LSU | E. J. Ourso College of Business
Department of Economics

DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

Are Normative Appeals Moral Taxes? Evidence from A Field Experiment on Water Conservation

Daniel A. Brent
Louisiana State University
Corey Lott
University of California, Santa Barbara
Michael Taylor
University of Nevada, Reno
Joseph Cook
University of Washington
Kim Rollins
University of Nevada, Reno
Shawn Stoddard
Truckee Meadows Water Authority

Working Paper 2017-07
http://faculty.bus.lsu.edu/workingpapers/pap17_07.pdf

Department of Economics
Louisiana State University
Baton Rouge, LA 70803-6306
http://www.bus.lsu.edu/economics/

Are Normative Appeals Moral Taxes? Evidence from a Field Experiment on Water Conservation

Daniel A. Brent*[a], Corey Lott[b], Michael Taylor[c], Joseph Cook[d], Kim Rollins[c], and Shawn Stoddard[e]

[a]Department of Economics, Louisiana State University
[b]Department of Economics, University of California Santa Barbara
[c]Department of Economics, University of Nevada Reno
[d]Evans School of Public Policy & Governance, University of Washington
[e]Truckee Meadows Water Authority

June 2017

Abstract

We investigate how normative appeals for water conservation drive behavioral change using a large-scale field experiment. Using a new social comparison that reduces the correlation between pre-treatment consumption and the difference from the peer group, we isolate the normative component of the message. The strength of the message, which we define as a household's performance relative to a peer group, is a primary driver of social comparisons' efficacy, consistent with social comparisons imposing a moral cost on excess consumption. Relative to a nudge highlighting financial savings, social comparisons generate less persistent water savings and are more dependent on multiple mailers.

Keywords: behavioral interventions, social norms, field experiments, water conservation, water demand
JEL Classification Numbers: D12, C93, H42, L95, Q21, Q25.

*Corresponding Author: Daniel Brent, dbrent@lsu.edu, 2300 Business Education Complex, Department of Economics, Louisiana State University, Baton Rouge, LA 70803.
The authors are grateful to the Truckee Meadows Water Authority (TMWA) for financial support and all TMWA staff who contributed to the research through information about water demand issues, data management, and logistical support. We especially appreciate the help of Laine Christman. Brent acknowledges from the Cooperative Research Centre for Water Sensitive Cities (CRC grant number 20110044).

1 Introduction

A number of high-quality, randomized experiments have established that information treatments that compare an individual household's electricity or water consumption with that of a peer group cause significant reductions in consumption.[1] In general, most of these social comparison studies have also found that the average treatment effects are driven largely by effects among customers with higher pre-treatment consumption (Allcott, 2011; Dolan and Metcalfe, 2015; Ito et al., 2015; Ferraro and Price, 2013; Ferraro and Miranda, 2013; Brent et al., 2015). Why and how do these comparisons work? One explanation is that these comparisons invoke social norms: a consumer sees that his water consumption is higher than similar neighbors and feels guilty about his "overconsumption. In response to such a message, the consumer may impose a "moral tax" on consumption. Conversely, consuming less than ones peers can lead consumers to impart a "moral subsidy" (Levitt and List, 2007; Ferraro and Price, 2013; Allcott and Kessler, 2015). Framing the norm as a tax is appealing to economists, because it can be incorporated into welfare analyses of various price and non-price approaches for achieving conservation in the face of scarcity (as in the case we discuss) or reductions in associated consumption externalities. It also has the logical extension that the behavioral response should vary according to the level of such a moral tax, which we refer to as the strength of the normative message. When the moral tax is higher (one is consuming *much* more than one's peers), the effect of receiving information highlighting this should be larger than for those receiving a smaller, or negative, moral tax.

A second behavioral explanation draws from the more conventional household production approach (Becker, 1965), and recognizes that social comparisons also provide households with financially-useful information. Households informed that their water use is 50% higher than similar neighbors might also receive a signal that they can achieve the same utility from using water or electricity at lower costs by installing efficient appliances or making behavioral changes Ferraro and Price (2013). This updating of beliefs and re-optimization more generally fits the emerging rubric of correcting "internalities", or failures of consumers to internalize all of the private costs of their actions (Allcott et al., 2014; Allcott and Sunstein, 2015; Allcott and Taubinsky, 2015). Prominent examples of internalities in the water and energy sectors include: imperfect information about the costs of water/energy consumption (Allcott and Taubinsky, 2015); dynamic inconsistencies in decision-making (Allcott et al., 2014); lack of salience of infrequent or automatic billing (Sexton, 2015; Wichman, 2016); and confusion about nonlinear price structures (Ito, 2014; Wichman, 2014). Which of these two behavioral explanations best fits the pattern of results found in social comparisons?

The answer to this question is important from a welfare perspective. Information

[1] See among others Allcott (2011); Allcott and Rogers (2014); Ayres et al. (2013); Costa and Kahn (2013) for energy and Ferraro et al. (2011); Ferraro and Price (2013); Brent et al. (2015) for water.

treatments to re-optimize and correct internalities are always Pareto-improving in markets without externalities, since some households can achieve the same utility at lower cost, and all other households are no worse off.[2] In markets with negative externalities, information treatments operating through re-optimization that cause an aggregate decrease in consumption, such as most water and energy social comparisons, will have even larger positive welfare gains due to reducing external costs. In contrast, if a social comparison works purely by imposing a moral tax, there may be a significant fraction of households for whom the level of their moral tax exceeds the correct Pigouvian tax. Social comparisons might thus induce over-abatement and reduce welfare overall. Allcott and Kessler (2015) demonstrate theoretically that the social welfare effects of a nudge can vary substantially depending on the prevailing behavioral mechanism.

Testing these behavioral mechanisms has been empirically challenging because of a design feature of existing social comparisons: customers with above-average pre-treatment consumption are much more likely to receive a message that they are consuming more than their comparison group. In fact, water use and the type of message received are perfectly correlated in studies where the comparison group is simply a utility-wide average consumption. These customers receive a higher moral tax, but they also receive a stronger signal to re-optimize consumption (i.e. their water bill is much higher than their neighbors bills). Why? In our setting in Reno, Nevada, as in many cities in arid climates, summer (May-Oct) household water use is driven predominantly by outdoor water use on lawns and gardens. Consumption in the summer is generally four times higher than in the winter. Thus, the most impactful water conservation decisions are related to landscape changes (i.e. xeriscaping) and improvements in irrigation efficiency. Optimizing the amount and timing of irrigation water can achieve the same green landscape with less water. If a financial "re-optimization" channel predominates, a signal that household water use is higher than similar neighbors is likely a signal that outdoor water use is driving this difference. A re-optimization of outdoor water use, for example by installing an irrigation timer, may lower one's bill substantially.[3] In other words, with a traditional social comparison, high water users are faced with a higher moral tax but are also very likely the ones who might save the most money at the lowest utility-cost when re-optimizing.

Our primary contribution is the introduction of a novel social comparison treatment that allows us to better isolate a moral tax behavioral channel by decoupling pre-treatment water usage from the type of message sent to households, and thus the level of moral tax imposed. We do this by framing the comparison not in levels of con-

[2]This argument neglects the typically small cost of generating the information, which could in theory be passed along in higher water or electricity rates.

[3]These are devices that, if used properly, can be set to apply the correct amount of water at the correct time and thus keep a lawn healthy without overwatering and unnecessarily increasing one's water bill.

sumption (i.e. thousands of gallons of water used), as in traditional social comparisons, but in percentage reductions from a prior baseline compared to the percentage reduction of similar neighbors. More concretely, the water utility in Reno (Truckee Meadows Water Authority, or TMWA) asked all customers to reduce their water consumption in summer 2015 by 10% compared to summer 2013 in response to a drought (discussed more below). Our novel information treatment informs customers by what percentage they have reduced water consumption compared to 2013 as compared to the reduction achieved by similar neighbors. We argue that this treatment still imposes a moral tax by signaling whether one is "doing their part" to manage the drought, but provides much less financially-useful information.[4]

We analyze this new social comparison in a utility-scale field experiment among TMWA customers that sent one of five different mailers to single-family customers in TWMA's service area. Approximately 4,300 households were included in each of the five treatment groups, with 21,552 in the control group. Two of these include normative appeals: one is a traditional social comparison in terms of total gallons used by the household relative to a peer group, while the other uses the social comparison in percentage terms as described above. The third treatment provides rate information and frames conservation in terms of households' expected monetary savings. All of the mailers in the study describe TMWA's goal for each household to use 10% less water for each month of the summer of 2015 relative to the summer of 2013 to cope with a temporary drought. In this paper, we focus on three of the five mailers.

Each of the three treatments generate statistically significant average treatment effects (ATEs) of roughly 1.5%. Although the individual ATEs are not statistically different from each other, we find several patterns in our results which are consistent with social comparisons operating *at least partly* by raising the moral cost of consumption. First, we find that the strength of the normative appeal, defined as the difference between a customer's consumption (or percentage reduction) and that of similar neighbors, is a strong driver of differences in responses to social comparisons. Importantly, this result holds when the strength of the normative appeal is decoupled from pre-treatment consumption. The conservation rate social comparison shows the same pattern of response for low and high water users based on pre-treatment consumption, although the effect is magnified for high users.[5] Second, linking our experimental results with survey data

[4]To see this, imagine you are told that you are using 6,000 gallons per month and that your neighbors with a similar lot size used 4,000 gallons. This might indicate that you could reduce water consumption by 2,000 gallons, and thus your bill, while being able to maintain a yard that looks similar to your neighbors. If, on the other hand, you are told that similar neighbors reduced their consumption by 8% compared to 2013, while you reduced by only 4%, you would not learn much to update your beliefs about whether you are optimizing how you use water: you don't know the baseline consumption level of the comparison household.

[5]Randomization was balanced by design on pre-treatment water consumption, but we could not know a household's conservation rate in advance. The conservation rate comparison treatment was not therefore guaranteed to be balanced on percentage reduction and conditional average treatment effects

collected after the conclusion of the field experiment, we find that pro-social households were more responsive to the social comparison treatments, similar to Costa and Kahn (2013) and Bolsen et al. (2014).

Third, by randomizing whether households received one or two treatment letters and examining treatment effects after mailers stop arriving, we find that framing conservation in terms of monetary savings leads to more persistent treatment effects than our normative social comparisons. The two normative treatments largely replicate the action and backsliding pattern found by Allcott and Rogers (2014) (though in less temporal granularity), where the initial effect of a mailer wanes over time before increasing upon the receipt of a new mailer. The financially-oriented treatment causes a persistent effect and a second mailer has no additional effect. This is consistent with the findings by Ito et al. (Forthcoming) that pecuniary incentives lead to more persistent effects than moral suasion. It is also consistent with consumers being more likely to re-optimize to address internalities when cued with monetary savings information, especially if this re-optimization involves capital investments (e.g. low-flow toilets or xeriscaping) or more permanent behavioral changes (e.g. irrigation timers). Consumers cued with normative appeals may draw on actions that are intent-oriented (Attari, 2014) but which are more likely to be transient (e.g. shorter showers). Our survey results also support this: consumers who reported previous conservation actions are less responsiveness to the monetary treatment, suggesting that those households had already re-optimized and exhausted any low-cost conservation opportunities.

Previous studies have attempted to investigate whether different types of nudges trigger different mechanisms, but none has conclusively distinguished between moral and financial motivations to normative appeals. Ferraro and Price (2013) compare social comparisons and generic pro-social appeals, both based on a moral motivation, to show that social comparisons generate greater conservation responses; however, they do not consider non-moral based mechanisms, such as addressing consumption internalities. Ferraro and Miranda (2013) use the traditional quantity-based social comparison and attempt to distinguish between financial versus moral motivations by comparing treatments responses by households just above and below thresholds in increasing block rate pricing structures, but fail to find that financial gain influences conservation response.[6] Pellerano et al. (2015) find weak evidence that adding information about financial savings to social comparisons crowds out moral motivations, as postulated by Gneezy et al. (2011). Allcott and Kessler (2015) find that 35% actually have negative willingness to pay for continuing

should be interpreted cautiously. As we demonstrate below, however, the treatment achieved remarkable balance in conservation rates among baseline water use. In other words, some low baseline water users reduced their consumption by a higher amount than their peers, and some high water users reduced by a smaller fraction than their peers.

