Spillovers from Behavioral Interventions: Experimental Evidence from Water and Energy Use[1]

Katrina Jessoe,[2] Gabriel E. Lade,[3] Frank Loge,[4] and Edward Spang[5]

June 18, 2020

Abstract

This paper provides experimental evidence that behavioral interventions spill over to untreated sectors by altering consumer choice. We use a randomized controlled trial and high-frequency data to test the effect of social norms messaging about residential water use on electricity consumption. Messaging appears to induce a small reduction in summertime electricity use. Empirical tests and household survey data support the hypothesis that this nudge alters electricity choices. An engineering simulation suggests that complementarities between appliances that use water and electricity explain roughly a quarter of the electricity reduction. Incorporating the cross-sectoral spillover increases the net-benefits of the intervention substantially.

Keywords: Social Norms; Spillovers; Randomized Controlled Trial; Energy Use; Water Use

JEL: C93, D91, L95, Q40

---

[1]We gratefully acknowledge the staff at WaterSmart and Burbank Water and Power for their important roles in developing and implementing this project. We thank Lucas Davis, Paul Ferraro, Meredith Fowlie, Garth Heutel, Ashley Langer, Kendra Olmos, Kevin Novan, Michael Price, Ivan Rudik, Aaron Smith, Casey Wichman, and seminar participants at Resources for the Future, UCI, the University of Illinois, Duke, NC State and the AEA, Advances with Field Experiments, AERE and AAEA conferences for valuable comments. Natalie Popovich and Jack Gregory provided excellent research assistance. Research support for this project was provided by the Hellman Foundation, Southern California Gas, and the E2e Project. All errors are our own.

[2]Department of Agricultural and Resource Economics, University of California, Davis, One Shields Ave, Davis, CA 95616; Phone: (530) 752-6977; Email: `kkjessoe@ucdavis.edu`

[3]Department of Economics, Macalester College, Saint Paul, MN 55105; Phone: (651) 696-6737; Email: `glade@macalester.edu`

[4]Department of Civil and Environmental Engineering, University of California, Davis, One Shields Ave, Davis, CA 95616; Phone: (530) 754-2297; Email: `fjloge@ucdavis.edu`

[5]Department of Food Science and Technology, University of California, Davis, One Shields Ave, Davis, CA 95616; Phone: (530) 754-5447; Email: `esspang@ucdavis.edu`

1 Introduction

Firms and governments increasingly rely on behavioral instruments that seek to alter consumer behavior without changing prices or choice sets (The Economist, 2017). The widespread adoption of behavioral policies is informed by an economics and social psychology literature documenting their ability to alter behavior. Studies demonstrate the importance of default settings as a lever to increase participation in retirement savings, organ donations, and dynamic electricity pricing programs (Fowlie et al., 2017; Johnson and Goldstein, 2003; Madrian and Shea, 2001). Altering price salience changes behavior. Individuals are less responsive to taxes excluded from posted prices, shipping fees on eBay, earning statements appearing earlier in the week, and automatic bill payments (Chetty, Looney, and Kroft, 2009; Brown, Hossain, and Morgan, 2010; DellaVigna and Pollet, 2009; Sexton, 2015). In applications ranging from personal savings decisions to energy conservation, commitment devices and social comparisons have been deployed to alter behavior (Ashraf, Karlan, and Yin, 2006; Allcott, 2011). Common to all these studies is a focus on the evaluation of an intervention on a *targeted* outcome or goal.

Less well understood is whether, and to what extent, behavioral interventions extend beyond the targeted outcome, affecting unexposed margins of behavior for treated individuals. The possibility that interventions spill over is not novel. Studies have investigated if exposure to conditional cash transfers, deworming, and therapy affect outcomes of untreated agents (Angelucci and Giorgi, 2009; Miguel and Kremer, 2003; Fletcher and Marksteiner, 2017). However, the literature has paid less attention to whether exposure to treatment affects untreated margins of behavior for treated individuals, a spillover we define as a "cross-sectoral" spillover. One recent exception is Carpenter and Lawler (2019), who show that mandatory middle school Tdap vaccination programs increase take-up of Tdap boosters and non-mandated HPV vaccines. If behavioral policies extend beyond their intended objective or sector, this meaningfully affects their scope, cost-effectiveness, welfare impacts, and potentially the sign of a simple cost-benefit analysis. This paper uses a randomized controlled trial to examine whether exposure to a nudge influences behavior outside of the targeted outcome, and explores potential mechanisms underpinning the presence of cross-sectoral spillovers.

Our empirical context – the effect of social norms messaging on residential water and electricity consumption – presents an ideal and policy-relevant setting to probe for cross-sectoral spillovers. Social norms are perhaps the most frequently deployed and studied nudge and have been shown to alter consumer behavior in charitable giving, retirement savings, and voting behavior (Beshears et al., 2015; Croson and Shang, 2008; Duflo and Saez, 2003; Frey and Meier, 2013; Gerber and Rogers, 2009). Nowhere has the study of social norms been more widespread than in the context of energy and water use (Allcott, 2011; Ayres, Raseman, and Shih, 2013; Costa and Kahn, 2013; Brent, Cook, and Olsen, 2015; Dolan and Metcalfe, 2015; Ferraro and Price, 2013). Public and private utilities have responded by incorporating social pressure as a critical component of their water and energy conservation efforts. A notably absent feature in the deployment and evaluation of these programs is the possibility that social norms spill over to choices beyond the targeted outcome.

This paper deploys a randomized controlled trial in a service territory with high-frequency water and electricity usage data to investigate the effects of home water reports (HWR) on electricity use. HWRs compare a household's water use to that of similar neighbors and provide conservation tips and information about water use. Importantly, the messaging neither targets nor mentions electricity use or conservation. In our experimental design, we randomly assigned approximately 4,500 households to receive a HWR, a treatment we refer to as "WaterSmart." The roll-out of the experiment in an area with high-frequency water and electricity use data provides unique opportunities to study hourly patterns in response to treatment and to test several hypotheses underpinning the presence of cross-sectoral spillovers.

A central result is that the water conservation instrument led to a short-lived, but economically meaningful reduction in electricity use. We find an electricity conservation effect of 1 to 2%, though the effect does not persist beyond the first four treatment months. The finding is novel not only because a water conservation instrument affects energy use, but because the magnitude of the response rivals that from the deployment of home energy reports (HERs) that focus exclusively on electricity conservation (Allcott, 2011; Ayres, Raseman, and Shih, 2013). A second noteworthy element is when the reductions occur. Our estimated treatment effects are concentrated in the summer between 3 PM and 7 PM, a period that includes peak electricity demand hours both for the households in our study and the electric grid as a whole. This increases the value of the spillovers since they occur during the hours when wholesale electricity prices and emissions are typically highest.

To understand the mechanisms underpinning this summertime cross-sectoral spillover, we first outline a model that illustrates two channels through which HWRs may impact household energy use, and then empirically test for the presence of these channels. One candidate mechanism is that the cross-sectoral spillover arises from a mechanical relationship (i.e., complementarity) between appliances, such as dishwashers, washers and dryers, and the operation of pools and spas, that use both energy and water. For example, if households respond to treatment by washing one less load of laundry, this leads to a reduction in electricity use along two margins – less energy is used running the washing machine and the dryer. An alternative or coincident explanation is that HWRs affect electricity use choices more generally. For example, HWRs may increase household attention to utility bills or change the 'moral utility' associated with water and energy use. This channel would affect electricity use beyond what would be expected from mechanical complementarities alone.

Results from three empirical tests and an engineering simulation refute the hypothesis that mechanical complementarities are the sole driver of electricity savings, and are consistent with a framework in which HWRs alter customer choice about water and electricity consumption. A first test posits that if a mechanical relationship explains the cross-sectoral spillover, then water conservation should occur when electricity conservation occurs. Taking advantage of hourly interval data on water and electricity use, we find that the time profile of electricity and water savings are mismatched, with reductions in energy use but no significant reductions in water use between the hours of 3 PM to 7 PM and 8 PM to 10 PM. A second test indirectly examines whether households exposed to HWRs alter their consumption of air conditioning, an action that requires electricity but not water as an input. Electricity conservation increases in temperature while water use treatment effects exhibit no temperature response gradient, pointing to the possibility that households respond to treatment by adjusting their thermostats. Third, results from a post-treatment survey highlight that exposure to HWRs is positively correlated with some actions that use electricity but not water as an input. Finally, results from a detailed, appliance-level engineering simulation calibrated to Southern California households imply that mechanical complementarities can explain only twenty-five percent of the estimated electricity savings.

Inference in our setting poses a challenge for two reasons. First, relative to many peer studies, our experiment is underpowered. We chose a setting where both high-frequency water and electricity use data were available. Given the dearth of utilities collecting high-frequency water data and the difficulty in obtaining data on water and electricity use that can be merged at the household unit, our sample is limited to a single utility. By comparison, recent work in the residential energy and water settings leans on samples that aggregate across utilities (Allcott, 2011; Brandon et al., 2017). The trade-off in obtaining granular data on water and electricity use is the limited sample and, as a result, a setting that may be underpowered to precisely detect spillover effects. This concern is amplified because when designing our experiment and performing power calculations, we failed to account for within-unit serial correlation (Burlig, Preonas, and Woerman, 2020). Limited sample sizes and power calculations that do not account for within-unit serial correlation ex-ante are issues likely shared with many experiments in environmental and resource economics. Recent advances provide researchers with a toolkit to address the latter issue (Burlig, Preonas, and Woerman, 2020). Second, while high-frequency data provide a platform to test when and how households respond to treatment, testing the impact of a single intervention on multiple outcomes raises concerns about the detection of false positives. We deploy multiple approaches to examine this, including reporting an array of p-values that adjust for multiple hypothesis testing (List, Shaikh, and Xu, 2019).

