents.

Figure 2 shows that the τ^d coefficients increase rapidly around October 24th, 2008, the date when the first report is predicted to arrive. Four days before the predicted arrival date, the ATE for the monthly group is -0.02 kWh/day, with a 90 percent confidence interval of (-0.11,0.07). By November 3rd, 10 days after the predicted arrival date, the ATE is -0.30 kWh/day, with a confidence interval of (-0.21, -0.40). This is equivalent to each treatment group household turning off five standard 60-watt lightbulbs for an hour, every day. The point estimates decay slightly in absolute value over the next two weeks, but this decay is small relative to the confidence intervals.

The monthly group's second report is predicted to arrive on November 21, 2008. From four days before that date until 10 days after, the treatment effect doubles: it increases in absolute value from -0.28 to -0.61 kWh/day. There are also jumps in the absolute value of the treatment effect - i.e. sudden decreases in treatment group consumption - after the third and fourth reports, but they are not nearly as noticeable as the first two.

The blue line on Figure 2 plots the daily ATEs for the quarterly group, which was randomly selected from the same population as the monthly group. Their first report also should have arrived on October 24th. Between four days before and 10 days after that date, the quarterly group's electricity use similarly decreases by a point estimate of 0.30 kWh/day. Between early November and early January, the treatment effect weakens by 0.1 to 0.2 kWh/day. In practical terms, perhaps half of the lightbulbs that were initially turned off are now back on. The quarterly group's second report arrives on the same day as the monthly group's fourth report: January 23rd, 2009. In the 14 days between January 19th and February 2nd, the point estimates of the quarterly group's treatment effect increase in absolute value from -0.25 to -0.58. These effects similarly decay away until mid-April, when the third report arrives. Around this and the fourth report, the effects similarly jump and decay, although these cyclical responses appear to become less pronounced.

This initial presentation of raw data makes clear the basic trends in households' responses to the treatment. However, the standard errors are wide, and the point estimates fluctuate on top of this basic potential pattern of jumps and decays. Holidays and weather are likely to influence the treatment effects. Collapsing across multiple report arrivals can reduce standard errors and smooth over idiosyncratic variations, and controlling for weather can both increase efficiency and, if correlated with report timing, remove bias.

We therefore estimate the treatment effects in "event time," meaning days before and after predicted report arrival. We index the 48-hour periods before and after report arrival by a and define an indicator variable A_t^a that takes value 1 if day t is a 48-hour periods after a report arrival date. We construct a vector $\mathbf{M}_{it}$ of functions of heating and cooling degree days that parsimoniously captures the typical relationship between these variables and the treatment effects.[3] Denote τ^a as the ATE for each period a. The event time regression is:

$$Y_{it} = \sum_a \tau^a A_t^a T_i + \boldsymbol{\beta}_1 T_i \mathbf{M}_{it} + \boldsymbol{\beta}_2 \mathbf{M}_{it} + \boldsymbol{\theta} \mathbf{Y}_i^b + \pi_t + \varepsilon_{it} \tag{2}$$

Figure 3a plots the τ^a coefficients and 90 percent confidence intervals for the monthly, quarterly, and bimonthly groups using data around each group's first four reports. The point of this graph

[3] $\mathbf{M}_{it}$ contains six variables: $1(CDD_{it}) > 0$, CDD_{it}, $1(0 < HDD_{it} \leq 5)$, $1(5 < HDD_{it} \leq 35)$, $HDD_{it} \cdot 1(5 < HDD_{it} \leq 35)$, and $1(HDD_{it} > 35)$. This were chosen based on inspection of the non-parametric relationship between ATEs and degree-days, as illustrated by Figure 5.

is to show the shape of the τ^a coefficients in event time, not the average level. Thus, the levels of all τ^a coefficients within each group have been shifted so that all three groups can be presented on the same graph. Within each group, the effects follow remarkably similar patterns in event time. The treatment group decreases consumption by about 0.2 kWh/day in the several days around the predicted arrival date. Some reports arrive and are opened before the predicted date, which causes consumption to decrease before $a = 0$. The absolute value of the treatment effect reaches its maximum eight to ten days after the report arrives. After that point, the treatment effect decays, as the treatment group's conservation efforts diminish. This decay is difficult to observe for the monthly group, because before much decay happens, the next report arrives, causing event time to re-start. For the quarterly group, the treatment effect decays by 0.2 kWh/day between 10 days and 80 days after the report arrival.

Figure 3b is analogous to Figure 3a, except that it uses data for all reports beginning with the fifth report. The treatment effects are very close to constant in event time. Coefficients for the bimonthly group are very imprecisely estimated because there are only six reports delivered to this group, meaning that this graph is estimated off of the event windows around only two reports.

3.2 Empirical Strategy

We now carry out formal econometric estimates of the patterns suggested in the figures. First, we estimate the magnitude of the increases in the absolute value of the treatment effect around the report arrival window. Second, we estimate the rate of decay in the treatment effect between reports.

Define S_t^0 as an indicator variable for the seven-day arrival period beginning three days before and ending three days after the predicted arrival date. S_t^{-1} is an indicator for the seven-day period before that, and S_t^1 is an indicator for the seven-day period after. Define $S_t^a = S_t^{-1} + S_t^0 + S_t^1$ as an indicator for all 21 days in that window. As before, $\mathbf{M}_{it}$ is the same functions of weather, $\mathbf{Y}_i^b$ is the three baseline usage controls, and π_t are day-specific dummies. The coefficient on τ^1 in the following regression reflects the change in the treatment effect in period S^1 relative to period S^{-1}:

$$Y_{it} = \left(\tau^a S_t^a + \tau^0 S_t^0 + \tau^1 S_t^1 + \tau\right) \cdot T_i + \boldsymbol{\beta}_1 T_i \mathbf{M}_{it} + \boldsymbol{\beta}_2 \mathbf{M}_{it} + \boldsymbol{\theta} \mathbf{Y}_i^b + \pi_t + \varepsilon_{it} \qquad (3)$$

To estimate the decay rate, we define an indicator variable S_t^w to take value 1 if day t is in a window beginning eight days after a predicted arrival date and ending four days before the earliest arrival of a subsequent report. The variable d_t is an integer reflecting the number of days past the beginning of that period, divided by 365. For example, for a t that is 18 days after a predicted arrival date, d takes value (18-8)/365. Thus, the coefficient on d_t, denoted δ, measures the decay of the treatment effect over the window S^w in units of kWh/day per year.

$$Y_{it} = (\tau^w + \delta d_t) \cdot T_i S_t^w + \boldsymbol{\beta}_1 T_i \mathbf{M}_{it} + \boldsymbol{\beta}_2 \mathbf{M}_{it} + \boldsymbol{\theta} \mathbf{Y}_i^b + \pi_t + \varepsilon_{it} \tag{4}$$

Our model predicts linear decay of the treatment effects as d_t increases. One might hypothesize that the decay process could be convex or concave, and it would seem unrealistic to extrapolate beyond the time when the predicted treatment effect reaches zero. However, our sample is not long enough the effects to return to zero, and it is not large enough to precisely estimate non-linearities. We therefore use the linear model for simplicity.

3.3 Results

Analogously to Figures 3a and 3b, we run the above regressions separately for the initial set of four reports and all reports after that initial set. Table 3a presents the estimates of Equation (3) for the windows around the first four reports. There are three pairs of columns, for the monthly, quarterly, and bimonthly groups. Within each pair, the regression on the right includes the degree-day controls $\boldsymbol{\beta}_1 T_i \mathbf{M}_{it}$ and $\boldsymbol{\beta}_2 \mathbf{M}_{it}$, while the left regression does not. Across the six regressions, the coefficient τ^1 on the TS^1 interaction ranges from -0.162 to -0.248 kWh/day. This means that between the week before the seven-day arrival windows and the week after those windows, the average household that receives a letter reduces consumption relative to counterfactual by the equivalent of three or four 60-watt lightbulbs for one hour. The coefficients do not change substantially when controlling for weather.

What's especially remarkable about the immediate consumption decreases after the initial four reports is that they add up to more than the average daily flow of savings across all days in the first year of treatment. Multiplying the above bounds on the estimated per-report effects $\widehat{\tau}^1$ for the initial four reports by four gives a total decrease of 0.65 to 0.99 kWh per day - the equivalent of turning off a standard 60-watt lightbulb for an additional 11 to 16 hours. By contrast, we will estimate (in Column 1 of Table 6) that the first-year ATE is -0.66 kWh per day. Of course, if the effects did not decay in the intervening days after these initial reports, this would not be mathematically possible.

What makes this possible is that, as we saw in Figures 2 and 3a, the treatment effects do decay between the initial four reports. Table 4a measures this formally using Equation (4). The estimates of δ vary across the three groups, but the coefficients are positive in all regressions and statistically positive in all but one. The quarterly and bimonthly estimates are more highly robust to weather controls. For the monthly group, the point estimates differ somewhat when weather is included. This is likely because relative to the quarterly and bimonthly groups, the monthly group has shorter event windows S^w that can be used to estimate $\widehat{\delta}$, and because the sample period is limited to the first four reports, there are fewer days that can be used to estimate the weather controls $\widehat{\beta}$.

To put the magnitudes of δ in context, focus on the estimates for the quarterly group, controlling for weather. A $\widehat{\delta}$ of 0.738 means that a treatment effect of -0.738 kWh/day would decay to zero in one year, if the linear decay continued to hold. Thus, the jump in treatment effects of $\widehat{\tau}^1 = -0.248$ from Column 4 of Table 3a would decay away fully within about four months. This never happens, because the next report arrives less than three months after the window S^w begins.

Tables 3b and 4b replicate Tables 3a and 4b for the remainder of the samples beginning with the fifth report. Table 3b shows that in the monthly and quarterly groups from the first wave, the τ^1 coefficients are still statistically significant, but they are only about one-quarter the magnitude of $\widehat{\tau}^1$ for the initial four reports. The coefficient for the bimonthly group, however, is relatively large. Because this is estimated off of only the fifth and sixth reports, it is difficult to infer much of a pattern. It could be that there are unobserved moderators of the treatment effects that coincide with these reports, or that the information included in these particular reports was different in a particularly compelling way.

Table 4b shows that there is no statistically significant decay of the effects after the first four reports. Interestingly, all of the point estimates of δ are positive, suggesting that there may still be some decay, but the event windows are not long enough for precise estimates. This highlights the importance of the next section, in which we exploit the discontinuation of reports to estimate a decay rate over a much longer period: two years instead of two to ten weeks.

Appendix Tables A1a-b and A2a-b replicate Tables 3a-b and 4a-b with two sets of additional robustness checks. The left column of each pair excludes outliers: all observations of Y_{it} greater than 300 kWh/day and all households i with average baseline usage greater than 150 kWh/day, which is five times the mean. Based on our inspection of the data, these observations do not appear to be measured with error. However, they implicitly receive significant weight in the OLS estimation, so a small number of high-usage households could in theory drive the results. Relative to Tables 3a-b and 4a-b, the coefficients change only slightly.

The right column of each pair in the appendix replicates the right column in each pair of regressions from the body of the paper, except controlling for the interaction of the treatment dummy with control group average usage on day t. Daily treatment effects are strongly correlated with control group usage, and these regressions control for any underlying patterns in electricity use that might be associated with report arrival times. While much of this correlation acts through the weather controls which are included in Tables 3 and 4, control group average usage is a slightly better predictor of the ATEs. Here again, the coefficients of interest are strikingly robust. The only coefficient that changes is $\widehat{\delta}$ for the monthly group's initial four reports, which shrinks in magnitude, making it statistically indistinguishable from the estimated $\widehat{\delta}$'s for the other five specifications in Table 4a.

All households in all treatment groups receive reports around the same day of the month, typically between the 19th and the 25th. One might worry that our results could somehow be spuriously driven by underlying monthly patterns in the treatment effect. Of course, these underlying patterns would have to take a very specific form: they would need to generate cycles in treatment effects that begin in October 2008 and eventually attenuate for the monthly and quarterly groups, then appear beginning in February 2011 for second wave households but do not re-appear for the monthly and quarterly groups. We can explicitly test for spurious monthly patterns by exploiting the differences in report frequencies to generate placebo report arrivals. We focus on the monthly vs. quarterly frequencies, because they are randomly assigned, and consider only the period after the first four reports, because before that, the quarterly ATE changes significantly in the time between reports.

If there were spurious day-of-month effects, the quarterly group's treatment effects would jump in absolute value at the times when the monthly group receives reports but the quarterly group does not. Appendix Table A3 shows that the $\widehat{\tau}^1$ coefficient for these placebo report arrival dates is statistically zero and economically small relative to the $\widehat{\tau}^1$ estimated in Tables 3a and 3b.

3.4 Household-Specific Repeated Effects

There are two basic models of why average treatment group consumption would repeatedly decrease upon arrival of the first few reports. One model is that each of the initial reports affects new and different households. Because not everyone reads and pays attention to unsolicited mail, only a fraction of households open each report. The repeated cycles in the treatment effect are caused by incremental households opening a report for the first time, and the cycles attenuate because eventually every household has experienced the initial "shock" of opening that first report. A second model reflects the opposite extreme: repeated action and backsliding by the same households.

While it is often difficult to say much about individual-level treatment effects as opposed to average or conditional average treatment effects, our setting offers a unique opportunity to test between these two models. Intuitively, the test is whether households that decrease consumption after one report arrives are more or less likely to decrease consumption after the next report arrives. To implement this formally, index the first four home energy reports by $h = \{1,2,3,4\}$ and denote S_{th}^{-1} and S_{th}^{1}, respectively, as the pre-arrival and post-arrival periods for report h. Define ΔY_{ih} as the difference in household i's consumption after vs. before report h arrives: $\Delta Y_{ih} = \overline{Y}_{it}|(S_{th}^{1} = 1) - \overline{Y}_{it}|(S_{th}^{-1} = 1)$. We wish to test whether ΔY_{ih} is positively correlated with ΔY_{ih-1}.

