Harvard Environmental Economics Program

DEVELOPING INNOVATIVE ANSWERS TO TODAY'S COMPLEX ENVIRONMENTAL CHALLENGES

May 2015
Discussion Paper 15-63

Rank and Response: A Field Experiment on Peer Information and Water Use Behavior*

Syon Bhanot
Ph.D. 2015, Harvard University, Public Policy

*This paper won the Harvard Environmental Economics Program's 2015 Enel Endowment Prize for the Best Paper by a Doctoral Student.

HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

heep@harvard.edu
http://heep.hks.harvard.edu

Rank and Response: A Field Experiment on Peer Information and Water Use Behavior

Syon Bhanot

Harvard University

The Harvard Environmental Economics Program

The Harvard Environmental Economics Program (HEEP) develops innovative answers to today's complex environmental issues, by providing a venue to bring together faculty and graduate students from across Harvard University engaged in research, teaching, and outreach in environmental, energy, and natural resource economics and related public policy. The program sponsors research projects, convenes workshops, and supports graduate education to further understanding of critical issues in environmental, natural resource, and energy economics and policy around the world. For more information, see the Program's website: http://heep.hks.harvard.edu.

Acknowledgements

The Enel Endowment for Environmental Economics at Harvard University provides major support for HEEP. The Endowment was established in February 2007 by a generous capital gift from Enel SpA, a progressive Italian corporation involved in energy production worldwide. HEEP receives additional support from the affiliated Enel Foundation.

HEEP also receives support from the Alfred P. Sloan Foundation, Bank of America, BP, Castleton Commodities International LLC, Chevron Services Company, Duke Energy Corporation, and Shell. HEEP enjoys an institutional home in and support from the Mossavar-Rahmani Center for Business and Government at the Harvard Kennedy School. HEEP collaborates closely with the Harvard University Center for the Environment (HUCE). The Center has provided generous material support, and a number of HUCE's Environmental Fellows and Visiting Scholars have made intellectual contributions to HEEP.

HEEP and the closely-affiliated Harvard Project on Climate Agreements are grateful for additional support from the Belfer Center for Science and International Affairs and the Hui Research Fund for Generating Powerful Ideas at the Ash Center for Democratic Governance and Innovation—both located at the Harvard Kennedy School; Christopher P. Kaneb (Harvard AB 1990); and the International Emissions Trading Association (IETA).

Previous sponsors of the Harvard Project on Climate Agreements include: ClimateWorks Foundation, the Doris Duke Charitable Foundation, and the James M. and Cathleen D. Stone Foundation.

Citation Information

Bhanot, Syon. "Rank and Response: A Field Experiment on Peer Information and Water Use Behavior." Discussion Paper 2015-63. Cambridge, Mass.: Harvard Environmental Economics Program, May 2015.

Rank and Response:

A Field Experiment on Peer Information and Water Use Behavior

Syon P. Bhanot*

April 17, 2015

Abstract

Perception of social rank, or how we perform relative to our peers, can be a powerful motivator. While research exists on the effect of social information on decision making, there is less work on how ranked comparisons with our peers influence our behavior. This paper outlines a field experiment conducted with 5,180 households in Castro Valley, California, which used household mailers with various forms of peer information and social rank messaging to motivate water conservation. The experiment tests the effect of a visible social rank on water use, and how the cooperative and competitive framing of rank information influences behavioral response. Difference-in-difference and matching methods reveal sizable treatment effects of the mailers on household water use (reductions of 13-17 gallons per day, depending on mailer version). However, households with relatively low or high water use in the pre-treatment period responded differently to information framing. We find that neutrally-framed rank information caused a "boomerang effect" (i.e., an increase in average water use) for low water use households, but this effect was eliminated by competitive framing. At the same time, competitively-framed rank information demotivated high water use households, increasing their average water use further. This result is supported by evidence that the competitive frame on rank information increased water use for households who ranked "last" in the peer group - a detrimental "last place effect" from competitive framing.