[6]Differences in savings between tiers tend to be small and not all consumers are aware of marginal changes in the price (Ito, 2014; Wichman, 2014).

to receive social comparison, indicating that moral costs are a likely mechanism. However, their study takes place after the households had already received one year's worth of social comparisons, so the novel information component that helps correct internalities may have already been exhausted.

2 Background & Experimental Design

In 2015, in response to low snowpack and expected drought conditions during the summer irrigation season, TMWA launched a major media campaign through print, radio, TV, social media, and billboard messages requesting each TMWA customer to use 10% less water from May through September 2015, relative to their water use during the same months in 2013. The comparison year was 2013 because TMWA had asked for a 10% reduction during the latter part of the prior summer (July through September of 2014). Such a conservation request was uncommon in the region prior to 2014, however; the last time TMWA had requested customers to reduce water consumption to address drought was in 1992. To complement the media campaign, we worked with TMWA to design a randomized control trial to test the effectiveness of five different personalized letters mailed to residential single family households.

The TMWA request for a 10% system-wide reduction was met and surpassed. The conservation request ended officially at the end of September. In November-December the utility surveyed a sample of its customers (with email addresses) with the primary goal of assessing the effectiveness of their media campaign. The survey included questions about what customers did to conserve water, whether their actions were considered to be impositions, what prevented them from doing more to conserve water, and what more they would have been willing to do if called upon to do so. The survey responses included 2,544 households that were in our experimental groups, approximately 10% of the sample size in our experiment. In their survey sample, TMWA included all households from our experimental sample that had provided TMWA with an e-mail address. We were able to use these data to provide corroborating evidence regarding customers' motivations and mechanisms of response to the treatment letters they received. We discuss relevant TMWA survey outcomes in this paper in the context of interpreting our main experimental results.[7]

2.1 Description of Treatments

This article focuses primarily on three of the five treatments (denoted T1 through T5): a rate information treatment focused on financial savings (T3) and two social comparison treatments (T4 and T5). However, since the treatments of interest contain components of the first two treatments we present a brief description of all five treatments (Table 1; the appendix includes example components of the five mailers). Every letter began: "Because of the extended drought in Northern Nevada, we are asking all of our customers

[7] A summary of selected survey results is available through the TMWA staff report (Christman, 2016).

to reduce water use by at least 10% this summer compared to summer 2013 - the last summer before TMWA started asking for summer water use reductions." All letters also included the statement: "Since TMWA customers use on average about four times more water in summer than in the winter, we expect that for most customers the easiest way to achieve this reduction is to adjust outdoor watering."

Treatment 1 provided households with six tips that the TMWA media campaign publicized for how to reduce outdoor water consumption, similar to Ferraro et al. (2011). This letter was not customized to report on individual household water use. The six tips were also printed on the reverse side of the other four mailers (T2-T5).

Treatment 2 augmented the generic tips with personalized information about the customer's water use, with a title introducing the letters that read: "Below is your customized water use report." The T2 letter included a figure that displayed the customer's water use in thousands of gallons (kgal) for May through September of 2013 and also their water use in 2015 for each month from May up to the last month billed before the letter was sent out (Appendix Figure A.1 shows the mailer). This figure and accompanying descriptive text was included with Treatments 2 through 5. Therefore all our treatments of interest include water conservation tips and personalized historical water use information.

Treatment 3 (T3) contained the same components as T2, with the additional message "Saving water saves you money", a figure displaying (a) the rate structure with tiers and price for each tier, (b) the customer's water use in kgal within TMWA's increasing-block rate structure for the last month billed in 2015, and (c) the upcoming month's target of 10% less water than the same month in 2013 within the rate structure. The letter also provided the monetary savings that the customer could expect from meeting this goal (see Appendix Figure A.2). We refer to this treatment as the rate treatment and financial treatment throughout the remainder of this paper.

Treatment 4 (T4) provided the same information as T2 with the additional message "How does your water use compare?" and a figure comparing the customer's total water use in thousands of gallons for the last billed month to the median water use of a peer group consisting of single-family residences in their neighborhood with similar yard size and number of bedrooms. This treatment essentially reproduces the standard social comparison used in the OPower studies on energy and the Cobb-County (Ferraro et al., 2011) and Watersmart (Brent et al., 2015) experiments in water (see Appendix Figure A.3).

Treatment 5 (T5) included the same information as T2 with the additional message and figure providing a similar comparison between households as in T4, but instead expressed in terms of relative percent performance towards achieving the 10% goal compared to 2013 water use (Appendix Figure A.4). This treatment decouples the strength of the normative appeal from the level of water use, since low water users may not have

Table 1: Information Included in the Five Treatments

Information Components Included in Mailers*	Treatment (Mailer)				
	T1	T2	T3	T4	T5
The message "Helping our region deal with drought: What you can do" with a sheet showing 6 low cost tips to reduce outdoor water use	X	X	X	X	X
The message "What is your 10% Goal?" with a water use bar graph showing in 1000's of gallons the home's: • 2013 water use for May through September, • Target water use for each month in 2015 as 10% less, • 2015 actual monthly water use, up to last month billed.		X	X	X	X
The message "Saving water saves you money" with rate structure ** graphic showing home's water use in 1000's of gallons by tier/price: • for last month billed in 2015, • for same month in 2013, • for target goal of 10% less water used relative to 2013			X		
The message "How does your water use compare?" with a comparison of customer's 2015 last month billed water use in 1000's of gallons with similar neighborhood homes.				X	
The message "Are you doing your part?" with a comparison of customer's 2015 last month billed water use in terms of % change from 2013 with similar neighborhood homes.					X

*See Figures A.1 - A.4 in the Appendix an example of each information component of each treatment. T2 through T5 Mailers included the title: Below is your customized water use report.
**TMWA had recently merged with two other small regional utilities, Washoe County and South Truckee Meadows Groundwater Irrigation District. In 2015 all customers were subject to the rate structures that they had before the merger. Customers receiving this treatment were shown the tariff structure relevant to them.

conserved water and high users may have conserved a large amount in percentage terms, enabling us to isolate the impact of the strength of the message after conditioning on water consumption. The comparison group was identified in the same way as Treatment 4.

We also include injunctive norms for T4 and T5 in the form of a message. The message "Keep up the good work" was included for residences that had met their 10% goal in the last month billed. Households that did not meet their 10% goal in the previous month received the message "As a reminder TMWA is asking all customers to do their best to save at least 10% this summer. Please do your part to help with drought." We did not use emoticons or "smiley faces" as in Schultz et al. (2007).

2.2 Randomization

Our sample frame included 42,703 eligible[8] single family homes. We then randomly assigned each of these households to either the control group or one of the five treatment groups, with randomization blocks defined by billing cycles, rate schedule and frequency of recorded meter data (i.e. monthly, daily, or hourly, though all customers only receive

[8]Specifically, we included homes that (i) had metered water service; (ii) used enough water during at least one month of the 2013 irrigation season to exceed the tier 1 limit (6,000 gals), indicating some outdoor water use; (iii) had lived at their current residence since April 2013, and therefore had summer 2013 bills for comparison; (iv) had a billing address that corresponded with the residential service address to eliminate rental occupants and other users who may not pay for water or have limited control over water use at the residence; (v) had a 2-inch service main or smaller, excluding unusually large water users; (vi) live within one of the targeted bill cycle regions (some regions were excluded because they had a low number of single-family households, see appendix); and (vii) had nonzero water use during each month of the 2013 irrigation season (May-September) and pre-treatment months during the 2015 irrigation season (May-July) to exclude homes that were unoccupied for an extended period of time.

Table 2: Total Treated Households by Month/Treatment

Treatment Type	July Only	August Only	Both Months	Total
T1: Tips Sheet Only	1,420	1,410	1,402	4,232
T2: Tips + Water Use History	1,414	1,411	1,412	4,237
T3: Tips + History + Rate Information	1,413	1,411	1,410	4,234
T4: Tips + History + Social Norms (Gallons)	1,419	1,416	1,396	4,231
T5: Tips + History + Social Norms (Percent)	1,420	1,403	1,394	4,217
Total	7,086	7,051	7,014	21,151

monthly usage totals). The Appendix provides more detail on our randomization procedure and the process of generating the mailers. In total, 21,151 treatment households were assigned to receive mailers (Table 1). In addition, we randomized whether households received one or two mailers. A total of 7,086 households were assigned to receive a single mailer in July (using June consumption as the last month billed), 7,051 received a single mailer in August, and 7,014 received mailers in both July and August.

Table 3 shows that, in aggregate, the randomization achieved very strong balance on observables. Additionally, Tables A.1-A.6 in the Appendix show that the experimental sample is balanced on pre-treatment consumption for each treatment, across treatments, and within deciles of pre-treatment consumption. Figure A.5 graphically displays the densities of pre-treatment consumption for the pooled treatment, each of the five individual treatments, and the control group. In addition to achieving balance on average pre-treatment consumption, Figure A.5 shows the treatments are balanced across the full distribution of pre-treatment consumption. The graphical evidence is formalized by non-parametric Kolmogorov-Smirnov tests (Table A.4 in the Appendix) that fail to reject the null of equality of distributions for pre-treatment consumption across the control and the pooled treatment as well as each treatment individually. Our sample is well balanced by design, which allows us to make valid inferences for the conditional average treatment effects within subgroups, particularly subgroups that are functions of the normative appeal, which depends on pre-treatment consumption.

Table 3: All Treatments Balance on Observables

	Control Mean	Treatment Mean	Difference	(p-value)
2013 Water	23.56	23.55	0.02	0.89
2015 Water	16.90	16.99	-0.09	0.37
Summer Water	21.63	21.70	-0.07	0.52
Winter Water	8.15	8.15	-0.01	0.88
Year Built	1,987.61	1,987.67	-0.05	0.77
Appraised Value	214.85	214.87	-0.02	0.99
Bedrooms	3.37	3.37	0.01	0.43
Lot Acre	0.27	0.27	-0.00	0.91
Yard Acre	0.22	0.22	-0.00	0.95
Build Sq. Ft.	1,985.10	1,991.07	-5.98	0.41
Bathrooms	2.19	2.20	-0.00	0.56

Note: 21,552 Control Observations and 21,151 treated observations, p-value is based on two-sided t-test

3 Methodology

The primary variable of interest is monthly household consumption, obtained from TMWA billing records, expressed in average gallons per day (GPD). We calculate GPD by dividing total billing cycle usage by the number of days in that billing period to avoid problems with billing periods of different lengths. The specification in our regression analysis uses "normalized GPD" as the main dependent variable; every customer's GPD is divided by the average control group consumption across the experimental period (July-September 2015) following Allcott (2011). This allows the regression coefficients to be interpreted as the average percent change in consumption, while preserving the treatment effect of very high water users, which the logarithmic transformation of consumption would dampen. Our specification is:

$$y_{it} = \alpha + \gamma_l T_{i,l} + \beta \mathbf{x_{it}} + \epsilon_{it} \tag{1}$$

where y_{it} is normalized GPD, $T_{i,l}$ is a dummy variable for the pooled treatment and each of the five treatment letters ($l = Pooled, 1, 2, ..., 5$), and x_{it} is a vector of control variables. We restrict our sample to the post-intervention period, which comprises the billing months of August, September, and October 2015. While treatment is exogenous by virtue of the randomization, including control variables increases the precision of the estimates. All regressions therefore include average consumption during irrigation seasons prior to the intervention, billing cycle and month fixed effects, and average daily temperature and days of precipitation during the billing cycle. We matched daily weather data from the NOAA weather station at Reno-Tahoe Airport to the exact dates of each customer's water bill to calculate the weather variables. Robust standard errors are clustered at the household level.