There are two nuances to this test. First, ΔY_{ih} is the true household-specific treatment effect τ^1 for report h plus a relatively large idiosyncratic error. Thus, the autocorrelation coefficient for ΔY_{ih} is not to be interpreted as the magnitude of autocorrelation in treatment effects, as this would suffer from attenuation bias. Instead, this test should be interpreted as a conservative test of whether the autocorrelation in household-specific treatment effects differs from zero.

The second nuance is that there could be natural underlying sources of positive or negative autocorrelation in ΔY. For example, mean reversion would mechanically generate negative autocorrelation: a household that goes on vacation as the first report arrives and returns as the second report arrives has a negative series of idiosyncratic errors ε_{it} over that period. This would give a negative ΔY_{i1} and a positive ΔY_{i2} as consumption drops and then reverts to normal. Thus, our empirical specification must control for the control group's natural underlying correlation in ΔY and test whether the correlation is relatively higher or lower in the treatment group.

We regress ΔY_{ih} on ΔY_{ih-1}, controlling for the treatment indicator T_i and report-specific intercepts ϕ_h:

$$\Delta Y_{ih} = \rho T_i \Delta Y_{ih-1} + \sigma \Delta Y_{ih-1} + \tau^1 T_i + \phi_h + \varepsilon_{ih} \tag{5}$$

If $\rho < 0$, this means that the first model is more common: the decreases in average treatment group consumption are more likely to be caused by households that have not previously decreased consumption. On the other hand, if $\rho > 0$, this means the decreases in average consumption are more likely to be caused by households that have previously decreased consumption.

Table 5 presents the results of this regression. Columns 1-3 present the results separately for the monthly, quarterly, and bimonthly groups. In order to increase precision, Column 4 combines all of the data and all of the controls from the first three columns. Column 5 adds interactions of T with ϕ, which allows for differential treatment effects for each report in each frequency group. Column 6 excludes outliers - households with baseline usage larger than 150 kWh per day and observations of ΔY larger than 100 kWh/day in absolute value.

The τ^1 coefficients on the T dummies are analogous to the τ^1 coefficients from Equation (3), which measure the treatment group's reduction in consumption between pre-arrival period S^{-1} and post-arrival period S^1. The coefficients will differ in general because the regressions are structured differently, and in particular because Equation (5) excludes the large consumption reduction associated with the first report because there is no lagged change ΔY_{ih-1} for that report. In the first three columns, the $\widehat{\tau}^1$ range from -0.053 to -0.127, slightly less than the $\widehat{\tau}^1$ from Equation (3) but consistent with the basic result of immediate reductions in energy use after reports arrive.

The $\widehat{\rho}$ coefficient is positive in all six regressions, although it is statistically indistinguishable from zero for the bimonthly group when considered in isolation in Column 3. This means that the initial reports repeatedly stimulate at least some of the same households into immediate conservation.

4 Long-Run Analysis

4.1 Graphical

For the long-run analysis, we collapse the same data to the household-by-month level to reduce computational burden and analyze the intervention's effects over four years. We analyze the monthly and quarterly groups together, and we focus on the first wave, as second wave households began only in February 2011.

We first plot the ATEs for each month of the sample for both the continued and dropped treatment groups. The variables D_i and E_i are indicator variables for whether household i was assigned to the dropped group and the continued group, respectively. Both variables take value 0 if the household was assigned to the control group which never received reports, meaning that $D_i + E_i = T_i$. In this regression, m indexes the 56 calendar months from the beginning to the end of the post-baseline sample, from January 2008 through August 2012. The sets of coefficients τ_m^D and τ_m^E are month-specific treatment effects for the dropped and continued groups, respectively. We include 56 month-specific controls for baseline usage, denoted $\theta_m Y_{im}^b$, where Y_{im}^b is household i's average usage in the same calendar month. The variables π_m are month-specific intercepts. The estimating equation is:

$$Y_{im} = \tau_m^D D_i + \tau_m^E E_i + \theta_m Y_{im}^b + \pi_m + \varepsilon_{im} \tag{6}$$

Figure 4 present the estimates of Equation (6). Other than the controls for baseline usage, which substantially improve efficiency, these graphs present unadulterated differences in means. As a result, they give a clear sense of what the data contain and what should be considered in the more formal analysis below. The y-axis is the treatment effect, which is negative because the treatment causes a reduction in energy use. The first vertical line indicates the date of first report generation for the treatment groups. The second vertical line marks the date when the last reports were generated for the dropped group.

To the left of the first vertical line, the intervention has not yet started, and the treatment effect is statistically zero. As we saw in the previous section, consumers respond immediately to the reports. The effects continue to increase in absolute value fairly rapidly over the first year, and the rate of growth in the effect slows after that. Until the second vertical line, both the continued and dropped groups receive the same treatment, and the effects are indistinguishable in the two groups, as would be expected due to random assignment. The treatment effects are stronger in the winter because shorter days require more lighting and colder temperatures require more heating. While nearly all households in the experiment primarily use natural gas for heat, these heating systems also use electricity to power fans, and people may also have portable electric heaters.

The second vertical line marks the beginning of the program's third year. After this point, effects continue to increase in absolute value for the group still receiving reports. By contrast, the effects in the dropped group decay slightly relative to what they had been while the intervention was ongoing. The ATEs in winter and summer 2012 are each about 0.1 kWh/day less than they had been in winter and summer 2011, respectively.

4.2 Empirical Strategy

In the long-run analysis, we ask two questions. First, how durable are the effects as long as the treatment continues? Second, how persistent are the effects after treatment is discontinued? When answering the second question, we can compare long-run decay rates to the short-run decay rates estimated in the previous section.

We define P_m^0, P_m^1, and P_m^2 as indicator variables for whether month m is in the pre-treatment period or the first or second year of treatment, respectively. P_m^3 is an indicator variable for whether month m is in the third or fourth year of treatment, which is the period after the dropped group has reports discontinued. The variable r_m is the negative of time in years until the end of the sample. In the long-run analysis, $\mathbf{M}_{im}$ is a vector of weather controls with two elements: average heating degrees and average cooling degrees for household i in month m. Our estimating equation is:

$$
\begin{aligned}
Y_{im} = & (\tau^0 P_m^0 + \tau^1 P_m^1 + \tau^2 P_m^2) \cdot T_i + \gamma E_i P_m^3 \\
& + \alpha D_i P_m^3 + \delta^{LR} r_m D_i P_m^3 \\
& + \psi_1 \mathbf{M}_{im} (T_i P_m^2 + D_i P_m^3) + \psi_2 \mathbf{M}_{im} (P_m^2 + P_m^3) \\
& + \theta_m Y_{im}^b + \pi_m + \varepsilon_{im}
\end{aligned} \tag{7}
$$

In the first line of this equation, the coefficients τ^0, τ^1, and τ^2 are ATEs for the treatment groups - i.e., both the continued and dropped groups - for the pre-treatment period and the first and second year, respectively. The γ coefficient measures the continued group's treatment effect in years 3 and 4. The second line parameterizes the treatment effect for the dropped group after treatment is discontinued. The coefficient δ^{LR} is the long-run decay rate of the treatment effect. Because r_m has units in years, the units on δ^{LR} will be the change in the treatment effect per year, i.e. kWh/day per year. As in the short-run analysis, we assume linear decay rates for simplicity, because the sample is still not long enough or large enough to reject this assumption. Because r_m increases to zero at the end of the sample, α reflects the fitted treatment effect for the dropped group at the end of the sample.

In the third line, controlling for the interaction of $T_i P_m^2$ and $D_i P_m^3$ with heating and cooling degrees $\mathbf{M}_{im}$ means that the τ^2, α, and δ coefficients reflect treatment effects in months when the mean temperature is 65 degrees. Thus, the δ coefficient reflects the decay in the treatment effect after controlling for weather-related fluctuations. We do not include $E_i P_m^3$ in the interaction because the continued group has a stronger ATE, and thus a potentially larger association between weather and the ATE, during P^3.

4.3 Statistical Results

Table 6 presents the estimates of Equation (7) and closely-related specifications. Column 1 estimates just the τ, γ, and α coefficients, omitting the decay rates and not conditioning on weather. The treatment effects closely map to the effects illustrated in Figure 4: effects increase in absolute value from statistically zero in the pre-treatment period to -0.452 and -0.660 kWh/day in the first and second years, respectively. The program's effects are highly durable: when continued in the third and fourth years, the estimated ATE is -0.842 kWh/day. When the program is discontinued, the effects are also remarkably persistent: the ATE is -0.612 kWh/day for the dropped group in the two years after treatment is discontinued.

Column 2 tests for whether the effects in the two groups increase or decrease in the post-drop period, relative to what they had been in the second year. To do this, we re-estimate Column 1 after substituting $\tau^{23}(P_m^2+P_m^3)\cdot T_i$ for $\tau^2 P_m^2$. The α and γ coefficients now reflect each group's difference in effects for years 3 and 4 relative to year 2. Column 3 repeats this specification including the weather controls from the third line of Equation (7). In each of these two columns, we see that as long as the reports continue over the third and fourth years, treatment group households continue to incrementally reduce energy use. The effects in the dropped group are smaller in absolute value, but this difference is not statistically significant from what it had been in the second year. This means that the decay of the treatment effect is slow enough that it cannot be picked up in this specification. As Figure 4 illustrates, however, the effect does appear to decay after reports are discontinued. Combining the third and fourth years into one period makes it difficult to detect the decay between the beginning and the end of that period.

Column 4 adds the linear decay term $\delta^{LR} r_m D_i P_m^3$ to the basic specification in Column 1. Relative to control, consumption in the dropped group increases by 0.131 kWh/day per year across the third and fourth years. Column 5 includes the weather terms, meaning that this is exactly the specification in Equation (7). Although they are not statistically significant, the ψ coefficients are both negative, which implies that as temperatures deviate more from 65 degrees, the treatment effect becomes stronger. The weather controls change $\widehat{\delta}^{LR}$ only slightly, to 0.117 kWh/day per year. The $\widehat{\alpha}$ coefficients show that the predicted treatment effect at the end of the first year is -0.45 kWh/day, which closely aligns with the illustration in Figure 4.

Column 6 repeats Column 5 including only the balanced panel, and the coefficients are all essentially unchanged. This means that the changes in the effects over time are not somehow due to consumers with different treatment effects differentially attriting from the sample as they move.

Allcott (2011) documents that the program causes more conservation by heavier baseline users, and monthly treatment causes more conservation than quarterly. Appendix Table 4 confirms these results for this site but documents that the standard errors are too large to infer much about whether decay rates differ along these dimensions.

In sum, the results show that when reports are discontinued after two years of treatment, about two-thirds of the effect remains two years later. In tangible terms, a treatment effect of -0.660 kWh/day for the program's second year means that the average treatment group household took actions equivalent to turning off a standard 60-watt lightbulb for about 11 hours each day. At the end of the sample, the average dropped group household took actions equivalent to turning off that lightbulb for 7.5 hours each day. Meanwhile, the average continued group household was doing twice as much as the average dropped household - the equivalent of turning off that lightbulb for about 15 hours each day. We can also compare the estimated long run decay rate $\widehat{\delta}^{LR}$ to the decay rate between each of the first four reports, the $\widehat{\delta}$ estimated in the previous section. In most specifications, this $\widehat{\delta}$ was between 0.75 and 1.5 kWh/day per year, which is six to 12 times faster than $\widehat{\delta}^{LR}$. In the next two sections, we discuss the implications of the differences between the initial decay rate δ and the long-run decay rate δ^{LR}.

5 Mechanisms

Concretely, what actions underlie the observed effects? In many behavioral interventions, this question can be difficult to answer. For example, if a program incentivizes people to lose weight, it may not be clear how much of the observed weight loss comes from exercise vs. reduced calorie intake, and within these categories, what form of exercise is the most useful and what foods have been cut out of their diets. In our setting, we can use utility program participation data and households' self-reported actions to shed some light on two questions. First, how much of the effect comes from large observed changes to physical capital stock? Second, what do treatment group households report that they are doing differently than control group households?

5.1 Utility Energy Efficiency Program Participation

Like many utilities across the country, the utility we study runs a series of other energy conservation programs that subsidize or directly install energy efficient capital stock. The utility maintains household-level program participation data, which are primarily used to estimate the total amount of energy conserved through each program. These household-level data are also useful in estimating whether the Opower intervention affects energy use through an increase in program participation.

Table 7 presents the program participation data for the experimental population for an example year, calendar year 2011. For each program, the utility has estimated the kilowatt-hours of electricity that a typical participant would save. The table lists all ten programs where any amount of electricity would be conserved and at least one household in the first wave experimental population participated in 2011. The most popular of these programs are a subsidy for new energy efficient clothes washers, installation of compact fluorescent lightbulbs, the removal of an old energy-inefficient refrigerator or freezer, and installation of low-flow showerheads.[4]

Column 1 of Table 7 shows the estimated flow of savings per participant, translated into kilowatt-hours per day to be consistent with the units in the rest of the paper. Column 3 shows the difference, also in kWh/day per household, in estimated savings between the continued and control groups. Column 4 reports the difference in estimated savings between the entire treatment group and the control group. Only one program, a program to replace traditional incandescent lightbulbs with energy-saving Compact Fluorescent Lightbulbs (CFLs), shows a statistically significant in savings between treatment and control. The standard errors are very tight, allowing us to rule out any economically significant differences. The CFL program, for example, appears to generate 0.00224 kWh/day incremental savings in the continued treatment group. Using the estimates in the bottom row, which combine the savings across all programs, the upper bound of the 90 percent confidence interval on savings is 0.006 kWh/day. By contrast, the continued group's treatment effect in the program's third year was -0.842 kWh/day, which was an increment of -0.181 compared to the year before.[5]

Aside from changing the rate at which households participate in energy efficiency programs, the Opower intervention could also change the timing of their participation. In other words, the

[4] The utility also runs "weatherization" programs, including installation of new insulation and re-sealing of heating and cooling system ducts. The utility deems that these programs reduce natural gas use but do not reduce electricity use. Regardless of whether this is exactly correct, participation rates are also statistically and economically identical between treatment and control.