*Harvard Kennedy School. Email: syon_bhanot@hksphd.harvard.edu. I would like to thank Richard Zeckhauser, Brigitte Madrian, and Michael Norton for their guidance and advice. Additionally, I want to thank Alberto Abadie, Hunt Allcott, Dan Ariely, Gary Charness, John List, Tim McCarthy, Duncan Simester, Monica Singhal, and seminar participants at the UK Behavioral Insights Team, UCSD, and Harvard for their feedback. A special thanks is also due to Ora Chaiken, Chad Haynes, and Peter Yolles, without whom this work would not be possible. Finally, I want to acknowledge Peter Hadar, Shahrukh Khan, and especially Vivien Caetano for their excellent work as research assistants at various stages during this project. Note that the most up-to-date version of this paper is always available through the following link: http://goo.gl/AlWbex

1 Introduction

Research in psychology and behavioral economics has consistently demonstrated that our social surroundings affect our behavior and that we are influenced by how we compare to our peers (Schultz et al., 2007; Grisevicius et al., 2008; Beshears et al., 2014; Fehr and Gintis, 2007). Traditionally, economists studying decision making have focused less on these social motivators and more on financial ones. However, financial incentives are unable to change behavior in some contexts, either because pricing is not salient or because behavioral elasticities are low (Ferraro et al., 2011; Allcott, 2011; Olmstead and Stavins, 2007). In these cases, it may be beneficial to stimulate behavior change with social motivators, like the desire to attain a high rank relative to our peers (Tran and Zeckhauser, 2009). In this paper, we present an experiment that tests the effect of social rank on behavioral response, and explores how the framing of peer information can influence this response.

Experimental researchers have studied how social information can alter behavior in a variety of contexts, including energy conservation, voting, and savings (Allcott, 2011; Gerber et al., 2008; Kast et al., 2012; Beshears et al., 2014).[1] Most interventions have provided individuals with information on the average performance of a broader social group, with mixed results. Allcott (2011), for example, finds that showing individuals how their energy use compares to the mean of both their most efficient neighbors and all of their neighbors reduces electricity consumption in the average household by over 2%. However, other research suggests that sharing peer information can lead to socially undesirable behavior. Beshears et al. (2014) find that the provision of peer information about savings for retirement can reduce savings rates by demotivating low-performers. John and Norton (2013) document a related phenomenon in the context of workplace exercise "walkstations." They find that people tend to converge to the bottom performer, exercising less at walkstations when given information about the low rates of use by others. Another set of studies on the use of social information to influence alcohol abuse on college campuses find no discernable effect of such messaging on overall outcomes (Wechsler et al., 2003; Clapp et al., 2003; Granfield, 2005).

One limitation of existing work is that it does not disentangle the various motivating and demotivating elements of social information, and does not isolate those that are central to behavioral responses. When social information works, is it because the information primes our competitive drive or because it stimulates our cooperative spirit? The literature also does not say a great deal about peer comparisons, explicit rank

[1] Beshears et al. (2014) identifies a number of additional studies in this space, noting that, "providing information about peers moves behavior towards the peer norm in domains such as entrée selections in a restaurant, contributions of movie ratings to an online community, small charitable donations, music downloads, towel re-use in hotels, taking petrified wood from a national park, and stated intentions to vote (Cai, Chen, and Fang, 2009; Chen et al., forthcoming; Frey and Meier, 2004; Salganik, Dodds, and Watts, 2006; Goldstein, Cialdini, and Griskevicius, 2008; Cialdini et al., 2006; Gerber and Rogers, 2009)." (Beshears et al., 2014, 1)

information, or heterogeneities in their motivational effects. How do ranked comparisons to specific people who are "like us" motivate us differently than aggregate social comparisons? We hope to offer insights into this question.

In this paper, we outline a natural, randomized field experiment that tests how peer information influences behavior.[2] The experiment was conducted with a partner firm in California, which works with local utilities to reduce water use at the household level through household mailers and other outreach campaigns. In the experiment, we used mailers with different forms of peer information and social rank messaging to motivate reductions in household water use. Through the experimental design, we can address existing theories about how peer information, peer comparison, social rank, and the framing of behavioral messaging can influence behavior. The goal of this study is to provide insights that contribute to an improved, coherent theory about social information and its potential for heterogenous effects.

