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, Paul Cheshire, Lucas Davis, Stefano DellaVigna, Xavier Gabaix, Francesca Gino, Uri Gneezy, Michael Greenstone, Judd Kessler, David Laibson, Katy Milkman, Sendhil Mullainathan, Karen Palmer, Charlie Sprenger, staff at the utilities we study, and a number of seminar participants for feedback and helpful conversations. Thanks also to Tyler Curtis, Lisa Danz, Rachel Gold, Arkadi Gerney, Marc Laitin, Laura Lewellyn, Elena Washington, 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. Opower provided the data analyzed in this paper to the authors under a nondisclosure agreement. The authors and Opower structured the agreement in a way that maintains the authors' independence. In particular, the agreement stipulates that Opower has the right to review the publication prior to public release solely for factual accuracy. 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, Revised January 2014
JEL No. D03,D11,L97,Q41

ABSTRACT

We document three remarkable features of the Opower program, in which social comparison- based home energy reports are repeatedly mailed to more than six million households nationwide. First, initial reports cause high-frequency "action and backsliding," but these cycles attenuate over time. Second, if reports are discontinued after two years, effects are relatively persistent, decaying at 10-20 percent per year. Third, consumers are slow to habituate: they continue to respond to repeated treatment even after two years. We show that the previous conservative assumptions about post-intervention persistence had dramatically understated cost effectiveness and illustrate how empirical estimates can optimize program design.

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

Across domains such as smoking, exercise, school performance, and many others, there is increasing interest in behavioral interventions that affect our choices in ways that might increase welfare. In some contexts, interventions such as simplified information provision, commitment contracts, appeals to the public good, and social comparisons have had at least short-term effects. Evidence on long-term effects, however, is much more limited, and some of the evidence that is available suggests that it can be difficult to achieve lasting changes in outcomes.[1]

We study a widely-implemented and highly-publicized behavioral intervention, the "home energy report" produced by a company called Opower. The Opower reports feature personalized energy use feedback, social comparisons, and energy conservation information, and they are mailed to households every month or every few months for an indefinite period. Utilities hire Opower to send the reports primarily because the resulting energy savings help to comply with state energy conservation requirements. There are now 6.2 million households receiving home energy reports at 85 utilities across the United States. It is already well-documented that social comparisons can cause consumers to reduce energy use (Nolan *et al.* 2008 Schultz *et al.* 2007, Allcott 2011, Ayres, Raseman, and Shih 2012) and can affect a variety of other outcomes.[2] In this paper, we ask two further questions that, as we shall see, provide deeper insight into human behavior and have important policy implications.

First, how persistent are effects after the intervention ends? One potential model is that the treatment acts by providing information. In this model, consumers update their information sets and maintain behavior at a re-optimized level. An alternative model is that the treatment acts as a "cue" that draws attention to energy use. As consumers discard or forget about the reports, their behavior could return quickly to its baseline state. Even in this latter model, however, repeated treatment could cause persistent effects as consumers begin to change their energy use habits or physical capital stock.

[1] See, for example, Cahill and Perera's (2009) review of the long-run effects of smoking cessation programs, as well as some of the studies of exercise, weight loss, school performance, and other behaviors that we cite later in the introduction.

[2] There is 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), water use (Ferraro and Miranda 2013, Ferraro and Price 2013) and charitable giving (Frey and Meier 2004, Shang and Croson 2009), as well as a broader literature in psychology on social norms, including Cialdini, Reno, and Kallgren (1990) and Cialdini *et al.* (2006).

Second, what is the incremental effect of continued treatment? By the end of our sample in 2013, some households have received home energy reports for 60 consecutive months, and one might wonder whether people have habituated to the reports after such a long time. This second question is mechanically connected to the first: if post-intervention effects are not persistent, then continued treatment is required for continued effects.

We study three Opower programs with four key features that make them well-suited to answer our questions. First, the programs are implemented as randomized control trials at a total of 234,000 households, allowing unbiased and precise estimates of effects on energy use. Second, these are the three longest-running Opower programs, having begun between early 2008 and early 2009. Third, treated households were randomly assigned to have treatment either discontinued after about two years or continued indefinitely. This allows us to measure both post-intervention persistence and the incremental effects of continued treatment relative to discontinuation. Fourth, while most utilities manually record residential electricity use on a monthly basis, one of our three utilities 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. At this site alone, we have 225 million observations of daily energy use.

Several aspects of the results are remarkable. At first, there is a pattern of "action and backsliding": consumers reduce electricity use markedly within days of receiving each of their initial reports, but these immediate efforts decay at a rate that might cause the effects to disappear after a few months if the treatment were not repeated. Over time, however, the cyclical pattern of action and backsliding attenuates. After the first four reports, the immediate consumption decreases after report arrivals are about five times smaller than they were initially.

For the groups whose reports are discontinued after about two years, the effects decay at about 10 to 20 percent per year - four to eight times slower than the decay rate between the initial reports. This difference implies that as the intervention is repeated, people gradually develop a new "capital stock" that generates persistent changes in outcomes. This capital stock might be physical capital, such as energy efficient lightbulbs or appliances, or "consumption capital" - a stock of energy use habits in the sense of Becker and Murphy (1988). Strikingly, however, even though the effects are relatively persistent and the "action and backsliding" has attenuated, consumers do not habituate fully even after two years: treatment effects in the third through fifth years are 50 to 60 percent stronger if the intervention is continued instead of discontinued.

What tangible actions do consumers take in response to the intervention? The only substantial differences between treatment and control on surveys of energy conservation actions relate to participation in utility-run energy efficiency programs. These typically involve improvements to large physical capital stock, such as insulation or refrigerators, that would mechanically generate persistent energy savings. We analyze administrative data from two utilities, which show that while the intervention does increase program participation, this explains only a small share of the effects on energy use. This implies that the intervention acts primarily through some combination of utilization habits and smaller unobserved changes to physical capital stock.

Although the field experiments were designed for program evaluation, not for distinguishing mechanisms, some models are more consistent with the results than others. One framework that is particularly useful is a cue-driven consumption model which embeds the Becker and Murphy (1993) persuasive advertising model in a multi-period framework. In such a model, which is a very simple analogue to Laibson (2001) or Bernheim and Rangel (2004), the intervention is an exogenous "cue" which temporarily lowers the marginal utility of energy consumption. As the cue is removed, consumers' energy use returns to its un-cued level. Tangibly, this is to say that the initial reports remind us to turn off the lights when we leave the house, but we lose motivation after a week or two. While the cues are active, consumers also gradually "invest" in capital stock changes, which cause persistent effects. For example, repeated home energy reports help us to get in the habit of turning the lights off, and if we eventually end up buying an air conditioner or washing machine, the reports induce us to buy Energy Star instead of the standard model.

Our results have concrete policy importance. Each year, electric and natural gas utilities spend billions of dollars on energy conservation programs in an effort to reduce energy use externalities and address other market failures that may reduce investment in energy efficient durable goods (Allcott and Greenstone 2012). Traditionally, one significant disposition of these funds has been to subsidize energy efficient physical capital 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.[3] The Opower programs are perhaps the most widely-implemented example of this approach. One of the foremost questions on practitioners' minds has been the extent to 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. In the absence of these empirical results, regulatory analysts had typically assumed zero persistence. We show that this assumption understates electricity cost savings over our four to five year samples by more than a factor of two, predicting \$41 to \$63 per household versus the observed figures of \$100 to \$149. Given the cost effectiveness of competing energy efficiency programs, the improved cost effectiveness from observed levels of persistence relative to the previous assumptions could change program adoption decisions for typical utilities.

We also show how understanding the timing of persistence and habituation can play an important role in designing behavioral interventions. In this context, it appears that program designers can improve cost effectiveness a factor of more than three relative to a one-shot intervention by initially repeating the intervention and then reducing treatment frequency as participants develop a new "capital stock" of habits or technologies. This highlights the importance of optimizing an intervention's timing and intensity, not just its content.

The paper proceeds as follows. The introduction concludes with a discussion of related literatures. Section 2 gives additional background on the program and describes the data. Section 3 presents the high-frequency analysis using daily data, while Section 4 presents the long-run analysis. Section 5 discusses physical and behavioral mechanisms, including the utility energy efficiency program participation data. Section 6 presents the cost effectiveness analysis and policy implications, and Section 7 concludes.

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

1.1 Related Literatures

Our study is 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.* 2013, Haselhuhn *et al.* 2012). Similarly, Gallagher (2013) shows that local homeowners are more likely to take up flood insurance immediately after a flood, but this effect steadily declines over time. The interpretation of home energy reports as a cue to save energy makes 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 illustrate one version of how people respond to repetition of similar cues: attention initially cycles, but people eventually become accustomed to the repeated reminders.

There are also studies of the medium- and long-run effects of interventions to affect exercise (Acland and Levy 2013, Charness and Gneezy 2009, Milkman, Minson, and Volpp 2013, Royer, Stehr, and Snydor 2013), 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, Ferraro and Price 2013), academic performance (Jackson 2010, Jensen 2010, Levitt, List, and Sadoff 2010, Walton and Cohen 2011), voting (Gerber, Green, and Shachar 2003), charitable donations (Landry *et al.* 2010, Shang and Croson 2009), job choices (Coffman, Featherstone, and Kessler 2013), labor effort (Gneezy and List 2006), and other choices. Compared to these studies, we document relatively persistent changes in outcomes over a relatively long time horizon. Furthermore, one unusual feature of our experiments is the random assignment to continued vs. discontinued treatment, which allows us to cleanly measure the incremental effect of continued treatment.

Finally, our paper is directly related to other studies of Opower and similar programs. 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 literature that studies Opower programs over shorter time horizons, including Allcott (2011), Ayres, Raseman, and Shih (2012), Costa and Kahn (2013), and a number of industry reports such as Ashby *et al.* (2012), Integral Analytics (2012), KEMA (2012), Opinion Dynamics (2012), Perry and Woehleke (2013), and Violette, Provencher, and Klos (2009). Allcott and Mullainathan (2012) show that the average treatment effects in the first one to two years across the first 14 Opower sites range from 1.4 to 2.8 percent of electricity use. Relative to this literature, our contributions are clear. First, we document consumers' "action and backsliding" using high-frequency data.[4] Second, we study Opower's three longest-running programs over a relatively long time horizon. Third, we exploit the continued vs. discontinued treatment groups to measure both habituation and post-intervention persistence. Fourth, we bring together the high-frequency and long-run analyses to analyze how persistence and habituation affect cost effectiveness and optimal program design.

2 Experiment Overview

2.1 The Home Energy Report

Figure 1 is a home energy report for an example utility. The first page features a Neighbor Comparison module, which compares the household's recent energy use to that of 100 neighbors with similar house characteristics. The second page includes personalized energy use feedback, which varies from report to report. This feedback might include comparisons to the household's usage in previous years or trends in usage compared to neighbors. The second page also includes an Action Steps module, which provides energy conservation tips. These are drawn from a large library of possible tips, and they vary with each report. Opower targets specific tips to different households: for example, a household with relatively heavy summer usage is more likely to see information about purchasing energy efficient air conditioners.

[4] Although we are the first to document it, this potential effect has been of previous interest. For example, Ayres, Raseman, and Shih (2012) test for what they call a "staleness effect" using monthly billing data for recipients of quarterly reports, but find no evidence that effects vary with the time since the last report. It would have been unlikely for Ayres, Raseman, and Shih (2012) or others to even find suggestive evidence in monthly billing data because the report arrival dates do not match up well with the monthly data reporting periods. In their 2009 working paper, Ayres, Raseman, and Shih had also discussed informal visual tests of what they called a "retrenchment effect" using weekly data for the first few months of an Opower program, but they removed these two sentences from the published version.

2.2 Experimental Design

Table 1 outlines experimental design and provides descriptive statistics for our three sites, which we have been asked not to identify directly. Site 1 is in the upper Midwest, with cold winters and mild summers, while Sites 2 and 3 are on the West coast. The initial experimental populations across the three sites comprise 234,000 residential electricity consumers. To be eligible for the program, households must be single-family homes, have at least one to two years of valid pre-experiment energy use data, and satisfy some additional technical conditions.[5] Site 1 is a relatively small utility, and its entire residential customer population was included. In Site 2, the utility decided to limit the program to the approximately 100,000 consumers in one county that purchase both electricity and natural gas. From this group, about 16,000 additional households were eliminated because they did not have enough comparable neighbors or because they used relatively little energy (less than the equivalent of 80 million British thermal units per year). In Site 3, Opower selected census tracts within the customer territory to maximize the number of eligible households.

The experimental populations were randomly assigned to treatment or control. In Site 3, which was Opower's first program ever, households were grouped into 952 geographically-contiguous "block batch" groups, each with an average of 88 households, which were randomly assigned to treatment or control. This was done because of initial concern over geographic spillovers: that people would talk with their neighbors about the reports. No evidence of this materialized, and all programs since then, including Sites 1 and 2, have been randomized at the household level. In Sites 1 and 2, treatment group households were randomly assigned to receive either monthly or quarterly reports. In Site 3, heavier users were assigned to receive monthly reports, while lighter users were assigned to quarterly.

The three experiments began between early 2008 and early 2009. After about two years, a subset of treatment group households were randomly selected to stop receiving reports. We call this group the "dropped group." The remainder of the treatment group, which we call the "continued group,"

[5] Typically, households in Opower's experimental populations need to have valid names and addresses, no negative electricity meter reads, at least one meter read in the last three months, no significant gaps in usage history, exactly one account per customer per location, and a sufficient number of neighbors to construct the neighbor comparisons. Households that have special medical rates or photovoltaic panels are sometimes also excluded. Utility staff and "VIPs" are sometimes automatically enrolled in the reports, and we exclude these non-randomized report recipients from any analysis. These technical exclusions eliminate only a small portion of the potential population.

is still receiving reports. In Sites 2 and 3, the entire continued group is still receiving reports at their original assigned frequency. In Site 1, the continued group was changed to biannual frequency at the beginning of 2012.

2.3 Data for Long-Run Analysis

In the "long-run analysis," we analyze monthly billing data from the three sites over the past four to five years. The three utilities bill customers approximately once a month, and our outcome variable is mean electricity use per day over a billing period. We therefore have about 12 observations per household per year, or 16.7 million total observations in the three sites.

In each site, we construct baseline usage from the earliest one-year period when we observe electricity bills for nearly all households.[6] In each site, average baseline usage is around 30 kilowatt-hours (kWh) per day, or between 11,000 and 11,700 kWh per year. These figures are comparable to the national average of 11,280 and to the average across all residential customers in each utility (U.S. Energy Information Administration 2011, 2013).

For context, one kilowatt-hour is enough electricity to run either a typical new refrigerator or a standard 60-watt incandescent lightbulb for about 17 hours. In the average American home, space heating and cooling are the two largest uses of electricity, comprising 26 percent of consumption. Refrigerators and hot water heaters use 17 and 9 percent of electricity, respectively, while lighting also uses about 9 percent (U.S. Energy Information Administration 2009). Appendix Figure A1 provides more detail on nationwide household electricity use.

The three utilities also have fairly standard pricing policies. The utility in Site 1 charges 10 to 11 cents/kWh, depending on the season. The utilities in Sites 2 and 3 have increasing block schedules, with marginal prices of 8 to 11 cents/kWh and 8 to 18 cents/kWh, respectively, depending again on the season.