[5] In other locations, the Opower program often does cause a statistically significant difference in utility program participation between treatment and control. As in the utility we study, however, these differences typically account for only a small portion of the treatment effect. See, for example, Opinion Dynamics (2012).

intervention could move forward investments that would otherwise have happened later. To test this, Column 5 reports estimated savings for calendar year 2011 only, pro-rating each participant's savings over only the part of the year after their recorded date of program participation. There are no statistically or economically significant differences.

Of course, these data only reflect households that participate in utility-sponsored programs. For large investments such as clothes washers, refrigerators, and insulation, the subsidies are large, so suppliers have strong incentives to report their customers' investment in order to collect the subsidy. Thus, these data are likely to be good measures of large changes to physical capital stock.

5.2 Surveys of Self-Reported Actions

Other than large changes to physical capital stock, there are many other ways to conserve energy. One way to attempt to measure these is through surveys of self-reported actions. In the past two years, Opower has surveyed about six thousand people in treatment and control groups in six sites nationwide, including 800 people in the utility we study in this paper. The surveys are conducted via telephone on behalf of the utility, and for practical reasons no effort is made to obscure the fact that the surveys are about energy use and conservation behaviors. Completion rates are typically between 15 and 25 percent of attempts. Because these are self-reported actions that reflect only a small share of the experimental population, we discuss these data only briefly, and they should be interpreted with great caution.

In these surveys, respondents are first asked if they have taken any steps to reduce energy use in the past 12 months. Those who report that they have are asked whether or not they have taken a series of specific actions in the past twelve months. The particular actions vary by survey, but many of the same actions are queried in multiple sites. In the utility we study, respondents were asked about 11 actions. We group actions into three categories: *repeated actions* such as switching off power strips and turning computers off at night, *physical capital changes* such as purchasing Energy Star appliances, and *intermittent actions* such as replacing air filters on air conditioning or heating systems.

Table 8 presents the survey data. The left three columns contain results from all six survey sites, while the right three columns contain results only from the site we study in this paper.

Within each set of three columns, the first presents the mean number of people who report taking the action in the past 12 months. The second column presents the difference between treatment and control in the probability of reporting the action. The third column presents the difference in probability after controlling for five respondent characteristics: gender, age, whether homeowner or renter, education, and annual income.

Table 8 shows that there is very little difference in self-reported actions across treatment and control groups. The only difference that is consistent across different surveys is that treatment group households are more likely to report having a home energy audit. Audits often include installation of new compact fluorescent lightbulbs, which typically last several years until they must be replaced, and also are typically required before moving forward with larger investments such as weather sealing or new insulation. One survey, at a utility in a warmer climate than the one we study, finds that the treatment group is more likely to report using fans to keep cool instead of running air conditioners. Across all sites, the treatment group is more likely to report participating in utility energy efficiency programs, but the difference is not statistically significant in the specification that conditions on observables. In fact, in the utility we study, the treatment group reports being *less* likely to have taken any steps to reduce energy use, although this is not the case in other sites, and the difference is also not statistically significant when including observable characteristics.

For each of the three categories, the first row (in bold) presents a test of whether the average probability of taking actions in each category differs between treatment and control. When aggregating in this way across actions and across sites, our standard errors are small enough to rule out with 90 percent confidence that the intervention increases the probability of reporting energy conservation actions by more than one to two percent. Throughout Table 8, this lack of statistical significance would only be further reinforced by adjusting the p-values for multiple hypothesis testing.

There are multiple interpretations of these results. First, the true probabilities of taking actions might differ between treatment and control, but the surveys might not pick this up due to demand effects, respondents' general tendency to systematically over-report, selected samples, and the fact that different respondents might interpret questions differently. Second, the treatment could cause small changes in the true probability of taking a wide variety of actions. Such changes could potentially add up to the observed effects on electricity use even though no one action accounts for much on its own. Third, it is possible that the intervention does not change whether or not people take specific actions to conserve energy, which is what Table 8 reports, but instead changes the intensity with which some people take actions they were already taking. In other words, an important impact of the intervention could be not to give information about new ways to conserve, but instead to increase attention and motivation to conserve in more of the same ways.

5.3 Summary: The Dynamics of Household Responses

Taken together, our analyses of short-run and long-run treatment effects, along with information on program participation and self-reported actions, start to paint a picture of how consumers respond to the Opower intervention. As the initial reports arrive, some consumers are immediately motivated to conserve. They take actions that are feasible within a short period of time, probably changing utilization choices by adjusting thermostats, turning off lights, and unplugging unused electronics. However, behaviors soon begin to "backslide" toward their pre-intervention levels. In this intervention, subsequent reports arrive before behaviors fully return to their pre-intervention state, and these additional reports repeatedly induce at least some of the same households to conserve. This process of action and backsliding is a repeated version of the phenomena documented by Agarwal *et al.* (2011) and Haselhuhn *et al.* (2012), who show that consumers learn to avoid credit card and movie rental late fees after incurring a fee, but they act as if this learning depreciates over time.

Several mechanisms might cause this repeated action and backsliding. Repeated learning and forgetting could generate this pattern, although it seems unlikely that people would forget this information at this rate. Alternatively, it could be that consumers learn that energy conservation actions are more difficult or save less energy than they had thought. However, to generate repeated cycles in energy use, consumers must be experimenting with different energy conservation actions after receiving each new report, and it is not clear whether this is the case. A third mechanism is what we call *malleable attention*: reports immediately draw attention to energy conservation, but attention gradually returns to its baseline allocation.

After the first few reports, this repeated action and backsliding attenuates, and decay of treatment effects can only be observed over a longer period after reports are discontinued for part of the treatment group. This means that in the intervening one to two years between the initial reports and the time when the reports are discontinued, consumers form some type of new "capital stock" that decays at a much slower rate. The program participation data shows that very little of this capital stock is large changes to physical capital such as insulation or major appliances. However, this capital stock might take the form of a wide variety of smaller changes, such as installing energy efficient Compact Fluorescent Lightbulbs.

Much of this capital stock may also reflect changes to consumers' utilization habits, which Becker and Murphy (1988) call "consumption capital." This stock of past conservation behaviors lowers the future marginal cost of conservation, because the behavior has become automatic and can be carried out with little mental attention (Oullette and Wood 1998, Schneider and Shiffrin 1977, Shiffrin and Schneider 1977). However, just as in the Becker and Murphy (1988) model and most models of capital stock, consumption capital also decays. This story is consistent with the results of Charness and Gneezy (2009), who show that financial incentives to exercise have some long-run effects after the incentives are removed, suggesting that they induced people to form new habits of going to the gym.

6 Cost Effectiveness and Program Design

In this section, we assess the importance of persistence for cost effectiveness and for program design. We use a simple measure of cost effectiveness: the dollar cost to produce and mail reports divided by the kilowatt-hours of electricity conserved. Although cost effectiveness is a common metric by which interventions are assessed, we emphasize that this is not the same as a welfare evaluation. In this context, consumers might experience additional unobserved costs and benefits from the intervention: they may spend money to buy more energy efficient appliances or spend time turning off the lights, and they might be more or less happy after learning how their energy use compares to their neighbors'. Furthermore, the treatment causes households to reduce natural gas use along with electricity; we have left this for a separate analysis. In addition, this measure does not take into account the fact that electricity has different social costs depending on the time of day when it is consumed.

In a first analysis, we calculate the cost effectiveness of the existing program using only the sample data, which allows us to demonstrate the importance of persistence with zero additional assumptions. Second, we calculate the cost effectiveness of different potential program designs using additional assumptions about discount rates and persistence.

6.1 In-Sample Cost Effectiveness

Table 9 presents the "in-sample cost effectiveness" estimates: total program costs divided by total electricity savings observed between the beginning and end of the sample. Calculating cost effectiveness only over the sample period allows us to demonstrate the importance of different assumptions about persistence without needing to predict future effects. If we were to extrapolate further into the future, assumptions about persistence would make even more of a difference.

For this stylized analysis, we assume that the cost per report is \$1 and that there are no fixed costs of program implementation. The savings estimates are simply the average treatment effects for each period estimated in column 1 of Table 3 multiplied by the length of each period. For example, scenario 1 reflects the observed results for the continued treatment group. The total electricity savings are (0.452 kWh/day)·365 days + (0.660 kWh/day)·(365 days) + (0.842 kWh/day)·(700 days)=995 kWh. Standard errors are calculated using the Delta method. For simplicity, there is no time discounting

Opower competes against other energy conservation programs, and cost effectiveness is one of the most important metrics for comparison. There are some benchmark numbers available, although they are controversial (Allcott and Greenstone 2012). The American Council for an Energy Efficient Economy estimates that in 14 states with aggressive energy conservation programs, the states' cost effectiveness estimates ranged from 1.6 to 3.3 cents per kilowatt-hour (Friedrich *et al.* 2009). Using nationwide data, Arimura *et al.* (2011) estimate cost effectiveness to be about 5.0 cents/kWh, assuming a discount rate of five percent.

For the continued treatment group in scenario 1, dividing a total cost of \$32.10 per household by the 995 kWh of observed savings gives cost effectiveness equal to 3.23 cents/kWh. Scenario 2 reflects the observed results for the dropped treatment group. Because the reports were only sent for two years, the costs of this treatment are substantially less, and because the effects are so persistent, the savings are almost as large. As a result, the in-sample cost effectiveness is significantly improved relative to the continued group: 2.15 cents per kilowatt-hour.

Of course, these estimates benefit from hindsight: up to sampling error, we now know exactly what the effects were. When deciding whether to adopt a program, policymakers must make a series of assumptions about efficacy during treatment and persistence after treatment is discontinued. Scenarios 3 and 4 calculate cost effectiveness under the two alternative extreme assumptions. The first assumption is zero persistence: after the treatment stops, the effects end immediately. While this assumption may seem unduly conservative in light of our empirical results, this is the implicit assumption under which many ongoing Opower programs have been evaluated in the absence of these results. The second assumption is full persistence over the remainder of the sample. Under zero persistence, cost effectiveness is 4.42 cents/kWh, while under full persistence, cost-effectiveness is 2.07. Thus, the alternative extreme assumptions about persistence make more than a 100 percent difference in cost effectiveness. Counter to the way the programs have often been evaluated, the empirical estimates over this sample period are closer to the full persistence assumption.

The bottom panel of Table 9 displays retail electricity cost savings. The average household in the continued and dropped groups, respectively, consumed \$100 and \$83 less electricity between October 2008 and the end of August 2012. If these treatments had been applied to all 78,887 households in the experimental population, the total retail electricity cost savings over the sample would be \$7.85 and \$6.58 million, respectively. If treatment were discontinued after two years and there had been zero persistence or full persistence over the sample period, total retail electricity cost savings would have been \$3.20 or \$6.85 million, respectively.

The point of this discussion is not the exact cost effectiveness estimates: the calculation is stylized, and the exact numbers will not generalize across locations. Instead, the point is to highlight two important and potentially-generalizable issues. First, alternative assumptions about persistence can make a large difference in cost effectiveness. Second, the extent of persistence can influence program adoption decisions: in this case, all of the cost effectiveness numbers are close to estimated ranges reported above for competing energy conservation programs.

6.2 Program Design

The in-sample estimates above demonstrate the importance of persistence with no additional assumptions, but they do not take account of the additional energy conservation that will be almost surely be observed in the future. In this subsection, we compare the cost effectiveness of different potential program designs while making out-of-sample assumptions about persistence. Because some of the electricity usage reductions are now much farther in the future, time discounting is important. All dollars costs and usage reductions are now discounted to the beginning of the program at a five percent discount rate.

Table 10 compares four different program designs. The first is a one-shot intervention - one report, with no future contact. We assume that there is an initial effect of 0.30 kWh/day, which decays linearly at 0.75 kWh/day per year, consistent with estimates for the quarterly group from Table 4. As a result, the effect is assumed to have fully disappeared in approximately 0.4 years. The next three designs are one, two, and four-year interventions. As in the previous table, effect sizes during treatment are taken directly from the parameter estimates in Table 3. Because Table 4 shows that there is no statistically significant decay between reports after the initial four, we assume the long-run decay rate δ^{LR} estimated in Table 5 for each of these programs that last one year or longer. With additional empirical experimentation, this assumption could be refined, and one might naturally hypothesize that δ^{LR} decreases continuously with the length of the intervention.

We conceptually distinguish two fundamental channels through which repeated intervention affects outcomes. First is the *durability effect*: repeated intervention can increase the cumulative change in outcomes, holding constant the decay of post-intervention effects. This happens both because the treatment period is mechanically extended and because repeated intervention can affect the intensity of people's responses. For example, receiving incremental reports might eventually cause consumers to turn off two lights each night instead of one. Second is the *persistence effect*: repeated intervention can change the composition of responses to involve changes in "capital stock," which generates more persistent post-intervention effects. For example, receiving one report might cause consumers to turn off the lights every night, until their attention wanes. Receiving multiple reports might eventually cause consumers to buy an automatic switch that turns off the lights for them every night.