3.1 Identifying the Effect of Difference from the Peer Group

If normative appeals work by increasing the moral cost of consumption, then we should see stronger treatment effects for households who received a "stronger" social norms and thus a higher moral cost. We define the magnitude of the moral cost as the difference between a household's performance (consumption level or percentage conservation) and the performance of the relevant comparison group. This is one of the cases nested within the consumer utility model of social comparisons presented in Allcott and Kessler (2015), where consumers face a constant marginal moral cost for "inappropriate" consumption. Households consuming further above the norm receive a stronger normative appeal.

As discussed in the introduction, a key feature affecting the interpretation of behavioral responses to social comparisons is that consumers with high pre-treatment consumption are more likely to receive a stronger normative (and negative) message, but they are also likely to have better opportunities to reduce water use at low welfare cost. Figure 1 illustrates the existing correlation between baseline water use and the strength of the social norm in traditional social comparisons (our Treatment 4), and demonstrates that our experimental design successfully decouples the two in our new treatment (T5). In each of the two panels (panel (a) for the comparison in gallons and panel (b) for the comparison in percentage reduction, or conservation rate), we partition households into quartiles of baseline water use (displayed on the x-axis). Within each of those quartiles, we further partition households into quartiles of the difference between a household's level of consumption (or conservation rate) and that of its peer group. Since comparisons are based on median consumption (or percent conservation) within the peer group, the first two quartiles (Q1:Much better and Q2:Better) are households who are doing better than their peer group, and the upper two quartiles (Q3:Worse and Q4:Much Worse) represent households who are doing worse.

Panel (a) of Figure 1 shows that for the traditional social comparison in gallons most low water users (Q1 on the x-axis) consume less water than their peer group: roughly 90% of the consumers with the lowest water consumption were informed that they used less than their neighbors (Q1 + Q2 of Difference from Peer group (kgal)).[9] Likewise, most high users (Q4 on the x-axis) received a message telling them they used more water than their peers.

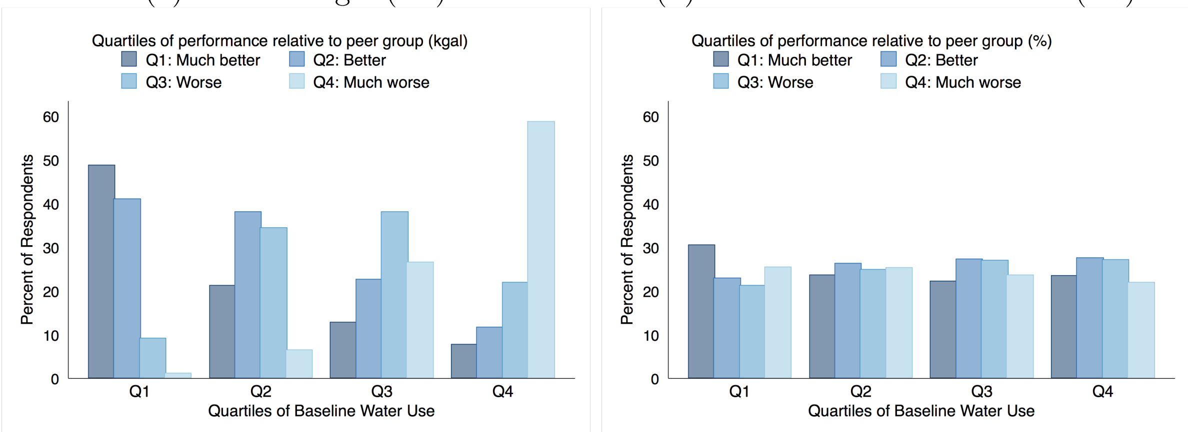
This is not the case for our social comparison in percentage terms (Treatment 5): a substantial fraction of low users conserved less than their peers and many high users

[9]The only reason why there are some low users who are above their peer group in the traditional social comparison (T4) is that the norm is based on a peer group - defined by households in the same meter route who have similar number of bedrooms and yard size (above/below the median). By comparison, Ferraro and Price (2013) compare household consumption to the full sample median, producing a treatment where the strength of the descriptive norm is perfectly correlated with pre-treatment consumption. Therefore a household with a high-water-use peer group can be above the median or 75th percentile of the sample-wide distribution of pre-treatment consumption, but still consume less than the peer group.

conserved more than their peers (Figure 1, Panel (b)). Even among households in the bottom quartile of pre-treatment water consumption, some reduced their water consumption by less than the median conservation rate in their peer group and thus received a strong normative appeal. Likewise, some households with high pre-treatment consumption reduced consumption by a larger percentage than their peer group. The distribution of norms within each quartile of pre-treatment consumption is remarkably balanced for the conservation rate comparison treatment. Figure 1 uses all data from the experimental sample, but the same general pattern of consumption holds if we restrict the sample to the treatment group, the control group, or any combination of treatment and control for individual months of the sample.

Figure 1: Strength of Normative Message by Quartiles of Pretreatment Consumption

(a) Norm in kgal (T4) (b) Norm in conservation rate (T5)



Note: The graph displays the percentage of respondents receiving messages divided up by quartiles of the performance relative to the peer group within each quartile of pre-treatment consumption. The x-axis displays the quartiles of pre-treatment consumption and the y-axis displays the percentage of respondents receiving a given message. The performances relative to the norm are designated by the different colored bars. The performance relative to the peer group is defined based on quartiles of the difference between a household's consumption (panel (a) - T4) or conservation rate (panel (b) - T5) and the peer group's consumption.

We incorporate the information content of the mailers by estimating conditional average treatment effects (CATEs), where we condition on the difference between pre-treatment consumption and the peer group median. The content of the mailers depends on recent water use, which can introduce endogeneity into the estimated treatment effects for treatment groups that receive multiple mailers. The first mailing received only depends on pre-treatment water use in the months immediately preceding the intervention, but additional mailers include information from the first month of treatment. For households randomly-selected to receive two letters, we therefore use only the first month of post-treatment data, before the second letter was received. Since treatment is randomized across the distribution of pre-treatment water use, CATEs provide valid inference - the results can be interpreted as causal treatment effects in the same style as studies that condition on pre-intervention consumption (Allcott, 2011; Ferraro and Miranda, 2013; Brent et al., 2015).

The CATE model is defined as

$$y_{it} = \alpha + \sum_{c=1}^{k} \gamma_{l,c} T_{i,l} \times C_{i,c} + \sum_{c=1}^{k} \theta_c C_{i,c} + \beta \mathbf{x_{it}} + \epsilon_{it} \tag{2}$$

In this model we are concerned with $\gamma_{l,c}$, which is the CATE for letter l in subgroup c. $T_{i,l}$ is an indicator for whether a household was treated with letter l and $C_{i,c}$ is an indicator for whether a household falls into subgroup c of the conditioning variable $C_{i,c}$. The presence of $C_{i,c}$ accounts for the sample-wide differences in consumption for subgroup c. When we condition based on the performance relative to the peer group, $C_{i,c}$ equals a set of dummies indicating the quartiles of the difference between a household's level of consumption (or percentage conservation) and that of its peers in the month directly preceding treatment. Since control households received no treatment, we calculate the normative appeal that they would have received had they been treated. This accounts for unobserved factors within each subgroup that are common to the treatment and control groups. In the case of the difference from the comparison group in kgal (T4), $C_{i,c}$ controls for the fact that high users generally receive comparisons above their peer group in the gallons comparison (T4) and issues such as mean reversion in the conservation rate comparison (T5). Since treated households only receive one realization of the difference from their peer group, based either on their July or August water consumption, we randomly assign each control household the difference from their peer group based on either July or August in the same proportion as the treated households.[10] We also run specifications where the difference variables for the control households change over time based on the realization in the previous month that produce very similar results.

4 Results

4.1 Base Results

We begin by reporting the average treatment effects pooling the three treatments of interest, and then briefly discuss each treatment individually. Column 1 of Table 4 shows that the average treatment effect (ATE) pooling all three treatments is slightly greater than a 1.5% reduction in consumption. For reference, the generic tips treatment (T1) had no statistically-significant impact on conservation, and the ATE of the tips plus historical information (T2) treatment was slightly less than 1% and statistically significant. Overall, our pooled ATE is a smaller effect than commonly reported for social

[10]One third of the treatment group received a single mailer in July, one third received a single mailer in August, and one third received mailers in July and August. Therefore two-thirds of the controls have a comparison based on July consumption and one-third based on August consumption. Moreover, we drop all observations after the first post-intervention month for the one-third of control households that are randomly assigned to stand in for the two-mailer treatments.

comparisons: Opower's interventions typically reduced energy consumption by about 2%, and both Ferraro and Miranda (2013) and Brent et al. (2015) find average reductions in consumption of approximately 5%. However, these results should be considered in the context of an extensive utility-wide water conservation campaign during the second year of a severe drought. Additionally, given that the aforementioned studies on water examine some of the first interventions using social comparisons for water conservation, the lower treatment effects are consistent with the findings of Allcott (2015) that initial sites often have higher average treatment effects than subsequent sites.

Column (2) breaks down the treatments individually. Each treatment generated statistically significant reductions in consumption and the point estimates are all very close to each other. Columns (3)-(5) reproduce the ATE for each letter in separate regressions using the individual treatment group and the control. This is simply to demonstrate that both the point estimates and the standard errors are almost identical whether we use the entire sample with three treatment dummies or restrict the sample to one treatment and the control. Restricting the sample simplifies the presentation of the results. All subsequent regressions also include controls for temperature, precipitation, bill cycle fixed effects, month fixed effects, and pre-treatment consumption.

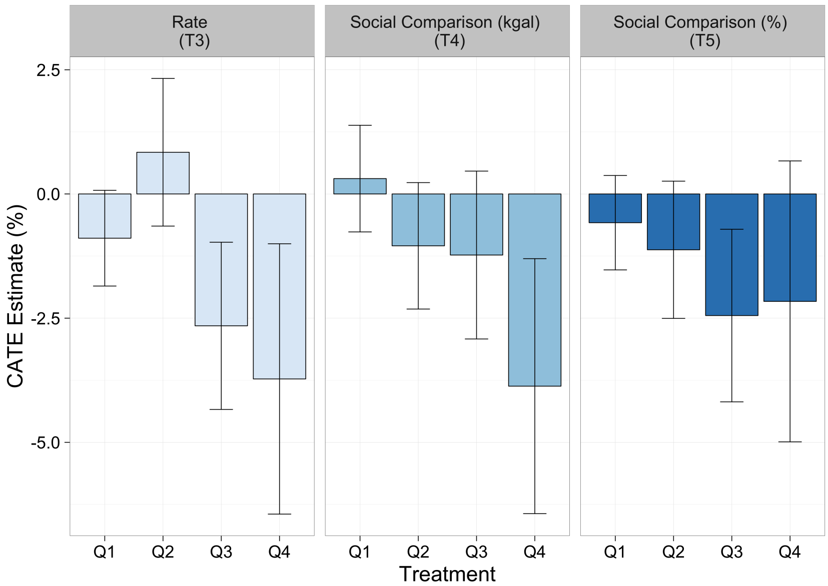
Table 4: Base Regression

	(1) Full Sample (T3-T5)	(2) Individual Treatments	(3) Rate (T3)	(4) Social Comp. kgal (T4)	(5) Social Comp. % (T5)
All Treatments	-1.550***				
	(0.304)				
Rate (T3)		-1.604***	-1.596***		
		(0.466)	(0.466)		
Social Comp. kgal (T4)		-1.455***		-1.451***	
		(0.446)		(0.446)	
Social Comp. % (T5)		-1.591***			-1.591***
		(0.473)			(0.473)
Weather Controls	Yes	Yes	Yes	Yes	Yes
Bill Cycle FEs	Yes	Yes	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes	Yes	Yes
Baseline Water	Yes	Yes	Yes	Yes	Yes
Households	33,937	33,937	25,532	25,529	25,510
Observations	96,759	96,759	74,530	74,494	74,449

Note: The dependent variable is normalized average daily water consumption; the coefficients can be interpreted as a percentage change in consumption. Robust standard errors clustered at the household level are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Next, we show that results for all of our treatments are consistent with previous research that finds the level of pre-treatment consumption is positively correlated with the conservation response to social comparisons. The results are based on the same model in equation 2 but the conditioning variables are quartiles of pre-treatment consumption. Figure 2 estimates CATEs by quartiles of pre-treatment consumption for each individual treatment. Each panel reflects the results of one regression of the CATEs; the shaded bars are the point estimates and the error bands are the 95% confidence intervals. Each of the three treatments primarily generates statistically significant savings among high users. Next, we analyze whether the treatment effect for the social comparison treatments (T4 and T5) is driven by pre-treatment consumption or the strength of the normative message.