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 any observations with more than 1500 kilowatt-

[6] As shown in Table 1, the 12-month baseline periods in the three sites begin 16 to 23 months before the first reports. The remaining four to 11 months before the interventions begin are used in Figure 4 and the first row of Table 4 to show that pre-treatment levels and trends do not differ between treatment and control. We have much higher power to detect potential spurious differences in levels and trends once we condition on baseline usage.

hours per day. In Site 2, for example, this is 0.00035 percent of observations. Table 1 documents that in all three sites, baseline energy usage is balanced between treatment and control groups, as well as between the dropped and continued groups within the treatment group.

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. Because heating and cooling are such important uses of electricity in the typical household, heating and cooling degrees are important correlates of electricity demand.

There is one source of attrition from the data: households that become "inactive," typically when they move houses. If a customer moves, he or she no longer receives reports after the inactive date, and in most cases we do not observe electricity bills. In our primary specifications, we do include the households that eventually become inactive, but we exclude any data observed after the inactive date. As Table 1 shows, 20 to 26 percent of households move in the four to five years after treatment begins, or about five percent per year. The table presents six tests of balanced attrition from moving: treatment vs. control and dropped vs. continued in each of the three sites. One of those six tests rejects equality: in Site 1, dropped group households are slightly more likely to move than continued households. For several reasons, we are not very concerned that this could bias the results: the two groups are balanced on pre-treatment usage, Figure 4a shows that the treatment effects during the joint treatment period are almost visually indistinguishable, and Table 5 confirms that the treatment effects are statistically indistinguishable during the first and second years of joint treatment.

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 these sites, about two percent of the treatment group has opted out since the programs began. 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 intent-to-treat estimates, where by the end of the sample, the Local Average Treatment Effect on the compliers who do not opt out is about 1/(1-0.02) larger than our reported ATE. Because the opt-out rate is so low, we do not make any more of this distinction. 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 High-Frequency Analysis

In Sites 1 and 3, each household's electricity meter is read each month by utility staff, who record the total consumption over the billing period. By contrast, Site 2 has advanced electricity meters which record daily electricity consumption. The "high-frequency analysis" exploits these daily data.

For the high-frequency analysis, it is useful to separately analyze the groups randomly assigned to monthly vs. quarterly frequencies. We also exclude the dropped group households in the monthly and quarterly groups after their reports are discontinued. This reduces the sample size somewhat after September 2010 but does not generate imbalance because these households were randomly selected.

There was also a "second wave" of about 44,000 households from a nearby suburb that began treatment in February 2011. The treatment group received a total of six bimonthly reports before their intervention was discontinued in mid-2012. Instead of random assignment, households were assigned to treatment and control using even vs. odd address numbers. This generated mild imbalance on baseline usage (0.69 kWh/day, SE=0.20 kWh/day). Although it appears that conditioning on season-specific baseline usage addressed potential biases, we have relegated these results to the online appendix. Results from this group are consistent with results from the monthly and quarterly groups. Between the monthly, quarterly, and bimonthly groups, there are 225 million household-by-day observations at 123,000 households.

All reports delivered in a given month to any household in Site 2 are generated and mailed on the same days. 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. We use these mailing dates and the U.S. Post Service estimates of delivery times to residences in Site 2 to predict report arrival dates.[7] Of course, reports may arrive before or after the predicted day, and people may not open the letters immediately.

3 High-Frequency Analysis

3.1 Graphical

Figure 2 plots the average treatment effects for each day of the first year of the Site 2 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 within a seven-day window around day d. We include a set of day-specific constants π_t, and we also 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). Here and everywhere else in the paper, superscripts always index time periods; we never use exponents. For each day d, the regression is:

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

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). Appendix Figure A2 replicates this figure but also includes standard errors, which average 0.067 and 0.095 kWh/day for the monthly and quarterly groups. Note that treatment effects are negative, indicating that the treatment causes a reduction in electricity use, and much of the apparently-idiosyncratic variation in treatment effects is within the confidence intervals.

[7] 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.

Figure 2 has two important features. First, households reduce energy use markedly within one to two weeks of the first few report arrival dates. The first report arrival, which occurred around October 24th, is the most stark: energy use decreases by 0.3 to 0.4 kWh/day between mid-October and November 3rd. This is 1 to 1.3 percent of average electricity use, and it is equivalent to each treatment group household turning off six standard 60-watt lightbulbs for an hour every day. When the second reports arrive in late November (for the monthly group) or late January (for the quarterly group), there is again a marked reduction in energy use. After the first few reports, however, it becomes harder to visually distinguish any immediate conservation effects after the predicted arrival dates. Note that these smoothed treatment effects begin to change slightly before the predicted report arrival dates because the seven-day bandwidths start to include some post-arrival days.

Figure 2's second key feature is that consumers appear to backslide on their immediate conservation actions. This is easiest to see for the quarterly group, as they have three times longer than the monthly group to backslide between reports. Between early November and early January, for example, the quarterly treatment effect weakens by about 0.2 kWh/day, meaning that about half of their initial conservation actions were abandoned within two months.

While we like the transparency of this simple presentation of raw data, collapsing across multiple report arrivals and analyzing effects in "event time" can both increase precision and smooth over idiosyncratic factors such as holidays. Furthermore, controlling for weather could be important to ensure that "action and backsliding" is caused by changes in conservation effort, not by changes in weather correlated with report arrivals. For example, if the second or third report happened to arrive in an extremely cold week, the treatment effects would likely have been stronger in that week even if the report had arrived a week later. If weather is systematically correlated with report arrivals, failing to control for weather might cause us to falsely interpret such treatment effect fluctuations as immediate cue-driven responses of conservation effort.

We therefore estimate a vector of event time treatment effects $\boldsymbol{\tau}^a$, where a indexes days before and after report arrivals. We include a vector of indicators $\boldsymbol{\phi}_t$ for the periods around each individual report arrival. The interaction of $\boldsymbol{\phi}_t$ with T controls for the fact that the treatment effect could be weaker or stronger over the entire window around each particular report, due to seasonality or other factors. The two-part vector $\mathbf{M}_{it}$ includes heating degrees and cooling degrees on day t at the weather station closest to household i. The event time regression is:

$$Y_{it} = \boldsymbol{\phi}_t T_i + \boldsymbol{\tau}^a T_i + \boldsymbol{\beta}_1 T_i \mathbf{M}_{it} + \boldsymbol{\beta}_2 \mathbf{M}_{it} + \boldsymbol{\theta} \mathbf{Y}_i^b + \pi_t + \varepsilon_{it} \quad (2)$$

To further increase precision, we graph the average of the daily ATEs over three or five day moving windows, with standard errors calculated using the Delta method.[8]

Figure 3a plots the treatment effects, using only the sample of days around the first four reports. The monthly and quarterly groups both follow the same striking pattern. There is no trend four to ten days before the arrival date. Effects start to appear one to three days before the predicted arrival date, both because the mail may arrive earlier than predicted and because the running mean bandwidth includes some post-arrival days. Treated households then reduce consumption by about 0.2 kWh/day in the several days after the predicted arrival date. Conservation effort reaches its peak about 10 days after report arrival, and consumers backslide after that point. The monthly group does not have much time for backsliding, because the next report soon arrives and cues additional action. For the quarterly group, the treatment effect decays by almost 0.2 kWh/day between 10 days and 80 days after the report arrival.

Figure 3b is analogous to Figure 3a, except that the sample begins with the fifth report. The cyclical action and backsliding effects, if any, are substantially attenuated relative to the first four reports. Consumers act as if they become accustomed to the reports and are no longer "surprised" and spurred into immediate action.

3.2 Empirical Strategy

We now formally quantify the "action and backsliding" patterns suggested by the figures. We first estimate the immediate conservation effects. Define S_t^0 as an indicator variable for the "Arrival Period": the seven days beginning three days before and ending three days after the predicted arrival date. S_t^1 is an indicator for the seven day period after that, which we call the "Post-

[8] The choice of bandwidth does not affect the basic shape of the graph. Based on visual inspection, we used three-day and five-day bandwidths for the monthly and quarterly groups, respectively. The omitted a categories are the first two days in each of the event windows, so these effects are fixed to zero, with zero standard errors.

Arrival Period," and S_t^{-1} is an indicator for the seven-day "Pre-Arrival Period" before. Define $S_t^a = S_t^{-1} + S_t^0 + S_t^1$ as an indicator for all 21 days in that window. As above, $\boldsymbol{\phi}_t$ is a vector of indicators for the window around each individual report, $\mathbf{M}_{it}$ is heating and cooling degrees on day t for the weather station nearest household i, $\mathbf{Y}_i^b$ is the three seasonal baseline usage controls, and π_t are day-of-sample dummies. The regression is:

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

τ^1 is our coefficient of interest: the change in the treatment effect in period S^1 relative to period S^{-1}.

We then estimate the rate at which the treatment effect decays between reports. 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 time (in years) past the beginning of the window. 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. The regression is:

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

For simplicity, this model assumes that treatment effects decay linearly over time. One might hypothesize that the decay process could be convex or concave, and it is almost certainly unrealistic to extrapolate beyond the time when the predicted treatment effect reaches zero. However, we do not have enough households or time between reports to test this.

3.3 Results

Table 2 presents the estimates of Equation (3). There are four columns, one pair each for the monthly and quarterly groups. Analogously to Figures 3a and 3b, we present separate estimates for the earliest four reports (the left column of each pair) and all later reports (the right column). Using four reports as the division into "Early" and "Later" was our initial judgment. It is intended as a discrete approximation to what is likely a gradual process through which the action and backsliding effect might attenuate. We would need many more programs with high-frequency data to reliably estimate the speed of this attenuation.

The formal estimates mirror the figures. For the first four reports, τ^1 is 0.185 and 0.197 kWh/day for the monthly and quarterly groups. This means that in the week after the seven-day arrival windows compared to the week before those windows, electricity consumption decreases by the equivalent of about three 60-watt lightbulbs used for one hour. After the first four monthly and quarterly reports, the τ^1 coefficients are still statistically significant, but they are less than one-fifth the magnitude of $\widehat{\tau}^1$ for the initial four reports.

These results can be used to highlight how much of consumers' responses to the intervention happen almost immediately after receiving the initial reports. First consider the monthly group. Multiplying the incremental post-arrival period effect $\widehat{\tau}^1$ by four gives a total decrease of 0.74 kWh/day - the equivalent of turning off a standard 60-watt lightbulb for an additional 12 hours. This means that if the intervention's only effect were to generate immediate action in the post-arrival period, and if that immediate action were sustained over time, the treatment effect after the first four reports would be -0.74 kWh/day. However, the average treatment effect just before the fifth report (the monthly group's $\widehat{\tau}^d$ estimated by Equation (1) for February 13, 2009) is -0.52. The reason for this potential "overestimate" is that the treatment group's immediate action is not sustained over time - the effects decay in the intervening days between the seven-day post arrival period S^1 and the arrival of the next report. It must be the case that consumers are backsliding, or the average treatment effects would need to be larger.

The story is the same for the quarterly group. Multiplying $\widehat{\tau}^1$ by four gives a total decrease of 0.79 kWh/day. Thus, if these immediate actions were sustained, the treatment effect after the first four reports would be -0.79 kWh/day. In contrast, the ATE just before their fifth report is -0.35. This difference is even larger than for the monthly group because the quarterly group has three times longer to backslide on its immediate actions.

Table 3 formally measures this backsliding using Equation (4). A positive δ implies that treatment group consumption increases in the windows between reports. This backsliding is statistically significant only for the initial four reports, and the point estimates are much larger than for the later period. To put the magnitudes of δ in context, focus on the estimates for the quarterly group. A $\widehat{\delta}$ of 0.708 means that a treatment effect of -0.708 kWh/day would decay to zero in one year, if the linear decay continued to hold. Thus, the jump in treatment effects of $\widehat{\tau}^1$ =-0.197 from column (3) of Table 2 would decay away fully within just over three months. This never happens, because the next report arrives less than three months after the window S^w begins.

After the initial four reports, the fact that the point estimates of δ are still positive suggests 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 to three years instead of two to ten weeks.

Appendix Tables A1 through A4 present robustness checks for Tables 2 and 3. The results are highly insensitive to excluding weather controls, using different weather controls, and excluding outliers. The only substantive difference is that when weather controls are excluded, the decay rate δ for the monthly group between the initial four reports becomes smaller and has a t-statistic of 1.07. This particular coefficient is relatively difficult to estimate because the monthly event windows S^w are so short and because the sample is limited to the first four reports. The results for the bimonthly group are similar to the monthly and quarterly results.[9]

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 consider only the period after the

[9] The one substantive difference is that the bimonthly group's τ^1 coefficient, which reflects the immediate conservation effect in the Post-Arrival Period, is larger in absolute value for the fifth and sixth reports than it is for the first four. This difference is not statistically significant, however, and because the coefficient is estimated off of only the fifth and sixth reports, it is difficult to infer much of a pattern. For example, there could have been other idiosyncratic factors that increased the treatment effects as these two reports arrived, or these reports could have presented information in a particularly compelling way. This also highlights that the action and backsliding effect likely attenuates gradually, not suddenly, and one might still expect some immediate action as the fifth and sixth reports arrive.

first four reports, because before that, the quarterly ATE decays 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 A5 shows that the $\widehat{\tau}^0$ and $\widehat{\tau}^1$ coefficients for these placebo report arrival dates are statistically zero and economically small relative to those estimated in Table 2.

4 Long-Run Analysis

For the long-run analysis, we analyze the household-by-month billing data at each of the three sites. Unlike in the high-frequency analysis, we combine the monthly and quarterly groups, as their differences are not useful in making our argument. We ask two questions. First, how persistent are effects for the dropped group after treatment is discontinued? Second, does treating the continued group cause incremental conservation, or have people fully habituated after two years of treatment?

4.1 Graphical

We first plot the time path of treatment effects over the sample for both the continued and dropped groups, for each of the three sites. Analogously to the high-frequency graphical analysis, we use a three month moving window to smooth over idiosyncratic variation. The variable Y_{itm} is household i's average daily electricity usage for the billing period ending on date t occurring in month-of-sample m. 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, with $D_i + E_i = T_i$. The coefficients τ_n^D and τ_n^E are the average treatment effects for the three-month window around month n for each group. We include month-by-year 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 during the baseline period. The π_m are month-by-year intercepts.

For each month n, the regression is:

$$Y_{itm} = \tau_n^D D_i + \tau_n^E E_i + \theta_m Y_{im}^b + \pi_m + \varepsilon_{itm}, \qquad \forall m \in [n-1, n+1] \tag{5}$$

In this regression, standard errors are clustered over time at the level of randomization, per Bertrand, Duflo, and Mullainathan (2004). We cluster by household in Sites 1 and 2 and by block batch in Site 3.

Figures 4a-4c present the results for Sites 1-3. The y-axis is the treatment effect, which is negative because the treatment causes a reduction in energy use. The three figures all illustrate the same basic story. To the left of the first vertical line, the intervention has not yet started, and the treatment effects are statistically zero. The effects grow fairly rapidly over the intervention's first year, after which the growth rate slows. Until the second vertical line, both the continued and dropped groups receive the same treatment, and the effects for the two groups are indistinguishable, as would be expected due to random assignment. The average treatment effects in the second year range from 0.7 to 1.0 kWh/day, or about three percent of average consumption. After the dropped group's last report, the effects begin to decay relative to what they had been during the intervention, but the effects are remarkably persistent. The dropped group ATEs seem to diminish by about 0.1 to 0.2 kWh/day each year.