Our work highlights the importance of moving beyond a partial equilibrium framework in program evaluation and contributes to a growing literature examining the cost-effectiveness of nudges and their welfare impacts (Allcott and Kessler, 2015). Economists have shown that social norms messaging is a cost-effective water and energy conservation policy (Allcott and Mullainathan, 2010; Brent, Cook, and Olsen, 2015; Ferraro and Price, 2013). Our results suggest that these cost-effectiveness estimates are understated for two reasons. First, electricity conservation yields private and social savings. Incorporating the private savings and social benefits from the electricity conservation spillover increases the net benefits of HWRs by 39% from \$2.91 to \$4.04/household. Second, electricity savings occur during peak hours when wholesale costs are highest, and the marginal emissions from electricity generation are larger. Accounting for when reductions in electricity use occur further increases the benefit to cost ratio to \$4.70/household. While these cost-effectiveness estimates demonstrate the importance of accounting for spillovers in policy analysis, this valuation metric should not be interpreted as a welfare analysis.[1]

As behavioral instruments grow in popularity and breadth of application, our results point to the possibility of cost-sharing and collaboration in their deployment. This is particularly relevant for the context of water and energy conservation in California. Policymakers and utilities in the state are investigating whether water conservation programs present an opportunity to save energy as well. To date, efforts to document a 'water-energy nexus' have focused on 'embedded' energy savings from water conservation, i.e., electricity savings from the conveyance and treatment of water. We depart from this literature by contributing the first causal data point on the end-use energy impacts of a water conservation program.[2]

While our paper finds evidence of a cross-sectoral electricity spillover in one utility during a severe drought, an outstanding question is whether this spillover generalizes to other locations and periods. External validity poses a challenge in our experiment, and arises from non-random site selection (Allcott, 2015).[3] Our experimental time frame and utility partner are distinct along two dimensions. First, we performed the experiment in a municipal water and electric utility that collects high-frequency data for water and electricity. Most cities in Southern California are served by an investor-owned electric utility and meter water on a

[1]For example, our analysis is silent to possible disutility households experience from the receipt of HWRs, the reduction in surplus attributable to inefficient pricing of electricity and water, and lost consumer surplus from conservation (Allcott and Kessler, 2015). For these reasons, the welfare gains from the deployment of HWRs are likely lower than the calculated savings from our cost-benefit framework.

[2]The "file drawer" problem may explain the absence of experimental evidence on energy spillovers (Rosenthal, 1979). If results in which the null hypothesis is rejected are more likely to be published, this will result in publication bias and difficulty in determining the true proportion of tests that reject the null (Kennedy, 2004; Christensen and Miguel, 2018).

[3]Recent work offers some strategies to account for external validity bias arising from unit selection into an experiment (Andrews and Oster, 2019). Unit selection leads to an experimental sample that may differ from the population of interest along both observables and unobservables. We refer the reader to Athey and Imbens (2017) and Peters, Langbein, and Roberts (2016) for a more comprehensive review of external validity.

monthly, not hourly, basis. Given that our utility looks systematically different from other cities in California and nationally, it is difficult to apply our results elsewhere. Second, the timing of our experiment coincided with a historic drought during which an urban water conservation mandate of 25% was in place. Given the unique timing of the experiment, it is difficult to know if the spillover effects we detect would have occurred in the absence of extreme drought. To evaluate bias from site selection, studies have implemented research designs that replicate the same experiment across a range of locations and time periods, compared outcomes in experimental sites to outcomes in non-experimental sites, and conducted meta-analyses (Bell et al., 2016; Allcott, 2015; Meager, 2019; Vivalt, 2019).

2 Research Design and Setting

We partnered with a municipally owned water and electric utility and the water technology vendor WaterSmart to evaluate the potential energy and water savings of non-pecuniary water conservation instruments. The study occurred in the jurisdiction served by Burbank Water and Power (BWP) and spanned March 2015 to May 2016. BWP serves roughly 18,500 single-family customers in the City of Burbank, an inland city in Los Angeles County that, like much of Southern California, is characterized by a subtropical Mediterranean climate. A unique feature of BWP is that since it is both a municipally owned water and electric utility, households receive water and electricity bills in the same billing statement. The timeline of the project includes the summer of 2015, a period that coincides with the worst drought in California's recorded history, and a relatively wet and cool winter.

Our sample consists of 7,341 single-family homes served by BWP. To be eligible for our study, a household needed to reside in a single-family home and have six months of meter readings before the launch of the experiment. Of the roughly 18,000 eligible homes, we randomly assigned using the "big stick" re-randomization method 4,559 accounts to the WaterSmart treatment and 2,782 to control.[4] Control households did not receive notification

[4]Our randomization protocol tested for and ensured covariate balance in the mean and variance of pre-treatment monthly water use across control and treatment accounts. Re-randomization into treatment and control continued until the p-value from a test of the joint significance of these covariates in explaining assignment to treatment exceeded a 0.25 threshold.

that they were in a pilot program.

Figure 1 provides an example of a WaterSmart report that treatment households received. The report includes a social comparison, water savings recommendations, information about the report, and information about BWP's water conservation programs. The social comparison provides a household with information on its current water use and compares its use to that of similar households and an "efficient" household. An injunctive norm accompanies the comparison, conveying pro-social behavior through the display of a smiling or frowning face depending on the household's water use relative to its neighbors. The report also includes individualized recommendations on ways to use water more efficiently. The bottom panel of Figure 1 provides an example of a "hard" recommendation – changing from grass to native plants could save 78 gallons per day and $242 per year – and a "soft" recommendation – upgrading irrigation timer settings could save 53 gallons per day and $148 per year. Importantly, no messages or water conservation tips target or mention energy conservation.

Between the treatment period spanning May 15, 2015 to May 31, 2016 households received bi-monthly reports. WaterSmart sent an introductory letter to all treatment households in March 2015, explaining the product and notifying them that they would be receiving HWRs over the subsequent year. The first home water report was mailed to treatment households in early May 2015, and five subsequent reports were sent to households over the treatment year.[5] WaterSmart mailed all treatment households the introduction letter and first HWR. However, the five subsequent reports were sent according to the following rule: households enrolled in electronic billing with BWP (roughly 2/3 of the sample) received electronic HWRs, and households enrolled in paper billing continued to receive physical reports in the mail.

Prior to the experiment, we performed power calculations using monthly water use data from BWP. Based on discussions with our partners, we assumed the sample size would include only 6,600 households with 50% of households assigned to treatment and control. Further, we assumed households would receive just five HWRs. Under these assumptions, our power calculations indicated a minimal detectable effect of 1.4% for water use with

[5]We exclude data from March 15 to May 15, 2015 from our analysis since households may have changed their behavior in response to the introductory letter.

80% power and 5% size. We likely miscalculated this for four reasons. First, recent work on power calculations using panel data suggests that we likely overstated the minimum detectable treatment effects. This is because, though we do account for within unit serial correlation ex-post, our power calculations did not account within unit serial correlation ex-ante. Best practices for designing experiments when using panel data, including a statistical package for performing these calculations, are the focus of Burlig, Preonas, and Woerman (2020). Second, when performing our calculations, we had yet to receive data on energy use. We, therefore, assumed power calculations on water use were transferable to electricity use, despite important differences between water and electricity use both within and across households. Third, we assumed that the effect of each HWR on use would be constant over time, though patterns of use and the margins for conservation differ by season. Lastly, we did not account ex-ante for multiple hypothesis testing.

3 Conceptual Motivation

To understand why cross-sectoral spillovers may exist, we outline a simple conceptual framework that illustrates two channels through which a water nudge may impact energy use. Consider a household that derives utility from consuming energy services (e.g., air conditioning), water services (e.g., a shower), and a composite good. Following Allcott and Kessler (2015) and Farhi and Gabaix (2017), we allow imperfect information and behavioral biases to influence demand for water and electricity. First, consumers may be misinformed about consumption quantities because demand for electricity and water are often derived (Jessoe and Rapson, 2014). For example, rather than directly choosing water and electricity usage, households choose how many loads of laundry to run or how often to use the dishwasher. Second, households may misperceive water and energy prices, responding to say average instead of marginal prices (Ito, 2013, 2014). To account for imperfect information about electricity and water prices, we assume customers make decisions based on perceived prices, $\tilde{\mathbf{p}}(\mathbf{p};\gamma)$, which are a function of actual prices $\mathbf{p}$ and a misperception factor γ. Third, since our experiment was deployed during the worst drought in California's history, consumers likely experience 'moral utility' from water consumption. Residents faced enormous conservation appeals from local and state governments and significant social pressure to conserve water. Following Levitt and List (2007), we introduce a moral tax on water and electricity use, denoted by $\boldsymbol{\mu}$.

Now, suppose that nudges about water usage are exogenously deployed. In this setting, the introduction of nudges denoted by χ may alter choices by influencing perceptions about prices or the moral tax. Perceived prices and moral taxes are now modeled as $\tilde{\mathbf{p}}(\mathbf{p};\gamma,\chi)$ and $\boldsymbol{\mu}(\chi)$. Demand for both goods is given by $\tilde{\mathbf{x}} = \mathbf{x}\left(\tilde{\mathbf{p}}(\mathbf{p};\gamma,\chi), \boldsymbol{\mu}(\chi), w\right)$, where w is the consumer's budget for water and energy.

The water conservation nudge may influence household energy demand through two channels: a mechanical (complementarities) channel, and a behavioral channel. First, water and electricity are complementary inputs in the use of many household appliances. For example, if households respond to a nudge by washing one less load of laundry per day, this leads to a reduction in electricity use along two margins: less energy is used running the washing machine, and the dryer is used less. The reduction in electricity use that occurs in response to the water nudge is a result of the Leontief nature of demand for this appliance, a channel we label "mechanical complementarities".

The second channel reflects the possibility that a nudge on water consumption may change customer perception about electricity prices or the moral tax placed on electricity consumption. We refer to this as the behavioral channel. This will occur if the vectors $\tilde{\mathbf{p}}(\mathbf{p};\gamma,\chi)$ and $\boldsymbol{\mu}(\chi)$ share some common components where a nudge influences the awareness and moral tax for both water and electricity.[6] If a behavioral channel exists, electricity use may either increase or decrease in response to nudges. For example, assume a water nudge increases the moral tax on electricity use. The will lead to a decrease in electricity use above and beyond the amount attributable to the mechanical spillover. Alternatively, if the nudge decreases the moral tax on electricity use, for example through moral licensing, then

[6] In our setting, a common component is that water and electricity bills appear in a single statement. When customers view their water bill and usage, they also see their electricity bill and usage. Shared billing may lead to an association between water and electricity use. Further, all households in our sample received home energy reports from OPower. Given the similarity between the reports, the receipt of a HWR may have made HERs and electricity use more front of mind.

this non-mechanical channel could induce households to increase electricity use (Jacobsen, Kotchen, and Vandenbergh, 2014; Harding and Rapson, 2014). Importantly, the sign of the behavioral spillover could be positive or negative.

4 Data and Quality of Randomization

High-frequency metering data on household electricity and water use serve as the primary data for our analysis. BWP provided hourly water and electricity data for every household in our sample for the treatment period, May 2015 to May 2016. Pre-treatment monthly billing data on water and electricity use were obtained for all households spanning the baseline period, March 2014 to February 2015. We also collected county assessor data on housing unit attributes and matched these at the address with the electricity and water use data.