Figure 5 illustrates the incremental effects of lengthening the intervention from one report to one year. A one-report intervention causes the average household to conserve 22 kWh. Extending the intervention to one year increases the ATE to 0.45 kWh per day, causing a durability effect of 194 kWh. Extending the intervention also reduces the decay rate from δ to δ^{LR}, which increases the present discounted value of post-treatment savings by 234 kWh.

The bottom panel of Table 10 quantifies the incremental effects of repeated intervention through these three channels. The largest change in cost effectiveness happens as the intervention is extended from one report to one year, which fully exploits the persistence effect. Given that we assume no change in δ after the first year, there is no persistence effect as the program is extended from one to four years. In this particular example, incremental intervention remains cost effective because the durability effect continues to generate more energy conservation.

The setup of this experiment makes it difficult to estimate with any precision how δ changes as a function of time with continued intervention, so Table 10 rests on a set of assumptions about that parameter. This should not obscure the central message: the persistence effect - inducing consumers to change "capital stock" - is a significant driver of improved long-run cost effectiveness. Thus, a crucial ingredient in designing behavioral interventions in many different contexts is to understand how long the intervention needs to be continued until the persistence effect has been exploited.

7 Conclusion

In this paper, we study part of a large and policy-relevant set of randomized control trials designed to induce people to conserve energy. Aside from the specific relevance of our results to policymakers and economists studying energy markets, we believe that there are four broader takeaways for behavioral scientists. First, our data clearly illustrate how people can repeatedly cycle through action and backsliding in response to repeated interventions. This suggests that attention is malleable, not static: an intervention can draw attention to one set of repeated behaviors, but that attention gradually returns to its baseline allocation. Second, our empirical results document how repeated intervention can eventually cause people to change the composition of their responses, which generates more persistent changes in outcomes. These persistent effects might result from habitual behavior change, or they may result from changes in physical capital or other technologies that change outcomes without additional action. As Charness and Gneezy (2009) and others have shown, the same effect translates to other contexts: for example, a one-time encouragement to lose weight might cause people to diet for a week, while a longer-term intervention is more likely to eventually encourage people to find a workout partner and habitually go to the gym.

Third, our cost effectiveness analysis shows how long-run persistence can materially change whether or not a program is cost-effective. This suggests that for other types of interventions where long-run results such as these are not available, it may sometimes be worthwhile to delay decisions about scaling up a program until long-run effects can be measured. Fourth, we conceptually distinguish and quantitatively measure two different reasons for ongoing intervention: the durability and persistence effects. In our example, we show that the greatest improvement in cost effectiveness happens as the intervention is continued long enough for the persistence effect to take hold. This suggests that an important part of the future research agenda on behavioral interventions is to more precisely identify when and how people form new "capital stocks" that cause persistent changes in outcomes.

References

[1] Abrahamse, Wokje, Linda Steg, Charles Vlek, and Talib Rothengatter (2005). "A Review of Intervention Studies Aimed at Household Energy Conservation." Journal of Environmental Psychology, Vol. 25, No. 3 (September), pages 273-291.

[2] Agarwal, Sumit, John Driscoll, Xavier Gabaix, and David Laibson (2011). "Learning in the Credit Card Market." Working Paper, Harvard University (April).

[3] Allcott, Hunt (2011). "Social Norms and Energy Conservation." *Journal of Public Economics*, Vol. 95, No. 9-10 (October), pages 1082-1095.

[4] Allcott, Hunt and Michael Greenstone (2012). "Is There an Energy Efficiency Gap?" *Journal of Economic Perspectives*, Vol. 26, No. 1 (Winter), pages 3-28.

[5] Allcott, Hunt, Sendhil Mullainathan, and Dmitry Taubinsky (2012). "Energy Policy with Externalities and Internalities." Working Paper, New York University (September).

[6] Allcott, Hunt, and Sendhil Mullainathan (2010a). "Behavior and Energy Policy." *Science*, Vol. 327, No. 5970 (March 5th).

[7] Allcott, Hunt, and Sendhil Mullainathan (2010b). "Behavioral Science and Energy Policy." Working Paper, New York University (February).

[8] Allcott, Hunt, and Todd Rogers (2012). "How Long Do Treatment Effects Last? The Persistence and Durability of a Descriptive Norms Intervention in Energy Conservation." Working Paper, Harvard University (October).

[9] Anderson, LA, T. Quinn, Karen Glanz, G Ramirez, LC Kahwati, DB Johnson, LR Buchanan, WR Archer, S Chattopadhyay, GP Kalra, DL Katz, and the Task Force on Community Preventive Services (2009). "The Effectiveness of Worksite Nutrition and Physical Activity Interventions for Controlling Employee Overweight and Obesity: A Systematic Review." *American Journal of Preventive Medicine*, Vol. 37, No. 4 (October), pages 340-357.

[10] Argote, Linda, Sara Beckman, and Dennis Epple (1990). "The Persistence and Transfer of Learning in Industrial Settings." *Management Science*, Vol. 36, No. 2 (February), pages 140-154.

[11] Arimura, Toshi, Shanjun Li, Richard Newell, and Karen Palmer (2011). "Cost-Effectiveness of Electricity Energy Efficiency Programs." Resources for the Future Discussion Paper 09-48 (May).

[12] Ayres, Ian, Sophie Raseman, and Alice Shih (2009). "Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage." NBER Working Paper 15386 (September).

[13] Becker, Gary, and Kevin Murphy (1988). "A Theory of Rational Addiction." *Journal of Political Economy*, Vol. 96, No. 4 (August), pages 675-700.

[14] Benkard, Lanier (2000). "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review*, Vol. 90, No. 4 (September), pages 1034-1054.

[15] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, Vol. 119, No. 1 (February), pages 249-275.

[16] Beshears, John, James Choi, David Laibson, Brigitte Madrian (2009). "How Does Simplified Disclosure Affect Individuals' Mutual Fund Choices?" NBER Working Paper No. 14859 (April).

[17] Beshears, John, James Choi, David Laibson, Brigitte Madrian, and Katherine Milkman (2012). "The Effect of Providing Peer Information on Retirement Savings Decisions." Working Paper, Stanford University (May).

[18] Bollinger, Brian, Phillip Leslie, and Alan Sorensen (2011). "Calorie Posting in Chain Restaurants." *American Economic Journal: Economic Policy*, Vol. 3, No. 1 (February), pages 91-128.

[19] Burke, LE, MA Styn, SM Sereika, MB Conroy, L Ye, Karen Glanz, MA Sevick, and LJ Ewing. "Using mHealth Technology to Enhance Self-Monitoring for Weight Loss: A Randomized Trial." *American Journal of Preventive Medicine*, Vol. 43, No. 1 (July), pages 20-26.

[20] Cahill, Kate, and Rafael Perera (2008). "Competitions and Incentives for Smoking Cessation." *Cochrane Database of Systematic Reviews*, No 3.

[21] Charness, Gary, and Uri Gneezy (2009). "Incentives to Exercise." *Econometrica*, Vol. 77, No. 3 (May), pages 909-931.

[22] Chetty, Raj, Adam Looney, and Kory Kroft (2009). "Salience and Taxation: Theory and Evidence." *American Economic Review*, Vol. 99, No. 4 (September), pages 1145-1177.

[23] Cialdini, Robert, Linda Demaine, Brad Sagarin, Daniel Barrett, Kelton Rhoads, and Patricia Winter (2006). "Managing Social Norms for Persuasive Impact." *Social Influence*, Vol. 1, No. 1, pages 3-15.

[24] Cialdini, Robert, Raymond Reno, and Carl 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*, Vol. 58, No. 6, pages 1015-1026.

[25] Costa, Dora, and Matthew Kahn (2010). "Energy Conservation Nudges and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment." NBER Working Paper No. 15939 (April).

[26] Cutler, David, Robert Huckman, and Mary Beth Landrum (2004). "The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery." *American Economic Review*, Vol. 94, No. 2 (May), pages 342-346.

[27] Davis, Lucas (2008). "Durable Goods and Residential Demand for Energy and Water: Evidence from a Field Trial." *RAND Journal of Economics*, Vol. 39, No. 2 (Summer), pages 530-546.

[28] Davis, Matthew (2011). "Behavior and Energy Savings." Working Paper, Environmental Defense Fund (May).

[29] DellaVigna, Stefano, and Ulrike Malmendier (2006). "Paying Not to Go to the Gym." *American Economic Review*, Vol. 96, No. 3 (June), pages 694-719.

[30] Dranove, David, and Ginger Zhe Jin (2010). "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature*, Vol. 48, No. 4 (December), pages 935-963.

[31] Dranove, David and Andrew Sfekas (2008) "Start Spreading the News: A Structural Estimate of the Effects of New York Hospital Report Cards." *Journal of Health Economics*, Vol. 27, No. 5, pages 1201–07.

[32] Duflo, Esther, and Emmanuel Saez (2003). "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *Quarterly Journal of Economics*, Vol. 118, No. 3 (August), pages 815-842.

[33] Ebbinghaus, Hermann (1885). Memory: A Contribution to Experimental Psychology. New York: Columbia University Teachers College.

[34] Ferraro, Paul, and Michael Price (2011). "Using Non-Pecuniary Strategies to Influence Behavior: Evidence from a Large Scale Field Experiment." NBER Working Paper No. 17189 (July).

[35] Ferraro, Paul, Juan Jose Miranda, and Michael Price (2011). "The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment." *American Economic Review*, Vol. 101, No. 3 (May), pages 318-322.

[36] Figlio, David, and Maurice Lucas (2004) “What’s in a Grade? School Report Cards and the Housing Market.” *American Economic Review*, Vol. 94, No. 3, pages 591-604.

[37] Frey, Bruno, and Stephan Meier (2004). "Social Comparisons and Pro-Social Behavior: Testing ‘Conditional Cooperation’ in a Field Experiment." *American Economic Review*, Vol. 94, No. 5 (December), pages 1717-1722.

[38] Friedrich, Katherine, Maggie Eldridge, Dan York, Patti Witte, and Marty Kushler (2009). “Saving Energy Cost-Effectively: A National Review of the Cost of Energy Saved through Utility-Sector Energy Efficiency Programs.” ACEEE Report No. U092 (September).

[39] Gabaix, Xavier (2012). "A Sparsity-Based Model of Bounded Rationality." Working Paper, NYU (September).

[40] Gabaix, Xavier, and David Laibson (2006). "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets.” *Quarterly Journal of Economics*, Vol. 121, No. 2, pages 505-540.

[41] Gerber, Alan, Donald Green, and Ron Shachar (2003). "Voting May Be Habit-Forming: Evidence from a Randomized Field Experiment." *American Journal of Political Science*, Vol. 47, No. 3 (July), pages 540-550.

[42] Gerber, Alan, and Todd Rogers (2009). “Descriptive Social Norms and Motivation to Vote: Everybody’s Voting and So Should You.” *Journal of Politics*, Vol. 71, pages 1-14.

[43] Gine, Xavier, Dean Karlan, and Jonathan Zinman (2010). "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." American Economic Journal: Applied Economics, Vol. 2, No. 4 (October), pages 213-235.

[44] Gneezy, Uri, and John List (2006). "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments." *Econometrica*, Vol. 74, No. 5 (September), pages 1365-1384.

[45] Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel (2011). "When and Why Incentives (Don’t) Work to Modify Behavior." *Journal of Economic Perspectives*, Vol. 25, No. 4 (Fall), pages 191-210.

[46] Greenstone, Michael, Paul Oyer, and Annette Vissing-Jorgensen (2006). “Mandated Disclosure, Stock Returns, and the 1964 Securities Acts Amendments.” *Quarterly Journal of Economics*, 121(2): 399–460.

[47] Haselhuhn, Michael, Devin Pope, Maurice Schweitzer, and Peter Fishman (2012). "The Impact of Personal Experience on Behavior: Evidence from Video-Rental Fines." *Management Science*, Vol. 58, No. 1 (January), pages 52-61.

[48] Hastings, Justine and Jeffrey M. Weinstein (2008) “Information, School Choice and Academic Achievement: Evidence from Two Experiments.” *Quarterly Journal of Economics*, Vol. 123, No. 4, pages 1373-1414.

[49] Jackson, Kirabo (2010). "A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program." *Journal of Human Resources*, Vol. 45, No. 3 (Summer), pages 591-639.

[50] Jin, Ginger Zhe, and Alan T. Sorensen (2006) “Information and Consumer Choice: The Value of Publicized Health Plan Ratings.” *Journal of Health Economics*, 25(2): 248–75.

[51] John, Leslie, George Loewenstein, Andrea Troxel, Laurie Norton, Jennifer Fassbender, and Kevin Volpp (2011). "Financial Incentives for Extended Weight Loss: A Randomized, Controlled Trial." *Journal of General Internal Medicine*, Vol. 26, No. 6 (June), pages 621-626.

[52] Joskow, Paul (2012). "Creating a Smarter U.S. Electricity Grid." *Journal of Economic Perspectives*, Vol. 26, No. 1 (Winter), pages 29-48.

[53] Joskow, Paul, and Catherine Wolfram (2012). "Dynamic Pricing of Electricity." American Economic Review, Vol. 102, No. 3 (May).

[54] Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman (2010). "Getting to the Top of Mind: How Reminders Increase Saving." Working Paper, Harvard University (October).

[55] KEMA (2012). "Puget Sound Energy's Home Energy Reports Program: Three Year Impact, Behavioral and Process Evaluation." Madison, Wisconsin: DNV KEMA Energy and Sustainability.

[56] Kling, Jeffrey, Sendhil Mullainathan, Eldar Shafir, Lee Vermeulen, and Marian Wrobel (2012). "Comparison Friction: Experimental Evidence from Medicare Drug Plans." *Quarterly Journal of Economics*, Vol. 127, No. 1 (February).

[57] Landry, Craig, Andreas Lange, John List, Michael Price, and Nicholas Rupp (2010). "Is a Donor in Hand Better than Two in the Bush? Evidence from a Natural Field Experiment." *American Economic Review*, Vol. 100 (June), pages 958-983.