The effects are highly seasonal. In all three sites, effects are stronger in the winter compared to the adjacent fall and spring. Although the great majority of households in the populations primarily use natural gas instead of electricity for heat, the fans for natural gas heating systems use electricity, and many homes also have portable electric heaters. In Sites 1 and 3, the effects are also stronger in the summer compared to the fall and the spring. This suggests that an important way in which people respond to the treatment is to reduce heating and cooling energy, either through reducing utilization or perhaps changing to more energy efficient physical capital stock. In Site 2, the average daily temperature in July is a mild 67 degrees, so air conditioner use is more limited, and the treatment effects are relatively weak in the summer. In Site 3, the monthly point estimates jump around more because of the block batch-level randomization, but they do not move more than we would expect given the confidence intervals and underlying seasonality.

The graphs also illustrate that the continued groups do not fully habituate to treatment: in all sites, continued treatment has incremental effects relative to the dropped group. Furthermore, in Sites 2 and 3 where treatment is continued at the same frequency, treatment effects continue to strengthen over time. In Site 1, the continued group's effects begin to diminish slightly as they begin to receive biannual instead of monthly or quarterly reports.

4.2 Empirical Strategy

For the formal long-run analysis, we break the samples into four periods. Period 0 is the pre-treatment period, period 1 is the first year of treatment, and period 2 runs from the beginning of the second year to the time when treatment is discontinued for the dropped group. Period 3 is the post-drop period: the remainder of the sample after the dropped group is discontinued. We denote P_m^p as indicator variables for whether month m is in period p. The variable r_t measures the time (in years) since the beginning of period 3. Analogous to the high-frequency analysis, $\mathbf{M}_{im}$ represents two weather controls: average heating degrees and average cooling degrees for household i in month m.

The primary estimating equation is:

$$\begin{aligned} Y_{itm} = \; & (\tau^0 P_m^0 + \tau^1 P_m^1 + \tau^2 P_m^2) \cdot T_i \\ & + (\alpha^0 P_m^0 + \alpha^1 P_m^1 + \alpha^2 P_m^2) \cdot E_i \\ & + (\tau^3 T_i + \alpha^3 E_i) \cdot P_m^3 \\ & + \left(\delta^{LR} r_t D_i + \rho r_t E_i + \omega r_t\right) \cdot P_m^3 \\ & + \mathbf{M}_{im}(P_m^2 + P_m^3) \cdot (T_i \psi_1 + \psi_2) \\ & + \theta_m Y_{im}^b + \pi_m + \varepsilon_{itm} \end{aligned} \tag{6}$$

The third and fourth lines parameterize the treatment effects for the continued and dropped groups in the post-drop period. The coefficient δ^{LR} captures the treatment effect decay rate for the dropped group, while ρ measures the trend in the continued group treatment effect. Because r_t has units in years, the units on δ^{LR} and ρ are kWh/day per year. The τ^3 and α^3 coefficients are intercepts: the fitted treatment effects for the day at the beginning of period 3.

The fifth line controls for the interaction of $\left(P_m^2 + P_m^3\right) \cdot T_i$ with heating and cooling degrees $\mathbf{M}_{im}$. When these controls are included, τ^2, τ^3, α^2, α^3, δ^{LR}, and ρ represent predicted effects and decay rates for a month in which the mean temperature each day is 65 degrees. These weather controls are important because if temperatures were more (less) mild later in the post-drop period, this would likely make the treatment effects weaker (stronger), which would otherwise load onto δ^{LR} and ρ. Such changes in the broader "economic environment" would confound our interpretation of the δ^{LR} parameter as reflecting a change in household behavior or capital stock.

4.3 Statistical Results

Table 4 presents estimates of Equation (6), excluding the fourth and fifth lines. This gives estimates of the dropped group treatment effects (τ) and the difference between continued and dropped group effects (α). The table contains two "placebo tests", both of which confirm the randomization's validity: effects are statistically zero in the pre-treatment period P^0, and effects do not differ between the dropped and continued groups while they both receive the same treatment in P^1 (the "1st Year") and P^2 ("2nd Year Until Drop").

The table demonstrates persistence: in all three sites, the dropped group still has a statistically non-zero treatment effect in the post-drop period. The τ^3 coefficients are very similar, ranging from -0.584 to -0.627. In tangible terms, a treatment effect of -0.6 kWh/day means that the average treatment group household took actions equivalent to turning off a standard 60-watt lightbulb for about 10 hours each day. Recalling that average usage is around 30 kWh/day, this corresponds to two percent of electricity use.

Table 4 also demonstrates that people do not fully habituate to the intervention, even after two years of repeated treatment. In all three sites, the continued group has a statistically significantly stronger treatment effect in the post-drop period relative to the dropped group. The point estimates of τ^3 and α^3 suggest that continuing the intervention increases the treatment effects in the post-drop period by a remarkable 50 to 60 percent.

Table 5 presents estimates of Equation (6), excluding the second line. The δ^{LR} and ρ coefficients are the bottom two coefficients in each column. The δ^{LR} parameters range from 0.09 kWh/day per year in Site 3 to 0.18 kWh/day per year in Site 1. If the linear trend continues, the effects would not return to zero until five to ten years after treatment was discontinued. If the linear model understates (overstates) persistence, our cost effectiveness projections later in the paper will be conservative (optimistic).

Compare these $\widehat{\delta}^{LR}$ parameters to the $\widehat{\delta}$ decay rate from the previous section between each of the first four reports. Our preferred estimate is the $\widehat{\delta}$ =0.708 the quarterly group, as this is the most statistically precise and is estimated off of the longest window between reports. This is four to eight times faster than $\widehat{\delta}^{LR}$. This implies that between the first four reports and the time when treatment is discontinued, the dropped group forms some kind of "capital stock" which causes substantially more persistence. In the next two sections, we discuss the potential causes and consequences of this process.[10]

The online appendix includes additional results. Appendix Table A6 tests whether the effects decay proportionally faster or slower for the different frequency groups or for heavier baseline users, but the standard errors are too wide for useful inference. Appendix Table A7 replicates Table 4, except excluding all data for households that move at any point. These balanced panel estimates are important because by the end of the sample, 20 to 26 percent of households have moved. Even though this is balanced between treatment and control, if the movers had systematically different treatment effects, this could cause the estimated treatment effects to change over time. Appendix Table A8 replicates Table 5, first excluding weather controls and then limiting to the balanced panel. The results are strikingly robust: every single coefficient is statistically and economically the same.

5 Physical and Behavioral Mechanisms

What actions underlie the observed effects? In particular, to what extent does the intervention change utilization habits vs. investments in physical capital stock? While this question is difficult to answer, we can provide some information from surveys of energy conservation actions and administrative data on participation in utility-run energy conservation programs. At the end of this section, we discuss potential behavioral mechanisms underlying these actions.

[10] We note that the long-run persistence is measured from one to four years after the period when the short run decay rate is measured. It is possible that changes in macroeconomic conditions or other time-varying factors might cause differences in these decay rates. Ultimately, however, we are not very concerned with this issue.

5.1 Utility Energy Efficiency Program Participation

We have analyzed surveys in which about six thousand consumers in six Opower sites were asked about a series of energy conservation actions. Because these are self-reported data, we relegate the results to the online appendix. The only substantive differences between treatment and control groups relate to participation in utility energy efficiency programs. Fortunately, these is precisely the area where additional data are available.

In this section, we analyze data on participation in utility energy efficiency programs in Sites 2 and 3. These data have three useful features. First, they are administrative data instead of self-reports, so they are comprehensive and consistent. Second, utilities estimate the energy conserved through each action, which makes it possible to translate percentage point effects into effects on energy use. Third, while these data cover only a small share of the ways that households can conserve, they are a good measure of the largest physical capital stock investments that save the most energy.

In Site 3, the utility offers rebates or low-interest loans in three categories: appliances, "home improvement," and HVAC (heating, ventilation, and air conditioning). For appliances, the utility mails \$50 to \$75 rebate checks to consumers who purchase energy efficient clothes washers, dishwashers, or refrigerators. To claim the rebate, the homeowner needs to fill out a one-page rebate form and mail it with the purchase receipt and a current utility bill to the utility within 30 days of purchase. For "home improvement," the utility offers up to \$5,000 in rebates for households that install better insulation or otherwise retrofit their homes in particular ways. For HVAC, the utility offers \$400 to \$2000 for energy efficient central air conditioning systems or heat pumps, or \$50 for energy efficient window air conditioners. Most home improvement and HVAC jobs are done by contractors. Some consumers probably buy energy efficient appliances and window air conditioners without claiming the utility rebates, and thus these capital stock changes might be unobserved in the data. However, because the home improvement and HVAC rebates are larger, and because the contractors coordinate with the utility and facilitate the rebate process, consumers who undertake these large physical capital improvements are very likely to claim the rebates and thus be observed in the administrative data.

The top panel of Table 6 presents Site 3's program participation statistics for the first two years of the program, from April 2008 through June 2010. In total, 3855 households in the experimental population participated in one of the three programs. Column (1) presents estimates of the savings that might accrue for the average participant.[11] Column (3) presents the difference in participation rates between treatment and control, in units of percentage points ranging from 0 to 100. The results confirm the qualitative conclusions from the household surveys: the treatment group is slightly (0.417 percentage points) more likely to participate in energy conservation programs. Participation rates are 44 out of every 1000 households in control and 48 out of every 1000 households in treatment.

How much of the treatment effect on energy use does this explain? Table 4 showed that by the program's second year, the treatment group is conserving about 860 Watt-hours/day (0.860 kilowatt-hours/day) relative to control. Column (4) of Table 6 multiplies the difference in participation rate by the savings estimates in Column (1), showing that the difference in program participation might cause energy use to decrease by 14 Watt-hours per day. Thus, while there are statistically significant changes in program participation, this explains less than two percent of the treatment effect.

The fact that treatment effects decay more slowly as the home energy report intervention continues suggests that it is especially important to test for capital stock formation later after treatment begins. Therefore, we also examine similar administrative data in Site 2 for 2011, the Opower program's third year. This utility offers a similar set of programs as in Site 3, except that the exact rebate amounts may vary, and some rebate forms can be submitted online instead of in the mail. In the Site 2 data, we observe more precisely the action that the consumer took, as well as the utility's estimate of the electricity savings.

The bottom panel of Table 6 presents the Site 2 data. The most popular programs are clothes washer rebates, insulation, removal of old energy-inefficient refrigerators and freezers, installation of low-flow showerheads, energy efficient windows, and compact fluorescent lightbulbs (CFLs). Savings in Column (1) are zero for insulation and duct sealing because for regulatory purposes, the utility deems that these programs reduce natural gas use but not electricity.

Columns (3) and (4) compare the takeup rates and implied electricity savings between the

[11] We do not have the utility's administrative estimates in Site 3. These are thus our estimates based on the administrative estimates for similar programs in Site 2.

control group and the continued treatment group, which is still receiving home energy reports during 2011. There is a statistically significant difference for only one program: CFL replacement, which generates 2.25 Watt-hours/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 about 6 Watt-hours/day. By contrast, the continued group's treatment effect in the post-drop period was (negative) 870 Watt-hours/day (0.870 kilowatt-hours/day), which was an increment of 181 Watt-hours/day compared to the year before. Thus, as in Site 3, only a small fraction of the savings are due to participation in utility energy efficiency programs.[12]

5.2 Behavioral Mechanisms

Although these experiments were not designed to provide sharp tests of behavioral models, some models are more likely than others to explain the results. For example, one potential model would have been that the energy conservation tips and social comparisons act purely through information provision. In a standard information provision model, consumers update information sets and permanently re-optimize consumption. If this were the only mechanism through which the intervention acted, it would be difficult to explain the observed backsliding.

As suggested in the introduction, one model consistent with these results is a multi-period model of persuasive advertising combined with long-run formation of capital stock. The reports are an exogenous "cue" which causes people to pay attention to energy conservation. This lowers the marginal utility of energy consumption (increases the marginal utility of energy conservation) and thus reduces energy use. The cue is removed as people discard the paper report, and as memory decays, the marginal utility of consumption returns to it un-cued state. This causes energy use to cycle with report arrivals. The fact that the cycles have relatively high frequency implies that the initial reports primarily affect utilization behaviors, such as adjusting thermostats, turning off lights, and unplugging unused electronics.

[12] Several recent consulting reports, including Integral Analytics (2012), KEMA (2012), Opinion Dynamics (2012), and Perry and Woehleke (2013), have also examined the intervention's effect on utility program participation at these sites and others. Their findings are very similar to ours: the Opower intervention sometimes causes increases in program participation, but this accounts for only a small fraction of the overall reduction in energy use.

However, the attenuation of these cycles after about the first four reports suggests that people become accustomed to the cues. This is consistent with psychological models of habituation such as those reviewed by Rankin *et al.* (2009) and Thompson and Spencer (1966). This result is different than the Laibson (2001) cue-theory model, in which cues affect marginal utility more powerfully over time as people increasingly associate the cue with a behavior. Laibson (2001) gives the example of Pavlov's dogs, who begin to salivate when they hear bells after repeated pairings of bells with food. In our case, the cue is already closely associated with behavior: a report about energy conservation naturally makes one think about ways to conserve energy. Thus, repeated cues are not needed to generate a conditioned response. Instead, people become accustomed to them, and eventually we are not "surprised" when the next cue arrives.

Of course, there are other models that could explain the observed "action and backsliding" and attenuation thereof. For example, the energy conservation tips could cause people to experiment with different energy conservation actions, which they discard after learning that the net benefits are not as high as expected. While some treated households may do this, three factors make this model seem less likely to be widely applicable. First, the initial research by Nolan *et al.* (2008) and Schultz *et al.* (2007) suggested that the most powerful feature of this type of intervention is the social comparison module, which makes energy use salient but gives no practical guidance on energy conservation actions. Second, the survey results on "repeated actions" in Appendix I imply that the treatment group is not experimenting with new actions. Instead, people appear to be increasing the effort devoted to actions that they were already taking. Third, the primary way in which consumers learn about the gross benefits of energy conservation is when they receive their energy bills. These bills, however, are calculated and sent with some delay, while the observed backsliding starts less than two weeks after the home energy report arrives.

One additional model is that consumers could literally learn and forget new energy conservation actions, as suggested by the Agarwal *et al.* (2013) phrase of "learning and backsliding" in the case of credit card fees. However, it seems unlikely that people would literally forget new information so quickly.

Simultaneous to this high-frequency cyclicality, there is also a long-run process of capital formation: the fact that the treatment effects decay more slowly after two years than between the initial reports means that consumers have formed some type of new "capital stock." The program participation data shows that very little of this capital stock is large changes to physical capital such as insulation or home energy retrofits. However, consumers may make other smaller changes to physical capital stock, such as installing energy efficient compact fluorescent lightbulbs or window air conditioners.