Table 1 reports summary statistics on pre-treatment electricity use, water use, and housing unit attributes, and examines the credibility of the randomization. The first two columns report means for households assigned to control and treatment, respectively, and the third column reports the difference in means.[7] A comparison of means across control and treatment shows no statistically significant differences in monthly water use, monthly electricity use, or summertime electricity use. For the households for whom we observe assessor data, housing unit characteristics are also balanced across control and treatment. We further investigate the quality of the randomization by comparing baseline electricity use across control and treatment in each month of the year preceding the experiment. As shown in Figure 2, there are no significant differences in electricity use across control and treatment in any calendar month, including the summertime months of May, June, July, and August 2014. These descriptive statistics provide a first layer of support for the integrity of the randomization. They also offer visual evidence that control households may use more, but not significantly

[7]Attrition occurs in our sample for two reasons. First, households move. This leads to the termination of an account and the omission of post-move household hours from our sample. This form of attrition is uncorrelated with assignment to treatment. The second source of attrition arises from approximately 50 households opting out of the WaterSmart treatment. This selective attrition leads to the presence of "never takers" in the treatment group and may compromise the experimental research design. To address this concern, we monitor water and electricity use for households that opted-out of HWRs and estimate intent-to-treat effects.

more, electricity than treatment households, particularly in the summer months. To account for the possibility that differences in post-treatment outcomes may be attributable to pre-treatment differences in electricity use, we follow Bruhn and McKenzie (2009) and control for pre-treatment annual, summer, and winter electricity use in our preferred specifications.

5 Empirical Approach and Results

This section discusses the empirical approach used to isolate the effect of HWRs on average electricity consumption and the results that follow. We then discuss the impacts of HWRs on the time profile of electricity usage.

5.1 Average Treatment Effects

To identify the causal effect of HWRs on energy use, we begin by comparing average hourly electricity use across treatment and control households conditional on weather controls and time fixed effects, estimating the equation:

$$y_{ih\tau} = \beta_0 + \beta_1 \mathrm{WS}_i + f(T_{h\tau}; \Theta) + \theta_p P_{h\tau} + \gamma_\tau + \gamma_h + \epsilon_{ih\tau}. \tag{1}$$

The dependent variable $y_{ih\tau}$ is the level of electricity use specified in kilowatt-hours per hour (kWh/hr) for household i during hour h of day τ.[8] The indicator variable WS_i denotes assignment to the WaterSmart treatment and equals one if customer i is assigned to treatment.

[8]We estimate our model in levels rather than logs for three reasons. First, similar to previous work, we are interested in the effect of assignment to treatment on the quantity of electricity savings, and want to understand the margins by which households respond to treatment (Boomhower and Davis, 2017; Novan and Smith, 2017; Fowlie et al., 2017). Given the substantial heterogeneity in hourly electricity consumption across hours of the day and months of the year, focusing on the percentage change in electricity use masks substantial differences in level changes across hours and months. For example, a 2% reduction in electricity use at 8 AM in the summer translates into very different behaviors than a 2% reduction at 8 PM in the winter. Second, level estimates are inputs in our engineering simulation and cost-effectiveness calculations. Using level estimates maintains consistency throughout the paper. Third, our analysis uses hourly electricity and water use as dependent variables, and there are many hourly intervals when recorded water use is zero. We could use the inverse hyperbolic sine of energy and water use to overcome this issue, but the transformations could affect our estimates. In Appendix Table A.3, we show that our results are qualitatively robust to a log transformation and inverse hyperbolic sine transformation of the dependent variable.

Calendar date and hour of day fixed effects, denoted by γ_τ and γ_h, control for seasonal and hourly patterns in electricity use. Weather controls include, $f(T_{h\tau};\Theta)$, a flexible function of hourly temperature ($T_{h\tau}$) with parameters Θ, and $P_{h\tau}$, an indicator variable denoting if precipitation was recorded in hour h of day τ. We specify $f(T_{h\tau};\Theta)$ as a series of 5°F temperature bins,

$$f(T_{h\tau};\Theta) = \begin{bmatrix} \theta_{60} \cdot \mathbf{1}(T_{h\tau} < 65°\text{F}) \\ \theta_{65} \cdot \mathbf{1}(65°\text{F} \leq T_{h\tau} < 70°\text{F}) \\ \vdots \\ \theta_{80} \cdot \mathbf{1}(80°\text{F} \leq T_{h\tau} < 85°\text{F}) \\ \theta_{85} \cdot \mathbf{1}(T_{h\tau} \geq 85°\text{F}) \end{bmatrix}$$

where $\mathbf{1}(\cdot)$ is an indicator function that equals one whenever the outdoor temperature in an hour lies within the specified range. Standard errors are clustered at the household.

Table 2 reports results for the effect of assignment to WaterSmart on average hourly electricity use.[9] Column (1) displays our results from estimating equation (1) during all hours of the treatment period spanning May 15, 2015 to May 31, 2016. Column (2) includes pre-treatment summer, winter, and annual electricity use to control for possible differences in baseline electricity use across households. In columns (3) and (4), we restrict our sample to Summer 2015 (May 15, 2015 to August 31, 2015). We focus on summer months for two reasons. First, system-wide electricity loads are highest and water demand peaks in the summer. Second, this period corresponds to the first 75 days of our pilot when HWRs may be most salient. Columns (5) and (6), restrict the sample to the peak electricity use hours (3 PM to 8 PM) in the summer.

Our interpretation of Table 2 is based primarily on the estimates that condition on pre-treatment electricity use. Our choice to concentrate on these results is guided by earlier work on evaluation in randomized controlled trials, and a comparison of the results presented in even and odd columns (Bruhn and McKenzie, 2009). The latter shows that the estimated treatment effects reduce in magnitude after controlling for pre-treatment electricity use, and

[9] We exclude AMI meter reads above 20 kWh/hr as they are likely errors. This restriction reduces the sample by 1,115 observations or 0.01%. We also remove outliers, excluding households if pre-treatment electricity use was in the 98.5 or 1.5 percentiles for more than four months in the year preceding treatment. Results are not sensitive to relaxing either restriction.

reflects our finding in Figure 2 that households assigned to treatment use less electricity in the pre-treatment period.

As shown in columns (1) and (2), over the duration of the year WaterSmart does not induce a change in electricity use. This average treatment effect, however, masks important heterogeneity in the time profile of the response to WaterSmart. To illustrate the time profile of treatment effects over the duration of the experiment, we interact assignment to treatment with month-of-year indicators to estimate monthly treatment effects conditional on baseline electricity use, weather controls, calendar date, and hour-of-day fixed effects. Figure 3 plots the estimated treatment effect during all hours (circles) and peak hours (diamonds), as well as the 95% confidence interval for peak hours. The figure demonstrates that significant reductions in electricity use occur, but are confined to June, July, and August. After the summertime, HWRs induce no change in electricity use. The difference between the summertime and annual response may occur because earlier reports are more salient, there are more levers to reduce electricity use in the summer, electricity conservation is more front of mind in the summer, or the summer coincided with a historic drought.

Zooming in on the summertime response highlights that assignment to treatment reduced summertime hourly electricity use by approximately 0.017 kWh/hour, or 1.35 percent (column 4). This amounts to every household ceasing use of one 15W light bulb over the summer. Alternatively, this is equivalent to every household reducing its use of an Energy Star dishwasher by one load every three days during the summer months. The estimated impact falls within the range of treatment effects reported from the deployment of home *energy* reports focused exclusively on energy consumption and conservation (Allcott, 2011). Despite the limited sample size and the challenges this poses for inference, for two main reasons it is unlikely that the summertime spillover effect is the result of sampling variability or Type I error. First, our coarse power calculations yield a minimum detectable effect of 1.4 percent with a sample of 6,600. Given that our study uses a long-panel, for a set minimum detectable effect, our power miscalculations likely resulted in an experiment where the probability of Type I error does not increase, but the probability of a false negative does (Burlig, Preonas, and Woerman, 2020). Framed differently, the missteps in our power calculations decrease our ability to precisely reject the null hypothesis when it should be rejected. Second, this result is insensitive to the inclusion of a richer set of time fixed effects (Table A.1) and monthly controls for baseline electricity use (Table A.2); is qualitatively robust to inverse hyperbolic sine transformations of the dependent variables (Table A.3), though the magnitude of the effect attenuates to 0.9% in the latter specification; and can be seen visually in Figure 3.

As shown in columns (5) and (6), the treatment effects are most pronounced during peak hours in the summer when HWRs induce a 1.5 to 2.5 percent reduction in electricity use. This provides the first piece of empirical evidence that, in addition to a conservation effect, the reports may yield savings via a reduction in peak electricity demand. However, we are cautious in drawing more than suggestive evidence of a reduction in peak electricity demand. While our results are insensitive to the inclusion of alternative time fixed effects and baseline controls for electricity use, under log and inverse hyperbolic sine transformations of electricity use, HWRs no longer induce a statistically significant reduction in use (Table A.3). To further examine if treatment induces a reduction in electricity use during peak hours, we take advantage of the temporal granularity in the data and formally estimate the effect on treatment on electricity use in each hour of the day.

5.2 Time Profile of Treatment Effects

Hourly interval data allow us to estimate treatment effects across hours of the day. The estimates provide further insight into the load management and environmental impacts of home water reports, as well as the levers by which households respond to treatment. To decompose the treatment effect, we augment equation (1) and estimate

$$\begin{aligned} y_{ih\tau} = & \sum_{j=1}^{23} \alpha_h \mathbf{1}[h=j] + \sum_{j=0}^{23} \beta_h \left(\mathbf{1}[h=j] \times \mathrm{WS}_i\right) \\ & + f(T_{h\tau};\Theta) + \theta_p P_{h\tau} + \Gamma \mathbf{X}_i + \gamma_\tau + \epsilon_{ih\tau}, \end{aligned} \tag{2}$$

where $\mathbf{1}(\cdot)$ is an indicator function that equals one when hour h equals hour-of-day j. Controls for monthly pre-treatment average electricity use in the twelve months, summer months, and winter months preceding treatment are denoted by $\mathbf{X}_i$. The coefficients α_h reflect the conditional average hourly electricity consumption for control households relative to the omitted hour, 12 AM. The coefficients of interest, β_h, capture the average effect of assignment to treatment for each hour of the day. Given our findings in Table 2 and Figure 3, we restrict the sample to the summer of 2015.