Figure 2: Conditional Average Treatment Effects by Quartiles of Pre-treatment Consumption



Note: Each bar graph represents the output of one regression where the dependent variable is normalized average daily water consumption. The bars are the point estimates of the CATEs for each quartile of pre-treatment consumption, and the error bars are 95% confidence intervals constructed from robust standard errors clustered at the household level. All regressions include controls for temperature, precipitation, bill cycle fixed effects, month fixed effects, and pre-treatment consumption.

4.2 Difference from the Peer-Group Descriptive Norm

Our first CATE specification defines strong normative appeals as mailers that informed households that they were *above* their peer group: either using more water than their peer group (T4) or a lower percentage reduction than their peer group (T5). Our second specification estimates CATEs based on quartiles of the difference from the peer group, providing more variation in the performance relative to the peer group.

Column (1) of Table 5 shows the CATEs based on being above or below the peer group for the social comparison in gallons. All of the savings essentially come from households who are informed that they are consuming more than their peers. The same pattern holds when we estimate CATEs within quartiles of the difference from the comparison group (Column (2) of Table 5). There is a monotonic relationship between the difference from the peer group and the estimated treatment effects.[11] As described above, it is

[11] Negative differences, which reflect households who are doing better than their peers, are associated with insignificant treatment effects that are either positive or small and negative. Positive differences,

impossible to determine whether these results are due to the moral cost imposed by the message or due to the fact that high users have more scope to correct internalities.

Columns (3)-(4) of Table 5 estimate the CATEs based on the difference from the comparison group framed in terms of percentage conservation, which separates the performance relative to the peer group from pre-treatment consumption. The pattern of results is similar: households treated with a message telling them they conserved less than their peers save water, whereas those who conserved more than their neighbors do not. In fact, the results in column (4) using quartiles of the difference from the peer group are even starker. Among households treated with the social comparison in percentage terms, those who were conserving much less than their peers before receiving the letter (Treat*Q4 Norm) saved over 5.7%. This is compared to a 1.3% increase (though not statistically significant) among households who reduced consumption by a much larger percentage than their peers (Treat*Q1 Norm). These results are consistent with social comparisons operating through an increase in the moral cost of consumption: larger increases in the moral cost lead to larger reductions in consumption. However, these results do not rule out the second behavioral channel: stronger messages indicate more potential to correct internalities.

To further disentangle the CATE's from the level of consumption, we divide the sample based on below-median baseline water users (column 5) and above-median users (column 6) and estimate the CATE's within each subsample. We do not perform the analysis for the social comparison in gallons due to the strong correlation between the difference from the peer group and pre-treatment consumption. The pattern of results is the same for both low users who likely have fewer low-cost opportunities to correct internalities and high users who likely have more opportunities. This is strong evidence that the message operates by increasing the moral cost of consumption.

Moreover, higher levels of pre-treatment consumption act to magnify the results. The savings for households in the upper two quartiles of the difference from peer group (worse performance than peers, Treat*Q3 Norm and Treat*Q4 Norm) are roughly twice as large for high users compared to low users. This suggests that there is a synergistic effect between the moral cost of consumption and the motivation to correct internalities in consumption. Alternatively, similar conservation actions, such as changing the irrigation controller, are scaled up for high users so that the same action leads to more conservation. There is no statistically-significant increase in consumption among those who outperform their peers. Overall, these results suggest that social comparisons appear to operate at least partly through increasing the moral cost of consumption for households performing worse than their peers.

---

which reflect households who are doing worse than their peers, are associated with significant and negative treatment effects, which increase with the difference from peer group.

Table 5: Difference from Peer Group

	Norm in Gallons (T4)		Norm in % (T5)			
	(1) All	(2) All	(3) All	(4) All	(5) Low Users	(6) High Users
Treat*Below Peer	0.205		0.184			
	(0.644)		(0.766)			
Treat*Above Peer	-2.208**		-3.768***			
	(0.914)		(0.907)			
Treat*Q1 Norm		0.347		1.267	1.423	1.310
		(0.901)		(1.166)	(0.980)	(2.119)
Treat*Q2 Norm		0.0579		-0.117	0.477	-0.624
		(0.874)		(0.918)	(0.884)	(1.521)
Treat*Q3 Norm		-1.692*		-2.900***	-1.637**	-3.877***
		(1.014)		(0.898)	(0.828)	(1.465)
Treat*Q4 Norm		-2.627*		-5.670***	-4.358***	-7.092***
		(1.528)		(1.368)	(1.313)	(2.349)
Observations	68,933	68,873	68,929	68,834	33,129	35,477

Note: The dependent variable is normalized average daily water consumption, and the sample is restricted to the month after a household's first mailer. CATEs are estimated based on the difference (above/below) of household consumption relative to the peer group. Column headers All represents the whole sample, and Low/High Users restrict the sample to households above or below median pre-treatment consumption. All regressions include controls for temperature, precipitation, bill cycle fixed effects, month fixed effects, and pre-treatment consumption. Robust standard errors clustered at the household level are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

4.3 Survey Evidence for Behavioral Channels

To provide additional evidence of the behavioral mechanisms of response to our treatments, we link our data to responses from a utility-sponsored survey conducted after the intervention (Christman, 2016). Approximately 1,500 households in our experimental sample answered the survey, and although the survey respondents are not representative of the greater service area (overall they used less water prior to treatment and lived in smaller, less expensive homes), the survey sample is balanced across treatments and the control group for key variables (Tables A.6-A.8 show balance statistics for the survey sample). We pool the two social comparison treatments to focus on how differences in attitudes influence the treatment effects for the rate information treatment, which frames conservation as a financial savings, compared to the social comparisons, which focus more heavily on moral suasion.

Table A.5 shows the results of regressions that interact the treatment effects with indicators for households that answered yes to questions related to their motivations for using water or capacity to conserve. In Table A.5 "Survey Variable" corresponds to a dummy indicating that the respondent answered yes to the question indicated in the column header. Summary statistics for the survey variables are available in the Appendix. The first column simply replicates the treatment effects for the survey sample; the point estimates are similar, but slightly higher and not statistically significant.[12] The next three columns (2-4) represent pro-social motivations. These households indicated that they

[12] Pooling the three treatments together does produce a statistically significant treatment effect in the survey sample.

would be willing to let their lawn go brown if the drought continued (column (2)); thought that water scarcity was a serious concern (column (3)); and indicated that they saved water to help the environment and/or their community (column (4)). The interaction terms for all three pro-social indicators are negative and often significant, indicating that pro-social households were more responsive to *both* types of treatments. Recall also that the financially-oriented treatment also contained a reminder about TMWA's 10% goal and an exhortation to do their part.

The next two columns (5-6) represent households who make their water decisions based on grounds other than water scarcity, environmental, or community concerns. These households indicated that they should be able to use water as they choose (columns (5)) or saved water in order to save money (column (6)). We argue that households that want the freedom to use water as they choose and save water for financial reasons will be less responsive to the social comparison treatments. This is indeed the case: while the results are not significant (nor significantly different) relative to the rate information treatment, the financially motivated households are less responsive to treatment. In the last column, we show results for households that have already conserved and likely have lower capacity to correct internalities. The rate information treatment is significantly less effective for households with previous conservation efforts, which suggests that these households might have previously over-perceived the marginal cost of water. By contrast, for households who have not already conserved, rate and monetary savings information leads to a 5% reduction in consumption that is significant at the 10% level. There is essentially no impact of prior conservation on the savings from social comparisons. Overall, these survey results indicate that pro-social households are more responsive to both treatments, which corroborates our main finding that social comparisons operate in part by imposing a moral cost on consumption. There is also evidence that the rate information treatment is relatively more effective at prompting households to correct internalities than the social comparisons, a point that we investigate further in the next section.

Table 6: Survey Evidence of Behavioral Channels

		Pro-social Motivations			Alternative Motivations		Capacity
	(1) Baseline	(2) Lawn Brown	(3) Water Scarcity	(4) Environment & Community	(5) Use How I Want	(6) Save Money	(7) Already Conserved
Rate (T3)	-1.965	-0.631	0.536	3.772	-1.216	-2.739	-5.134*
	(1.920)	(2.091)	(2.460)	(5.230)	(2.271)	(3.243)	(2.633)
Social Comp. (T4-T5)	-2.573	-1.556	0.00760	5.433	-2.812	-5.448*	-2.421
	(1.571)	(1.636)	(2.095)	(4.472)	(1.804)	(2.807)	(1.992)
Survey Variable		4.139	1.472	0.0162	0.847	-2.152	-0.0708
		(3.856)	(1.784)	(2.134)	(2.007)	(1.826)	(1.768)
T3*Survey Variable		-13.52***	-6.496*	-6.696	-3.135	1.183	7.285*
		(5.065)	(3.925)	(5.628)	(4.204)	(4.046)	(3.840)
T4-T5*Survey Variable		-12.71**	-6.709**	-9.323*	1.489	4.285	-0.433
		(5.427)	(3.087)	(4.780)	(3.511)	(3.403)	(3.126)
Households	1,536	1,536	1,536	1,536	1,536	1,536	1,536
Observations	4,371	4,371	4,371	4,371	4,371	4,371	4,371

Note: The dependent variable is normalized average daily water consumption, and the sample is restricted to survey respondents. The first column replicates the ATE for the rate information treatment and pooled social comparisons. Columns (2) - (7) interact the treatment variable with dummy variables if the respondent answered that survey question positively: "Survey Variable" refers to the dummy variable indicated by the column headers. All regressions include controls for temperature, precipitation, bill cycle fixed effects, month fixed effects, and pre-treatment consumption. Robust standard errors clustered at the household level are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

4.4 Persistence & Additionality

To further investigate the mechanisms of response to our treatments, we analyze the variation in the persistence of the treatment effects. We posit that increasing the moral cost of consumption has a temporary impact on consumers. This is consistent with previous studies that the magnitude of the treatment effect decays as time from the last mailer increases, although treatment effect are often detectable long after treatment is discontinued (Ferraro and Miranda, 2013; Bernedo et al., 2014; Allcott and Rogers, 2014; Brandon et al., 2017). Moreover, consumers who experience an increased moral cost may seek to adjust consumption by making temporary behavioral changes, whereas consumers seeking to correct internalities are more likely to undertake investments that will have more lasting conservation effects.[13] Under this reasoning, treatments that operate through increasing moral cost should have less persistent treatment effects than treatments that prompt consumers to correct internalities. Treatments that generate similar patterns of savings over time are likely to operate through similar behavioral mechanisms. To this end, we compare the persistence of the treatment effects for the treatment focused on rate information and monetary savings (T3) and the social comparisons (T4-T5). The rate information treatment directly prompts consumers to address internalities in consumption in order to save money, whereas our previous results suggest that social norms operate in part through a moral motivation. We also examine whether there is a differential effect of sending an additional letter on these treatments. We expect to see less of an effect of additionality on nudges that operate primarily through prompting consumers to correct

[13]This is not to say that moral cost cannot induce investments in efficiency. Brandon et al. (2017) provide evidence that OPower treatments induce savings after the initial tenants who received the reports leave, which is consistent with investments in energy efficiency.

internalities in consumption. If a consumer receives useful information that helps them re-optimize, seeing similar information a second time should not generate additional savings. In order to understand the implications of the timing and number of treatments we run the following regression,

$$
\begin{aligned}
y_{it} = \alpha + &\gamma_{1,l}1\{First\ Mailer, First\ Month_{it,l}\} + \\
&\gamma_{2,l}1\{First\ Mailer, Second\ Month_{it,l}\} + \\
&\gamma_{3,l}1\{First\ Mailer, Third\ Month_{it,l}\} + \\
&\gamma_{4,l}1\{Second\ Mailer_{it,l}\} + \beta \mathbf{x_{it}} + \epsilon_{it}
\end{aligned}
\tag{3}
$$

where the treatments are now broken down by the months since the first letter was received in addition to a dummy for whether there was a second mailer ($\gamma_{4,l}$). We analyze three months from the first letter since this corresponds to the end of the summer demand season (October). Therefore, γ_2 and γ_3 test whether the treatment was persistent and γ_4 represents the additionality of the second mailer.[14] Table 7 presents the results for the rate and social comparison treatments.