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 in environments that are stable over time (Oullette and Wood 1998, Schneider and Shiffrin 1977, Shiffrin and Schneider 1977). This is consistent with the results of Charness and Gneezy (2009), who show that financial incentives to exercise have some long-run effect after the incentives are removed, suggesting that they induce people to form new habits of going to the gym. In Becker and Murphy (1988), consumption capital also depreciates, which is consistent with the finding that treatment effects decay even after two years of the intervention.

6 Implications: Cost Effectiveness and Program Design

In this section, we assess the importance of persistence for cost effectiveness and for program design. We define cost effectiveness as the cost to produce and mail reports divided by the kilowatt-hours of electricity conserved.[13]

[13] We assume that the cost per report is $1 and ignore fixed costs. Although cost effectiveness is a common metric by which interventions are assessed, we emphasize several of the reasons why this is not the same as a welfare evaluation. First, 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'. Second, the treatment also causes households to reduce natural gas use, which we do not study here. Third, 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. Of course, this distinction between the observed outcome and welfare is not unique to this domain: with the exception of DellaVigna, Malmendier, and List (2012), most studies of weight loss, smoking, charitable contributions, and other behaviors are only able to estimate effects on behaviors, not on welfare. In our setting, however, the focus on cost effectiveness is still relevant: regulators mandate that utilities run cost-effective energy conservation programs, without explicit regard for welfare.

6.1 Persistence Matters for Cost Effectiveness

When assessing the cost effectiveness of Opower home energy reports and other "behavioral" energy conservation programs, most utilities have implicitly or explicitly assumed zero persistence. These programs are often evaluated in one-year cycles, where the program costs for that year are compared to econometric estimates of energy conserved in that year. This conservatively ignores the possibility that reports delivered during a given year will also cause additional conservation in future years. In contrast, utilities typically evaluate traditional programs to replace air conditioners, lightbulbs, and other physical capital changes by summing all expected future savings over assumed capital stock lifetimes. The reason for this difference is that until now, it was an open question whether behavioral interventions like the home energy reports would cause persistent savings. When evaluating interventions still in progress, academic studies such as Ayres, Raseman, and Shih (2012) and our own past work (Allcott 2011) have similarly calculated cost effectiveness by considering only the costs accrued and energy savings up to a given date.

Zero persistence would almost certainly be wrong, as it was the most conservative possible assumption. But how wrong was it? Table 7 presents electricity savings and cost effectiveness for the programs delivered to the dropped group in each site, using empirical estimates from Section 4.[14] To keep the results transparent and avoid extrapolating out of sample, we assume no time discounting and limit the time horizon only to the observed sample period. Of course, extrapolating into the future only magnifies the importance of persistence, and Appendix Table A9 re-creates Table 7 with linearly-extrapolated decay rates.

The top panel of Table 7 shows that under the zero persistence assumption, electricity savings are 405 to 628 kWh per household, compared to the 1004 to 1487 kWh/household actually observed. At benchmark electricity prices of \$0.10/kWh, the observed savings amount to \$100 to \$149. Under the zero persistence assumption, cost effectiveness ranges from 3.20 to 4.44 cents/kWh-hour. By contrast, the observed persistence over the sample implies a cost effectiveness of 1.35 to 1.79 cents/kWh. If applied to all households in the dropped groups, total retail electricity cost savings over the sample would be between \$470,000 and \$760,000 assuming zero persistence, whereas the

[14] The electricity savings estimates are simply the average treatment effects for each period multiplied by the length of each period. For example, post-treatment savings under observed persistence in Site 2 are ($\widehat{\tau}^3 = 0.584$ kWh/day)·(910 days).

true numbers to date are \$1.16 to \$1.80 million. These simple calculations underscore the importance of our empirical results: in each site the intervention is more than twice as effective as had often been assumed.

One reason why assumptions about persistence are so important is that they can impact whether utilities adopt behavioral interventions or other energy conservation programs. There are some benchmark cost effectiveness estimates for traditional programs, although they are controversial (Allcott and Greenstone 2012). Using nationwide data, Arimura *et al.* (2011) estimate average cost effectiveness to be about 5.0 cents/kWh when they assume a five percent discount rate. The American Council for an Energy Efficient Economy (ACEEE) 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). Under the conservative zero persistence assumption, the two-year programs are better than Arimura *et al.*'s estimates but tend to be worse than ACEEE's. This suggests that at least for some utilities, alternative energy conservation programs might be preferred. Allowing for the observed persistence, however, the two-year programs are about as good as the most optimistic estimates from the literature. This example suggests that empirical estimates of persistence could make an important difference in policymakers' program adoption decisions.

6.2 Persistence Matters for Program Design

Table 8 shows the cost effectiveness of incremental intervention at each site. The top panel shows the costs and energy savings from a one-shot intervention. The estimates are the same for each site because they are all based on the initial effect size and decay rates for the Site 2 quarterly group in Tables 2 and 3. The middle panel shows the incremental cost effectiveness of a two-year program relative to the one-shot intervention, using the treatment effects and decay rates for each site's dropped groups, as estimated in Tables 4 and 5. The bottom panel shows the incremental effects of a four-year program relative to the two-year, using each site's continued group treatment effects from Table 4 and assuming the same post-intervention decay rate as observed for the dropped treatment.[15] Because we are now considering longer interventions than in Table 7, we count the full horizon of effects until the predicted savings decay to zero. All dollar costs and electricity savings are now discounted to the beginning of the program at a five percent discount rate.

Our high-frequency estimates suggest that a one-shot intervention would have had a cost effectiveness of 4.31 cents/kWh. Extending the intervention to two years has two effects. First, more energy is saved during treatment, both because the treatment effect (the "flow" of daily savings) increases and mechanically because that flow accrues over more days. Second, more energy is saved after treatment, because the effects decay at a slower rate due to "capital stock" formation. The middle panel shows that across the three sites, these two forces contribute roughly equally to the incremental savings. The two-year intervention is much more cost effective than the one-shot intervention, both because people have not habituated after the first report and because the capital stock formation process takes time.

Extending the intervention to four years has different results. In Site 1, the continued group received biannual instead of monthly or quarterly reports, so the incremental cost is very low. The incremental savings are still substantial, and thus the incremental cost effectiveness of this reduced-intensity program design is extremely good: 0.69 cents per kilowatt-hour. In Sites 2 and 3, the continued groups' treatment intensity was unchanged over these four years. Given the assumption that the post-intervention decay rate is the same as for the two-year intervention, no additional savings accrue through this channel, and the total incremental savings are thus more limited. Extending the intervention with the same report frequency is likely to reduce cost effectiveness relative to the two-year intervention. However, it is remarkable how little cost effectiveness decreases after two years, suggesting strikingly little habituation.

These assumptions suggest a result that would be remarkable if it is true. Typically one might model an intervention as having concave effects, i.e. decreasing marginal effects. These results suggest that some additional reports are complementary to the first report, by reinforcing effects on capital stock formation, and thus have improved cost effectiveness relative to a one-shot intervention. This generates increasing marginal effects until habituation eventually causes marginal effects to decrease. However, these results rely on linearly-extrapolated decay. Since the linear de-

[15] One might hypothesize that the decay rate is slower after four years than after two, but we do not have any data that allows us to improve on our assumption.

cay model predicts that the two-year intervention is between 2.5 and 4.2 times more cost effective than the one-shot intervention, the linear model would have to substantially overstate decay for the one-shot intervention relative to the two-year intervention for the "result" to be incorrect. This is certainly an interesting question for a future experiment, either in energy conservation, exercise, or some other domain, which would randomly assign people to be discontinued from an intervention at many more points in time.

These calculations highlight how measuring the dynamics of habituation and persistence can help to optimize program design. Although further experimentation and long-term measurement will clearly be useful in refining these calculations, the basic principle suggested by Table 8 is to repeat an intervention to induce consumers to form new capital stock, and reduce treatment intensity after this has happened.[16]

7 Conclusion

We study the three longest-running sites of a large and policy-relevant behavioral intervention, the Opower home energy report. There are several striking empirical regularities. First, we show how the intervention spurs immediate energy conservation, but consumers' efforts begin to decay relatively quickly. This could be explained by multiple models, including a simple model in which the reports are "cues" that change the marginal utility of consumption, but utility returns to its un-cued state after the cue is removed. Second, the cyclical pattern of action and backsliding diminishes as people become accustomed to receiving reports. Third, we show how effects become more persistent as the intervention continues, implying that consumers gradually change their capital stock of habits or physical technologies. If the intervention stops after two years, the effects decay at only 10 to 20 percent per year. Fourth, even after two years of treatment, consumers have not fully habituated, and continued treatment still has substantial incremental effects.

There are two main policy implications. First, we demonstrate how long-run persistence can

[16] It would also be useful to vary the content of the intervention to test what generates more persistent effects. In this context, marketing weatherization programs or providing more tips about energy efficient appliances might induce additional households to make long-lasting changes to physical capital stock. In the context of exercise, Royer, Stehr, and Snydor (2013) show that combining incentives with commitment contracts causes more persistent changes in gym attendance than incentives alone.

materially change cost effectiveness, which in some cases could affect whether a policymaker should or should not adopt a program. In this case, many policymakers had made assumptions that we now see were far too conservative. Second, we show how empirical estimates of persistence and habituation can be used to optimize program design. In this setting, the optimal program design may be to continue the intervention for long enough for people to develop some new capital stock, then reduce treatment intensity. This suggests that an important part of the future research agenda on behavioral interventions is to more precisely identify when and why people form a new "capital stock" that causes persistent long-run effects.

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] Acland, Dan, and Matthew Levy (2013). "Habit Formation, Naivete, and Projection Bias in Gym Attendance." Goldman School of Public Policy Working Paper GSPP13-002 (March).

[3] Agarwal, Sumit, John Driscoll, Xavier Gabaix, and David Laibson (2013). "Learning in the Credit Card Market." Working Paper, Harvard University (April).

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

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

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

[7] 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.

[8] 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).

[9] Ashby, Kira, Hilary Forster, Bruce Ceniceros, Bobbi Wilhelm, Kim Friebel, Rachel Henschel, and Shahana Samiullah (2012). "Green with Envy: Neighbor Comparisons and Social Norms in Five Home Energy Report Programs." http://www.aceee.org/files/proceedings/2012/data/papers/0193-000218.pdf

[10] Ayres, Ian, Sophie Raseman, and Alice Shih (2012). "Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage." *Journal of Law, Economics, and Organization*, Vol. 29, No. 4, (August).

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

[12] Becker, Gary, and Kevin Murphy (1993). "A Simple Theory of Advertising as a Good or Bad." *Quarterly Journal of Economics*, Vol. 108, No. 4 (November), pages 941-964.

[13] Bernheim, B. Douglas, and Antonio Rangel (2004). "Addiction and Cue-Triggered Decision Processes." American Economic Review, Vol. 94, No. 5 (December), pages 1558-1590.

[14] 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.

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

[16] 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).

[17] 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.

[18] Cahill, Kate, and Rafael Perera (2009). "Competitions and Incentives for Smoking Cessation." *Cochrane Database of Systematic Reviews*, No 3.

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

[20] 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.

[21] 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.

[22] Coffman, Lucas, Clayton Featherstone, and Judd Kessler (2013). "Can a Social Information Nudge Affect Job Choice? Experimental Evidence from Teach For America." Working Paper, University of Pennsylvania (July).

[23] Costa, Dora, and Matthew Kahn (2013). "Energy Conservation Nudges and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment." *Journal of the European Economic Association*, forthcoming.

[24] DellaVigna, Stefano, Ulrike Malmendier, and John List (2012). "Testing for Altruism and Social Pressure in Charitable Giving." Quarterly Journal of Economics, Vol. 127, No. 1 (February), pages 1-56.

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

[26] Ferraro, Paul, and Michael Price (2013). "Using Non-Pecuniary Strategies to Influence Behavior: Evidence from a Large Scale Field Experiment." Review of Economics and Statistics, Vol. 95, No. 1 (March), pages 64–73.

[27] 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.

[28] Ferraro, Paul, and Juan Jose Miranda (2013). "Heterogeneous Treatment Effects and Causal Mechanisms in Non-Pecuniary, Information-Based Environmental Policies: Evidence from a Large-Scale Field Experiment." *Resource and Energy Economics*, Vol. 35, pages 356-379.

[29] 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.

[30] 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).

[31] Gallagher, Justin (2013). "Learning about an Infrequent Event: Evidence from Flood Insurance Take-up in the US." Working Paper, Case Western Reserve University (May).

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] Integral Analytics (2012). "Sacramento Municipal Utility District Home Energy Report Program." http://www.integralanalytics.com/ia/Portals/0/FinalSMUDHERSEval2012v4.pdf

[38] 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.

[39] Jensen, Robert (2010). "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics*, Vol. 125, No. 2, pages 515-548.

[40] 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.

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

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

[43] 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).

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

[45] Laibson, David (2001). "A Cue-Theory of Consumption." *Quarterly Journal of Economics*, Vol. 116 , No. 1, pages 81-119.

[46] 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.

[47] 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.

[48] 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.

[49] Milkman, Katherine, Julia Minson, and Kevin Volpp (2013). "Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling." Working Paper, University of Pennsylvania (August).

[50] 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.

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

[52] 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.

[53] Perry, Michael, and Sarah Woehleke (2013).“Evaluation of Pacific Gas and Electric Company’s Home Energy Report Initiative for the 2010-2012 Program.” Freeman, Sullivan, and Company (April).

[54] Rankin, Catherine, Thomas Abrams, Robert Barry, Seema Bhatnagar, David Clayton, John Colombo, Gianluca Coppola, Mark Geyer, David Glanzman, Stephen Marsland, Frances McSweeney, Donald Wilson, Chun-Fang Wu, and Richard Thompson. (2009). "Habituation Revisited: An Updated and Revised Description of the Behavioral Characteristics of Habituation." *Neurobiology of Learning and Memory*, Vol. 92, No. 2 (September), pages 135–138.

[55] Royer, Heather, Mark Stehr, and Justin Snydor (2013). "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company." Working Paper, University of California at Santa Barbara (March).

[56] 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.

[57] 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.

[58] 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.

[59] Shang, Jen, and Rachel Croson (2009). "A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods." The Economic Journal, Vol. 119, No. 540 (October), pages 1422–1439.

[60] 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.

[61] Thompson R., and W. Spencer (1966). "Habituation: A Model Phenomenon for the Study of Neuronal Substrates of Behavior." *Psychological Review*, Vol. 73, pages 16–43.

[62] U.S. Energy Information Administration (2009). "Table 2. Residential Consumption of Electricity by End Use, 2001." Available from http://www.eia.gov/emeu/recs/recs2001/enduse2001/enduse2001.html#table2

[63] 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

[64] 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.

[65] U.S. Energy Information Administration (2013). "Form EIA-861 Data Files." Available from http://www.eia.gov/electricity/data/eia861/

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

[67] 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.

[68] Walton, Gregory, and Geoffrey Cohen (2011). "A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students." *Science*, Vol. 331, No. 6023 (March 18), pages 1447-1451.