Figure 4(A), which plots the effect of assignment to treatment on electricity use for each hour of the day, reveals substantial heterogeneity in the time profile of our average treatment effect. From 12 AM to 5 AM, we observe visual, but not statistically significant, evidence that treatment households use less electricity. No discernible differences occur between 5 AM and 11 AM. Around 11 AM, treatment households begin to reduce electricity use relative to control households, with this effect growing in magnitude and significance over the afternoon. Significant reductions start to occur at 3 PM. The treatment effects are largest from 3 PM to 7 PM, peaking at 4 PM and 5 PM. They persist until 11 PM, with the effect slowly declining after 7 PM.[10]

Heterogeneity in the timing of the treatment effects has direct implications for pollution reductions, and wholesale energy savings attributable to HWRs. The former occurs because the marginal source of electricity generation, and thus the pollutants generated from this source, vary hourly. The largest estimated treatment effects also coincide with peak demand hours for both the households in our study and the electric grid as a whole.[11] This suggests that, in addition to a conservation effect, the reports may lead to savings because of a reduction in peak load. These gains occur because of the disparity between the marginal cost to supply electricity and the retail price of electricity. In Section 7, we weigh in on the importance of this heterogeneity.

While high-frequency data provide an opportunity to investigate patterns by which agents respond to treatment, our hourly treatment effects may be susceptible to multiple hypothesis

[10]Figure A.2, which presents hourly treatment effects using the log and inverse hyperbolic sine of electricity as the dependent variable, highlights a more uniform response to treatment across hours of the day. During the hours spanning 8 PM to 4 AM treatment leads to a weakly significant reduction in electricity use, where this uniform percentage reduction is largely driven by differences in baseline use across hours of the day.

[11]The California ISO classifies 'peak net load hours on the California electricity grid as 4 PM to 9 PM, with 'super peak occurring at the same hours in July and August each year. See http://www.caiso.com/Documents/CaliforniaISO_Time_UsePeriodAnalysis.pdf.

testing. As discussed in Appendix B, in our setting, the main insights we can draw from classic adjustments for multiple hypothesis testing are limited. This is because these tests are designed for cross-sectional data and analysis, and our study uses a long-panel and deploys a rich set of time fixed effects to account for much of the variation in electricity use. Still, we collapse our panel data, create cross-sectional measures of electricity use, and use the methods in List, Shaikh, and Xu (2019), Bonferroni (1935), and Holm (1979) to account for multiple hypothesis testing. Table B.1 presents adjusted p-values for the average treatment effects reported in Table 2, and highlights that under these adjustments treatment no longer impacts electricity use. It remains an open question as to what we learn from these corrections since a simple t-test of mean post-treatment electricity use across control and treatment also fails to reject the null. The loss of precision is largely driven by the limited sample, 7,341 as compared to millions of observations, and our inability to control for much of the temporal variation in electricity use.

6 Mechanisms: Empirical Evidence

Our finding that the water conservation instrument may have induced electricity conservation leads to several questions on the mechanisms underpinning the result. One leading explanation for cross-sectoral spillovers is mechanical – reductions in energy use are due to reductions in water-consuming activities that also use electricity. An alternative hypothesis is that treatment alters consumer choice about resource use more generally, encouraging households to reduce electricity use. We now implement three empirical tests and use a household engineering simulation to gauge the plausibility and relative importance of these two channels.

6.1 Timing of Water and Energy Conservation

In a first empirical test, we hypothesize that if a mechanical relationship explains the cross-sectoral spillover, then reductions in electricity use should be accompanied by reductions in water use. One indication that energy conservation is driven exclusively by reductions in water use is if the observed timing of water and electricity conservation coincide. To examine this, we compare the estimated hour of day treatment effects from Figure 4(A) with the corresponding hour of day treatment effects for water use.

As a first step, we present results for the mean hourly effect of assignment to Water Smart on water use. Table 2 Panel B replicates the specifications from Panel A using hourly water use as the dependent variable. The results make clear that HWRs induced large water savings. The intent to treat effects range from 0.45 to 0.625 gallons per hour. Our results correspond to a 4.4 and 2.9 percent reduction in water use over the whole year and in the summertime, respectively. Unlike our findings in Panel A, the annual treatment effects are larger in magnitude than the summertime effects, suggesting that water and electricity use may not be entirely linked. When we restrict the sample to the peak summer hours of 3 PM to 8 PM, we find that the estimated water reductions increase in levels and percentages.

We next compare hour-of-day treatment effects for electricity use to the corresponding hour-of-day treatment effects for water use. Panel B of Figure 4 plots the treatment effects from estimating equation (2) using hourly water use as the dependent variable. This comparison provides our first piece of empirical evidence that a mechanical relationship between actions that use both water and electricity is not the sole driver of our observed cross-sectoral spillover. The time profile of electricity reductions runs counter to the time profile of water savings during some hours. While moderate reductions in electricity use occur from 3 PM to 11 PM, there are no statistically significant reductions in water usage over this time interval except for a large decrease at 7 PM. Under the mechanical hypothesis, if electricity reductions occur, water reductions should also occur. From 3 PM to 7 PM and 8 PM to 10 PM this is not the case, suggesting that HWRs affect at least some electricity use through a non-mechanical channel.

6.2 Response to Temperature

An ideal exercise to tease out the importance of the behavioral channel in explaining the cross-sectoral spillover would test whether electricity choices that do not require water as an input are affected by assignment to treatment. If treatment households make systematically different electricity consumption decisions, this suggests that assignment to treatment impacts electricity use through a behavioral channel. While the data needed to examine this hypothesis directly are not available, we propose and implement an indirect empirical test to investigate its plausibility.

We examine the hypothesis along an important and salient dimension of household electricity use: cooling. To do this, we first test whether electricity use for treated households responds differently to increases in outdoor temperature relative to control households. We then test whether water conservation from assignment to treatment also increases in temperature. The rationale guiding this test is that cooling comprises a significant share of household summertime electricity use, its use increases in outdoor temperature, and it requires electricity but not water as an input. A differential response to high temperatures for electricity but not for water suggests that households exposed to HWRs alter their consumption of air conditioning. We implement this empirical test by augmenting equation (1) and estimating

$$y_{ih\tau} = f(T_{h\tau}; \Theta) + g(\mathrm{WS}_i \times T_{h\tau}; \Lambda) + \theta_p P_{h\tau} + \Gamma \mathbf{X}_i + \gamma_\tau + \epsilon_{ih\tau}, \tag{3}$$

where all variables are defined as before. We specify $g(\mathrm{WS}_i \times T_{h\tau}; \Lambda)$ as the interaction of each of the temperature bins in $f(T_{h\tau}; \Theta)$ with the indicator variable for assignment to treatment. As before, we limit our sample to summer 2015.

Our empirical strategy identifies the differential impact of a given temperature bin on electricity (and water) use across treatment and control households. The coefficients θ_j estimate the conditional average electricity (water) use for temperatures in bin j relative to the excluded bin, temperatures less than 65°F. The coefficients λ_j estimate the difference in average electricity and water use when temperatures fall in bin j between treatment and control households. If treatment households respond to HWRs by reducing their air conditioning use, λ_j will increase in temperature for electricity but not for water use.

We face an important challenge in using this empirical strategy – the water consumption response to treatment may increase in temperature as well.[12] Households may respond to treatment by shifting outdoor irrigation from the afternoon to the morning, or by reducing outdoor watering altogether on hot days. A unique feature of our setting allows us to imperfectly control for the relationship between temperature and outdoor irrigation. Due to the ongoing drought, BWP imposed and enforced outdoor watering restrictions and households were only permitted to water on two exogenously determined days of the week. As such, we split our sample into utility-wide watering and non-watering days. We hypothesize that on watering days, the water response to WaterSmart is likely increasing in temperature. In contrast, on non-watering days, the relationship between temperature and response to treatment should be substantially attenuated. If our cooling hypothesis holds, then (i) the electricity response to treatment should be increasing in temperature but should not vary substantially across watering and non-watering days, and (ii) the water response to treatment should be relatively constant across temperatures on non-watering days.

Our results highlight that differences in electricity use across treatment and control households are increasing in temperature. Columns (1) and (2) of Table 3 report the results from the estimation of equation (3) on non-watering and watering days, respectively. On non-watering days the magnitude of the treatment effect jumps from 0.002 to 0.027 kWh per hour as temperature increases from below 65 F to above 85 F. On watering days, the response increases from around 0 to 0.036 kWh per hour. The findings highlight that the relationship between electricity use and treatment is increasing in temperature and is nearly identical across watering and non-watering days, suggesting that it is unaffected by watering restrictions.[13] In contrast to the electricity results, the water response to treatment does not increase in temperature on watering days or non-watering days. We report these results in columns (3) and (4) of Table 3. On both watering and non-watering days, the largest treatment effects occur during the relatively mild temperature hours of 70F or cooler. This highlights that the water use response to treatment exhibits no temperature gradient, suggesting that factors uncorrelated with temperature are driving water conservation.

[12]The literature on the relationship between ambient air temperature and residential water demand highlights that water demand increases with temperature (Balling and Gober, 2006; Gato, Jayasuriya, and Roberts, 2007).

[13]P-values testing whether the temperature bin coefficients $<$65F and $\geq$85F are significantly different from each other are 0.18 and 0.06 for columns (1) and (2), and 0.63 and 0.44 for columns (3) to (4).

Our findings provide indirect evidence that households exposed to HWRs alter electricity use through decreased cooling. A comparison of the electricity and water response to treatment at different temperatures makes explicit that while the electricity response to treatment is increasing in temperature, the water response to treatment is not. We find that the water use response to treatment is not correlated with temperature increases, and exhibits the largest response in both magnitude and significance during hours with relatively mild temperatures.

6.3 Household Survey of Electricity Use Behavior

To explore the plausibility of our hypothesis that HWRs affect customer choices about electricity use, we designed and sent a post-treatment household survey on electricity and water use decisions to 2,400 households. In Appendix C we detail the deployment of the survey, highlight empirical challenges in drawing inference from survey data, and report results. We view a comparison of survey responses across control and treatment households as offering additional empirical insight into whether assignment to treatment is correlated with electricity conservation actions. Our results, which are reported in Table C.2, indicate that for every energy conservation question asked in the survey, households assigned to treatment are (weakly) more likely to respond that they participated in that action. We also find that even with our small sample, households assigned to the WaterSmart treatment are significantly more likely to report turning off the lights or turning off the TV. These comparisons provide another line of evidence consistent with our behavioral hypothesis.

6.4 Simulation to Bound Mechanical Spillovers

We develop an engineering model of household water and energy use to understand the potential magnitude of mechanical complementarities in explaining the cross-sectoral electricity spillover. The model simulates expected changes in yearly electricity use assuming all electricity reductions come from reductions in end uses that use both water and electricity as inputs. The model makes several restrictive assumptions, but can be viewed as an approximation of the expected electricity savings from the deployment of WaterSmart if the behavioral channel was "shut off."