The rate information treatment is very persistent with similar magnitude treatment effects in each of the three months following the first letter, and a second letter generates no additional savings. By contrast, the social comparison treatments (T4 and T5) are not persistent; however these treatments achieve significant additionality from the second mailer. The savings from the first treatment decreases by roughly 50% or more per month. Sending another mailer increases conservation by almost 2%. The results for the social comparisons are consistent with the Allcott and Rogers (2014) results where the pattern of action and backsliding stems from a model of consumption cues, and indicate similar behavioral motivations for response to social comparisons of energy consumption. The same pattern holds when we estimate regressions for each of the social comparisons individually.

We also divide the sample based on whether the household was above or below the median of pre-treatment consumption to compare the persistence of treatment for low and high users. We see high users within both treatments have a pattern of savings that is consistent with the overall treatment effect pattern. High users drive the treatment effect, which is not surprising, since most of the water savings in all of our five treatments come from high users. The fact that social comparisons are not persistent and require additional mailers even among high users with greater potential to correct internalities, provides further evidence that households are responding to social comparisons due to increased moral costs.

The differences in persistence and additionality across treatments can be interpreted

[14] The second mailer affects consumption in two months: September and October.

in the framework of intent-oriented actions and impact-oriented actions developed by Attari (2014). Intent-oriented actions include behavioral changes such as turning off the water when brushing teeth, while impact-oriented actions include investments such as replacing a toilet. Households perform intent-oriented actions with the intention to help the environment, while impact-oriented actions are focused on achieving more substantial conservation. The results suggest that nudges that operate by prompting consumers to correct internalities (T3) lead to impact-oriented actions, while nudges that increase the moral cost of consumption (T4 and T5) lead to intent-oriented actions.

Indeed TMWA survey results corroborate these findings. A series of questions asked households what types of actions they took to reduce water consumption. We focused on responses to questions that correspond to our treatment period. The types of actions can be divided into two general categories: investments in efficiency such as repairing leaks or replacing lawn with xeriscaping and behavioral changes such as taking shorter showers. We generate indicator variables for whether a household engaged in any investment or any behavioral change and use these as the dependent variables in a linear probability model estimated with ordinary least squares. The results, presented in Table 8, show that households treated with social comparisons are significantly less likely to invest in efficiency compared to households that received the rate information treatment. There is no statistical difference in the relative propensity to engage in behavioral changes. This provides further evidence of the different mechanisms that are operating for the rate and social comparison treatments. The results are consistent with households correcting internalities in response to rate information and making short-term cuts in consumption in response to social comparisons.

Table 7: Persistence & Additionality of Second Mailers

	All		Below Median		Above Median	
	(1) Rate	(2) Social Comp.	(3) Rate	(4) Social Comp.	(5) Rate	(6) Social Comp.
1st Letter, 1st Month	-1.395**	-1.474***	-0.00465	-0.536	-2.680***	-2.371***
	(0.542)	(0.403)	(0.565)	(0.375)	(0.903)	(0.692)
1st Letter, 2nd Month	-1.899***	-1.108**	-0.241	-0.839**	-3.326***	-1.374*
	(0.592)	(0.449)	(0.565)	(0.417)	(0.998)	(0.762)
1st Letter, 3rd Month	-1.499**	-0.304	-0.305	-0.773	-2.850**	0.149
	(0.739)	(0.554)	(0.677)	(0.547)	(1.250)	(0.917)
2nd Letter	0.158	-1.922***	-0.453	-0.321	0.743	-3.456***
	(0.759)	(0.558)	(0.694)	(0.522)	(1.296)	(0.951)
Households	25,532	29,722	12,305	14,311	13,227	15,411
Observations	74,530	85,586	35,923	41,206	38,607	44,380

Table 8: Mechanisms of Water Conservation: Investment and Changes in Behavior

	(1) Invested in Efficiency	(2) Behavioral Changes
Rate (T3)	0.0528	0.0256
	(0.0354)	(0.0240)
Social Comparisons (T4-T5)	-0.0759***	-0.00339
	(0.0228)	(0.0188)
HOA w/ Landscape Restrictions	Yes	Yes
Already Invested in Traditional Landscape	Yes	Yes
Already Invested in Water Efficiency	Yes	Yes
Observations	1,536	1,536

4.5 Robustness

We run several robustness tests to assure that our results are not actually being driven by spurious correlation or the discrete effects of being above or below the norm. The treatment period started in August 2015 utilizing July 2015 data on the mailers. Our first robustness check runs falsification tests by using the content generated by June 2015 data. We generate the difference from the peer group using June 2015 for both treated and control households to see if the types of households that were above or below the peer group were actually driving the results as opposed to the actual content of the mailers. This is a valid falsification test of our results due to the fact that variation in seasonal consumption leads to variation in household performance relative to a peer group. Some households can be below their peer group earlier in the summer, but have much higher peak water use than the peer group later in the summer. Table 9 shows results using the simulated mailer content. The results are substantially different than Table 5. Households above and below the false peer group have statistically significant CATEs in the conservation rate information treatment (column (3)), and neither of the CATEs are statistically significantly different from each other. The CATE for households below the peer group in the traditional social comparison is not statistically significant, but this is likely due to the fact that they are low users and their real letter also indicated that they were below their peer group. The specification with quartiles of the norm in percentage terms (column (4)) shows that three out of the four CATEs generate savings over 2% and are significant at the 5% level. In general, households above the false peer group do save more water, but this is likely due to serial correlation in water consumption; households above the peer group in June are more likely to be above the peer group in July and August.

The regressions presented in Table 5 have different numbers of observations for each household. Households treated with one mailer in July have three observations after treatment (Aug, Sep, and Oct), households treated with one mailer in August have two observations (Sep and Oct), and households treated with two mailers only have the observation immediately following their initial letter. In order to test if the difference in the number of observations impacts the results we drop all households treated with two

Table 9: **Falsification Test for Difference from Peer Group**

	Norm in Gallons (T4)		Norm in % (T5)	
	(1)	(2)	(3)	(4)
Treat*Below Peer	-0.660		-1.184	
	(0.636)		(0.786)	
Treat*Above Peer	-1.794*		-2.181**	
	(0.938)		(0.934)	
Treat*Q1 Norm		-0.148		-2.274**
		(0.910)		(1.093)
Treat*Q2 Norm		-1.166		-0.238
		(0.881)		(1.090)
Treat*Q3 Norm		-2.197**		-1.629*
		(1.108)		(0.990)
Treat*Q4 Norm		-1.290		-2.708*
		(1.512)		(1.615)
Observations	68,933	68,930	68,929	68,896

Note: The dependent variable is normalized average daily water consumption, and the sample is restricted to the month after a household's first mailer. CATEs are estimated based on the quartiles of the difference of household consumption relative to the peer group. Robust standard errors clustered at the household level are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

mailers and only use the first two observations after the first mailer. The results are very similar to our preferred specification, which incorporates all valid observations, and are available in Table A.9 in the Appendix. We also run the same set of regressions limiting the sample to the first month after a household receives a mailer. This sample includes observations immediately after a household's first mailer for households treated with two mailers. The results, presented in Table A.10 in the Appendix, also display the same general pattern where the magnitude of the treatment effect depends on the difference from the peer group.

Since the mailers also present information on historical consumption and reference the utility-wide 10% goal we also run regressions for quartiles of the difference between an household's conservation rate and the 10% goal. The distance from the 10% goal does affect the CATEs, but this is at least in part due to the correlation with pre-treatment water use and the conservation rate. The results, presented in Table A.11 in the Appendix, do not show the monotonic relationship that we observe for CATEs based on the difference from the peer group, and they vary across treatments. Lastly, we test for the discrete effects for moving above a peer group or failing to meet the 10% goal in a regression discontinuity design. We find no effect of moving above the peer group in either gallons or percentage terms, nor do we find any effect of moving slightly above the 10% goal. This is consistent with the findings of Allcott (2011) for social comparisons in energy and with unpublished results from the project reported in Brent et al. (2015). The Appendix describes the regression discontinuity design in more detail and presents both graphical evidence and the regression discontinuity estimates based on Calonico et al. (2015).

5 Conclusions

We designed and implemented an experiment using multiple treatments in a policy environment where a municipal water utility was running a well-publicized campaign asking every customer to reduce their monthly consumption by 10%. We find that two versions of a social comparison and a third treatment providing information on monetary savings lead to a 1.5% reduction in water consumption relative to the control group. This reduced consumption was *in addition* to the overall system-wide reduction achieved of over 15%. While there may be some concern that the 10% goal reduces the external validity of our experimental results, calls for uniform reductions are a common utility practice as evidenced by Governor Brown's 2015 call for reducing water consumption in California by 25%. We also note that there was a utility-wide appeal for conservation in the well-known water conservation experiments in Cobb County examined by Ferraro and Price (2013). The treatments generated effects that were similar in magnitude and statistically indistinguishable from each other.

We introduced a new normative appeal method that decouples pre-treatment consumption from the difference from the peer group by framing savings in terms of percentage achievement to the target conservation goal of 10%. The strength of the normative message is a strong driver of variation in the treatment effect over low and high consumption levels for our new conservation rate social comparison treatment, although treatment effects are magnified for larger water users. These results are consistent with an interpretation that social comparisons impose a moral cost on consumption, and the size of the moral tax depends on the magnitude of the difference between household performance and that of a peer group. Both social comparison treatments display the pattern of "action and backsliding" shown by Allcott and Rogers (2014) indicating that they operate through a similar behavioral mechanism; however, we find a different pattern over time for the treatment that highlights financial savings. In contrast to the social comparisons, a single mailer for the financial savings treatment is persistent over the study period, with no additional conservation generated by an additional mailer. This is consistent with households responding to the social comparisons with intent-oriented actions that fade over time, whereas the financial treatment may lead to re-optimizing water use to correct internalities, as in a household production framework, where responses are more persistent (Attari, 2014). Survey evidence shows that households in the financial savings information treatment are more likely to invest in water efficiency relative to households treated with social comparisons.

The behavioral mechanisms underlying the response to social comparisons are important when evaluating economic welfare effects of programs designed to promote voluntary reduction in water consumption. Use of customized information in mailings that prompt consumers to re-optimize consumption to address internalities will generally produce higher welfare gains than those that impose a moral cost on consumption (Allcott and Kessler, 2015). Individuals operating with moral motivations may act to reduce existing internalities, thereby limiting welfare-reducing effects of normative appeals. However, there is likely variation in the available internalities across regions due to differences in whether irrigation is necessary to maintain healthy landscape as well as variation in the extent to which utilities have promoted investment in efficient fixtures and appliances. In regions similar to the area serviced by TMWA, which experiences seasonal outdoor water use that is 4 to 5 times higher than indoor use alone, consumers may have ample opportunities to correct internalities by investing in irrigation efficiency and water efficient landscape. Similarly, households in areas which have experienced fewer instances of calls for water conservation are likely to have room to exploit internality correction, since consumers have not been previously prompted to invest in efficient appliances. Thus, prior experience with conservation is likely to impact the welfare effects of normative appeals for water and energy conservation. Understanding the behavioral mechanisms also helps policymakers select interventions that meet specific objectives. For urban water managers there are important distinctions between interventions that temporarily reduce water consumption during a drought and permanently reduce consumption after water supplies have recovered. Therefore, understanding different nudges prompt different types of actions expands the toolbox for policymakers to address the different forms of water scarcity that varies across regions.