Tables

Table 1: Descriptive Statistics

Site	1	2	3
Region	Upper Midwest	Northwest	Southwest
Average January Heating Degrees	46.9	25.4	19.3
Average July Cooling Degrees	5.6	2.2	8.9
Narrative			
Baseline period begins	October, 2007	January, 2007	April, 2006
First reports generated	January and February, 2009	October, 2008	March to May, 2008
Last report generated for dropped group	January, 2011	September, 2010	June, 2010
End of sample	April, 2013	March, 2013	March, 2013
Frequency			
	60% Monthly 40% Quarterly (Randomly assigned); Continued group changed to to Biannual in 2011	72% Monthly 28% Quarterly (Randomly assigned)	71% Monthly (Heavier users) 29% Quarterly (Lighter users)
Number of Households			
Treatment: Continued	26,262	23,399	21,630
Treatment: Dropped	12,368	11,543	12,117
Control	33,524	43,945	49,290
Total	72,154	78,887	83,037
Number of Electricity Bill Observations	4,931,925	5,418,250	6,393,523
Average Usage in 2007 (kWh/day)			
(For all residential customers at the utility)	29.9	32.3	24.2
Baseline Usage (kWh/day)			
(For experimental population)			
Mean	30.1	30.3	32.1
Standard deviation	16.7	13.5	15.6
Treatment - Control	0.024	0.044	-0.450
(Standard error)	(0.124)	(0.097)	(0.51)
Dropped - Continued	-0.074	0.062	0.026
(Standard error)	(0.182)	(0.154)	(0.17)
Attrition due to Moving			
Share of households that move	0.20	0.23	0.26
Treatment - Control	-0.0043	0.0021	0.0109
(Standard error)	(0.0030)	(0.0030)	(0.0069)
Dropped - Continued	0.011	-0.0074	0.0032
(Standard error)	(0.0044)	(0.0048)	(0.0051)
Opt-Out Rate	0.020	0.019	0.026

Table 2: Effects Immediately After Report Arrival

	(1)	(2)	(3)	(4)
	Monthly Early	Monthly Later	Quarterly Early	Quarterly Later
1(Treated) · 1(Post-Arrival Period)	-0.185	-0.033	-0.197	-0.038
	(0.027)***	(0.009)***	(0.035)***	(0.022)*
1(Treated) · 1(Arrival Period)	-0.062	-0.017	-0.070	-0.005
	(0.024)***	(0.007)**	(0.028)**	(0.019)
1(Treated)	-0.451	-0.706	-0.420	-0.509
	(0.086)***	(0.059)***	(0.084)***	(0.095)***
N	8,515,691	75,217,587	19,333,058	52,418,516

Notes: This table presents estimates of Equation (3) for the monthly and quarterly groups. Within each group, the "Early" column presents estimates for the first four reports, and the "Later" column presents estimates for all reports after that. 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 3: Decays Between Reports

	(1)	(2)	(3)	(4)
	Monthly Early	Monthly Later	Quarterly Early	Quarterly Later
1(Treated) · 1(Window) · Time	4.082	0.393	0.708	0.023
	(1.302)***	(0.315)	(0.187)***	(0.140)
1(Treated)	-0.098	-0.682	-0.338	-0.532
	(0.095)	(0.058)***	(0.084)***	(0.091)***
N	8,515,691	75,217,587	19,333,058	52,418,516

Notes: This table presents estimates of Equation (4) for the monthly and quarterly groups. Within each group, the "Early" column presents estimates for the first four reports, and the "Later" column presents estimates for all reports after that. 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: Long-Run Effects

	(1)	(2)	(3)
	Site 1	Site 2	Site 3
1(Treated) · 1(Pre-Treatment)	0.016	0.004	-0.004
	(0.080)	(0.052)	(0.071)
1(Treated) · 1(1st Year)	-0.549	-0.438	-0.642
	(0.060)***	(0.062)***	(0.094)***
1(Treated) · 1(2nd Year Until Drop)	-0.852	-0.638	-0.840
	(0.073)***	(0.075)***	(0.104)***
1(Treated) · 1(Post-Drop)	-0.591	-0.584	-0.627
	(0.085)***	(0.089)***	(0.123)***
1(Continued) · 1(Pre-Treatment)	-0.079	-0.038	-0.007
	(0.085)	(0.057)	(0.055)
1(Continued) · 1(1st Year)	-0.024	-0.018	0.002
	(0.062)	(0.067)	(0.066)
1(Continued) · 1(2nd Year Until Drop)	-0.022	-0.032	-0.039
	(0.075)	(0.081)	(0.080)
1(Continued) · 1(Post-Drop)	-0.339	-0.286	-0.380
	(0.086)***	(0.096)***	(0.093)***
N	3,294,294	4,435,689	5,063,949

Table 5: Long-Run Decay Parameters

	(1)	(2)	(3)
	Site 1	Site 2	Site 3
1(Treated) · 1(1st Year)	-0.565	-0.450	-0.641
	(0.042)***	(0.043)***	(0.084)***
1(Treated) · 1(2nd Year Until Drop)	-0.925	-0.584	-0.756
	(0.062)***	(0.062)***	(0.107)***
1(Treated) · 1(Post-Drop)	-0.840	-0.631	-0.590
	(0.090)***	(0.091)***	(0.134)***
1(Continued) · 1(Post-Drop)	-0.190	-0.174	-0.305
	(0.096)**	(0.102)*	(0.114)***
1(Dropped) · 1(Post-Drop) x Time	0.178	0.113	0.090
	(0.053)***	(0.047)**	(0.046)*
1(Continued) · 1(Post-Drop) x Time	0.087	-0.061	-0.082
	(0.041)**	(0.039)	(0.036)**
N	3,294,294	4,435,689	5,063,949

Notes: Table 4 presents estimates of Equation (6), omitting the fourth and fifth lines, while Table 5 presents estimates of the same equation, omitting the second line. The outcome variable is monthly average electricity use, in kilowatt-hours per day. Standard errors are robust and clustered by household in Sites 1 and 2 and by block batch in Site 3. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table 6: Program Participation

Measure	(1) Average Savings (kWh/day)	(2) Number of Households	(3) Treatment-Control Participation Rate (%)	(4) Treatment - Control Savings (Wh/day)
Site 3				
New Appliance	1	1590	0.23 **	2.3 **
			(0.098)	(1.0)
Heating, Ventilation, and Air Conditioning	5	1481	0.028	1.4
			(0.093)	(4.7)
Insulation, Air Sealing, and other "Home Improvement"	5	890	0.18 **	9.0 **
			(0.074)	(3.7)
All Programs	3.4	3855	0.417 ***	14.2 ***
			(0.149)	(5.1)
Site 2				
Clothes Washer	0.35	1357	0.11	0.38
			(0.11)	(0.43)
Insulation	0	271	0.040	0
			(0.049)	(0)
Refrigerator Decommissioning	1.37	215	0.045	0.41
			(0.043)	(0.59)
Showerhead	0.15	214	0.025	0.09
			(0.043)	(0.10)
Duct Sealing	0	213	-0.021	0
			(0.042)	(0)
Compact Fluorescent Lightbulbs	2.27	204	0.161 ***	2.25 **
			(0.046)	(1.14)
Water Heater	1.36	144	0.035	-0.99
			(0.035)	(1.16)
Freezer Decommissioning	1.52	99	0.020	0.31
			(0.028)	(0.43)
Heat Pump	1.77	41	0.010	-0.19
			(0.019)	(0.38)
New Refrigerator	1.75	6	-0.007	-0.10
			(0.007)	(0.13)
Windows	6.69	5	0.008	0.89
			(0.008)	(0.82)
Conversion to Gas Heat	28.08	1	-0.002	-0.64
			(0.002)	(0.64)
All Programs	0.70	2481	0.36 **	2.40
			(0.14)	(2.19)

Notes: This table presents data on participation in energy conservation programs in Site 3 for April 2008 through June 2010, and in Site 2 for calendar year 2011. For readability, the coefficients in column (3) are in percentage points, ranging from 0 to 100, and the coefficients in column (4) are in Watt-hours per day instead of kilowatt-hours per day. Standard errors are robust. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table 7: In-Sample Cost Effectiveness for the Dropped Group Design

Site	1	2	3
Program cost ($/household)	17	18	20
Electricity Savings (kWh/household)			
Savings during treatment	523	405	628
(Standard Error)	(25)	(25)	(52)
Post-treatment savings	709	600	859
(Standard Error)	(43)	(47)	(91)
Total savings	1232	1004	1487
(Standard Error)	(50)	(53)	(105)
Cost Effectiveness (cents/kWh)			
Zero Persistence Assumption	3.31	4.44	3.20
(Standard Error)	(0.16)	(0.27)	(0.26)
Observed Persistence	1.40	1.79	1.35
(Standard Error)	(0.06)	(0.09)	(0.1)
Dropped Group Electricity Cost Savings ($millions)			
Zero Persistence Assumption	0.65	0.47	0.76
(Standard Error)	(0.03)	(0.03)	(0.06)
Observed Persistence	1.52	1.16	1.80
(Standard Error)	(0.06)	(0.06)	(0.13)

Notes: This table shows the results of the interventions received by the dropped groups in each site, under different assumptions about post-treatment persistence. Standard errors are calculated using the Delta method.

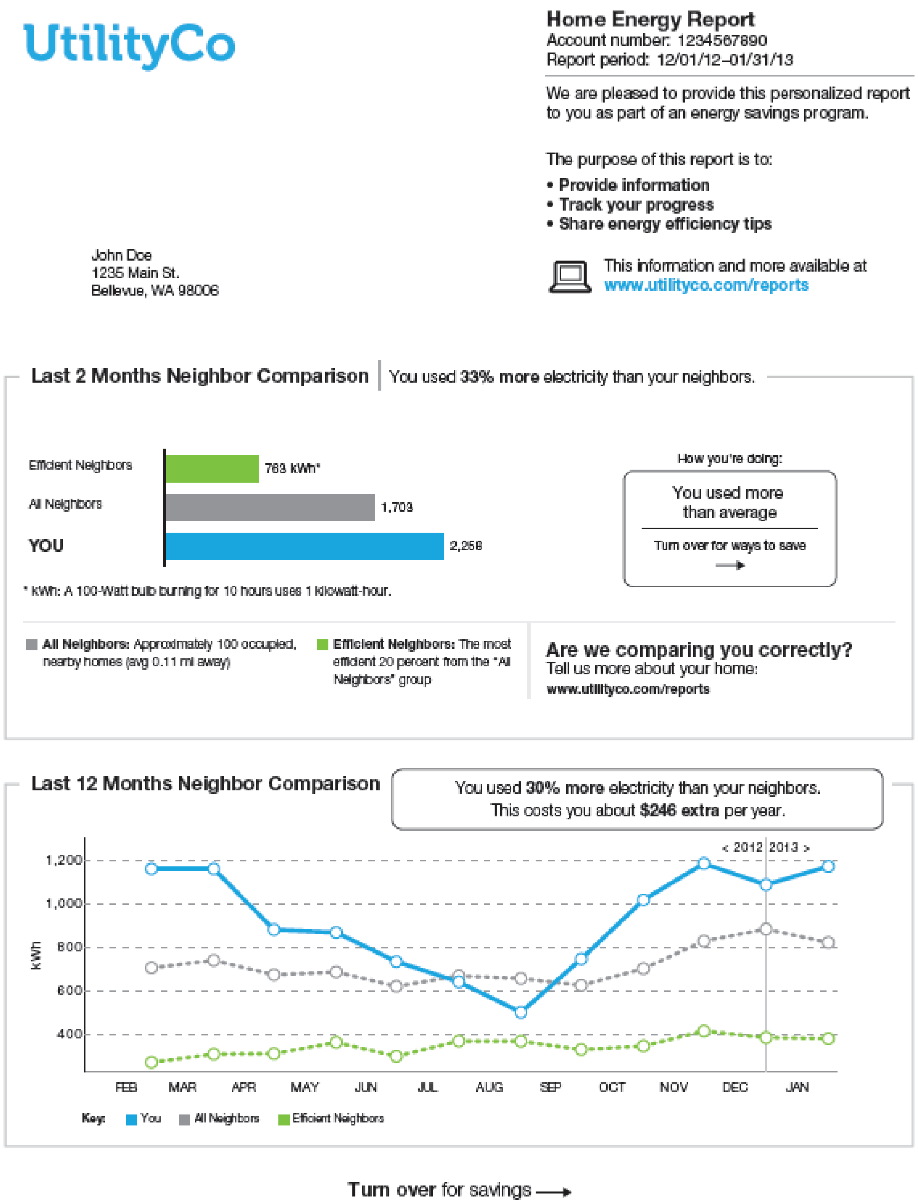
Table 8: Cost Effectiveness of Incremental Treatment

Site	1	2	3
One-Shot Intervention			
Cost ($/household)	1.00	1.00	1.00
Savings (kWh/household)	23	23	23
Cost Effectiveness (cents/kWh)	4.31	4.31	4.31
Two-Year Intervention			
Incremental Cost ($/household)	15.89	16.55	17.90
Incremental Savings (kWh/household)	1108	967	1727
Due to slower decay (kWh/household)	495	498	1057
Due to effects during treatment (kWh/household)	613	469	670
Incremental Cost Effectiveness (cents/kWh)	1.43	1.71	1.04
Overall Cost Effectiveness	1.49	1.77	1.08
Four-Year Intervention			
Incremental Cost ($/household)	4.35	15.92	17.14
Incremental Savings (kWh/household)	631	885	902
Due to slower decay (kWh/household)	0	0	0
Due to effects during treatment (kWh/household)	631	885	902
Incremental Cost Effectiveness (cents/kWh)	0.69	1.80	1.90
Overall Cost Effectiveness	1.21	1.78	1.36

Notes: This table shows the cost effectiveness of different program designs in each site. See text for details.

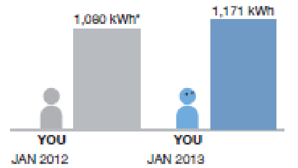
Figures

Figure 1: Home Energy Report, Front and Back



Personal Comparison

How you're doing compared to last year:



* kWh: A 100-Watt bulb burning for 10 hours uses 1 kilowatt-hour.

So far this year, you used **8% MORE** electricity than last year.

Looking for ways to save? Visit **www.utilityco.com/reports**

Action Steps | **Personalized tips chosen for your home**

Smart Purchase
An affordable way to save more

☐ **Program your thermostat**

A programmable thermostat can automatically adjust your heat or air conditioning when you're away, then return to your preferred temperature when you're home to enjoy it.

If you don't already have a programmable thermostat, look for one at your local home improvement store. For comfort and convenience, be sure to program your thermostat with energy-efficient settings.

If you need help installing or programming your thermostat, consult your manual or call the manufacturer for assistance.

SAVE UP TO $80 PER YEAR

Smart Purchase
An affordable way to save more

☐ **Check your air filters every month**

You can improve the energy efficiency of your heating and cooling systems and improve your indoor air quality by checking your filters monthly.

First, remove the filter — it usually slides right out. Next, hold the filter up to a light to see if it is clogged.

You can find an inexpensive replacement for a clogged disposable filter at your local hardware store. Check your manual for cleaning instructions if you have a permanent filter.

SAVE UP TO $45 PER YEAR

Smart Purchase
An affordable way to save more

☐ **Seal air leaks**

Gaps and cracks between the inside and outside of your home can allow heated or cooled air to escape. This forces your heating or cooling system to work harder, increases energy costs, and decreases comfort.

To find leaks, follow drafts to their source. Check where materials meet, like between the foundation and walls, the chimney and siding, and where gas and electricity lines exit your house.