We simulate the expected change in yearly electricity use from the WaterSmart intervention assuming that HWRs only impact electricity use through mechanical complementarities. We use individual response data from the 2009 California Statewide Residential Appliance Saturation Study (RASS) to parameterize an appliance-level model of a representative household's water and electricity use. The RASS data include detailed information on appliance ownership and estimates of yearly, appliance-level electricity use for over 25,000 homes in California. We restrict our sample to around 8,500 single-family homes in Southern California to ensure the RASS respondents better resemble the households in our study. From the list of appliances covered in the RASS, we identify the following end uses as directly or indirectly consuming water and energy: water heating, clothes washing, clothes drying, dishwashing, groundwater pumping, evaporative cooling, and pool/spa operation. We use the average ownership rates, the distribution of annual energy use for each appliance, and the average water treatment effect from column (4) of Table 2 Panel B to simulate the distribution of average electricity use reductions under four scenarios.[14]

Table 4 reports our results. In our baseline scenario, we assume that water use reductions are distributed equally across all appliances. Our additional scenarios take more extreme stances on appliance use and ownership patterns. Our intention is to gauge under each of these implausible assumptions, the extent to which mechanical complementarities could explain our estimated electricity reductions. In our baseline scenario – in which the profile of appliance ownership mirrors that observed in Southern California and reductions in water use are distributed equally across appliances – household electricity consumption decreases by approximately 39 kWh per year. For comparison, the estimated treatment effect from column (2) of Table 2 Panel A is 149 kWh per year. In this scenario, we can attribute about 25 percent of the observed electricity savings from the WaterSmart intervention to mechanical complementarities.

[14]Appendix D discusses the RASS data and simulation procedure in more detail.

The second scenario assumes that all water reductions come exclusively from decreases in the use of indoor appliances that require both water and electricity as inputs. We assume that the quantity of water savings estimated in column (2) of Table 2 Panel B occurs from a reduction in the use of clothes washers, electric dryers, and dishwashers only. This would require, on average, dishwasher and clothes washer use to decrease by nearly 40 percent in response to treatment. Under this assumption, household electricity use decreases by 140 kWh per year, explaining 93% of the electricity spillover. Framed differently, if households respond to WaterSmart by reducing their use of clothes washers, electric dryers, and dishwashers by 20% (50 fewer clothes washer and 25 fewer dishwasher loads per year), we could explain only 45% of the estimated reduction in electricity use.

Our third scenario assumes that water use reductions are distributed equally across appliances and that every household uses an electric water heater. While ownership of electric water heaters is well below 10 percent in our area of study, this scenario allows us to gauge the importance of water heating in explaining mechanical complementarities. Even under this extreme assumption, the simulated electricity savings explain only 80 percent of the estimated electricity spillover. If we assume that 10% of households have electric water heaters and all the estimated water savings occur within these households, then these households would need to reduce water use by almost 45% for the electricity savings to be explained by mechanical complementarities.

Our last scenario casts light on one last lever that could explain the cross-sectoral spillover: pools and spas. If we assume that all households with pools and spas reduce their use of pool filters and spas by 100 percent, mechanical complementarities lead to electricity savings far exceeding our estimated treatment effect. Phrased differently, if around twenty percent of the households in our sample that own pools and spas discontinued their use of pools and spas entirely, mechanical complementarities could explain our estimated cross-sectoral spillover. We can assess the extent to which pool users drive our results. To do this, we replicate our results from Table 2 using only the sample of households that do not own pools. Results reported in Table D.2 make clear that households with pools do not drive the estimated electricity spillover.

7 Valuing Electricity Conservation Spillovers

Our finding that social norms messaging about water consumption induces electricity conservation implies that existing cost-effectiveness estimates of HWRs are understated. In this section, we set forth a cost-benefit framework to quantify the net benefits from home water reports. The analysis departs from existing cost-effectiveness measures in the residential water and energy settings (Allcott, 2011; Brent, Cook, and Olsen, 2015; Ferraro and Price, 2013). Previous estimates calculate ratios of program costs to the energy or water savings, yielding a dollar per kWh or dollar per gallon metric. Our framework uses dollars as the common unit to aggregate water and energy savings, and incorporates savings from reductions in local and global pollutants. In addition to reporting average savings, our framework takes advantage of granular data on electricity use to account for when these reductions occur, since the timing of electricity conservation affects both wholesale electricity costs and pollution emissions.

An ideal valuation exercise would move beyond a benefit-cost framework and contribute to a growing literature that quantifies the welfare effects of nudges (Handel, 2011; Bernheim, Fradkin, and Popov, 2015). Of particular relevance is recent work that conducts a follow-on experiment to elicit consumer willingness to pay for HERs, and provides a first estimate of the welfare impacts of nudges in the residential electricity sector. Our research setting does not provide an opportunity to do this effectively. Instead, we outline a cost-benefit framework that provides a transparent and practical metric to evaluate the relative importance of electricity spillovers.

Our cost-benefit exercise aggregates water savings, electricity savings, and savings from reductions in local and global pollutants into an annual, dollars per household metric. We measure the average annual net benefits per household from the deployment of HWRs as:

$$\text{Net Benefit} = P^{\text{Ret,w}}\Delta w + P^{\text{Ret,e}}\Delta e + \phi^{e}\Delta e + \Delta\Pi. \tag{4}$$

The first two terms estimate the average reduction in water and electricity bills from the intervention, where $P^{\text{Ret,w}}$ and $P^{\text{Ret,e}}$ denote the water and electricity retail prices, respectively. Households in BWP's territory face an increasing block rate pricing structure for both water and electricity, complicating our determination of marginal prices. To approximate the marginal price, we calculate each household's cumulative water and electricity use for each calendar month, and assign each household the marginal price associated with the final unit of consumption. The term $\phi^e \Delta e$ accounts for social benefits from electricity conservation. We focus on changes in SO_2, NO_x, $PM_{2.5}$, and CO_2 emissions that occur because of a reduction in upstream generation. To value the avoided pollution from the intervention, we use estimates from Holland et al. (2016) on hourly marginal damages.[15]

Our net benefits measure also includes the net change in utility revenue, $\Delta\Pi$, from the deployment of HWRs. HWRs affect utility revenue along three dimensions: a reduction in water and electricity sales to consumers; a decrease in wholesale water and electricity acquisition costs; and an increase in costs to pay for HWRs. We calculate the change in utility revenue as,

$$\Delta\Pi = (P^{\mathrm{Wh,e}} - P^{\mathrm{Ret,e}})\Delta e + (P^{\mathrm{Wh,w}} - P^{\mathrm{Ret,w}})\Delta w - C,$$

where $P^{\mathrm{Wh,j}}$ are wholesale costs for $j = e, w$; C is the annual cost per household to supply bi-monthly HWRs; and all other terms are as defined in equation (4).[16]

If we substitute the measure of net utility revenue ($\Delta\Pi$) into equation (4), customer bill savings are fully offset by utility revenue losses,

$$\text{Net Benefit} = P^{\mathrm{Wh,e}}\Delta e + P^{\mathrm{Wh,w}}\Delta w + \phi^e \Delta e - C. \tag{5}$$

We choose to measure the savings and losses in retail expenditures because their inclusion makes explicit how our calculation differs from a welfare analysis. In a welfare analysis, we would replace household bill savings with the net change in consumer surplus attributable to

[15] CO_2 damages are valued at the EPA's social cost of carbon, \$41/ton, and marginal local pollutant damages are estimated with the AP2 model, an extension of the model developed by Muller and Mendelsohn (2009), using an \$8.1 million value of a statistical life.

[16] Based on utility audit reports, we assume the price incurred by the utility to purchase a marginal unit of water is \$2.072/HCF. We use hourly, day-ahead wholesale electricity prices for a large aggregation point near our utility in Southern California as our measure of wholesale electricity costs, and an average annual cost per household of \$10 to supply bi-monthly HWRs.

HWRs (Allcott and Kessler, 2015). The implication of this is that the revenue losses incurred by the utility from a reduction in sales no longer falls out of the net benefits calculation.

Panel A of Table 5 presents our results, itemizing each term from equation (4). In columns (1a-3a) of this panel, we assume that the reductions in water and electricity use are uniform across all hours and months. In columns (2b-3b), we assume that the reductions in electricity use are uniform across hours but occur exclusively during the summer of 2015, defined as May to August. All columns report results from the estimation of a modified version of equation (1) in which (i) we condition on baseline electricity or water use and (ii) the dependent variable is the outcome of interest listed in each row.[17] In the first column we restrict our attention to the water savings and revenue losses from the deployment of HWRs. Columns (2a) and (2b) focus exclusively on the net benefits from energy conservation, including the social benefits from reductions in local pollutants and greenhouse gas emissions. Columns (3a) and (3b) aggregate the water and energy savings and losses for each scenario.

The table makes clear the importance of incorporating spillovers into a benefit-cost analysis. The net benefits increase by around 40% from \$2.91 per household to \$4.04 when we account for private and social benefits of the electricity savings over the year, and by 120% to \$6.47 when we consider the electricity savings only in the Summer 2015. On average, households save \$22 and \$3–\$8 on their water and electricity bills, respectively, from the deployment of HWRs though lost utility revenues perfectly offset these savings. Net benefits accrue from a reduction in the cost to purchase water and electricity. Wholesale water savings amount to \$13 per household per year, and electricity savings total at \$0.86–\$1.94 per household per year. The value of the CO_2 reductions from the HWRs is \$0.16-\$0.71 per household per year, and the value from reductions in local pollutants amounts to \$0.11–\$0.91 per household per year.

Our valuation of the electricity spillover in Panel A does not allow for temporal heterogeneity in when these treatment effects occur. Accounting for heterogeneity may substantially affect the value of spillovers because the wholesale price of electricity varies considerably across months of the year and hours of the day, and the marginal source of generation differs

[17]Tables E.1 to E.3 in the Appendix present the regression results.

in emissions rates. To incorporate this form of heterogeneity into our net benefits framework, we follow Novan and Smith (2017) and Boomhower and Davis (2017) and estimate treatment effects for each month of the year and hour of the day. As shown in Panel B of Table 5, accounting for variation in the timing of when electricity reductions occur further increases the net benefits of HWRs by 16% from 4.04 to 4.70. This increase in savings aligns with our expectations because treated households reduced electricity use during the hours of the day when electricity is most expensive to supply, and the marginal source of generation is dirtier.

We view these net benefit estimates as a starting point to quantify the importance of electricity spillovers, and caution against interpreting them as the welfare impacts. For a number of reasons, the net benefits calculated using our framework likely exceed those that would be calculated using a welfare analysis. First, our analysis does not account for potential costs incurred from efforts to conserve water and electricity, or an increase in the moral costs attached to water or electricity consumption. Second, the disparity between marginal prices for electricity and water, and the social marginal costs to supply these goods, indicates that there may be reductions in consumer surplus from conservation. A welfare analysis would account for differences between marginal prices and the social marginal cost of production, something that our framework does not do. Third, even if pricing was efficient, we do not incorporate the lost consumer surplus from a reduction in water and energy consumption. Even with these caveats in mind, our benefit-cost framework provides a relevant and valuable launchpad with which to incorporate spillovers into policy evaluation.