[58] Levitt, Steve, John List, and Sally Sadoff (2010). "The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment." Working Paper, University of Chicago.

[59] Macharia, W., G. Leon, B. Rowe, B. Stephenson, and R. Haynes (1992). "An Overview of Interventions to Improve Compliance with Appointment Keeping for Medical Services." *Journal of the American Medical Association*, Vol. 267, No. 13 (April 1), pages 1813-1817.

[60] Mathios, Alan (2000). "The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market." *Journal of Law and Economics,* Vol. 43, No. 2 (October), pages 651-678.

[61] Mullainathan, Sendhil (2002). "A Memory-Based Model of Bounded Rationality." *Quarterly Journal of Economics*, Vol. 117, No. 3 (August), pages 735-774.

[62] Nolan, Jessica, Wesley Schultz, Robert Cialdini, Noah Goldstein, and Vladas Griskevicius (2008). "Normative Influence is Underdetected." *Personality and Social Psychology Bulletin*, Vol. 34, pages 913-923.

[63] Opinion Dynamics (2012). "Massachusetts Three Year Cross-Cutting Behavioral Program Evaluation Integrated Report." Waltham, MA: Opinion Dynamics Corporation.

[64] Oullette, Judith, and Wendy Wood (1998). "Habit and Intention in Everyday Life: The Multiple Processes by Which Past Behavior Predicts Future Behavior." *Psychological Bulletin*, Vol. 124, No. 1, pages 54-74.

[65] Pope, Devin (2009). "Reacting to Rankings: Evidence from "America's Best Hospitals." *Journal of Health Economics*, Vol. 28, No. 6, pages 1154-1165.

[66] Reis, Ricardo (2006a). "Inattentive Consumers." *Journal of Monetary Economics*, Vol. 53, No. 8, pages 1761-1800.

[67] Reiss, Peter, and Matthew White (2008). "What Changes Energy Consumption? Prices and Public Pressure." *RAND Journal of Economics*, Vol. 39, No. 3 (Autumn), pages 636-663.

[68] Rubin, David, and Amy Wenzel (1996). "One Hundred Years of Forgetting: A Quantitative Description of Retention." *Psychological Review*, Vol. 103, No. 4 (October), pages 734-760.

[69] Sallee, James (2011). "Rational Inattention and Energy Efficiency." Working Paper, University of Chicago (June).

[70] Scanlon, Dennis, Michael Chernew, Catherine McLaughlin, and Gary Solon (2002). "The Impact of Health Plan Report Cards on Managed Care Enrollment." *Journal of Health Economics*, Vol. 21, No. 1, pages 19–41.

[71] Schneider, Walter, and Richard Shiffrin (1977). "Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention." *Psychological Review*, Vol. 84, No. 1, pages 1-66.

[72] Schultz, Wesley, Jessica Nolan, Robert Cialdini, Noah Goldstein, and Vladas Griskevicius (2007). "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science*, Vol. 18, pages 429-434.

[73] Sexton, Steven (2011). "Automatic Bill Payment, Price Salience, and Consumption: Evidence from Residential Electricity Consumption. Working Paper, UC Berkeley (November).

[74] Shiffrin, Richard, and Walter Schneider (1977). "Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending and a General Theory." *Psychological Review*, Vol. 84, No. 1, pages 127-190.

[75] Sumi, David, and Brian Coates (1989). "Persistence of Energy Savings in Seattle City Light's Residential Weatherization Program." *Energy Program Evaluation: Conservation and Resource Management: Proceedings of the August 23-25, 1989 Conference*, pages 311-316. Chicago: Energy Program Evaluation Conference.

[76] U.S. Energy Information Administration (2005). "Table US 14: Average Consumption by Energy End Uses, 2005." Available from: http://www.eia.gov/consumption/residential/data/2005/c&e/summary/pdf/tableus14.pdf

[77] U.S. Energy Information Administration (2011). "Table 5A. Residential Average Monthly Bill by Census Division, and State." Available from http://www.eia.gov/electricity/sales_revenue_price/html/table5_a.html.

[78] van Dulmen, Sandra, Emmy Sluijs, Liset van Dijk, Denise de Ridder, Rob Heerdink, and Jozien Bensing (2007). "Patient Adherence to Medical Treatment: A Review of Reviews." *BMC Health Services Research*, Vol. 7, No. 55,

[79] Violette, Daniel, Provencher, Bill, and Mary Klos (2009). "Impact Evaluation of Positive Energy SMUD Pilot Study." Boulder, CO: Summit Blue Consulting.

[80] Volpp, Kevin, Andrea Troxel, Mark Pauly, Henry Glick, Andrea Puig, David Asch, Robert Galvin, Jingsan Zhu, Fei Wan, Jill deGuzman, Elizabeth Corbett, Janet Weiner, and Janet Audrain-McGovern (2009). "A Randomized, Controlled Trial of Financial Incentives for Smoking Cessation." *New England Journal of Medicine*, Vol. 360 (February 12), pages 699-709.

Tables

Table 1: Descriptive Statistics

Location	
Region	West
Average January Heating Degrees	25.0
Average July Cooling Degrees	2.2
Narrative	
Baseline period begins	January 1, 2007
First reports generated	October 8, 2008
Last report generated for dropped group	September 10, 2010
End of sample	August 31, 2012
Number of Households	
Treatment: Continued	23,399
Treatment: Dropped	11,543
Control	43,945
Total	78,887
Number of Observations	4,988,798
Baseline Usage (kWh/day)	
Mean	30.3
Standard deviation	13.50
Treatment - Control	0.045
(Standard error)	(0.097)
Dropped - Continued	0.062
(Standard error)	(0.154)
Inactive Households	
Share of Households	0.20
Treatment - Control	0.00071
(Standard error)	(0.0029)
Dropped - Continued	-0.00605
(Standard error)	(0.0045)
Opting Out of Treatment	
Share of treatment group that opts out	0.018
Share of continued group that opts out after October 2010	0.0055

Table 2: Descriptive Statistics for Short-Run Analysis

Narrative			
Wave	1	1	2
Start Date	October 8, 2008	October 8, 2008	February, 2011
Frequency	Monthly	Quarterly	Bimonthly
Number of Households			
Treatment	24,851	9,923	21,970
Control	33,003	10,995	21,891
Total	57,854	20,918	43,861
Number of Observations	102,285,975	36,716,378	61,081,187

Table 3: Immediate Effects at Arrival Window

Table 3a: First Four Reports

	Monthly	Monthly DD's	Quarterly	Quarterly DD's	Bimonthly	Bimonthly DD's
	(1)	(2)	(3)	(4)	(5)	(6)
TS^a	0.248	0.247	0.151	0.146	0.027	0.106
	(0.028)***	(0.028)***	(0.031)***	(0.03)***	(0.052)	(0.034)***
TS^0	-.079	-.058	-.079	-.095	-.049	-.047
	(0.024)***	(0.023)**	(0.029)***	(0.028)***	(0.033)	(0.029)
TS^1	-.192	-.220	-.218	-.248	-.162	-.188
	(0.03)***	(0.028)***	(0.041)***	(0.034)***	(0.038)***	(0.035)***
T	-.556	-.542	-.393	-.469	-.362	-.472
	(0.065)***	(0.119)***	(0.067)***	(0.091)***	(0.059)***	(0.098)***
$T{\cdot}1(CDD > 0)$				-.00007		-.045
				(0.039)		(0.052)
$T{\cdot}CDD$				0.019		0.008
				(0.01)*		(0.014)
$T{\cdot}1(0 < HDD \leq 5)$				0.042		0.101
				(0.04)		(0.038)***
$T{\cdot}1(5 < HDD \leq 35)$		0.274		0.036		0.355
		(0.166)*		(0.075)		(0.083)***
$T{\cdot}HDD \cdot 1(5 < HDD \leq 35)$		-.013		0.002		-.017
		(0.004)***		(0.006)		(0.009)*
$T{\cdot}1(HDD > 35)$				0.314		-.140
				(0.21)		(0.322)
Obs.	8515691	8515691	1.93e+07	1.93e+07	9610563	9610563

Table 3b: After First Four Reports

	Monthly	Monthly DD's	Quarterly	Quarterly DD's	Bimonthly	Bimonthly DD's
	(1)	(2)	(3)	(4)	(5)	(6)
TS^a	0.102	0.095	0.057	0.042	0.02	0.039
	(0.014)***	(0.013)***	(0.023)**	(0.021)**	(0.068)	(0.068)
TS^0	-.031	-.034	-.017	-.014	-.070	-.162
	(0.008)***	(0.007)***	(0.02)	(0.02)	(0.047)	(0.05)***
TS^1	-.049	-.051	-.058	-.042	-.263	-.277
	(0.01)***	(0.009)***	(0.023)**	(0.024)*	(0.06)***	(0.061)***
T	-.815	-.778	-.627	-.498	-.755	2.115
	(0.059)***	(0.062)***	(0.092)***	(0.117)***	(0.096)***	(0.315)***
$T{\cdot}1(CDD > 0)$		-.011		-.041		-9.438
		(0.022)		(0.05)		(1.109)***
$T{\cdot}CDD$		0.004		0.006		0.998
		(0.007)		(0.012)		(0.298)***
$T{\cdot}1(0 < HDD \leq 5)$		0.042		-.017		-2.273
		(0.024)*		(0.055)		(0.278)***
$T{\cdot}1(5 < HDD \leq 35)$		0.108		0.066		-2.247
		(0.046)**		(0.087)		(0.312)***
$T{\cdot}HDD \cdot 1(5 < HDD \leq 35)$		-.009		-.012		-.031
		(0.004)**		(0.006)**		(0.009)***
$T{\cdot}1(HDD > 35)$		-.121		-.082		-3.313
		(0.128)		(0.188)		(0.405)***
Obs.	7.00e+07	7.00e+07	4.37e+07	4.37e+07	9352415	9352415

Notes: Tables 3a and 3b present the estimates of Equation (3) for the first four reports and all remaining reports, respectively. In each pair of estimates, the left column does not control for degree days, while the right column does. The outcome variable is electricity use, in kilowatt-hours per day. Standard errors are robust, clustered by household. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table 4: Decays Between Reports

Table 4a: First Four Reports

	Monthly	Monthly DD's	Quarterly	Quarterly DD's	Bimonthly	Bimonthly DD's
	(1)	(2)	(3)	(4)	(5)	(6)
TS^w	-.110	-.203	-.061	-.033	-.055	-.084
	(0.034)***	(0.031)***	(0.036)*	(0.035)	(0.047)	(0.046)*
dTS^w	1.168	4.221	0.774	0.738	1.610	1.445
	(1.265)	(1.310)***	(0.196)***	(0.191)***	(0.423)***	(0.408)***
T	-.416	-.459	-.399	-.498	-.403	-.469
	(0.064)***	(0.119)***	(0.072)***	(0.093)***	(0.066)***	(0.103)***
T$\cdot 1(CDD > 0)$				-.004		-.030
				(0.039)		(0.051)
T$\cdot CDD$				0.018		0.008
				(0.011)*		(0.014)
T$\cdot 1(0 < HDD \leq 5)$				0.036		0.105
				(0.04)		(0.039)***
T$\cdot 1(5 < HDD \leq 35)$		0.382		0.03		0.338
		(0.168)**		(0.074)		(0.085)***
T$\cdot HDD \cdot 1(5 < HDD \leq 35)$		-.015		0.003		-.015
		(0.004)***		(0.006)		(0.009)*
T$\cdot 1(HDD > 35)$				0.283		-.084
				(0.21)		(0.324)
Obs.	8515691	8515691	1.93e+07	1.93e+07	9610563	9610563

Table 4b: After First Four Reports

	Monthly	Monthly DD's	Quarterly	Quarterly DD's	Bimonthly	Bimonthly DD's
	(1)	(2)	(3)	(4)	(5)	(6)
TS^w	0.024	0.019	0.039	0.053	-.485	-.340
	(0.012)**	(0.011)*	(0.036)	(0.036)	(0.105)***	(0.079)***
dTS^w	0.511	0.503	0.112	0.114	0.249	0.434
	(0.33)	(0.321)	(0.149)	(0.15)	(0.542)	(0.534)
T	-.783	-.748	-.661	-.541	-.563	2.115
	(0.056)***	(0.061)***	(0.09)***	(0.118)***	(0.092)***	(0.315)***
T$\cdot 1(CDD > 0)$		-.015		-.042		-9.440
		(0.022)		(0.05)		(1.109)***
T$\cdot CDD$		0.005		0.005		0.997
		(0.007)		(0.012)		(0.298)***
T$\cdot 1(0 < HDD \leq 5)$		0.044		-.024		-2.272
		(0.024)*		(0.055)		(0.278)***
T$\cdot 1(5 < HDD \leq 35)$		0.11		0.067		-2.377
		(0.046)**		(0.088)		(0.31)***
T$\cdot HDD \cdot 1(5 < HDD \leq 35)$		-.009		-.013		-.019
		(0.004)**		(0.006)**		(0.008)**
T$\cdot 1(HDD > 35)$		-.108		-.096		-3.009
		(0.127)		(0.187)		(0.391)***
Obs.	7.00e+07	7.00e+07	4.37e+07	4.37e+07	9352415	9352415

Notes: Tables 4a and 4b present the estimates of Equation (4) for the first four reports and all remaining reports, respectively. In each pair of estimates, the left column does not control for degree days, while the right column does. The outcome variable is electricity use, in kilowatt-hours per day. Standard errors are robust, clustered by household. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table 5: Household-Specific Repeated Effects