Seal any small cracks you find with caulk and larger ones with polyurethane foam.

SAVE UP TO $215 PER YEAR

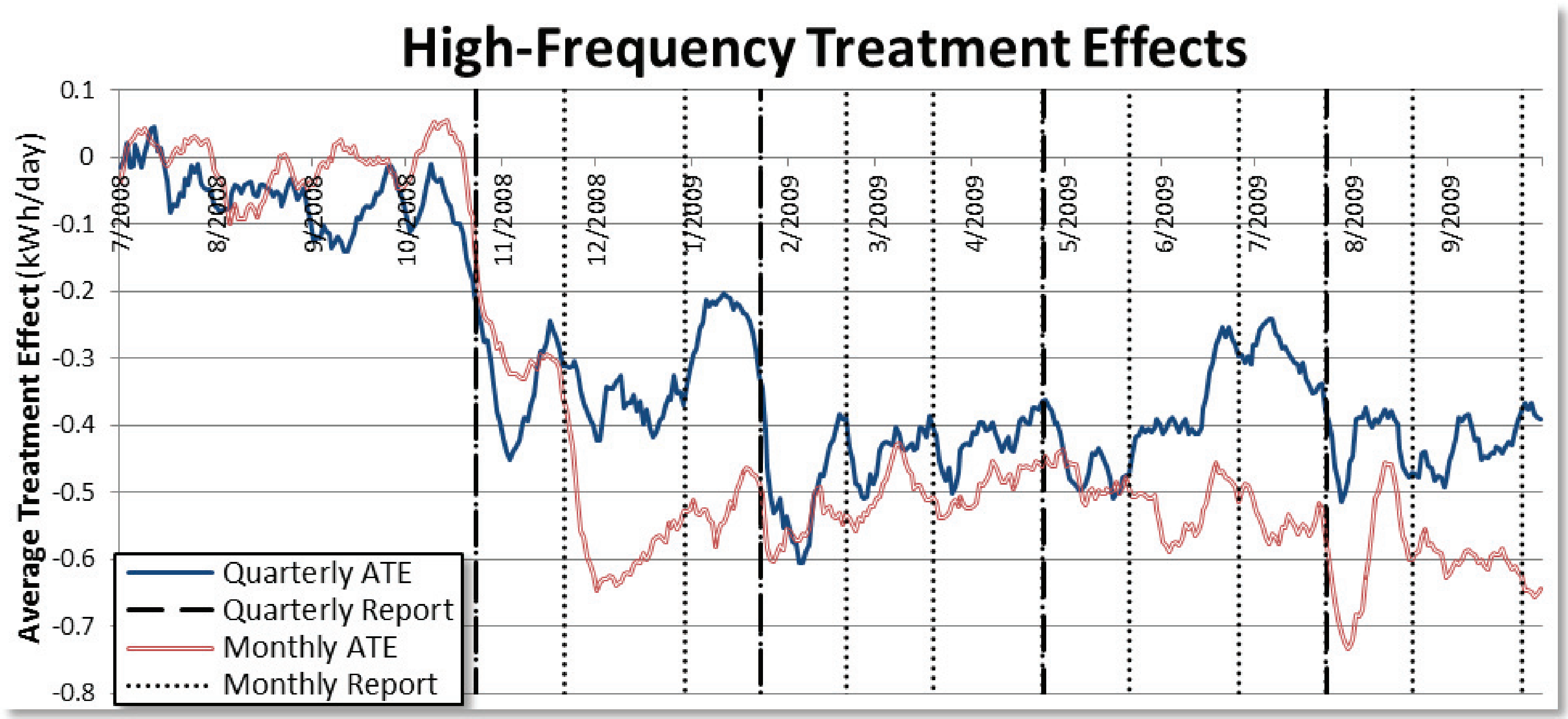
UtilityCo

runs on OPOWER®

www.utilityco.com/reports | (555) 555-5555 | energyreports@example.com

Printed on 10% post-consumer recycled paper using water-based inks.

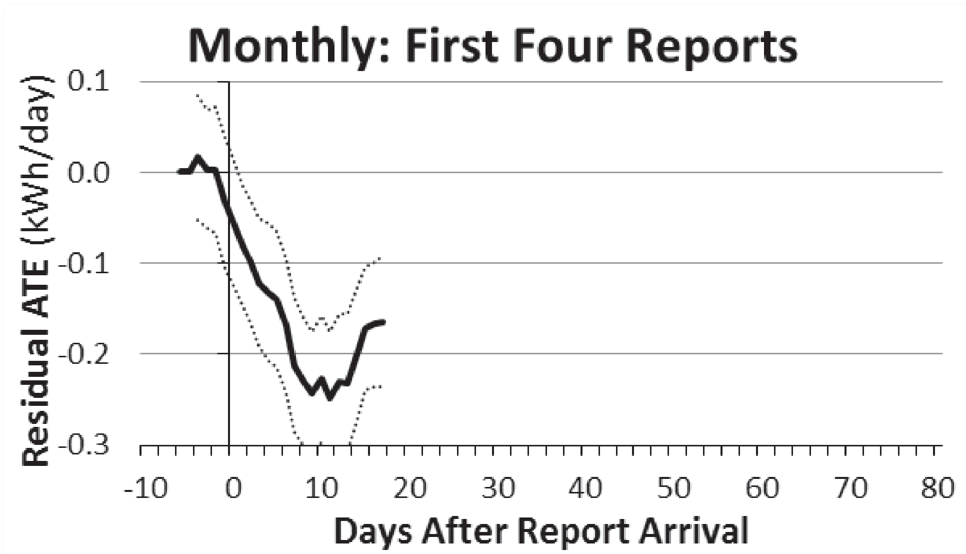
Figure 2: High-Frequency Treatment Effects



Notes: This figure plots the seven day running mean treatment effects for each day of the first year of treatment for the monthly and quarterly treatment groups, as estimated by Equation (1).

Figure 3: High-Frequency Effects in Event Time

Figure 3a: First Four Reports



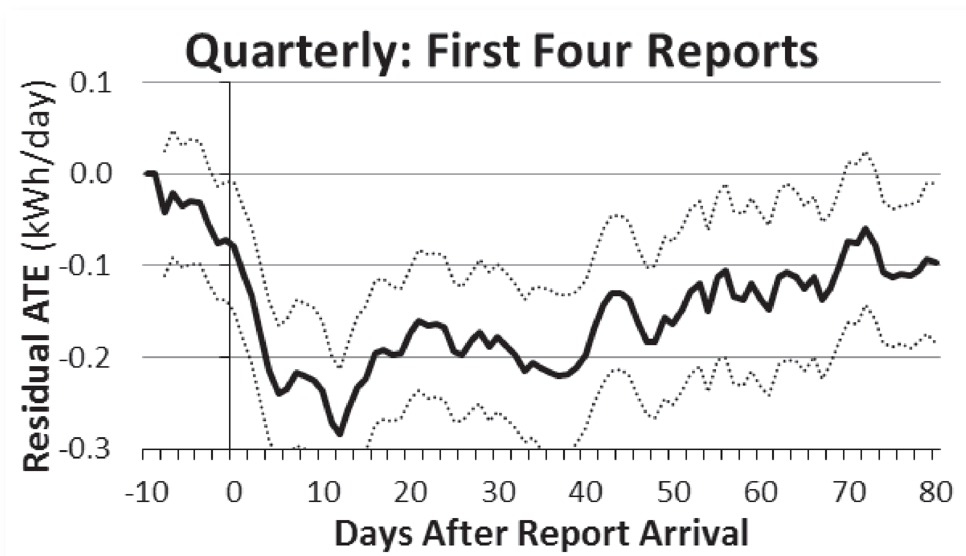
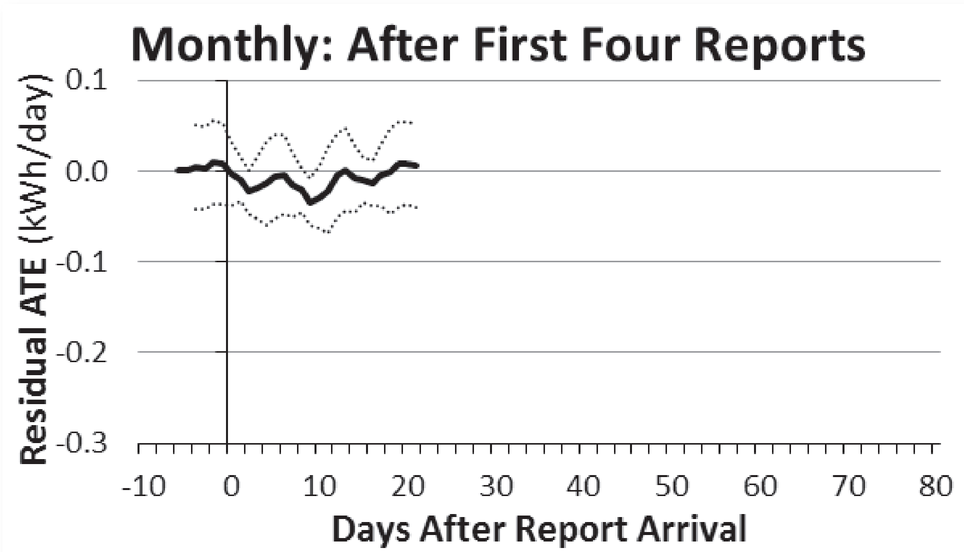
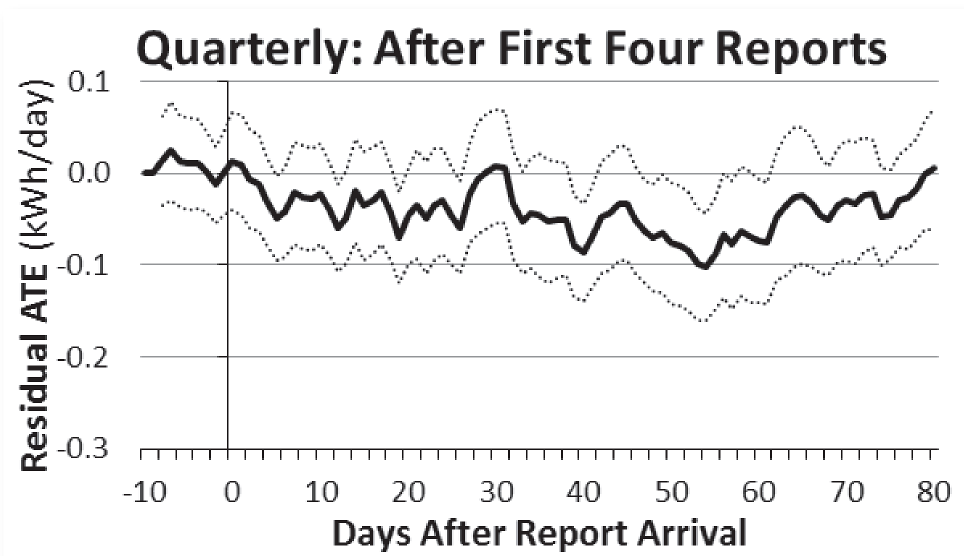


Figure 3b: After First Four Reports





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). "Residual ATE" refers to the fact that these ATEs are residual of the association of weather and report-specific controls with the treatment effect. The dotted lines reflect 90 percent confidence intervals, with robust standard errors clustered by household.

Figure 4: Long-Run Effects

Figure 4a: Site 1

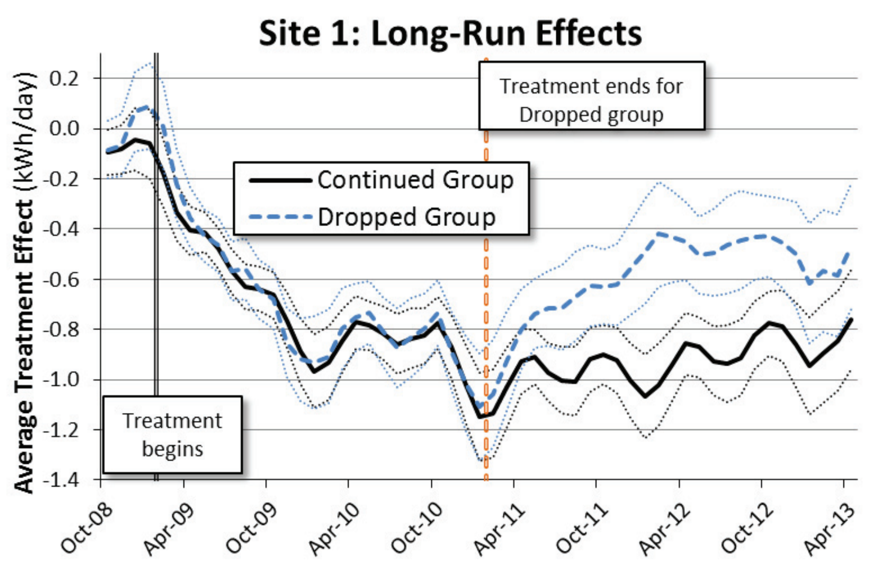


Figure 4b: Site 2

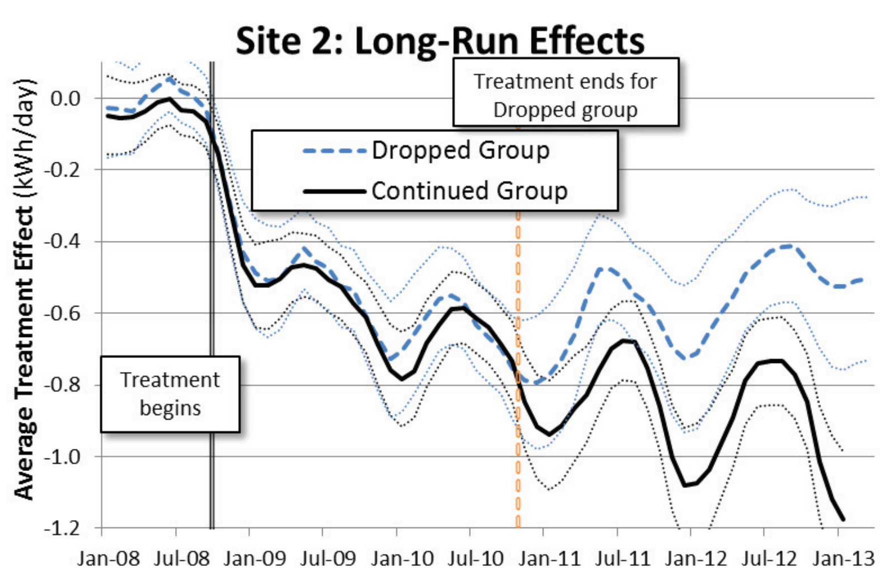
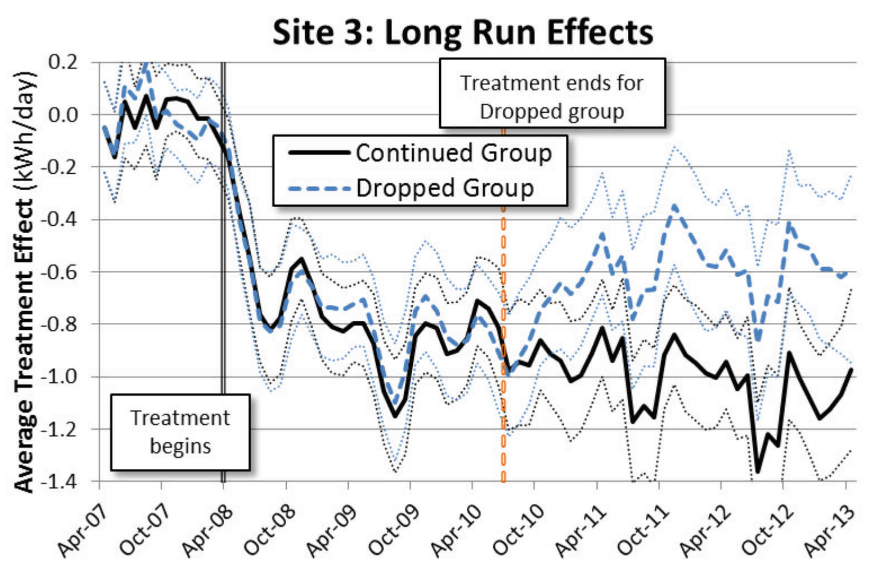


Figure 4c: Site 3



Notes: These figures plots the ATEs for each month of the sample for the continued and dropped groups, estimated by Equation (5). The dotted lines reflect 90 percent confidence intervals, with robust standard errors clustered by household in Sites 1 and 2 and by block batch in Site 3.