A practical contribution of our work is a more sophisticated understanding of the motivations that underlie responses to normative appeal campaigns to improve welfare outcomes of future campaigns. Utilities use normative messaging campaigns to reduce energy or water consumption ostensibly to improve welfare in situations where unpriced consumption externalities cause the social marginal cost to be greater than the retail price. Normative messaging campaigns that also improve welfare for a household whose current consumption does not in fact maximize their own welfare, by cuing the correction of internalities, generates both private and external benefits. Conservation campaigns that focus on addressing internalities, as opposed to imposing a moral cost, have the potential to increase the welfare benefits of these tools more than reliance on moral appeals alone, along with generating more persistent treatment effects.

References

Allcott, Hunt, "Social norms and energy conservation," *Journal of Public Economics*, 2011, *95* (9-10), 1082–1095.

_ , "Site Selection Bias in Program Evaluation," *Quarterly Journal of Economics*, 2015, *130* (3), 1117–1165.

_ **and Cass R Sunstein**, "Regulating internalities," *Journal of Policy Analysis and Management*, 2015, *34* (3), 698–705.

_ **and Dmitry Taubinsky**, "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market," *American Economic Review*, 2015, pp. 1–80.

_ **and Judd B. Kessler**, "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons," *NBER Working Paper Series*, 2015, *21671.*

_ **and Todd Rogers**, "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," *American Economic Review*, 2014, *104* (10), 3003–3037.

_ **, Sendhil Mullainathan, and Dmitry Taubinsky**, "Energy policy with externalities and internalities," *Journal of Public Economics*, apr 2014, *112*, 72–88.

Attari, Shahzeen Z, "Perceptions of water use," *Proceedings of the National Academy of Sciences*, 2014, *111* (14), 5129–5134.

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

Becker, Gary S., "A Theory of the Allocation of Time," *Economic Journal*, 1965, pp. 493–517.

Bernedo, Maria, Paul J. Ferraro, and Michael K. Price, "The Persistent Impacts of Norm-based Messaging and their Implications for Water Conservation," *Journal of Consumer Policy*, 2014, *37*, 437–452.

Bolsen, Toby, Paul J. Ferraro, and Juan Jose Miranda, "Are Voters More Likely to Contribute to Other Public Goods? Evidence from a Large-Scale Randomized Policy Experiment," *American Journal of Political Science*, jan 2014, *58* (1), 17–30.

Brandon, Alec, Paul J. Ferraro, John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer, "Do The Effects of Social Nudges Persist? Theory and Evidence from 38 Natural Field Experiments," Working Paper 23277, National Bureau of Economic Research March 2017.

Brent, Daniel A., Joseph H. Cook, and Skylar Olsen, "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*, 2015, *2* (4), 597–627.

Calonico, Sebastian, Matias D. Catteneo, and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

_ , _ , and _ , "Optimal Data-Driven Regression Discontinuity Plots," *Journal of the American Statistical Association*, 2015, *110* (512), 1753–1769.

Christman, Laine, "Presentation on the 2015 Summer Drought Campaign Survey results," Technical Report 2016.

Costa, Dora L. and Matthew E. Kahn, "Energy Conservation Nudges and Environmentalist Ideology: Evidence From a Randomized Residential Electricity Field Experiment," *Journal of the European Economic Association*, jun 2013, *11* (3), 680–702.

Dolan, Paul and Robert Metcalfe, "Neighbors, Knowledge, and Nuggets: Two Natural Field Experiments on the Role of Incentives on Energy Conservation," 2015.

Ferraro, Paul J. and Juan José Miranda, "Heterogeneous Treatment Effects and Mechanisms in Information-Based Environmental Policies: Evidence from a Large-Scale Field Experiment," *Resource and Energy Economics*, 2013, *35* (3), 356–379.

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

_ , Juan Jose Miranda, and Michael K. Price, "The Persistence of Treatment Effects with Norm-Based Policy Instruments : Evidence from a Randomized Environmental Policy Experiment," *American Economic Review*, 2011, *101* (3), 318–322.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel, "When and Why Incentives (Don't) Work to Modify Behavior," *Journal of Economic Perspectives*, nov 2011, *25* (4), 191–210.

Imbens, Guido W. and Thomas Lemieux, "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, feb 2008, *142* (2), 615–635.

Ito, Koichiro, "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," *American Economic Review*, February 2014, *104* (2), 537–563.

_ , Takanori Ida, and Makoto Tanaka, "The Persistence of Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand," *E2e Project Working Paper Series*, 2015, *017*.

_ , _ , and _ , "Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand," *American Economic Journal: Economic Policy*, Forthcoming.

Lee, David S. and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48* (June), 281–355.

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

Pellerano, José A, Michael K Price, Steven L Puller, and Gonzalo E Sánchez, "Do Extrinsic Incentives Undermine Social Norms? Evidence from a Field Experiment in Energy Conservation," *Environmental and Resource Economics*, 2015, pp. 1–16.

Schultz, P Wesley, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, and Vladas Griskevicius, "The constructive, destructive, and reconstructive power of social norms," *Psychological Science*, 2007, *18* (5), 429.

Sexton, Steven, "Automatic bill payment and salience effects: Evidence from electricity consumption," *Review of Economics and Statistics*, 2015, *97* (2), 229–241.

Wichman, Casey J., "Perceived price in residential water demand: Evidence from a natural experiment," *Journal of Economic Behavior & Organization*, 2014, pp. 1–16.

_ , "Information Provision and Consumer Behavior: A Natural Experiment in Billing Frequency," *Resources for the Future Discussion Paper*, 2016, *DP 15-35-REV.*

Appendix

Randomization and Implementation

The randomization used a procedure of quasi-pairwise matching within blocking groups. This method first defines a set of blocks within which the randomization occurs. The blocking procedure ensures that assignment to a treatment is balanced within certain groups of interest. Blocks were defined by billing cycles, rate schedule and frequency of recorded meter data (i.e. monthly, daily, or hourly, though all customers only receive monthly usage totals). Within each block we ordered all observations on average water consumption in summer 2013 in sets of five households. We randomly assigned each household to one of five experimental samples that correspond to the five treatments (regardless of the ultimate assignment to treatment group vs. control group). This ensures a similar distribution of 2013 water consumption within each of the five experimental samples.

Next, within each of the five experimental samples we repeated the procedure to assign households to one of three possible timing treatments (single letter in July, single letter in August, or two letters repeated in July and August), or the control group. The same blocking structure was used within each experimental sample and then households were re-ordered based on summer 2013 water consumption and in sets of 12: two households are randomly assigned to each of the three timing treatments and six households are assigned to the control.[15]

The process for generating and mailing letters was as follows:

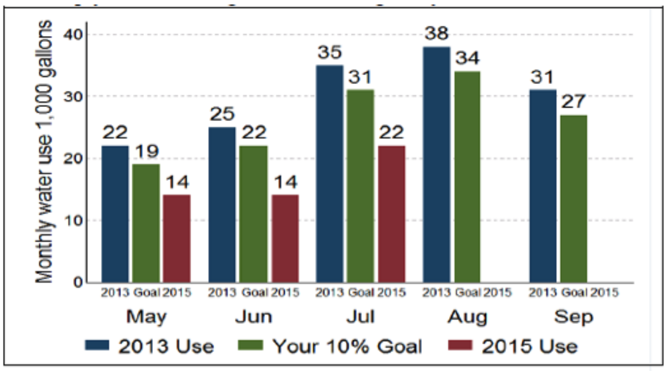
- 1-2 days after the most recent month's consumption data is loaded into the billing system we pull this information into Stata and using a set of pre-programmed routines use it to generate the graphics and data for the mail merge.
- A mail merge is performed in Microsoft Word using the generated data and graphics.
- PDF's of the letters are emailed to Digiprint
- Digiprint prints and ships the the letters within 1-3 days of receiving the electronic files.

The average time from the data upload to letters shipment was 2 days with a maximum of 8 days during this study. We had attrition during the study of about 1.5 percent of the treatment customers; 142 customers dropped out of the study in July and 211 customers dropped in August. This attrition was likely due to customers closing accounts or billing data (meter reading) errors. Furthermore, the mailers did not generate a very large increase in call center volume; out of the 23,213 customers we attempted to reach with this pilot we estimate that only 43 contacted the call center. Most of the customers who called the call center just wanted to ask clarification questions about the information in their letter; only 26 wanted further assistance beyond what the call center representatives could provide; and only five customers ended being truly upset by the pilot program.

[15]Due to the unequal size of the blocking groups, some timing treatments were oversampled, thereby creating some balance issues. We corrected these by identifying and dropping the oversampled observations after the conclusion of the field experiment (3,677 households: 2,025 control, 1,652 treatment). All balance tables and regression results reflect the corrected sample.

Treatment Figures

Figure A.1: Treatment 2 - Historical water use information



What is *your* 10% goal? The graph to the left shows your household's total water use in the past few months (red) compared to your water use in summer 2013 (blue). Water use on your July 2015 bill was **37% lower** than on your July 2013 bill. To do your part to help the community conserve water this summer, your total water use for August and September should be **at least 10%** lower than in 2013 (green). Your August goal amounts to saving **133 gallons per day** compared to how you used water in August 2013.

Figure A.2: Treatment 3 - Rate structure information

Saving water saves you money. The graph on the right shows how you **could save $13.00** if you reduce your water use on your August bill by 10% compared to 2013. The price you pay for each gallon of water increases as total water use increases. For most homes, indoor use falls within the first tier at the lowest price. Summer outdoor water use brings homes into the higher prices associated with tiers 2 and higher. The graph shows where your water use from your July 2015 bill (red) falls in the price schedule. It also shows your August 2013 water use (blue) and your 10% reduction goal for August 2015 (green).

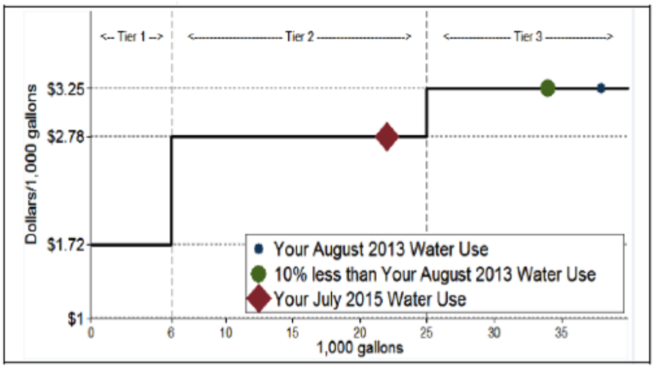


Figure A.3: Treatment 4 - Social comparison, reported in thousands of gallons

How does your water use compare?
The graph on the right shows your water use from your July bill compared to similar properties in your area. You used **1,000 gallons less** than your neighbors with similar properties.

You saved 35% on your July bill compared to 2013.
Keep up the good work!

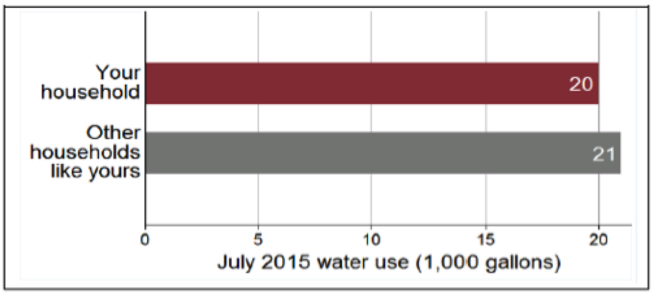
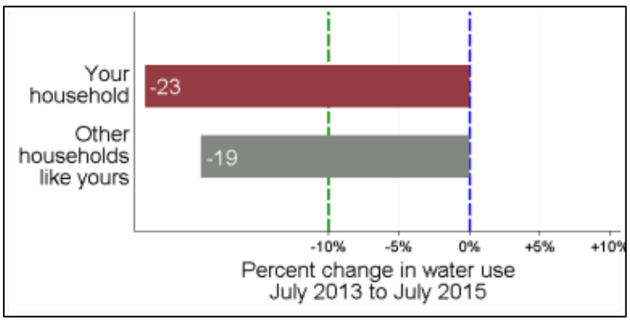


Figure A.4: Treatment 5 - Social comparison, reported as progress towards TMWA's 10% conservation goal

Are you doing your part? The graph on the right shows your change in water use form July 2013 to July 2015 compared to similar properties in your neighborhood. Your neighbors used 19% less water last month compared to 2013.