	Monthly	Quarterly	Bimonthly	Combined	Controls	Exclude Outliers
	(1)	(2)	(3)	(4)	(5)	(6)
$T\Delta Y_{h-1}$	0.022	0.033	0.011	0.02	0.025	0.023
	(0.006)***	(0.014)**	(0.015)	(0.006)***	(0.007)***	(0.005)***
ΔY_{h-1} (Monthly)	-.106			-.105	-.107	-.105
	(0.004)***			(0.004)***	(0.005)***	(0.004)***
ΔY_{h-1} (Quarterly)		-.045		-.041	-.042	-.041
		(0.009)***		(0.008)***	(0.008)***	(0.007)***
ΔY_{h-1} (Bimonthly)			0.052	0.047	0.045	0.038
			(0.01)***	(0.008)***	(0.008)***	(0.006)***
T (Monthly)	-.112			-.115		-.090
	(0.043)***			(0.043)***		(0.042)**
T (Quarterly)		-.127		-.132		-.118
		(0.064)**		(0.064)**		(0.064)*
T (Bimonthly)			-.053	-.039		-.040
			(0.043)	(0.039)		(0.036)
Obs.	178959	91277	115631	385867	385867	385472

Notes: This table presents the estimates of Equation (5). The outcome variable is change in electricity use after vs. before report arrival, in kilowatt-hours per day. Standard errors are robust, clustered by household. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table 6: Long-Run Effects

	Levels	Changes	Weather	Trends	Controls	Balanced
	(1)	(2)	(3)	(4)	(5)	(6)
TP^0	-.025	-.025	-.025	-.025	-.025	-.023
	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.033)
TP^1	-.452	-.452	-.452	-.452	-.452	-.466
	(0.043)***	(0.043)***	(0.043)***	(0.043)***	(0.043)***	(0.044)***
TP^2	-.660			-.660	-.617	-.657
	(0.051)***			(0.051)***	(0.068)***	(0.071)***
$\mathrm{T}\cdot(P^2 + P^3)$		-.660	-.595			
		(0.051)***	(0.068)***			
EP3	-.842	-.181	-.247	-.842	-.842	-.837
	(0.068)***	(0.053)***	(0.071)***	(0.068)***	(0.068)***	(0.07)***
DP3	-.612	0.049	0.053	-.489	-.456	-.448
	(0.087)***	(0.076)	(0.075)	(0.101)***	(0.102)***	(0.102)***
DrP^3				0.131	0.117	0.127
				(0.058)**	(0.056)**	(0.051)**
$\mathrm{HDD}\cdot(TP^2 + DP^3)$			-.004		-.003	-.002
			(0.004)		(0.004)	(0.004)
$\mathrm{CDD}\cdot(TP^2 + DP^3)$			-.012		-.008	0.005
			(0.025)		(0.025)	(0.027)
Obs.	4042155	4042155	4042155	4042155	4042155	3526102

Notes: This table presents the estimates of Equation (7). The outcome variable is monthly average electricity use, in kilowatt-hours per day. Standard errors are robust, clustered by household. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table 7: Program Participation

	Savings (kWh/day)	Number Installed	Continued - Control (kWh/day)	Treatment - Control (kWh/day)	First Year: Continued-Control kWh/day
	(1)	(2)	(3)	(4)	(5)
Clothes Washer	0.35	1606	0.00035	0.00024	0.00024
			(0.0004)	(0.00035)	(0.00023)
Compact Fluorescent Lightbulbs	2.23	260	0.00224 **	0.00126	0.00051
			(0.00113)	(0.00096)	(0.00038)
Refrigerator Decommissioning	1.32	250	0.0004	0.00038	-0.00013
			(0.00058)	(0.00051)	(0.00022)
Showerhead	0.22	187	0.00008	0.00007	0.00003
			(0.00009)	(0.00007)	(0.00003)
Freezer Decommissioning	1.52	102	0.00037	0.00057	0.00018
			(0.00044)	(0.00039)	(0.00017)
Heat Pump	1.61	54	-0.00019	-0.00014	-0.00003
			(0.00037)	(0.00036)	(0.00016)
Water Heater	8.20	28	-0.00099	-0.00091	-0.00099
			(0.00116)	(0.00105)	(0.00065)
New Refrigerator	1.80	7	-0.0001	-0.00013	-0.00004
			(0.00013)	(0.00011)	(0.00004)
Windows	12.2	5	0.00089	0.00053	0.00007
			(0.00082)	(0.00056)	(0.00007)
Conversion to Gas Heat	28.1	1	-0.00064	-0.00064	-0.00009
			(0.00064)	(0.00064)	(0.00009)
All		2500	0.00241	0.00123	-0.0002
			(0.00219)	(0.00191)	(0.00094)

Notes: This table presents data on participation in the utility's energy conservation programs for calendar year 2011. Standard errors are robust. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table 8: Self-Reported Actions

	All Sites			This Site		
	Mean	T-C	T-C$\mid$X	Mean	T-C	T-C$\mid$X
"In the past twelve months, have you..."						
Taken any steps to reduce energy use?	0.77	0.010	-0.001	0.81	-0.055	-0.035
		(0.012)	(0.015)		(0.030)*	(0.035)
Repeated Actions	0.62	0.005	0.011	0.59	0.004	0.011
		(0.008)	(0.010)		(0.027)	(0.030)
Adjusted your thermostat settings?	0.63	0.012	0.007			
		(0.015)	(0.019)			
Unplugged devices and chargers?	0.65	-0.020	-0.013	0.65	-0.020	-0.013
		(0.039)	(0.044)		(0.039)	(0.044)
Switched off power strips or appliances when unused?	0.59	0.002	0.011	0.51	0.013	0.022
		(0.014)	(0.018)		(0.041)	(0.047)
Turned off lights when unused?	0.96	0.005	0.006			
		(0.009)	(0.010)			
Hung laundry to dry?	0.42	0.010	0.001			
		(0.024)	(0.027)			
Used energy saving or sleep features on your computer?	0.56	0.008	0.021			
		(0.021)	(0.029)			
Turned off computer at night?	0.65	-0.034	-0.030	0.60	0.018	0.025
		(0.023)	(0.026)		(0.040)	(0.046)
Used fans to keep cool?	0.80	0.072	0.086			
		(0.034)**	(0.039)**			
Physical Capital Changes	0.55	-0.002	0.002	0.54	-0.003	0.014
		(0.008)	(0.010)		(0.017)	(0.019)
Replaced incandescent light bulbs with LEDs?	0.70	0.013	0.016	0.70	0.013	0.016
		(0.038)	(0.043)		(0.038)	(0.043)
Purchased Energy Star appliances?	0.74	0.002	0.019	0.77	0.012	0.063
		(0.016)	(0.022)		(0.035)	(0.039)
Disposed of a second refrigerator or freezer?	0.26	-0.001	0.030	0.16	0.013	0.030
		(0.015)	(0.019)		(0.029)	(0.033)
Installed light timers or sensors?	0.30	-0.018	-0.014	0.30	-0.018	-0.014
		(0.038)	(0.043)		(0.038)	(0.043)
Replaced incandescent light bulbs with CFLs?	0.81	0.000	-0.017			
		(0.013)	(0.017)			
Added insulation or replaced windows?	0.54	-0.039	-0.055			
		(0.024)	(0.029)*			
Had a home energy audit?	0.19	0.057	0.058			
		(0.022)***	(0.026)**			
Installed a programmable thermostat?	0.79	-0.033	-0.025	0.79	-0.033	-0.025
		(0.032)	(0.037)		(0.032)	(0.037)
Intermittent Actions	0.62	0.006	0.007	0.56	-0.005	0.000
		(0.012)	(0.017)		(0.031)	(0.034)
Tuned up your AC system?	0.63	-0.016	-0.014	0.61	-0.032	-0.050
		(0.018)	(0.024)		(0.040)	(0.044)
Used a programmable thermostat?	0.59	0.009	0.076			
		(0.028)	(0.047)			
Added weather-stripping or caulking around windows?	0.60	0.008	-0.009	0.51	0.022	0.050
		(0.018)	(0.025)		(0.041)	(0.047)
Cleaned or replaced heating or AC system air filters?	0.70	0.017	0.010			
		(0.038)	(0.044)			
Participated in any utility energy effiency programs?	0.19	0.018	0.010	0.61	0.007	0.028
		(0.010)*	(0.013)		(0.040)	(0.046)
N	5856			800		

Notes: This table presents survey data on self-reported energy conservation actions. The three columns at left present aggregated results across six sites, while the three on the right present results from the utility we study in the rest of the paper. Standard errors are robust. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table 9: In-Sample Cost Effectiveness

Scenario	1	2	3	4
Discontinue Reports?	No	Yes	Yes	Yes
Assumed Persistence	-	Observed	Zero	Full
Electricity Savings and Costs				
Total electricity savings during treatment (kWh)	995	406	406	406
(Standard Error)	(53.5)	(24.3)	(24.3)	(24.3)
Total electricity savings after treatment	0	428.4	0	462
(Standard Error)	(0)	(60.9)	(0)	(35.7)
Total savings (kWh)	995	834	406	868
	(53.5)	(65.6)	(24.3)	(43.2)
Total cost ($)	32.1	18.0	18.0	18.0
Cost Effectiveness				
Cost Effectiveness (cents/kWh)	3.23	2.15	4.42	2.07
(Standard Error)	(0.17)	(0.17)	(0.26)	(0.1)
Retail Electricity Cost Savings				
Per household electricity savings ($)	100	83	41	87
(Standard Error)	(5.4)	(6.6)	(2.4)	(4.3)
Total population electricity savings ($millions)	7.85	6.58	3.20	6.85
(Standard Error)	(0.43)	(0.52)	(0.19)	(0.34)

Notes: See text for details.

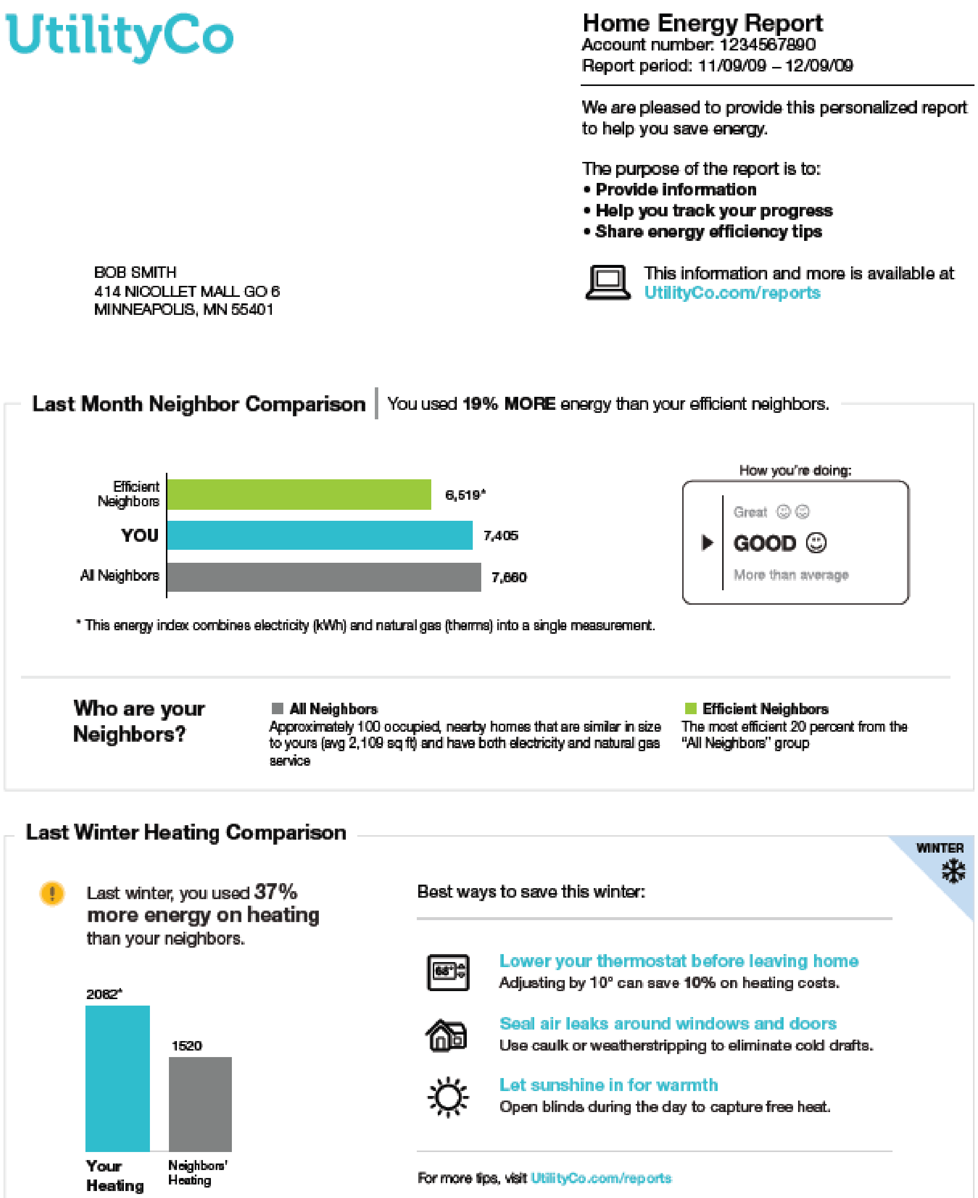
Table 10: Cost Effectiveness and Program Design

Design	1	2	3	4
Length of hypothetical program	One report	One year	Two years	Four years
Savings and Costs				
Savings during treatment (kWh PDV)	0	167	398	948
|ATE| during treatment (kWh/day)	0.30	0.45	0.66	0.84
Decay rate (kWh/day per year)	0.75	0.12	0.12	0.12
Years from end of treatment to zero effect	0.40	3.8	5.5	7.0
Savings after treatment (kWh PDV)	22	284	559	813
Total savings (kWh PDV)	22	450	958	1,760
Total cost ($ PDV)	1.0	9.1	17.7	30.4
Cost Effectiveness (cents/kWh)	4.61	2.01	1.85	1.72
Incremental Effects				
Durability Effect (kWh PDV)		194	507	803
Persistence Effect (kWh PDV)		234	0	0
Incremental savings (kWh PDV)	22	429	507	803
Incremental cost ($ PDV)	1.0	8.1	8.6	12.7
Incremental cost-effectiveness (cents/kWh)	4.61	1.88	1.70	1.58

Notes: See text for details.