Appendix: For Online Publication

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

Hunt Allcott and Todd Rogers

Appendix I: Surveys of Self-Reported Actions

During the past three years, Opower has surveyed about six thousand people in treatment and control groups in six sites nationwide, including 800 people in Site 2. These are telephone surveys, and completion rates are typically between 15 and 25 percent. Respondents are asked if they have taken a series of specific actions to reduce energy use in the past 12 months. We group these actions into three major 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 AI-1 presents the results, combining data across all sites where respondents were asked about an action. Column (1) presents the share of respondents that report taking the action in the past 12 months. For many of the physical capital changes and intermittent actions, the means are too high. While our focus is on the differences between treatment and control, not the means, this does generate concern about whether the surveys yield meaningful responses.

Column (2) shows that there is little difference between treatment and control for the vast majority of actions, and the standard errors are tight enough to detect differences of two to four percentage points. There are three differences: treated households are more likely to use fans to keep cool, have a home energy audit, and participate in utility energy efficiency programs. The latter two actions involve physical capital stock changes. Audits, which are typically offered as part of the utility's energy conservation programs, often include direct installation of new compact fluorescent lightbulbs and can be gateways to other utility programs. Other utility programs often feature subsidies for energy efficient physical capital such as appliances, heating and cooling systems, and insulation. Fortunately, these are the two areas where additional administrative data are available, and we analyze these administrative data in the body of the paper.

For each of the three major categories, the first row (in bold) presents a test of whether the average probability of taking all actions in that category differs between treatment and control. This aggregation across actions gives standard errors tight enough to detect differences of one to two percentage points, but treated households are still not different in any of these three tests. Throughout Table AI-1, the failure to reject equality between treatment and control would only be further reinforced by adjusting the p-values for multiple hypothesis testing.

There are multiple interpretations of these results. First, the intervention might increase the true probabilities of taking actions, but the surveys might not pick this up if demand effects, over-reporting, non-response, or some other factor differed systematically between treatment and control. However, while the survey results should be interpreted cautiously, it is not obvious what would cause the treatment group to systematically report that they do not take actions. Second, the treatment could cause small changes in the true probabilities of taking a wide variety of actions, none of which are statistically detectable. 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 affect the "extensive margin" reported in Table AI-1, which is whether or not people take a given action, but instead changes the intensity with which people take actions they were already taking. In other words, an important impact of the intervention could to increase attention and motivation to conserve in the same ways that people were already conserving, instead of giving information about new ways to conserve.

Table AI-1: Self-Reported Actions

	(1)	(2)
"In the past twelve months, have you..."	Mean	Treatment-Control
Taken any steps to reduce energy use?	0.77	0.010 (0.012)
Repeated Actions	0.62	0.005 (0.008)
Adjusted your thermostat settings?	0.63	0.012 (0.015)
Unplugged devices and chargers?	0.65	-0.020 (0.039)
Switched off power strips or appliances when unused?	0.59	0.002 (0.014)
Turned off lights when unused?	0.96	0.005 (0.009)
Hung laundry to dry?	0.42	0.010 (0.024)
Used energy saving or sleep features on your computer?	0.56	0.008 (0.021)
Turned off computer at night?	0.65	-0.034 (0.023)
Used fans to keep cool?	0.80	0.072 (0.034)**
Physical Capital Changes	0.55	-0.002 (0.008)
Replaced incandescent light bulbs with LEDs?	0.70	0.013 (0.038)
Purchased Energy Star appliances?	0.74	0.002 (0.016)
Disposed of a second refrigerator or freezer?	0.26	-0.001 (0.015)
Installed light timers or sensors?	0.30	-0.018 (0.038)
Replaced incandescent light bulbs with CFLs?	0.81	0.000 (0.013)
Added insulation or replaced windows?	0.54	-0.039 (0.024)
Had a home energy audit?	0.19	0.057 (0.022)***
Installed a programmable thermostat?	0.79	-0.033 (0.032)
Intermittent Actions	0.62	0.006 (0.012)
Tuned up your AC system?	0.63	-0.016 (0.018)
Used a programmable thermostat?	0.59	0.009 (0.028)
Added weather-stripping or caulking around windows?	0.60	0.008 (0.018)
Cleaned or replaced heating or AC system air filters?	0.70	0.017 (0.038)
Participated in any utility energy effiency programs?	0.19	0.018 (0.010)*
Total number of surveys	5856	

Notes: This table presents survey data on self-reported energy conservation actions. Robust standard errors. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Appendix Tables

Notes for Tables A1-A4

Tables A1-A4 present alternative estimates of Tables 2 and 3. Tables A1 and A2 present alternative estimates of Equation (3) for the first four reports and all later reports, respectively. Tables A3 and A4 similarly present alternative estimates of Equation (4) for the first four reports and all later reports. All tables include the bimonthly group as well as the monthly and quarterly groups.

Within each table, there are two panels. In the first, the left column excludes weather controls, while the right column exactly replicates the estimates in the body of the paper, also reporting the estimated weather coefficients. In the second, the left column 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 high-usage observations appear to be correct, not measurement errors. However, they implicitly receive significant weight in the OLS estimation, so a small number of high-usage households could in theory drive the results. The right column replaces the original $\mathbf{M}_{it}$ with 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 function was based on inspection of the relationship between ATEs and degree days for this site.

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 A1: Robustness Checks for Table 2, First Four Reports

	(1)	(2)	(3)	(4)	(5)	(6)
	Monthly Base	Monthly Weather	Quarterly Base	Quarterly Weather	Bimonthly Base	Bimonthly Weather
1(Treated) · 1(Post-Arrival Period)	-0.172	-0.185	-0.201	-0.197	-0.129	-0.152
	(0.030)***	(0.027)***	(0.041)***	(0.035)***	(0.038)***	(0.036)***
1(Treated) · 1(Arrival Period)	-0.062	-0.062	-0.067	-0.070	-0.047	-0.043
	(0.024)***	(0.024)***	(0.029)**	(0.028)**	(0.033)	(0.033)
1(Treated)	-0.534	-0.451	-0.391	-0.420	-0.366	-0.276
	(0.065)***	(0.086)***	(0.067)***	(0.084)***	(0.059)***	(0.106)***
1(Treated) · Heating Degrees		-0.004		0.002		-0.009
		(0.004)		(0.005)		(0.008)
Heating Degrees		0.038		0.020		0.083
		(0.016)**		(0.014)		(0.011)***
1(Treated) · Cooling Degrees				0.000		-0.031
				(0.010)		(0.019)*
Cooling Degrees				0.281		0.016
				(0.019)***		(0.027)
N	8,515,691	8,515,691	19,333,058	19,333,058	9,609,303	9,609,303

	(1)	(2)	(3)	(4)	(5)	(6)
	Monthly Outliers	Monthly Full M	Quarterly Outliers	Quarterly Full M	Bimonthly Outliers	Bimonthly Full M
1(Treated) · 1(Post-Arrival Period)	-0.183	-0.185	-0.190	-0.193	-0.154	-0.160
	(0.027)***	(0.027)***	(0.036)***	(0.035)***	(0.034)***	(0.036)***
1(Treated) · 1(Arrival Period)	-0.061	-0.061	-0.070	-0.069	-0.032	-0.052
	(0.024)**	(0.023)***	(0.028)**	(0.028)**	(0.031)	(0.029)*
1(Treated)	-0.430	-0.580	-0.413	-0.408	-0.228	-0.379
	(0.086)***	(0.114)***	(0.083)***	(0.091)***	(0.101)**	(0.103)***
1(Treated) · Heating Degrees	-0.004		0.002		-0.011	
	(0.004)		(0.005)		(0.007)	
Heating Degrees	0.039		0.021		0.089	
	(0.016)**		(0.014)		(0.010)***	
1(Treated) · Cooling Degrees			0.000		-0.018	
			(0.010)		(0.016)	
Cooling Degrees			0.279		-0.016	
			(0.019)***		(0.022)	
N	8,514,078	8,515,691	19,330,176	19,333,058	9,589,391	9,609,303

Table A2: Robustness Checks for Table 2, Later Reports

	(1) Monthly Base	(2) Monthly Weather	(3) Quarterly Base	(4) Quarterly Weather	(5) Bimonthly Base	(6) Bimonthly Weather
1(Treated) · 1(Post-Arrival Period)	-0.032	-0.033	-0.045	-0.038	-0.211	-0.230
	(0.009)***	(0.009)***	(0.022)**	(0.022)*	(0.060)***	(0.061)***
1(Treated) · 1(Arrival Period)	-0.015	-0.017	-0.010	-0.005	-0.025	-0.129
	(0.007)**	(0.007)**	(0.019)	(0.019)	(0.047)	(0.049)***
1(Treated)	-0.801	-0.706	-0.657	-0.509	-0.645	-0.048
	(0.058)***	(0.059)***	(0.092)***	(0.095)***	(0.089)***	(0.143)
1(Treated) · Heating Degrees		-0.006		-0.010		-0.034
		(0.003)**		(0.005)**		(0.008)***
1(Treated) · Cooling Degrees		-0.017		-0.007		-0.050
		(0.007)**		(0.013)		(0.028)*
Heating Degrees		0.004		0.007		0.082
		(0.010)		(0.011)		(0.013)***
Cooling Degrees		0.090		0.023		0.463
		(0.012)***		(0.015)		(0.029)***
N	75,217,587	75,217,587	52,418,516	52,418,516	19,554,914	19,554,914

	(1) Monthly Outliers	(2) Monthly Full M	(3) Quarterly Outliers	(4) Quarterly Full M	(5) Bimonthly Outliers	(6) Bimonthly Full M
1(Treated) · 1(Post-Arrival Period)	-0.030	-0.032	-0.036	-0.035	-0.217	-0.233
	(0.009)***	(0.009)***	(0.022)	(0.022)	(0.060)***	(0.061)***
1(Treated) · 1(Arrival Period)	-0.016	-0.017	-0.004	-0.006	-0.129	-0.137
	(0.007)**	(0.007)**	(0.019)	(0.019)	(0.049)***	(0.050)***
1(Treated)	-0.696	-0.762	-0.509	-0.555	-0.041	-0.144
	(0.059)***	(0.063)***	(0.094)***	(0.115)***	(0.138)	(0.141)
1(Treated) · Heating Degrees	-0.007		-0.009		-0.034	
	(0.003)**		(0.005)*		(0.008)***	
1(Treated) · Cooling Degrees	-0.017		-0.007		-0.042	
	(0.007)**		(0.013)		(0.026)	
Heating Degrees	0.004		0.007		0.082	
	(0.010)		(0.011)		(0.012)***	
Cooling Degrees	0.089		0.022		0.444	
	(0.012)***		(0.015)		(0.028)***	
N	75,201,504	75,217,587	52,409,856	52,418,516	19,513,453	19,554,914

Table A3: Robustness Checks for Table 3, First Four Reports

	(1)	(2)	(3)	(4)	(5)	(6)
	Monthly Base	Monthly Weather	Quarterly Base	Quarterly Weather	Bimonthly Base	Bimonthly Weather
1(Treated) · 1(Window) · Time	1.356	4.082	0.706	0.708	1.012	0.948
	(1.265)	(1.302)***	(0.195)***	(0.187)***	(0.439)**	(0.426)**
1(Treated)	-0.413	-0.098	-0.346	-0.338	-0.408	-0.242
	(0.064)***	(0.095)	(0.071)***	(0.084)***	(0.067)***	(0.104)**
1(Treated) · Heating Degrees		-0.013		-0.000		-0.011
		(0.004)***		(0.004)		(0.008)
Heating Degrees		0.042		0.021		0.085
		(0.016)***		(0.014)		(0.011)***
1(Treated) · Cooling Degrees				-0.004		-0.028
				(0.012)		(0.019)
Cooling Degrees				0.282		0.015
				(0.019)***		(0.027)
N	8,515,691	8,515,691	19,333,058	19,333,058	9,609,303	9,609,303

	(1)	(2)	(3)	(4)	(5)	(6)
	Monthly Outliers	Monthly Full M	Quarterly Outliers	Quarterly Full M	Bimonthly Outliers	Bimonthly Full M
1(Treated) · 1(Window) · Time	4.061	4.476	0.674	0.697	0.744	0.884
	(1.290)***	(1.309)***	(0.185)***	(0.187)***	(0.392)*	(0.417)**
1(Treated)	-0.083	-0.544	-0.325	-0.319	-0.222	-0.342
	(0.094)	(0.120)***	(0.083)***	(0.089)***	(0.101)**	(0.101)***
1(Treated) · Heating Degrees	-0.014		-0.001		-0.011	
	(0.004)***		(0.004)		(0.007)	
Heating Degrees	0.043		0.021		0.089	
	(0.016)***		(0.014)		(0.010)***	
1(Treated) · Cooling Degrees			-0.004		-0.010	
			(0.012)		(0.016)	
Cooling Degrees			0.280		-0.020	
			(0.019)***		(0.022)	
N	8,514,078	8,515,691	19,330,176	19,333,058	9,589,391	9,609,303

Table A4: Robustness Checks for Table 3, After First Four Reports

	(1)	(2)	(3)	(4)	(5)	(6)
	Monthly Base	Monthly Weather	Quarterly Base	Quarterly Weather	Bimonthly Base	Bimonthly Weather
1(Treated) · 1(Window) · Time	0.333	0.393	0.017	0.023	0.022	0.134
	(0.322)	(0.315)	(0.141)	(0.140)	(0.539)	(0.536)
1(Treated)	-0.777	-0.682	-0.606	-0.532	-0.551	-0.080
	(0.056)***	(0.058)***	(0.087)***	(0.091)***	(0.089)***	(0.141)
1(Treated) · Heating Degrees		-0.007		-0.006		-0.030
		(0.003)**		(0.004)		(0.008)***
1(Treated) · Cooling Degrees		-0.014		-0.005		-0.044
		(0.008)*		(0.013)		(0.029)
Heating Degrees		0.004		0.007		0.080
		(0.010)		(0.011)		(0.013)***
Cooling Degrees		0.089		0.023		0.460
		(0.012)***		(0.015)		(0.029)***
N	75,217,587	75,217,587	52,418,516	52,418,516	19,554,914	19,554,914