You saved 23% on your July water bill compared to 2013.
Keep up the good work!



Additional Balance Tests

Table A.1: Balance on Pretreatment Consumption for each Letter

	Control Mean	Treatment Mean	Difference	(p-value)
Letter 1	21.69	21.81	-0.11	0.66
Letter 2	21.56	21.69	-0.12	0.63
Letter 3	21.63	21.66	-0.03	0.90
Letter 4	21.62	21.66	-0.04	0.86
Letter 5	21.65	21.70	-0.05	0.84

Note: p-values are based on two-sided t-tests

Table A.2: Balance on Pretreatment Consumption across Letters

	Mean 1	Mean 2	Difference	(p-value)
Letter 1 vs Letter 2	21.75	21.62	0.13	0.49
Letter 1 vs Letter 3	21.75	21.65	0.10	0.58
Letter 1 vs Letter 4	21.75	21.64	0.11	0.55
Letter 1 vs Letter 5	21.75	21.67	0.08	0.68
Letter 2 vs Letter 3	21.62	21.65	-0.02	0.89
Letter 2 vs Letter 4	21.62	21.64	-0.02	0.93
Letter 2 vs Letter 5	21.62	21.67	-0.05	0.78
Letter 3 vs Letter 4	21.65	21.64	0.01	0.96
Letter 3 vs Letter 5	21.65	21.67	-0.03	0.89
Letter 1 vs Letter 5	21.64	21.67	-0.03	0.85

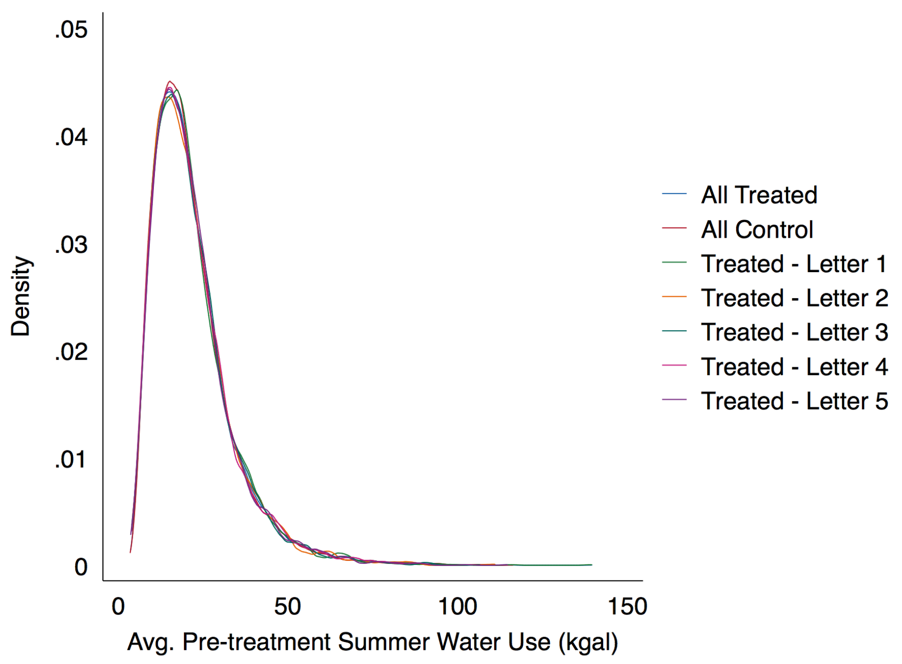
Note: p-values are based on two-sided t-tests

Table A.3: Balance within Deciles of Pretreatment Consumption

	Control Mean	Treatment Mean	Difference	(p-value)
Decile 1	7.96	7.91	0.05	0.20
Decile 2	11.06	11.06	-0.00	0.97
Decile 3	13.50	13.45	0.04	0.02
Decile 4	15.70	15.68	0.02	0.32
Decile 5	17.95	17.94	0.01	0.59
Decile 6	20.36	20.38	-0.02	0.41
Decile 7	23.25	23.23	0.02	0.38
Decile 8	26.81	26.83	-0.02	0.62
Decile 9	32.39	32.32	0.07	0.27
Decile 10	48.24	48.13	0.11	0.77

Note: p-values are based on two-sided t-tests

Figure A.5: Distributions of Pretreatment Consumption Across Treatment Status



Note: The lines are kernel density estimates of pre-treatment consumption for all treated households, all control households, and households within each of the five treatment groups.

Table A.4: Kolmogorov-Smirnov Tests

	D-statistic	(p-value)
All Treatment vs Control	0.01	0.88
Letter 1 vs Control	0.01	0.89
Letter 2 vs Control	0.01	0.87
Letter 3 vs Control	0.01	0.86
Letter 4 vs Control	0.01	0.99
Letter 5 vs Control	0.01	0.76

Note: p-values are based on the combined D-statistic

Survey Information

Table A.5: Summary Statistics for Survey Variables

	N	Mean	Standard Deviation
Lawn Brown	1949	0.08	0.28
Not Enough Water	1949	0.38	0.49
Environment/Community	1949	0.85	0.35
Use How Want	1949	0.26	0.44
Save Money	1949	0.66	0.47
Already Conserved	1949	0.38	0.49
Invest Water Efficiency	1949	0.22	0.41
Behavioral Change	1949	0.88	0.33
Invest Traditional Landscape	1949	0.31	0.46
Already Invested Efficiency	1949	0.50	0.50

Note: All variables are binary indicators.

Table A.6: Balance on Pretreatment Consumption across Letters within Survey Sample

	Control Mean	N	Treatment Mean	N	Difference	p-value
Rate (T3) vs.						
Social Comparison kgal (T4)	21.26	361	22.36	361	-1.09	0.18
Rate (T3) vs.						
Social Comparison % (T5)	21.26	361	21.23	361	0.03	0.97
Social Comparison kgal (T4) vs.						
Social Comparison % (T5)	22.36	426	21.23	426	1.13	0.18
Rate (T3) vs.						
Social Comparisons (T4 and T5)	21.26	361	21.82	361	-0.56	0.46

Note: Survey Sample, p-value is based on two-sided t-test

Table A.7: Balance on Pretreatment Consumption for Survey Respondents vs. Non-Respondents

	Non-Survey Mean	Survey Mean	Difference	p-value
2013 Water	23.58	22.86	0.72	0.05
2015 Water	17.00	15.48	1.53	0.00
Summer Water	21.71	20.66	1.05	0.00
Winter Water	8.15	7.47	0.68	0.00
Year Built	1,987.68	1,986.20	1.47	0.00
Appraised Value	215.30	200.68	14.62	0.00
Bedrooms	3.37	3.32	0.05	0.02
Lot Acre	0.27	0.24	0.03	0.00
Yard Acre	0.23	0.20	0.03	0.00
Build Sq. Ft.	1,989.19	1,958.02	31.17	0.13
Bathrooms	2.19	2.17	0.02	0.18

Note: Non-survey sample of 41,110, Survey sample of 1,664; p-value is based on two-sided t-test.

Table A.8: Balance on Pretreatment Consumption for Treatment and Control within Survey Sample

	Control Mean	Treatment Mean	Difference	p-value
2013 Water	24.04	23.57	0.46	0.45
2015 Water	16.42	16.14	0.28	0.54
Summer Water	21.76	21.43	0.33	0.54
Winter Water	8.82	7.66	1.16	0.25
Year Built	1,989.42	1,988.34	1.08	0.19
Appraised Value	225.01	222.67	2.34	0.75
Bedrooms	3.38	3.31	0.07	0.05
Lot Acre	0.28	0.27	0.01	0.44
Yard Acre	0.23	0.23	0.01	0.45
Build Sq. Ft.	2,039.40	2,020.15	19.25	0.56
Bathrooms	2.20	2.20	0.00	0.99

Note: Survey Sample of 954 Controls and 994 Treatments; p-value is based on two-sided t-test.

Robustness for Difference from Peer Group

Table A.9: Difference from Peer Group - First Two Months for Single Mailers

	Norm in Gallons (T4)		Norm in % (T5)			
	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	All	All	Low Users	High Users
Treat*Below Peer	-2.140***		-0.294			
	(0.817)		(0.971)			
Treat*Above Peer	-1.505		-3.151***			
	(1.177)		(1.179)			
Treat*Q1 Norm		-2.770**		0.907	0.694	1.562
		(1.109)		(1.490)	(1.050)	(2.793)
Treat*Q2 Norm		-1.801		-0.703	0.520	-2.030
		(1.125)		(1.134)	(1.200)	(1.828)
Treat*Q3 Norm		-2.174		-2.682**	-1.546	-3.758*
		(1.362)		(1.194)	(1.138)	(1.946)
Treat*Q4 Norm		-0.630		-4.715***	-3.263	-6.651**
		(1.898)		(1.804)	(2.165)	(2.768)
Observations	66,159	66,099	66,157	66,063	31,811	34,036

Note: The dependent variable is normalized average daily water consumption, and the sample is restricted to the first two months after a mailer for households that only receive one mailer. CATEs are estimated based on the difference (above/below) of household consumption relative to the peer group. Column headers All represents the whole sample, and Low/High Users restrict the sample to households above or below median pre-treatment consumption. All regressions include controls for temperature, precipitation, bill cycle fixed effects, month fixed effects, and pre-treatment consumption. Robust standard errors clustered at the household level are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table A.10: Difference from Peer Group - First Month After Mailer

	Norm in Gallons (T4)		Norm in % (T5)			
	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	All	All	Low Users	High Users
Treat*Below Peer	-0.424		-1.248*			
	(0.561)		(0.710)			
Treat*Above Peer	-2.299***		-2.408***			
	(0.841)		(0.761)			
Treat*Q1 Norm		0.0149		-1.123	-0.177	-2.106
		(0.855)		(1.068)	(0.704)	(2.045)
Treat*Q2 Norm		-0.665		-0.978	-0.661	-1.079
		(0.712)		(0.881)	(0.720)	(1.488)
Treat*Q3 Norm		-1.431*		-1.672**	-0.902	-2.195*
		(0.861)		(0.800)	(0.888)	(1.241)
Treat*Q4 Norm		-3.253**		-3.420***	-1.554	-5.406**
		(1.446)		(1.246)	(1.243)	(2.186)
Observations	46,603	46,568	46,581	46,525	22,416	23,956

Note: The dependent variable is normalized average daily water consumption, and the sample is restricted to the first month after a mailer. CATEs are estimated based on the difference (above/below) of household consumption relative to the peer group. Column headers All represents the whole sample, and Low/High Users restrict the sample to households above or below median pre-treatment consumption. All regressions include controls for temperature, precipitation, bill cycle fixed effects, month fixed effects, and pre-treatment consumption. Robust standard errors clustered at the household level are reported in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Effects of the community-wide 10% Goal

As a robustness check we analyze how the treatment effect varies based on whether households met their personal 10% savings goal. All three treatments (T3-T5) notify the household whether they met their personal 10% savings goal for the previous month. We estimate CATEs to based on how far away the household was from their 10% savings goal. The specification is based on equation 2 where the conditioning variables are indicators for the quartiles of difference between a household and its 10% goal in the month before treatment. One key distinction between quartiles based on the difference from the peer group and the difference from the personal goal is the interpretation within each quartile. The peer group is based on the median consumption or conservation rate so the quartiles are centered at zero. However, the median households conserved by *more* than 10% so the third quartile in the difference from the goal contains some households than saved more than 10% and some households that saved less than 10%. We examine the discrete effect of moving just below the 10% goal with a regression discontinuity below. For this analysis the conservation rate for households within: *Q1 Goal* $\gg$ 10%, *Q2 Goal* $>$ 10%, *Q3 Goal* $\lesseqgtr$ 10%, *Q4 Goal* $<$ 10%. Similar to the analysis on the difference from the peer group, we assign control households values for the conditioning variables even though they did not receive a letter, and we drop all observations following a household's second mailer. Whether a household met the goal is balanced across treatment and control since it is based on pre-treatment consumption.[16] Table A.11 shows that there is heterogeneity based on the 10% goal, but the patterns are different than the CATEs based on the strength of the social comparison message. For example, in the rate information treatment in column (1) three out of the four quartiles are negative and statistically significant with the largest effects in the third quartile. The pattern most closely mirrors the CATEs for the difference from the peer group for the social comparison in percentage terms (3). This is because the difference from the peer group is highly correlated with the difference from the 10% when the comparison metric is the percentage savings.