8 Conclusions

This paper evaluates whether behavioral interventions spill over into unintended sectors through the lens of urban water conservation instruments and energy use. To do this, we designed and implemented a field experiment that allows us to measure the effect of HWRs on residential electricity use. We find that water conservation instruments induce conservation beyond the water sector, leading to short-lived reductions in summertime electricity use.

High-frequency data allow us to explore heterogeneity in the timing of the treatment effect and reveal that electricity conservation is most pronounced during peak hours in the summer when electricity is most expensive to provide and marginal emissions from generation are higher.

We formalize two channels through which HWRs could affect electricity use: mechanical complementarities and behavioral choices. Under the former, electricity reductions are explained by actions that use both water and energy, such as doing a load of laundry. While mechanical complementarities account for some of the savings, empirical tests, household survey data, and simulation results all point to the likelihood that HWRs also alter consumer choices about electricity consumption more generally. Treatment households reduce electricity consumption during hours of the day when no water conservation occurs; increase electricity conservation as ambient temperatures increase; and report a higher frequency of engagement in energy conservation actions. Results from an engineering simulation imply that, under plausible scenarios, only 25% of the electricity savings can be explained by mechanical complementarities.

To date, economists have primarily evaluated behavioral interventions using a partial equilibrium framework. The presence of cross-sectoral spillovers points to the importance of broadening the focus in the direction of a general equilibrium framework, both when considering the net benefits of these programs as well as their welfare impacts. In our setting, reductions in electricity use augment the net benefits of the water conservation instruments by 62% from \$2.91 per household to \$4.70 per household. We focus exclusively on one possible cross-sectoral spillover, but spillovers from this intervention may extend beyond energy. Moving forward, we should improve our understanding of the conditions and circumstances under which one would expect cross-sectoral spillovers to occur.

References

Allcott, Hunt. 2011. "Social norms and energy conservation." *Journal of Public Economics* 95 (9):1082–1095.

———. 2015. "Site selection bias in program evaluation." *Quarterly Journal of Economics* 130 (3):1117–1165.

Allcott, Hunt and Judd Kessler. 2015. "The welfare effects of nudges: A case study of energy use social comparisons." Working Paper.

Allcott, Hunt and Sendhil Mullainathan. 2010. "Behavior and energy policy." *Science* 327 (5970):1204–1205.

Andrews, Isaiah and Emily Oster. 2019. "A simple approximation for evaluating external validity bias." *Economics Letters* 178:58–62.

Angelucci, Manuela and Giacomo De Giorgi. 2009. "Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption?" *American Economic Review* 99 (1):486–508.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying odysseus to the mast: Evidence from a commitment savings product in the Philippines." *The Quarterly Journal of Economics* 121 (2):635–672.

Athey, Susan and Guido W. Imbens. 2017. "The state of applied econometrics: Causality and policy evaluation." *Journal of Economic Perspectives* 31 (2):3–32.

Ayres, Ian, Sophie Raseman, and Alice Shih. 2013. "Evidence from two large field Experiments that peer comparison feedback can reduce residential energy usage." *Journal of Law, Economics, and Organization* 29 (5):992–1022.

Balling, Robert and Patricia Gober. 2006. "Climate variability and residential water use in the city of Phoenix, Arizona." *Journal of Applied Meterology and Climatology* 46:1130–1138.

Bell, Stephen H., Robert B. Olsen, Larry L. Orr, and Elizabeth A. Stuart. 2016. "Estimates of external validity bias when impact evaluations select sites nonrandomly." *Educational Evaluation and Policy Analysis* 38 (2):318–335.

Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov. 2015. "The welfare economics of default options in 401(k) plans." *American Economic Review* 105 (9):2798–2837.

Beshears, John, James Choi, David Laibson, Brigitte Madrian, and Katherine Milkman. 2015. "The effects of providing peer information on retirement savings decisions." *The Journal of Finance* 70 (3):1161–1201.

Bonferroni, C.E. 1935. "Il calcolo delle assicurazioni su gruppi di teste." *Tipografia del Senato* .

Boomhower, Judson and Lucas W. Davis. 2017. "Do energy efficiency investments deliver at the right time?" Working Paper 23097, National Bureau of Economic Research.

Brandon, Alec, Paul J. Ferraro, John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer. 2017. "Do The effects of social nudges persist? Theory and evidence from 38 natural field experiments." NBER Working Paper No. 23277.

Brent, Daniel, Joseph Cook, and Skylar Olsen. 2015. "Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials." *Journal of the Association of Environmental and Resource Economists* 2 (4):597–627.

Brown, Jennifer, Tanjim Hossain, and John Morgan. 2010. "Shrouded attributes and information suppression: Evidence from the field." *Quarterly Journal of Economics* 125 (2):859–876.

Bruhn, Miriam and David McKenzie. 2009. "In pursuit of balance: Randomization in practice in development economics." *American Economic Journal: Applied Economics* 1 (4):200–232.

Burlig, Fiona, Louis Preonas, and Matt Woerman. 2020. "Panel data and experimental design." *Journal of Development Economics* .

Carpenter, Christopher S and Emily C Lawler. 2019. "Direct and spillover effects of middle school vaccination requirements." *American Economic Journal: Economic Policy* 11 (1):95–125.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and taxation: Theory and evidence." *American Economic Review* 99 (4):145–1177.

Christensen, Garret and Edward Miguel. 2018. "Transparency, reproducibility, and the credibility of economics research." *Journal of Economic Literature* 56 (3):920–980.

Costa, Dora and Matthew Kahn. 2013. "Energy conservation "nudges" and environmental ideology: Evidence from a randomized residential electricity field experiment." *Journal of the European Economic Association* 1 (3):680–702.

Croson, Rachel and Jen Shang. 2008. "The impact of downward social information on contribution decisions." *Experimental Economics* 11 (3):221–233.

DellaVigna, Stefano and Joshua Pollet. 2009. "Investor inattention and Friday earnings announcements." *The Journal of Finance* 64 (2):709–749.

Dolan, Paul and Robert Metcalfe. 2015. "Neighbors, knowledge, and nuggets: Two natural field experiments on the role of incentives on energy conservation." Working Paper.

Duflo, Esther and Emmanuel Saez. 2003. "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment." *Quarterly Journal of Economics* 118 (3):815–842.

Farhi, Emmanuel and Xavier Gabaix. 2017. "Optimal taxation with behavioral agents." Working Paper.

Ferraro, Paul and Michael Price. 2013. "Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment." *Review of Economics and Statistics* 95 (1):64–73.

Fletcher, Jason and Ryne Marksteiner. 2017. "Causal spousal health spillover effects and implications for program evaluation." *American Economic Journal: Economic Policy* 9 (4):144–166.

Fowlie, Meredith, Catherine Wolfram, C Anna Spurlock, Annika Todd, Patrick Baylis, and Peter Cappers. 2017. "Default effects and follow-on behavior: evidence from an electricity pricing program." Tech. rep., National Bureau of Economic Research.

Frey, Bruno and Stephan Meier. 2013. "Social comparisons and pro-social behavior: Testing 'conditional cooperation' in a field experiment." *American Economic Review* 94 (5):1717–1722.

Gato, Shirley, Niranjali Jayasuriya, and Peter Roberts. 2007. "Temperature and rainfall thresholds for base use urban water demand modelling." *Journal of Hydrology* 337:364–376.

Gerber, Alan and Todd Rogers. 2009. "Descriptive social norms and motivation to vote: Everybody's voting and so should you." *Journal of Politics* 71 (1):1–14.

Handel, Benjamin. 2011. "Adverse selection and inertia in health insurance markets: When nudging hurts." *American Economic Review* 103 (7):2643–2682.

Harding, Matthew and David Rapson. 2014. "Do voluntary carbon offsets induce energy rebound? A conservationist's dilemma." Working Paper.

Holland, Stephen, Erin Mansur, Nicholar Muller, and Andrew Yates. 2016. "Are there environmental benefits from driving electric vehicles? The importance of local factors." *American Economic Review* 106 (12):3700–3729.

Holm, Sture. 1979. "A simple sequentially rejective multiple test procedure." *Scandinavian journal of statistics* 6 (2):65–70.

Ito, Koichiro. 2013. "How do consumers respond to nonlinear pricing? Evidence from household water demand." Working Paper.

———. 2014. "Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing." *American Economic Review* 104 (2):537–563.

Jacobsen, Grant, Matthew Kotchen, and Michael Vandenbergh. 2014. "The behavioral response to voluntary provision of an environmental public good: Evidence from residential electricity demand." *European Economic Review* 56 (1):946–960.

Jessoe, Katrina and David Rapson. 2014. "Knowledge is (less) power: Experimental evidence from residential energy usage." *American Economic Review* 104 (4):1417–1438.

Johnson, Eric J. and Daniel G. Goldstein. 2003. "Do defaults save lives?" *Science* 302 (5649):1338–1339.

Kennedy, Donald. 2004. "The old file-drawer problem." *Science* 305 (5683):451.

Levitt, Steven and John List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *The Journal of Economic Perspectives* 21 (2):153–174.

List, John, Azeem Shaikh, and Yang Xu. 2019. "Multiple hypothesis testing in experimental economics." *Experimental Economics* 22:773–793.

Madrian, Brigitte and Dennis Shea. 2001. "The power of suggestion: Inertia in 401(k) participation and savings behavior." *The Quarterly Journal of Economics* 116 (4):1149–1187.

Meager, Rachael. 2019. "Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments." *American Economic Journal: Applied Economics* 11 (1):57–91.

Miguel, Edward and Michael Kremer. 2003. "Worms: Identifying impacts on education and health in the presence of treatment externalities." *Econometrica* 72 (1):159–217.

Muller, Nicholar and Robert Mendelsohn. 2009. "Efficient pollution regulation: Getting the prices right." *American Economic Review* 99 (5):1714–1739.

Novan, Kevin and Aaron Smith. 2017. "The incentive to overinvest in energy efficiency: Evidence from hourly smart-meter data." *Working Paper* .

Peters, Jörg, Jörg Langbein, and Gareth Roberts. 2016. "Policy evaluation, randomized controlled trials, and external validity – A systematic review." *Economic Letters* 147:51–54.

Rosenthal, Robert. 1979. "The "file drawer problem" and tolerance for null results." *Psychological Bulletin* 86 (3):638–641.

Sexton, Steve. 2015. "Automatic bill payment and salience effects: Evidence from electricity consumption." *Review of Economics and Statistics* 97 (2):709–749.