	(1)	(2)	(3)	(4)	(5)	(6)
	Monthly Outliers	Monthly Full M	Quarterly Outliers	Quarterly Full M	Bimonthly Outliers	Bimonthly Full M
1(Treated) · 1(Window) · Time	0.312	0.449	0.037	0.023	-0.115	0.275
	(0.313)	(0.312)	(0.140)	(0.141)	(0.526)	(0.528)
1(Treated)	-0.672	-0.725	-0.532	-0.509	-0.070	-0.144
	(0.058)***	(0.061)***	(0.091)***	(0.104)***	(0.136)	(0.141)
1(Treated) · Heating Degrees	-0.007		-0.005		-0.031	
	(0.003)**		(0.004)		(0.008)***	
1(Treated) · Cooling Degrees	-0.014		-0.004		-0.037	
	(0.008)*		(0.013)		(0.026)	
Heating Degrees	0.004		0.007		0.080	
	(0.010)		(0.011)		(0.012)***	
Cooling Degrees	0.088		0.021		0.441	
	(0.012)***		(0.015)		(0.028)***	
N	75,201,504	75,217,587	52,409,856	52,418,516	19,513,453	19,554,914

Table A5: Placebo Report Arrivals

	(1)	(2)
	Base	Weather
1(Treated) · 1(Arrival Period)	-0.001	-0.007
	(0.016)	(0.015)
1(Treated) · 1(Post-Arrival Period)	0.011	-0.004
	(0.019)	(0.019)
1(Treated)	-0.671	-0.489
	(0.095)***	(0.093)***
1(Treated) · Heating Degrees		-0.012
		(0.005)**
1(Treated) · Cooling Degrees		-0.008
		(0.013)
Heating Degrees		0.007
		(0.011)
Cooling Degrees		0.023
		(0.015)
N	52,418,516	52,418,516

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 A6: Persistence by Subgroup

	(1)	(2)	(3)	(4)	(5)	(6)
	Site 1 Levels	Site 1 Decays	Site 2 Levels	Site 2 Decays	Both Sites Levels	Both Sites Decays
1(Dropped)	-0.601	-0.832	-0.650	-0.805	-0.626	-0.812
	(0.090)***	(0.097)***	(0.101)***	(0.108)***	(0.068)***	(0.072)***
1(Dropped) · 1(Quarterly)	0.077	0.324	0.233	0.293	0.169	0.290
	(0.188)	(0.202)	(0.177)	(0.190)	(0.131)	(0.139)**
1(Dropped) · Baseline Usage	-0.283	-0.477	-0.632	-0.561	-0.476	-0.495
	(0.163)*	(0.184)***	(0.142)***	(0.154)***	(0.107)***	(0.119)***
1(Dropped) · 1(Post-Drop) x Time		0.211		0.131		0.164
		(0.057)***		(0.054)**		(0.040)***
Quarterly Decay Difference		-0.232		-0.050		-0.109
		(0.122)*		(0.093)		(0.075)
Baseline Usage Decay Difference		0.183		-0.061		0.017
		(0.111)*		(0.081)		(0.067)
N	956,848	956,848	1,387,473	1,387,473	2,344,321	2,344,321

Notes: This table presents the estimates of Equation (6), allowing γ 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 sample is limited to the post-drop period and includes only dropped and control group households. 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 A7: Table 4 with Balanced Panel

	(1)	(2)	(3)
	Site 1	Site 2	Site 3
1(Treated) · 1(Pre-Treatment)	0.033	-0.039	-0.010
	(0.088)	(0.055)	(0.073)
1(Treated) · 1(1st Year)	-0.565	-0.515	-0.652
	(0.065)***	(0.065)***	(0.093)***
1(Treated) · 1(2nd Year Until Drop)	-0.882	-0.701	-0.859
	(0.076)***	(0.078)***	(0.102)***
1(Treated) · 1(Post-Drop)	-0.605	-0.554	-0.618
	(0.088)***	(0.093)***	(0.126)***
1(Continued) · 1(Pre-Treatment)	-0.045	0.023	-0.038
	(0.093)	(0.060)	(0.055)
1(Continued) · 1(1st Year)	-0.018	0.070	-0.110
	(0.067)	(0.071)	(0.066)*
1(Continued) · 1(2nd Year Until Drop)	0.005	0.045	-0.070
	(0.078)	(0.085)	(0.075)
1(Continued) · 1(Post-Drop)	-0.329	-0.299	-0.418
	(0.089)***	(0.100)***	(0.095)***
N	2,924,939	3,800,809	4,226,607

Notes: This table presents the estimates of Equation (6), omitting the third and fourth lines, with the sample limited to households that never move. It replicates Table 4, except with a balanced panel. The outcome variable is monthly average electricity use, in kilowatt-hours per day. Standard errors are robust and clustered by household in Sites 1 and 2 and by block batch in Site 3. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

Table A8: Robustness Checks for Table 5

	(1)	(2)	(3)	(4)	(5)	(6)
	Site 1 No Weather	Site 1 Balanced	Site 2 No Weather	Site 2 Balanced	Site 3 No Weather	Site 3 Balanced
1(Treated) · 1(1st Year)	-0.565	-0.578	-0.450	-0.469	-0.641	-0.722
	(0.042)***	(0.045)***	(0.043)***	(0.045)***	(0.084)***	(0.083)***
1(Treated) · 1(2nd Year Until Drop)	-0.867	-0.923	-0.659	-0.597	-0.865	-0.783
	(0.053)***	(0.065)***	(0.052)***	(0.065)***	(0.092)***	(0.106)***
1(Treated) · 1(Post-Drop)	-0.786	-0.826	-0.718	-0.595	-0.725	-0.551
	(0.090)***	(0.093)***	(0.095)***	(0.091)***	(0.129)***	(0.130)***
1(Continued) · 1(Post-Drop)	-0.243	-0.188	-0.087	-0.183	-0.163	-0.422
	(0.091)***	(0.098)*	(0.105)	(0.103)*	(0.106)	(0.108)***
1(Dropped) · 1(Post-Drop) x Time	0.176	0.161	0.114	0.106	0.086	0.080
	(0.053)***	(0.050)***	(0.047)**	(0.044)**	(0.045)*	(0.044)*
1(Continued) · 1(Post-Drop) x Time	0.091	0.067	-0.062	-0.067	-0.079	-0.044
	(0.041)**	(0.039)*	(0.039)	(0.036)*	(0.034)**	(0.034)
N	3,294,294	2,924,939	4,435,689	3,800,809	5,063,949	4,226,607
Weather Controls	No	Yes	No	Yes	No	Yes

Notes: This table presents the estimates of Equation (6), omitting the second line. These are robustness checks for Table 5. Within each site, the left column excludes weather controls, while the right column limits the sample to households that never move. The outcome variable is monthly average electricity use, in kilowatt-hours per day. Standard errors are robust and clustered by household in Sites 1 and 2 and by block batch in Site 3. *, **, ***: Statistically significant with 90, 95, and 99 percent confidence, respectively.

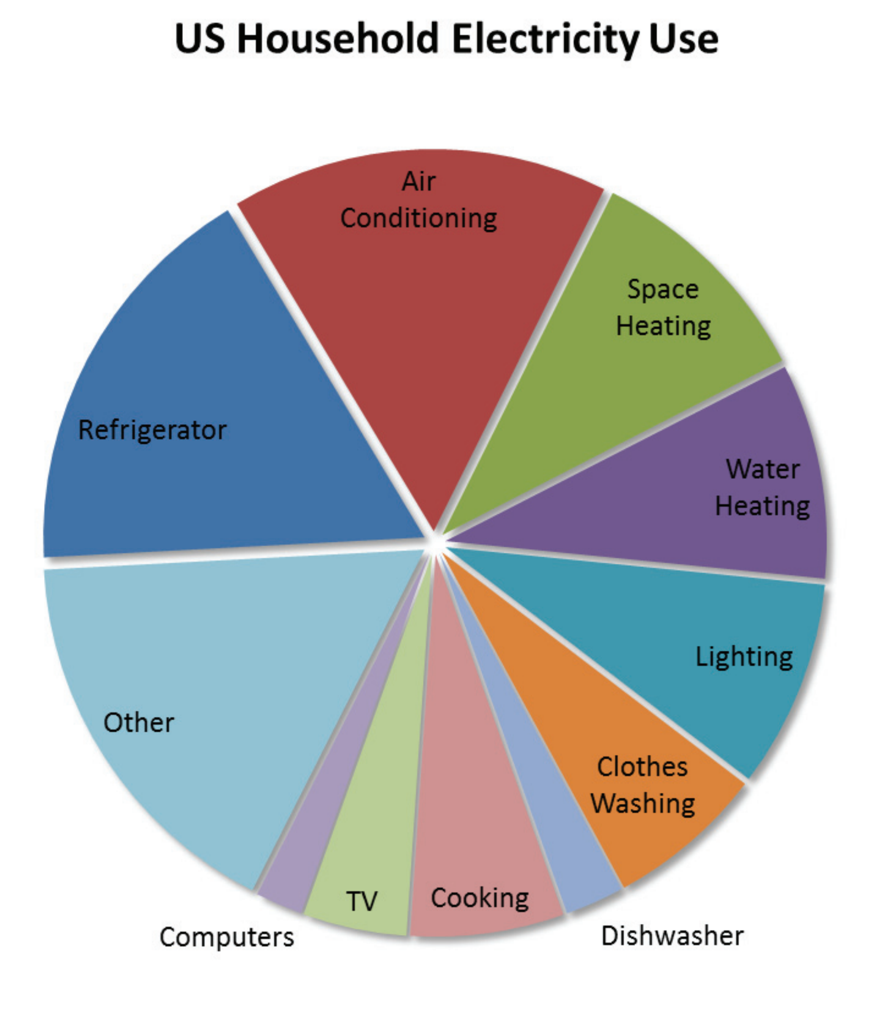
Table A9: Lifetime Extrapolated Cost Effectiveness for the Dropped Group Design

Site	1	2	3
Program cost ($/household)	17	18	20
Electricity Savings (kWh/household)			
Savings during treatment	523	405	628
(Standard Error)	(25)	(25)	(52)
Post-treatment savings	889	798	1491
(Standard Error)	(54)	(63)	(159)
Total savings	1412	1203	2119
(Standard Error)	(60)	(68)	(167)
Cost Effectiveness (cents/kWh)			
Zero Persistence Assumption	3.31	4.44	3.20
(Standard Error)	(0.16)	(0.27)	(0.26)
Observed Persistence	1.23	1.49	0.95
(Standard Error)	(0.05)	(0.08)	(0.07)
Dropped Group Electricity Cost Savings ($millions)			
Zero Persistence Assumption	0.65	0.47	0.76
(Standard Error)	(0.03)	(0.03)	(0.06)
Observed Persistence	1.75	1.39	2.57
(Standard Error)	(0.07)	(0.08)	(0.2)

Notes: This re-creates Table 7 over the projected lifetime of effects. Savings are extrapolated using the estimated linear decay parameter $\widehat{\delta}^{LR}$. Standard errors are calculated using the Delta method.

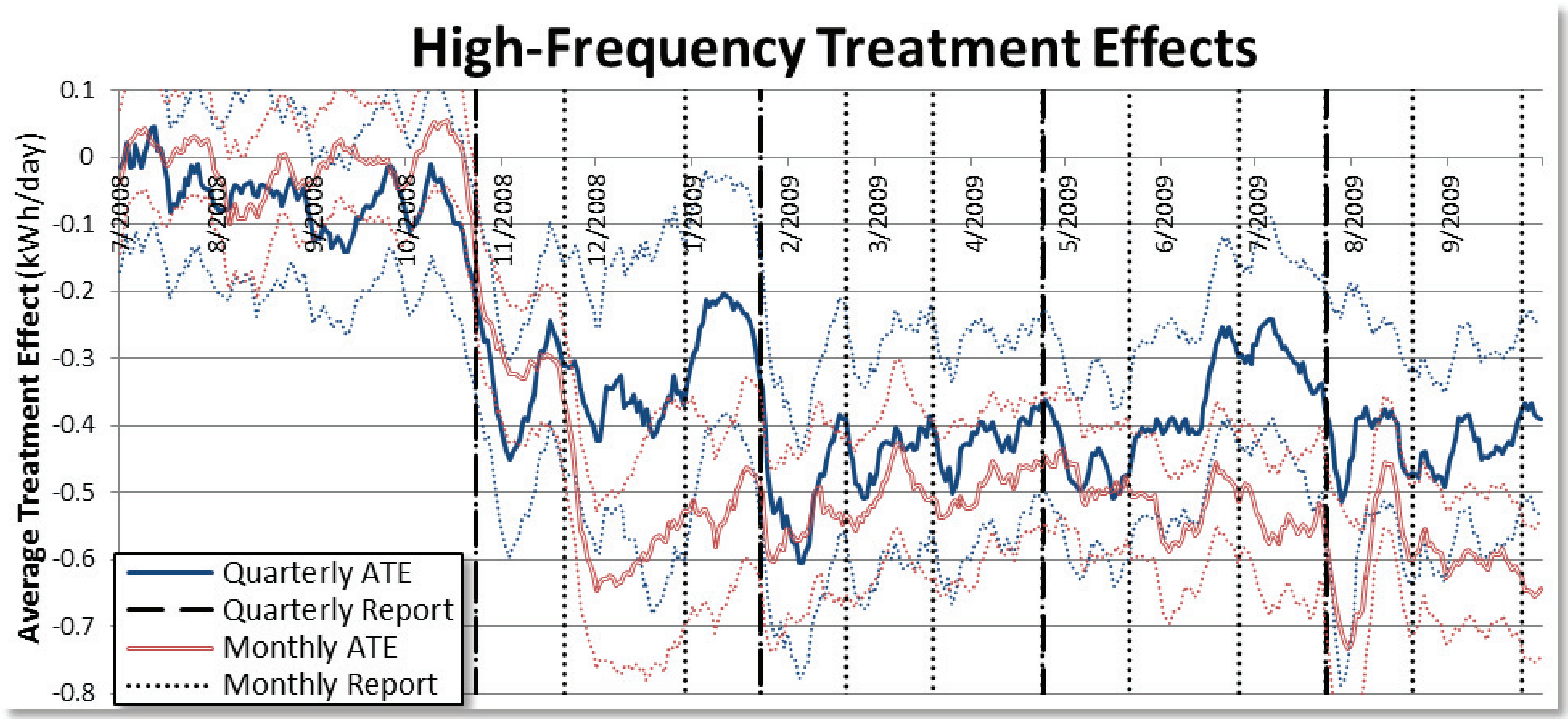
Appendix Figures

Figure A1: Breakdown of Household Electricity Use



Notes: This figure shows the breakdown of electricity use for the average American household in 2001, the most recent year for which detailed figures are available. Source: U.S. Energy Information Administration (2009).

Figure A2: High-Frequency Effects with Standard Errors



Notes: This figure plots the seven day running mean treatment effects for each day of the first year of treatment for the monthly and quarterly treatment groups, as estimated by Equation (1).This replicates Figure 2 but also includes standard errors.