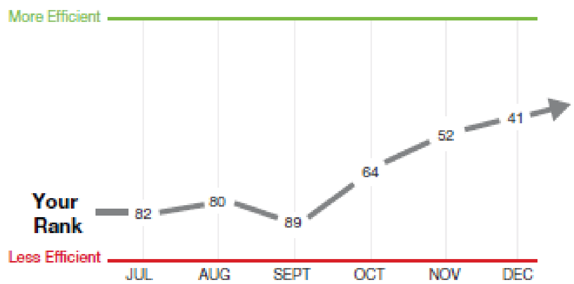
Figures

Figure 1: Home Energy Report, Front and Back



Neighbor Efficiency Rank

Your energy efficiency rank out of 100 neighbors:



Your rank dates correspond to your billing periods.
Your neighbors are nearby, occupied, similar-sized homes.

Your Rank Last Month

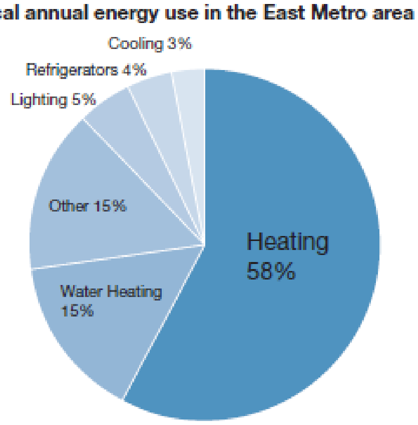
#41 out of 100 neighbors
#1 is the most efficient

★ Good work, your rank is improving!
Find more tips and ways to save:
www.UtilityCo.com/reports

Understanding Energy Use

Heating is the largest use of energy for a typical household in the East Metro area, accounting for more than 50% of total energy use. To maximize your savings, focus on the biggest users first.

Typical annual energy use in the East Metro area



Other appliances and electronics include dishwashers, washing machines, dryers, computers, TVs and entertainment systems.

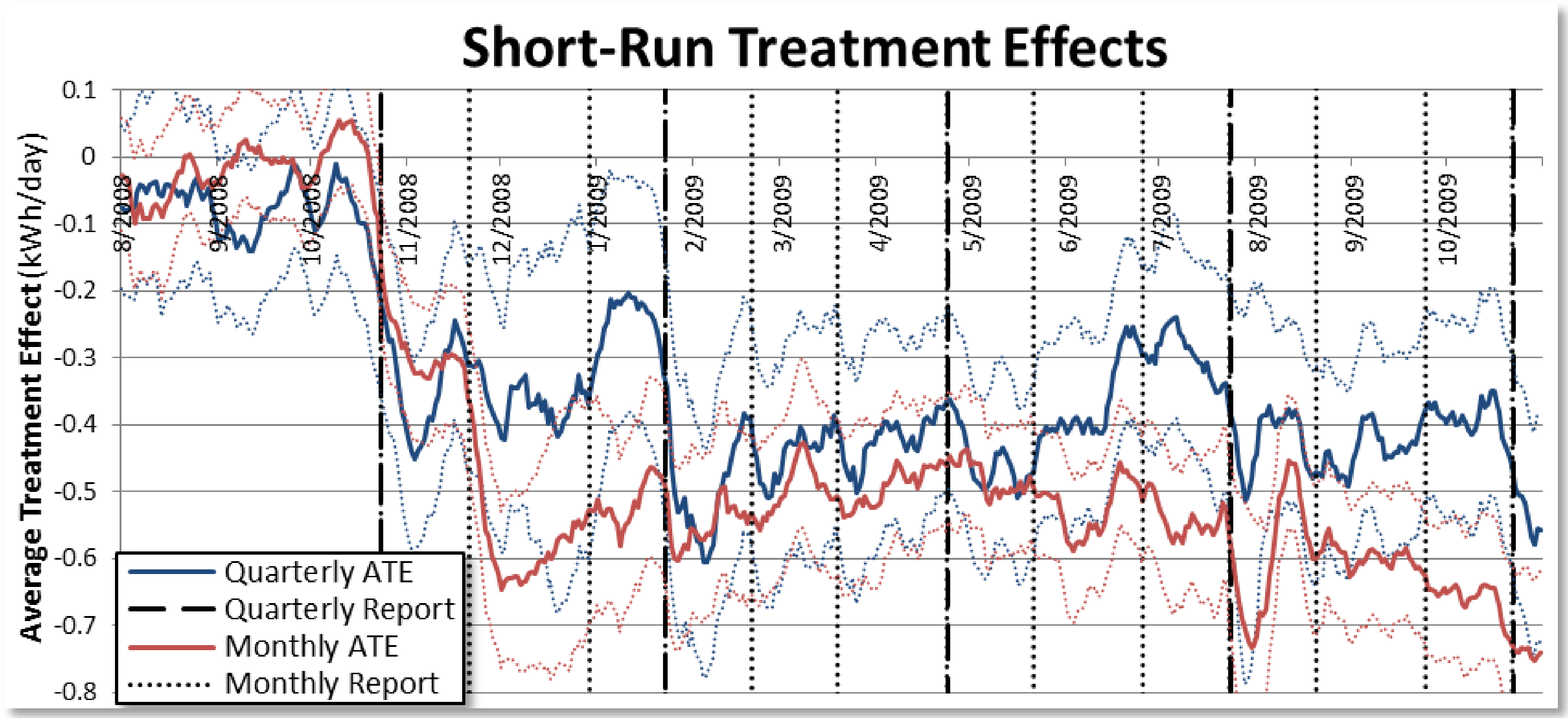
Based on a typical household with gas heating & water heating.

Top Tips For Saving	Save up to
☐ **Look for the ENERGY STAR® label** Next Steps: Look for the ENERGY STAR label when shopping for appliances and electronics.	$600/yr
☐ **Improve insulation and seal air leaks** Next Steps: Start with the places easiest to access, such as an attic.	$305/yr
☐ **Seal leaky ducts** Next Steps: Use mastic (a special adhesive) or duct tape to seal all accessible duct joints.	$170/yr
☐ **Recycle your second refrigerator** Next Steps: Try rearranging your main fridge to fit everything from your second fridge.	$145/yr
☐ **Turn off computer at night** Next Steps: Program your computer to automatically turn off after periods of inactivity.	$75/yr
☐ **Set your thermostat wisely** Next Steps: Set your thermostat 10 degrees off from your preferred setting when you're away or sleeping.	$65/yr
☐ **Install efficient showerheads** Next Steps: Get a new efficient showerhead and bathroom faucet aerator for free! Visit xcelenergy.com/energyreport for details.	$45/yr

Already do these tips?
Find more ways to save online

For energy-saving tips visit
www.UtilityCo.com/reports

Figure 2: Short-Run Treatment Effects



Notes: This figure plots the smoothed ATEs for each day of the first year of treatment for the monthly and quarterly treatment groups, as estimated by Equation (1). The dotted lines reflect 90 percent confidence intervals, with robust standard errors clustered by household.

Figure 3: Short-Run Effects in Event Time

Figure 3a: First Four Reports

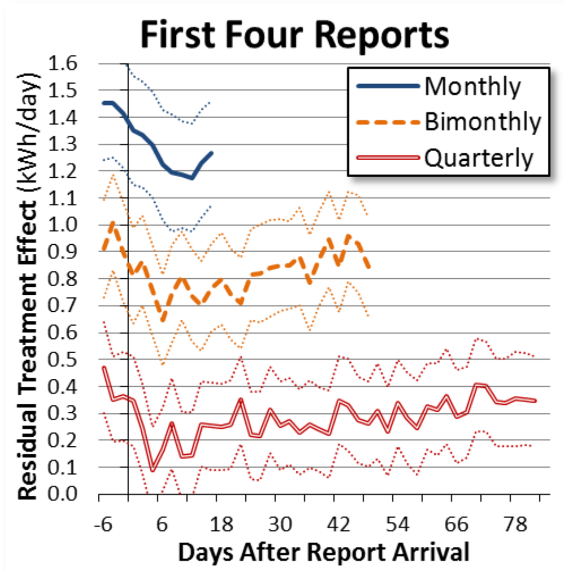
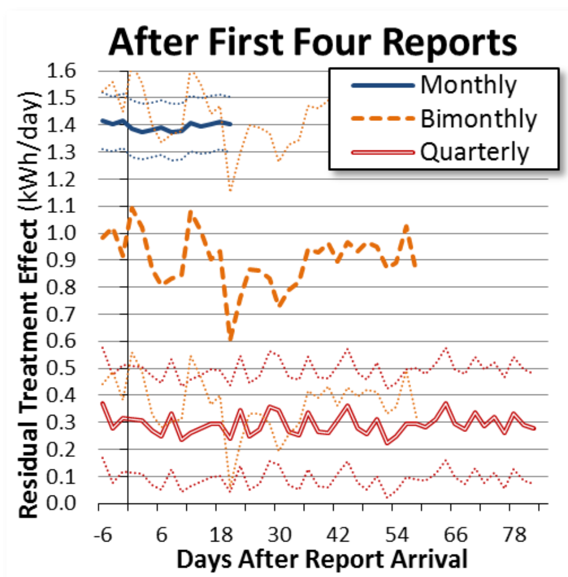
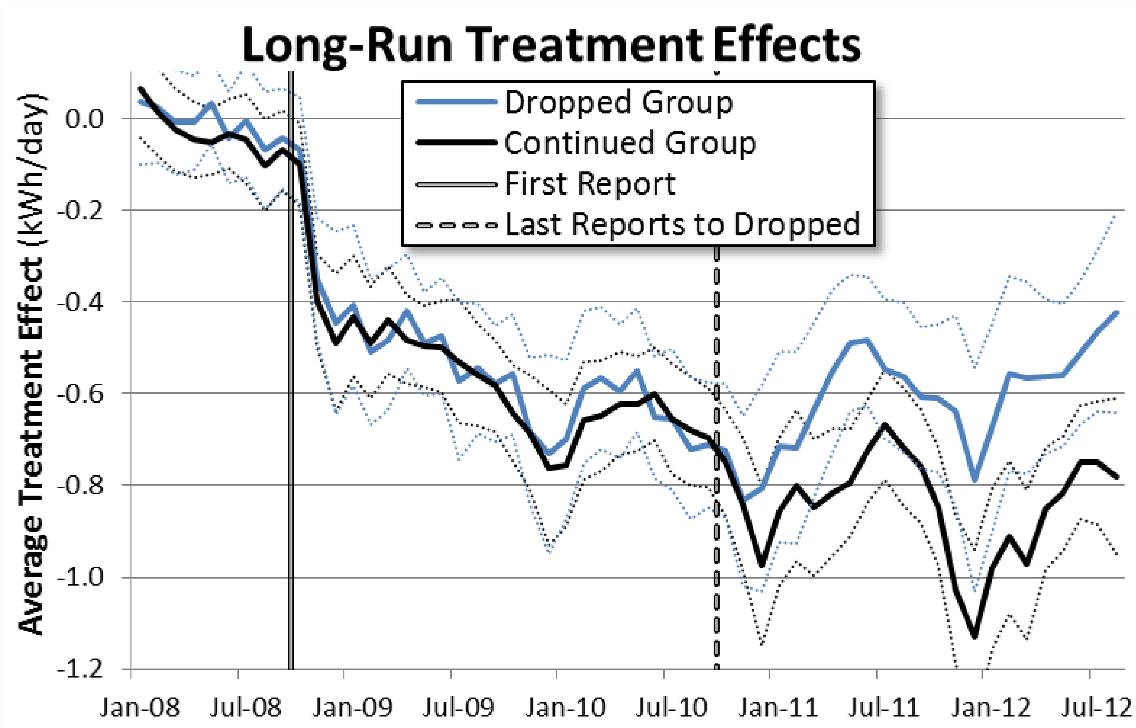


Figure 3b: After First Four Reports



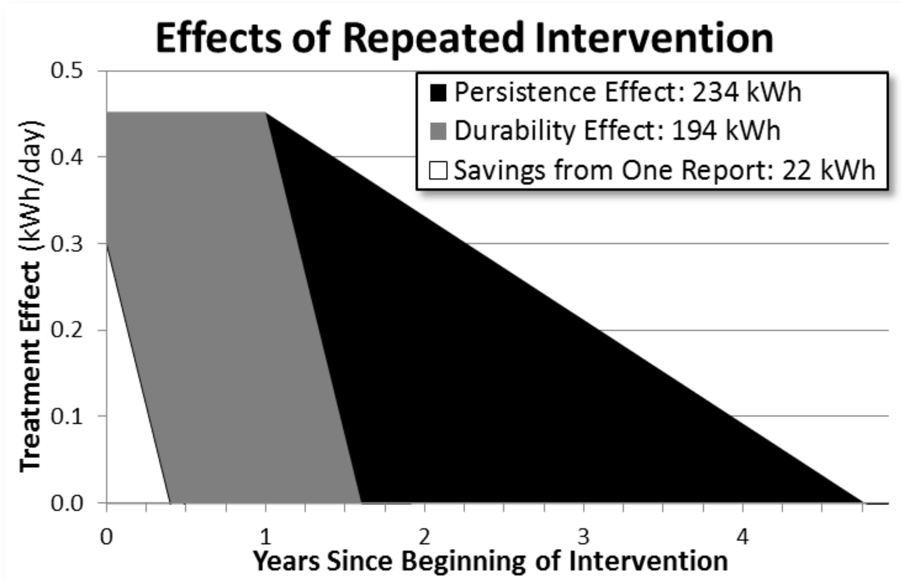
Notes: Figures 3a and 3b plot the ATEs in event time for the first four reports and all remaining reports, respectively, as estimated by Equation (2). The dotted lines reflect 90 percent confidence intervals, with robust standard errors clustered by household.