The Economist. 2017. "Nudge comes to Shove." *The Economist*. May 18, 2017.

Vivalt, Eva. 2019. "How much can we generalize from impact evaluations?" Working paper.

Table 1: Balance Tests

	Control	Treatment	Difference	Households (Total)
Elec Use (kWh/mth)	735.11	729.07	6.03	7,341
	(6.72)	(5.29)	(8.57)	
Summer Elec Use (kWh/mth)	1020.76	1010.81	9.95	7,341
	(9.58)	(7.59)	(12.26)	
Water Use (gals/mth)	12480.36	12333.84	146.52	7,341
	(132.39)	(109.53)	(174.23)	
Year Built	1944.91	1944.87	0.03	7,105
	(0.27)	(0.21)	(0.34)	
Bedrooms	2.91	2.89	0.02	7,105
	(0.02)	(0.01)	(0.02)	
Bathrooms	1.92	1.93	-0.01	7,105
	(0.02)	(0.01)	(0.02)	
Square feet	1621.35	1620.14	1.21	7,105
	(12.66)	(10.15)	(16.31)	
Total Value	391,670	394,493	-2,823.40	7,105
	(4915.66)	(3938.27)	(6329.74)	
Pool (Indicator)	0.23	0.22	0.00	7,105
	(0.01)	(0.01)	(0.01)	

Notes: The first two columns report means for control and treatment households with standard deviations in parentheses below. The third column displays the difference in means between treatment and control, with the standard error reported in parentheses below. The last column displays the total number of households included in the balance test. The assessor data on housing unit attributes are missing for 236 households. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2: Intent-to-Treat Effects

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Electricity Use (kWh/hr)						
WaterSmart	-0.010	-0.002	-0.028**	-0.017**	-0.049**	-0.029**
	(0.011)	(0.004)	(0.014)	(0.008)	(0.023)	(0.015)
Observations	60,071,977	59,922,674	14,403,499	14,374,525	3,637,588	3,630,243
Mean Control Group Use	1.00	1.00	1.26	1.26	1.99	1.98
Panel B: Water Use (gals/hr)						
WaterSmart	-0.585***	-0.528***	-0.571***	-0.448***	-0.722***	-0.624***
	(0.179)	(0.110)	(0.216)	(0.132)	(0.256)	(0.233)
Observations	63,203,481	62,819,470	18,487,941	18,373,022	4,623,663	4,594,925
Mean Control Group Use	12.9	12.9	15.2	15.2	14.1	14.1
Weather Controls	Yes	Yes	Yes	Yes	Yes	Yes
Hour of Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Electricity Use	No	Yes	No	Yes	No	Yes
Hours	All	All	All	All	Peak	Peak
Sample	5/15-5/16	5/15-5/16	5/15-8/15	5/15-8/15	5/15-8/15	5/15-8/15

Notes: The table reports intent-to-treat results from an OLS regression of hourly electricity use (Panel A) and water use (Panel B) on assignment to the treatment. Columns (1) and (2) include all observations from May 15, 2015 to May 31, 2016. Columns (3) and (4) restrict the sample to the summer of 2015 (May 15 to August 30). Columns (5) and (6) further limit the sample to include only peak demand hours (3 PM to 8 PM). Pre-treatment electricity use controls include mean monthly electricity use in the summer, winter and year preceding treatment. Standard errors are clustered at the household. *, **, *** denote significance at the 10%, 5%, and 1% level.

Table 3: Treatment Effect Heterogeneity and Outdoor Temperature
(Dependent Variable: Electricity and Water Use)

	Electricity Use		Water Use	
	(1)	(2)	(3)	(4)
WaterSmart (<65F)	-0.002	-0.000	-0.502**	-0.244
	(0.017)	(0.017)	(0.206)	(0.390)
WaterSmart (65F-70F)	-0.012	-0.011	-0.363*	-0.823**
	(0.009)	(0.009)	(0.185)	(0.353)
WaterSmart (70F-75F)	-0.017**	-0.020***	-0.490***	-0.473
	(0.008)	(0.008)	(0.177)	(0.313)
WaterSmart (75F-80F)	-0.019**	-0.022**	-0.387**	-0.349
	(0.009)	(0.010)	(0.172)	(0.283)
WaterSmart (80F-85F)	-0.019	-0.017	-0.396**	-0.330
	(0.012)	(0.013)	(0.193)	(0.266)
WaterSmart (85F-90F)	-0.027	-0.036*	-0.488**	-0.405
	(0.019)	(0.019)	(0.193)	(0.262)
Observations	8,422,375	5,952,150	10,616,892	7,756,130
Calendar Date FE	Yes	Yes	Yes	Yes
Day of Week	No Watering	Watering	No Watering	Watering
Sample	5/15-8/15	5/15-8/15	5/15-8/15	5/15-8/15

Notes: The table reports intent-to-treat effects across 5F temperature bins. Columns (1-2) and (3-4) report results for electricity and water use, respectively. The data include the period spanning May 15, 2015 to August 31, 2015. 'Day of Week' indicates the days included in our sample; 'No Watering' restricts the sample to days of the week when outdoor watering was banned, and 'Watering' restricts the sample to days when outdoor watering was allowed. All regressions include controls for the temperature bins 65F-70F, 70F-75F, 75F-80F, 80F-85F and $\geq$ 85F, hourly outdoor precipitation, and calendar date fixed effects. Pre-treatment controls include mean monthly use in the summer, winter and year preceding treatment. Standard errors in parentheses are clustered at the household. *, **, *** denote significance at the 10%, 5%, and 1% level.

Table 4: Household Energy-Water Use Appliance Model:
Electricity Savings (kWh/year)

Scenario	Mean	5th Percentile	95th Percentile	% TE Explained
All appliances	39.09	37.85	40.36	26.23
Clothes washer, dishwasher, and dryers only	139.73	135.86	143.68	93.78
All appliances: Electric water heater	119.63	117.95	121.33	81.43
Pool and spa use only	766.12	731.96	800.74	514.17

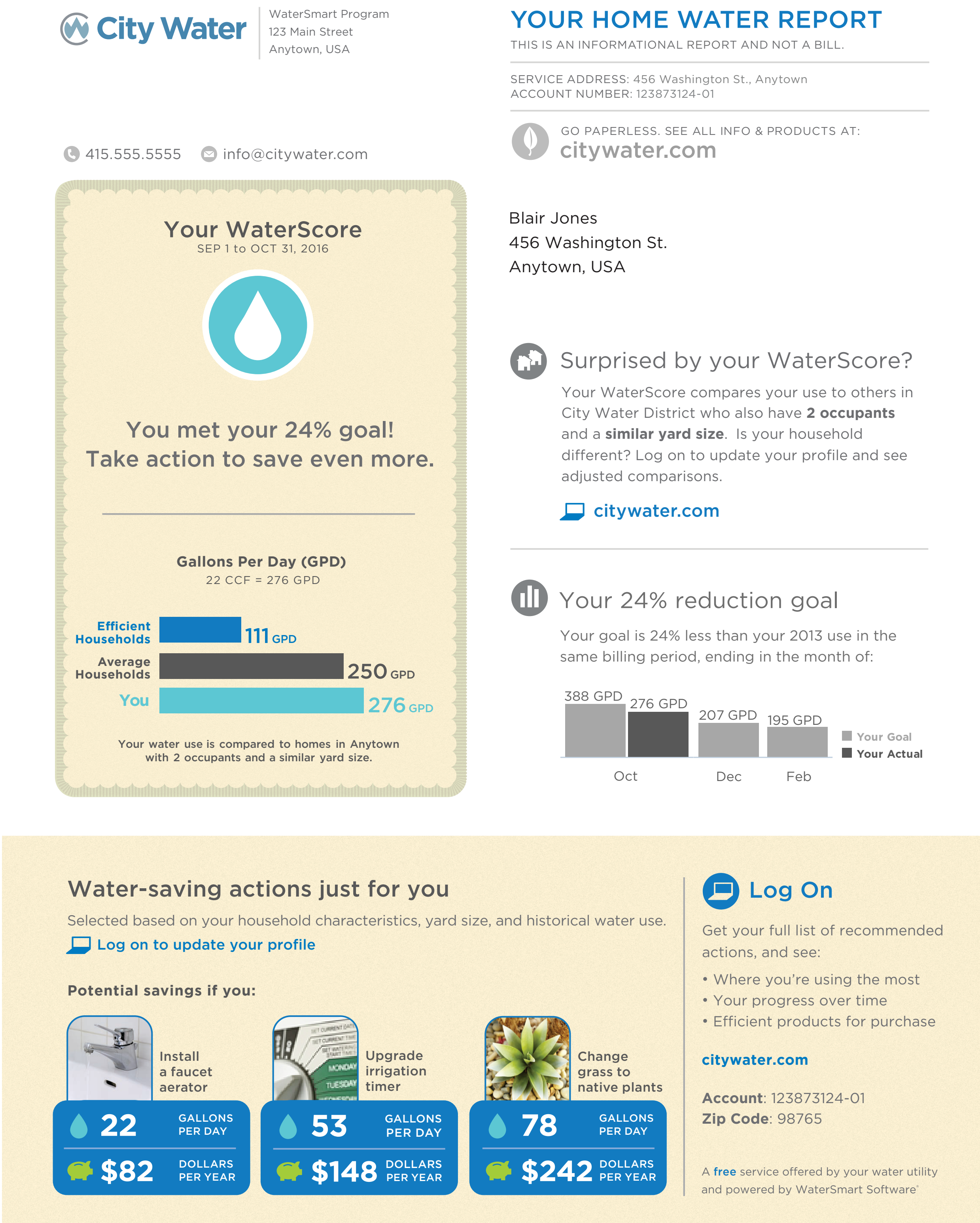
Notes: The table presents the mean, 5th percentile, and 95th percentile of the estimated electricity savings for each scenario in kWh/year. The last column reports the percent of the estimated electricity treatment effect (TE) explained by the mean electricity savings in each scenario. The estimated electricity TE is 149 kWh/year on average for each household based on our estimate in column (4) of Table 2.