[16]The p-values based on the two-sided t-tests for differences in the July goal and August goals across treatment status are 0.54 and 0.35 respectively.

Table A.11: Conditional Average Treatment Effects: Quartiles of Difference from the 10% Goal

	(1)	(2)	(3)
	Rate (T3)	Social Comp. kgal (T4)	Social Comp. % (T5)
Treat*Q1 Goal	-0.737	-1.984	0.896
	(1.260)	(1.236)	(1.366)
Treat*Q2 Goal	-1.689*	-2.447***	-3.456***
	(0.888)	(0.909)	(0.975)
Treat*Q3 Goal	-4.465***	-1.762*	-3.087***
	(0.854)	(0.978)	(0.842)
Treat*Q4 Goal	-2.754*	-4.752***	-6.499***
	(1.425)	(1.045)	(1.238)
Observations	68,860	68,840	68,834

Note: The dependent variable is normalized average daily water consumption, and the sample is restricted post treatment data dropping all observations after the second mailer. The CATEs are based on quartiles of the difference from the 10% goal. All regressions include controls for temperature, precipitation, bill cycle fixed effects, month fixed effects, and pre-treatment consumption. Robust standard errors clustered at the household level are reported in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Discrete Effects Moving Above the Peer Group and 10% Goal

It is important to distinguish the continuous difference from peer group consumption from the discrete injunctive norm defining appropriate behavior (Schultz et al., 2007). We consider two separate effects. First, we test whether performing slightly worse than one's peers has an effect on consumption, and second we test for the discrete effect of just barely missing the 10% conservation goal. In our setting the descriptive injunctive norm is based on whether a household met the 10% goal. Therefore if a household less than 10% it received the message, "Please do your part to help with the drought.", while a household that saved more than 10% was told, "Keep up the good work!". It is also possible that the household considered their performance relative to their peer group as an additional categorical norm. The results in Table 5 contain *both* the effect of the discrete injunctive norm and the continuous descriptive norm.

To isolate the effect of the injunctive norm, we employ a regression discontinuity (RD) design (Imbens and Lemieux, 2008; Lee and Lemieux, 2010), analyzing behavior on either side of the injunctive category similar to Allcott (2011).[17] In the RD analysis we restrict the sample to households treated with a social comparison (T4 and T5) and examine the effect of being just above the peer group. The running variable is the difference in performance between a household and the peer group in gallons for Treatment 4 and in percentage reduction for Treatment 5. The dependent variable in both is residual normalized consumption based on a regression of normalized consumption on weather, month fixed effects, and household fixed effects, following the approach of (Allcott, 2011). The RD estimation assumes that factors varying with the difference from the peer group, such as pre-treatment water and the strength of the descriptive norm, are the same for households just above and below their peer group. Since some households were above their peer group but saved more than 10%. Similarly some households were below their peer and saved less than 10%. We repeat the analysis dropping these households and the results are very similar.

We begin with graphical evidence of differences in consumption near the peer group (Figure A.6), as is standard in RD approaches.[18] For both social comparisons the graphical evidence in Figure A.6 suggests that moving above the peer group does not affect consumption.

The graphical evidence is corroborated in the RD estimates for moving above the peer group. We use three different RD estimators developed by Calonico et al. (2014): the conventional, bias-corrected, and bias corrected with robust standard errors. In all specifications the impact for moving above the peer group is small and not statistically significant (Table A.12). The RD

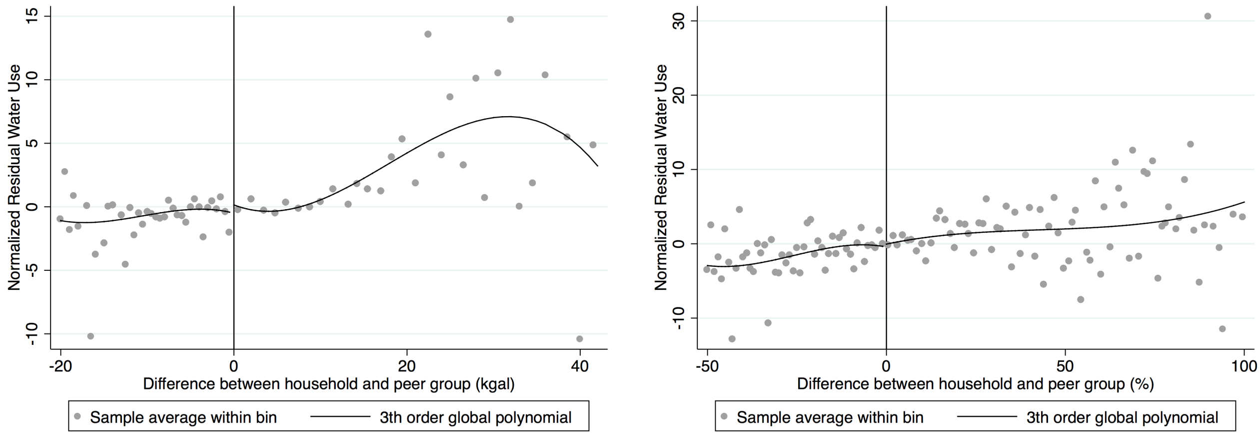
[17]Allcott (2011) showed little impact of moving into one of the three distinct categories in the Home Energy Report ("Great", "Good", or "Below Average") in a regression discontinuity design. In that study a household is assigned the category "Great" if they consume below the 20th percentile of peer consumption, "Good" if they consume below the average of peer consumption, or "Below Average" if they consume above the average of peer consumption.

[18]The graphs are generated with a data-driven approach using spacing estimators to generate the bin sizes in the plots (Calonico et al., 2015). The points on the graph are the average normalized residual consumption within each bin, and the lines are the fitted values of separate third-order polynomial regressions on either side of the distance threshold (zero).

estimates show that the effects in Table 5 are driven by the *distance* from the peer group as opposed to simply being above or below the peer group. There is either no effect associated with adding a negative injunctive norm.

Figure A.6: Effect of Moving Above the Peer Group

(a) Norm in kgal (T4) (b) Norm in percentage reduction (T5)



Note: The dependent variable is residual normalized consumption and the units are percentage terms. The discontinuity is based on the moving above the peer group's conservation rate.

Table A.12: Regression Discontinuity Estimates of the Moving Above the Peer Group

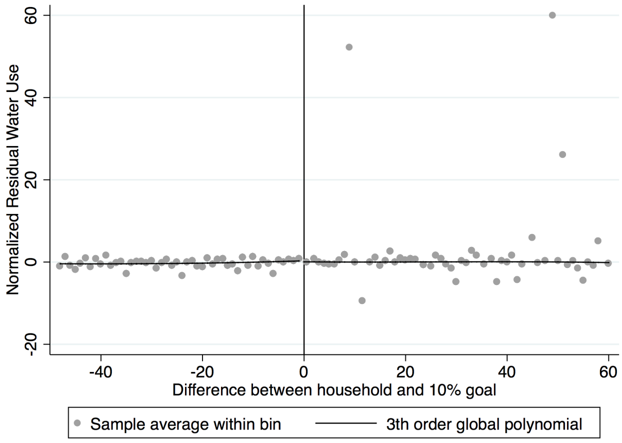
	(1) Kgal (T4)	(2) % (T5)
Conventional	1.326	-0.159
	(0.897)	(1.037)
Bias-corrected	1.329	-0.212
	(0.897)	(1.037)
Robust	1.329	-0.212
	(1.063)	(1.234)
Observations	3,842	2,869

Note: The rows are three separate RD estimators and the appropriate standard errors according to Calonico et al. (2014). The dependent variable is residual normalized consumption. The discontinuity in column (1) is based on the moving above the peer group's consumption (T4), and in column (2) is based on the moving above the peer group's year-on-year change in consumption (4).

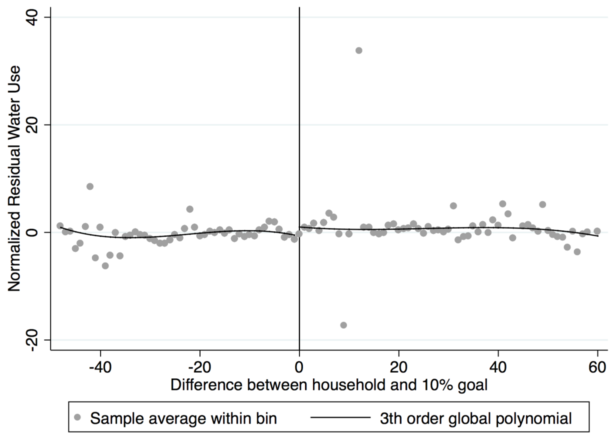
We repeat this exercise to see if moving slightly below above the 10% goal influences consumption. The analysis is the same as reported above except the running variable is the year-on-year percentage change in water consumption and the threshold is the -10%. To be consistent with the analysis in the main text we subtract 10% off the running variable and such that the threshold is at zero and year-on-year changes of less than 10% are positive and more than 10% are negative. Figure A.7 graphs residual consumption on the y-axis with the year-on-year percentage change in water consumption (minus 10%) on the x-axis. There is no visual evidence of a change in consumption right at the threshold. This is corroborated with the RD estimates (Table A.13) for each of the treatments.

Figure A.7: Effect of the Failing to Meet the 10% Goal

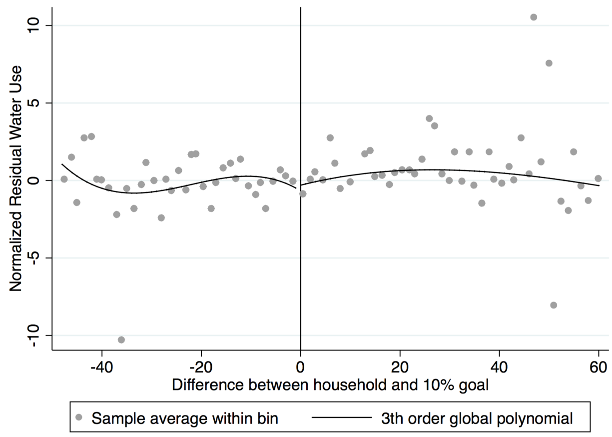
(a) Rate (T3)



(b) Social Comparison kgal (T4)



(c) Social Comparison (T5)



The dependent variable is residual normalized consumption and the units are percentage terms. The discontinuity is based on the moving above the consumption threshold that constitutes the household's 10% goal. Data from all treatments are included.

Table A.13: Regression Discontinuity Estimates of the Failing to Meet the 10% Goal

	(1) History (T2)	(2) Rate (T3)	(3) Social kgal (T5)	(4) Social % (T5)
Conventional	0.238	-0.308	1.022	-0.047
	(1.104)	(1.193)	(1.253)	(1.146)
Bias-corrected	0.338	-0.387	1.262	0.058
	(1.104)	(1.193)	(1.253)	(1.146)
Robust	0.338	-0.387	1.262	0.058
	(1.312)	(1.415)	(1.443)	(1.338)
Observations	3,472	3,568	3,409	3,269

Note: The rows are three separate RD estimators and the appropriate standard errors according to Calonico et al. (2014). The dependent variable is residual normalized consumption. The discontinuity in column (1) is based on moving above the consumption threshold that constitutes the household's 10% goal. The columns show the pooled treatment and each of the individual treatments.