Figure 4: Long-Run Treatment Effects



Notes: This figure plots the ATEs for each month of the sample for the continued and dropped groups, estimated by Equation (6). The dotted lines reflect 90 percent confidence intervals, with robust standard errors clustered by household.

Figure 5: Effects of Repeated Intervention



Notes: This figure shows the channels of incremental electricity conservation when the intervention is lengthened from one report to one year. Electricity savings are kilowatt-hours (kWh) per household, discounted to the present at five percent. The assumptions used to calculate these effects are detailed in the text, and the numbers match those in Table 10.

Appendix: For Online Publication

The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation

Hunt Allcott and Todd Rogers

8 Appendix Tables

Table A1: Short-Run Effects at Arrival Window

Table A1a: First Four Reports

	M Outliers	M YControl	Q Outliers	Q YControl	B Outliers	B YControl
	(1)	(2)	(3)	(4)	(5)	(6)
TS^a	0.244	0.245	0.142	0.152	0.107	0.083
	(0.028)***	(0.028)***	(0.03)***	(0.03)***	(0.033)***	(0.034)**
TS^0	-.059	-.069	-.095	-.093	-.047	-.037
	(0.023)***	(0.023)***	(0.028)***	(0.028)***	(0.028)*	(0.029)
TS^1	-.217	-.207	-.241	-.257	-.189	-.161
	(0.028)***	(0.029)***	(0.034)***	(0.036)***	(0.034)***	(0.035)***
T	-.566	0.545	-.463	-.636	-.427	0.5
	(0.117)***	(0.331)*	(0.09)***	(0.294)**	(0.094)***	(0.326)
$T{\cdot}1(CDD > 0)$			-.0002	0.002	-.025	-.035
			(0.039)	(0.039)	(0.049)	(0.052)
$T{\cdot}CDD$			0.019	0.015	0.002	0.008
			(0.01)*	(0.011)	(0.013)	(0.014)
$T{\cdot}1(0 < HDD \leq 5)$			0.039	0.044	0.104	0.097
			(0.04)	(0.04)	(0.036)***	(0.038)**
$T{\cdot}1(5 < HDD \leq 35)$	0.308	-.191	0.035	0.056	0.329	0.149
	(0.163)*	(0.162)	(0.075)	(0.074)	(0.079)***	(0.066)**
$T{\cdot}HDD \cdot 1(5 < HDD \leq 35)$	-.014	-.003	0.002	-.0004	-.017	0.01
	(0.004)***	(0.004)	(0.006)	(0.006)	(0.009)*	(0.006)*
$T{\cdot}1(HDD > 35)$			0.264	0.204	-.103	0.792
			(0.202)	(0.188)	(0.309)	(0.222)***
$T{\cdot}\overline{Y}^{T==0}$		-.025		0.006		-.042
		(0.009)***		(0.01)		(0.012)***
Obs.	8514078	8515691	1.93e+07	1.93e+07	9590651	9610563

Table A1b: After First Four Reports

	M Outliers	M YControl	Q Outliers	Q YControl	B Outliers	B YControl
	(1)	(2)	(3)	(4)	(5)	(6)
TS^a	0.095	0.076	0.042	0.042	0.018	0.155
	(0.013)***	(0.012)***	(0.021)**	(0.021)**	(0.065)	(0.068)**
TS^0	-.032	-.034	-.012	-.014	-.161	-.001
	(0.007)***	(0.007)***	(0.02)	(0.02)	(0.049)***	(0.049)
TS^1	-.048	-.039	-.040	-.041	-.264	-.195
	(0.009)***	(0.009)***	(0.024)*	(0.024)*	(0.06)***	(0.06)***
T	-.771	-.029	-.500	-.524	2.166	4.458
	(0.062)***	(0.181)	(0.116)***	(0.304)*	(0.299)***	(0.443)***
$T{\cdot}1(CDD > 0)$	-.012	-.026	-.037	-.041	-9.597	-9.415
	(0.022)	(0.022)	(0.05)	(0.05)	(1.079)***	(1.109)***
$T{\cdot}CDD$	0.004	0.019	0.006	0.006	1.030	1.011
	(0.007)	(0.007)***	(0.012)	(0.012)	(0.296)***	(0.298)***
$T{\cdot}1(0 < HDD \leq 5)$	0.043	0.033	-.012	-.016	-2.321	-2.225
	(0.024)*	(0.024)	(0.055)	(0.055)	(0.26)***	(0.278)***
$T{\cdot}1(5 < HDD \leq 35)$	0.112	0.018	0.065	0.069	-2.291	-2.653
	(0.045)**	(0.045)	(0.087)	(0.091)	(0.295)***	(0.308)***
$T{\cdot}HDD \cdot 1(5 < HDD \leq 35)$	-.010	0.003	-.012	-.013	-.031	0.033
	(0.004)**	(0.004)	(0.006)**	(0.006)**	(0.008)***	(0.006)***
$T{\cdot}1(HDD > 35)$	-.138	0.241	-.069	-.095	-3.343	-1.139
	(0.125)	(0.119)**	(0.186)	(0.173)	(0.382)***	(0.371)***
$T{\cdot}\overline{Y}^{T==0}$		-.029		0.001		-.101
		(0.007)***		(0.01)		(0.012)***
Obs.	7.00e+07	7.00e+07	4.37e+07	4.37e+07	9332423	9352415

Notes: Tables A1a and A1b present the estimates of Equation (3) for the first four reports and all remaining reports, respectively. In each pair of estimates, the left column excludes outliers, while the right column controls for the interaction of the treatment effect with control group average usage. The outcome variable is electricity use, in kilowatt-hours per day. Standard errors are robust, clustered by household. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table A2: Short-Run Effects Between Reports

Table A2a: First Four Reports

	M Outliers	M YControl	Q Outliers	Q YControl	B Outliers	B YControl
	(1)	(2)	(3)	(4)	(5)	(6)
TS^w	-.200	-.174	-.027	-.034	-.081	-.057
	(0.031)***	(0.033)***	(0.035)	(0.035)	(0.043)*	(0.046)
dTS^w	4.118	2.699	0.705	0.737	1.217	1.094
	(1.297)***	(1.294)**	(0.188)***	(0.191)***	(0.375)***	(0.395)***
T	-.485	0.661	-.494	-.532	-.417	0.491
	(0.117)***	(0.327)**	(0.093)***	(0.286)*	(0.097)***	(0.325)
$T{\cdot}1(CDD > 0)$			-.004	-.003	-.014	-.022
			(0.039)	(0.039)	(0.048)	(0.051)
$T{\cdot}CDD$			0.019	0.017	0.003	0.009
			(0.011)*	(0.012)	(0.013)	(0.014)
$T{\cdot}1(0 < HDD \leq 5)$			0.033	0.036	0.107	0.099
			(0.04)	(0.04)	(0.037)***	(0.039)**
$T{\cdot}1(5 < HDD \leq 35)$	0.415	-.101	0.029	0.034	0.314	0.137
	(0.164)**	(0.166)	(0.075)	(0.073)	(0.082)***	(0.07)*
$T{\cdot}HDD \cdot 1(5 < HDD \leq 35)$	-.015	-.004	0.003	0.002	-.016	0.011
	(0.004)***	(0.005)	(0.006)	(0.005)	(0.009)*	(0.006)*
$T{\cdot}1(HDD > 35)$			0.234	0.26	-.053	0.837
			(0.203)	(0.186)	(0.309)	(0.228)***
$T{\cdot}\overline{Y}^{T==0}$		-.026		0.001		-.042
		(0.009)***		(0.01)		(0.012)***
Obs.	8514078	8515691	1.93e+07	1.93e+07	9590651	9610563

Table A2b: After First Four Reports

	M Outliers	M YControl	Q Outliers	Q YControl	B Outliers	B YControl
	(1)	(2)	(3)	(4)	(5)	(6)
TS^w	0.021	0.019	0.052	0.053	-.312	-.088
	(0.011)*	(0.011)*	(0.036)	(0.036)	(0.077)***	(0.074)
dTS^w	0.422	0.535	0.108	0.116	0.152	1.160
	(0.319)	(0.32)*	(0.149)	(0.148)	(0.524)	(0.533)**
T	-.741	0.027	-.541	-.530	2.165	4.417
	(0.061)***	(0.181)	(0.117)***	(0.304)*	(0.299)***	(0.436)***
$\text{T}\cdot 1(CDD > 0)$	-.016	-.030	-.037	-.042	-9.600	-9.414
	(0.022)	(0.022)	(0.05)	(0.051)	(1.079)***	(1.109)***
$\text{T}\cdot CDD$	0.005	0.021	0.006	0.006	1.029	1.011
	(0.007)	(0.007)***	(0.012)	(0.012)	(0.296)***	(0.298)***
$\text{T}\cdot 1(0 < HDD \leq 5)$	0.045	0.034	-.019	-.024	-2.320	-2.224
	(0.024)*	(0.024)	(0.055)	(0.056)	(0.26)***	(0.278)***
$\text{T}\cdot 1(5 < HDD \leq 35)$	0.114	0.014	0.066	0.065	-2.424	-2.620
	(0.045)**	(0.045)	(0.088)	(0.092)	(0.293)***	(0.307)***
$\text{T}\cdot HDD \cdot 1(5 < HDD \leq 35)$	-.010	0.004	-.012	-.013	-.018	0.031
	(0.004)**	(0.004)	(0.006)**	(0.006)**	(0.007)**	(0.007)***
$\text{T}\cdot 1(HDD > 35)$	-.124	0.269	-.083	-.090	-3.053	-1.153
	(0.125)	(0.118)**	(0.186)	(0.173)	(0.369)***	(0.371)***
$\text{T}\cdot\overline{Y}^{T==0}$		-.030		-.0004		-.099
		(0.007)***		(0.01)		(0.011)***
Obs.	7.00e+07	7.00e+07	4.37e+07	4.37e+07	9332423	9352415

Notes: Tables A2a and A2b present the estimates of Equation (4) for the first four reports and all remaining reports, respectively. In each pair of estimates, the left column excludes outliers, while the right column controls for the interaction of the treatment effect with control group average usage. The outcome variable is electricity use, in kilowatt-hours per day. Standard errors are robust, clustered by household. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table A3: Placebo Report Arrivals

	Unconditional	Weather
	(1)	(2)
TS^a	0.059	0.05
	(0.021)***	(0.021)**
TS^0	-.026	-.025
	(0.016)	(0.016)
TS^1	-.015	-.019
	(0.02)	(0.02)
T	-.638	-.508
	(0.093)***	(0.118)***
$T{\cdot}1(CDD > 0)$		-.037
		(0.05)
$T{\cdot}CDD$		0.005
		(0.012)
$T{\cdot}1(0 < HDD \leq 5)$		-.022
		(0.055)
$T{\cdot}1(5 < HDD \leq 35)$		0.07
		(0.087)
$T{\cdot}HDD \cdot 1(5 < HDD \leq 35)$		-.013
		(0.006)**
$T{\cdot}1(HDD > 35)$		-.099
		(0.189)
Obs.	4.37e+07	4.37e+07

Notes: This table presents the estimates of Equation (3) for the quarterly group, for reports that the monthly group received but the quarterly group did not. The sample includes the period after the quarterly group's first four reports. The left column does not control for degree days, while the right column does. The outcome variable is electricity use, in kilowatt-hours per day. Standard errors are robust, clustered by household. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table A4: Persistence by Subgroup

	Heterogeneous Effects	Trends	Weather
	(1)	(2)	(3)
D	-.683	-.797	-.701
	(0.099)***	(0.116)***	(0.124)***
Dr^2		0.127	0.094
		(0.068)*	(0.066)
D·(Quarterly Frequency)	0.254	0.26	0.26
	(0.173)	(0.208)	(0.208)
Dr·(Quarterly Frequency)		-.007	-.007
		(0.121)	(0.121)
$D \cdot \widetilde{Y}^b$	-.596	-.560	-.392
	(0.138)***	(0.173)***	(0.191)**
$Dr^2 \cdot \widetilde{Y}^b$		-.040	-.028
		(0.11)	(0.095)
$D \cdot HDD \cdot \widetilde{Y}^b$			-.008
			(0.01)
$D \cdot CDD \cdot \widetilde{Y}^b$			-.246
			(0.168)
$D \cdot HDD$			-.005
			(0.005)
$D \cdot CDD$			0.036
			(0.077)
$HDD \cdot \widetilde{Y}^b$			0.021
			(0.03)
$CDD \cdot \widetilde{Y}^b$			0.209
			(0.205)
HDD			-.005
			(0.017)
CDD			-.113
			(0.067)*
Obs.	1084738	1084738	1084738

Notes: This table presents the estimates of Equation (7), allowing α and δ^{LR} to differ for monthly vs. quarterly groups and as a function of $\widetilde{Y}^b$, which is baseline usage normalized to mean 0, standard deviation 1. The variable r^2 is analogous to r, but it is defined as the time in years since reports were discontinued at the end of the intervention's second year. The sample is limited to the third and fourth years after the intervention begins, including only the control group and the dropped group. The outcome variable is monthly average electricity use, in kilowatt-hours per day. Standard errors are robust, clustered by household. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.