Table 5: Net Benefits of Home Water Reports ($/Household)

		All Year		Summer 2015	
	Water	Electricity	Total	Electricity	Total
	(1)	(2a)	(3a)	(2b)	(3b)
Panel A: Year (Summer) Average Treatment Effects					
(+) Customer Bill Savings	21.65	3.17	24.82	8.25	29.87
(-) Utility Revenue Loss	21.65	3.17	24.82	8.25	29.87
(+) Utility Wholesale Expenditure Savings	12.91	0.86	13.77	1.94	14.85
(+) Local Externalities Benefits	–	0.11	0.11	0.91	0.91
(+) Global Externalities Benefits	–	0.16	0.16	0.71	0.71
(-) HWR Cost	10	–	10	–	10
(=) Private Net Benefits	2.91	–	3.77	–	4.85
(=) Private & Public Net Benefits	2.91	–	4.04	–	6.47
Panel B: Monthly-by-Hour Average Treatment Effects					
(+) Customer Bill Savings	21.65	4.53	26.18	–	–
(-) Utility Revenue Loss	21.65	4.53	26.18	–	–
(+) Utility Wholesale Expenditure Savings	12.91	1.26	14.17	–	–
(+) Local Externalities Benefits	–	0.26	0.26	–	–
(+) Global Externalities Benefits	–	0.27	0.27	–	–
(-) HWR Cost	10	10	10		
(=) Private Net Benefits	2.91	1.26	4.17	–	–
(=) Private & Public Net Benefits	2.91	1.79	4.70	–	–

Notes: Panel A presents results from estimating equation (1), replacing the dependent variable with each term in equation (4) and scaling up to reflect the benefits/costs on an annual, per household basis. We present two results for electricity in Panel A. In 2a, we estimate and scale up the impact over the entire year of the intervention. In 2b we use only data from May to August 2015, scaling up the impact only by those four months, and assume the electricity impact is zero thereafter. Panel B presents results where we estimate the impacts of each by hour-of-day and month-of-year to account for temporal variation in the treatment effect and value of external benefits.

Figure 1: Home Water Report



City Water

WaterSmart Program
123 Main Street
Anytown, USA

415.555.5555 info@citywater.com

YOUR HOME WATER REPORT

THIS IS AN INFORMATIONAL REPORT AND NOT A BILL.

SERVICE ADDRESS: 456 Washington St., Anytown
ACCOUNT NUMBER: 123873124-01

GO PAPERLESS. SEE ALL INFO & PRODUCTS AT:
citywater.com

Blair Jones
456 Washington St.
Anytown, USA

Your WaterScore

SEP 1 to OCT 31, 2016

You met your 24% goal!
Take action to save even more.

Gallons Per Day (GPD)
22 CCF = 276 GPD

Efficient Households 111 GPD
Average Households 250 GPD
You 276 GPD

Your water use is compared to homes in Anytown with 2 occupants and a similar yard size.

Surprised by your WaterScore?

Your WaterScore compares your use to others in City Water District who also have **2 occupants** and a **similar yard size**. Is your household different? Log on to update your profile and see adjusted comparisons.

citywater.com

Your 24% reduction goal

Your goal is 24% less than your 2013 use in the same billing period, ending in the month of:

388 GPD 276 GPD 207 GPD 195 GPD
Oct Dec Feb
Your Goal
Your Actual

Water-saving actions just for you

Selected based on your household characteristics, yard size, and historical water use.

Log on to update your profile

Potential savings if you:

Install a faucet aerator: 22 GALLONS PER DAY, $82 DOLLARS PER YEAR

Upgrade irrigation timer: 53 GALLONS PER DAY, $148 DOLLARS PER YEAR

Change grass to native plants: 78 GALLONS PER DAY, $242 DOLLARS PER YEAR

Log On

Get your full list of recommended actions, and see:

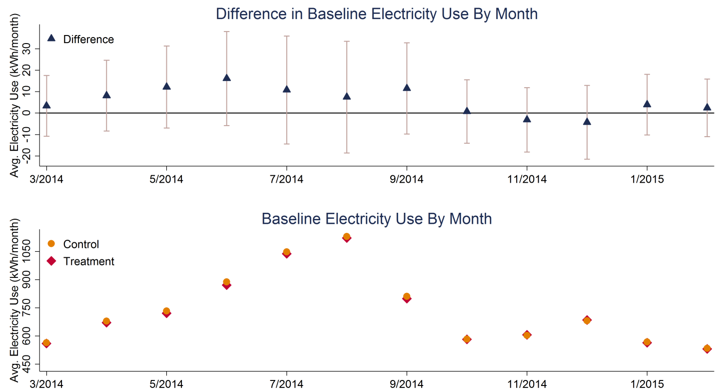
- Where you're using the most
- Your progress over time
- Efficient products for purchase

citywater.com

Account: 123873124-01
Zip Code: 98765

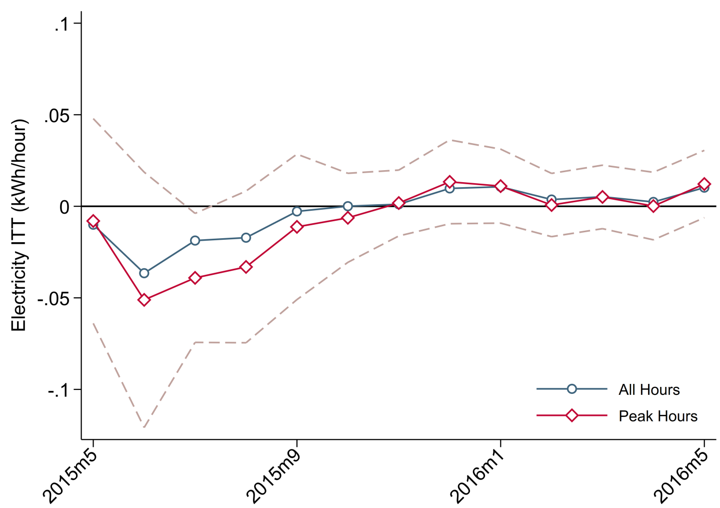
A free service offered by your water utility and powered by WaterSmart Software

Figure 2: Balance: Pre-treament Electricity Use by Month (kWh/mth)



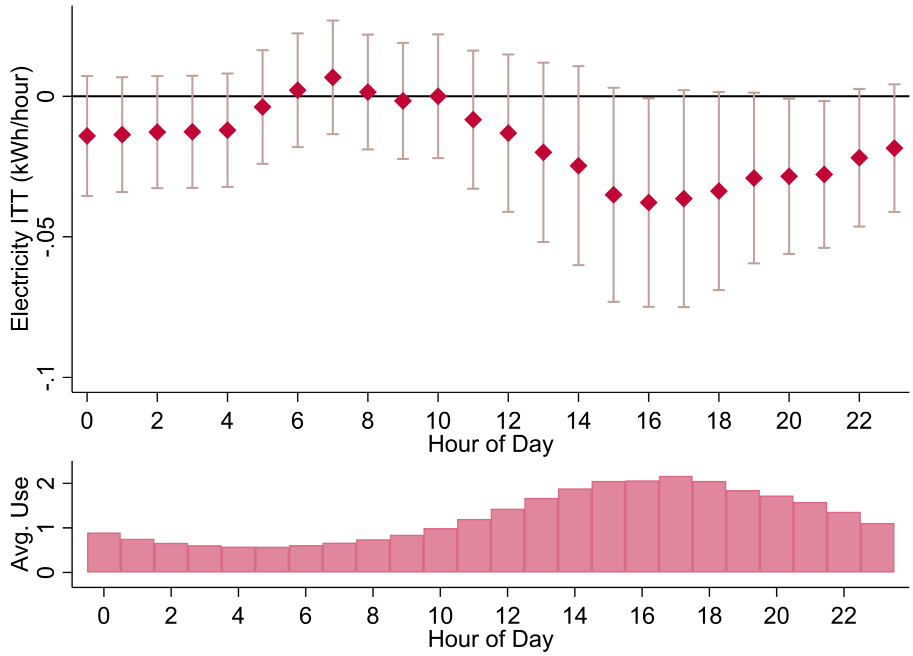
Notes: The upper portion of the figure plots the difference in mean monthly electricity use across control and treatment. The vertical lines are the 95% confidence intervals. The lower portion of the figure plots mean electricity use in a given month for control and treatment households.

Figure 3: Electricity Intent-to-Treat Effects Over Time

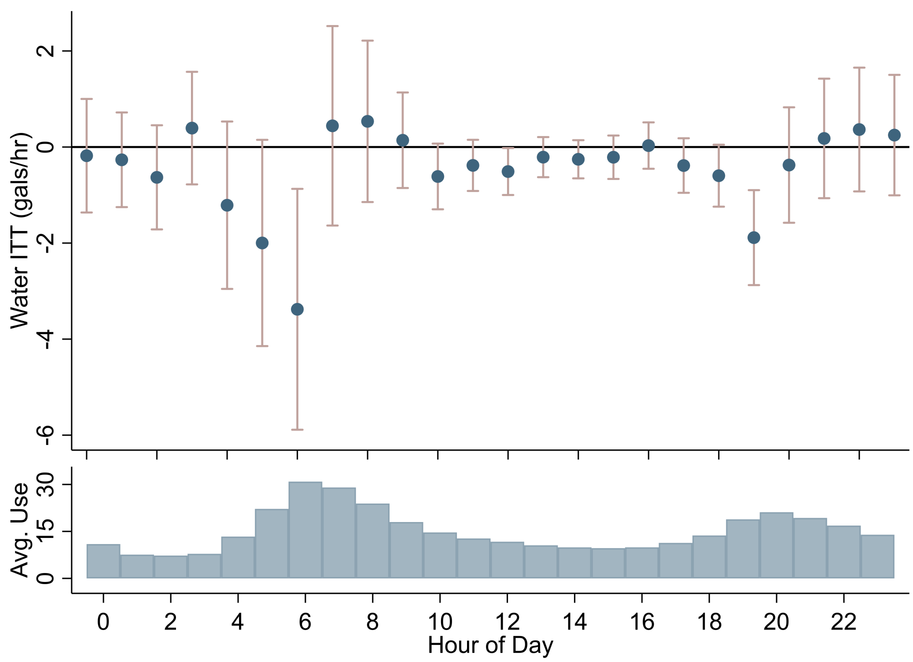


Notes: The figure plots intent-to-treat electricity effects (kWh/hr) for each treatment month. The blue circles and red diamonds plot estimates using all hours of the day and peak hours, respectively. The dashed lines are 95% confidence intervals for the peak hours treatment effects. All regressions include weather controls, hour of day fixed effects, calendar date fixed effects, and household pre-treatment electricity use controls.

Figure 4: Heterogeneity by Hour of Day



(A) Electricity Intent-to-Treat Effects (kWh/hr)



(B) Water Intent-to-Treat Effects (gals/hr)

Notes: The figure plots hourly intent-to-treat electricity effects in panel (A) and water effects in panel (B) from assignment to the WaterSmart treatment. The estimated treatment effects are denoted as diamonds in panel (A) and circles in panel (B). The 95% confidence intervals and mean hourly use for control households are plotted as well. All regressions include weather controls, calendar date fixed effects, and mean monthly electricity (water) use in the summer, winter and year preceding treatment.