Our results suggest that while social information can reduce water use, peer ranking and framing can have detrimental impacts on behavior. Specifically, we find overall water use reductions in the range of 13-17 gallons per day in response to a four-piece mailer campaign containing different peer information and social rank content. However, we find evidence of heterogeneity in treatment effects from rank information. In particular, households that were low water users prior to the experiment showed a "boomerang effect" (i.e., an increase in water use) from rank information, except when a competitive frame was included. This result is consistent with Garcia et al. (2006), who posit that recieving high rankings can spur competitiveness in a way that makes people less likely to "boomerang." However, the competitive frame had detrimental effects on the behavior of households that were high water users prior to the experiment, demotivating them and increasing their water use on average. Further analysis of rankings suggests the existence of a "last place effect," whereby competitively-framed rank information led to an increase in water use by the worst performer in the peer group - a movement away from the social norm. We believe this stems from the potentially demotivating power of peer information, in line with the results on peer information and savings in Beshears et al. (2014).

The paper proceeds as follows. Section 2 provides background on existing work, related theories, and the policy context for this intervention. Section 3 provides details on the experiment. Section 4 presents the

[2] We refer to four forms of behavioral messaging in the paper: social information, peer information, peer comparison and social rank. We define and distinguish between them as follows. Social information is the broadest category, referring to any messaging containing information about the behavior of others. Peer information is a subset of social information, referring to messaging that conveys information about a given individual's behavior and information about their peers (people "like them"). This term encompasses information conveyed either at the aggregate level ("here's how your peers performed on average") or in a more detailed manner ("here's how you performed relative to a similar household"). Peer comparison is a subset of peer information, referring to the display of the specific outcomes of an individual and their peers, provided explicitly at an individual level. Finally, social rank refers to messaging that informs individuals of their hierarchical position among peers.

empirical methods used to analyze the data from the experiment. Section 5 provides results. Section 6 provides a brief discussion and concludes.

2 Background

Water conservation provides an important context to test the effects of social information, since individual water use behavior is both important to change and difficult to influence. Water leaks provide a useful illustrative example. In a 1999 study, the American Water Works Association Research Foundation found that nearly 14% of household water use comes from leaks.[3] The leak problem is diffuse - the EPA estimates that roughly 10% of homes have leaks that waste 90 or more gallons of water each day.[4] Fixing a leak can significantly reduce a household's water use. Yet few households seem to be fixing their leaks.

The absence of such efforts can be partly explained by the human tendency to only process salient information.[5] Water leaks are generally invisible, requiring professional assistance to find and fix.[6] This salience problem is further compounded by the relatively low price of water - the average family in the United States spends only 0.5% of household income on water and sewage bills.[7] Given the lack of salience and the low price, it is not surprising that the price elasticity of water is low. Olmstead and Stavins (2007) estimate a water price elasticity of -0.33 and point to an earlier analysis by Espey et al. (1997) that placed 90 percent of all estimates between 0 and -0.75. More recent estimates in California in Lee and Tanverakul (2015) find elasticities in the -0.2 to -0.5 range, suggesting that water prices would have to increase significantly in our experiment's target areas to have a major impact on water conservation decisions.[8]

Households are also unlikely to change their water use without knowledge of what constitutes "good" and "bad" water consumption behavior in their community. Generally, households do not receive such information. However, social information interventions may offer a solution, by providing a reference point for individuals to evaluate their water use.[9] There is evidence that providing social information in this way can be a simple and low-cost means of changing behavior (Jessoe and Rapson, 2014; Dickerson et al., 1992).

[3] Mayer and DeOreo (1999)

[4] WaterSense (2015)

[5] This concept is known as "bounded attention" in the behavioral sciences and was popularized by Daniel Kahneman (see **?**).

[6] American Water has a 'Leak Detection Kit' online that outlines common indoor leaks in detail and how to find them. American Water is in partnership with EPA's WaterSense program, and their guide to finding and fixing leaks illustrates how difficult finding and fixing such leaks can be.

[7] United States Environmental Protection Agency and Water (2009)

[8] In Allcott (2011), providing social norms and information decreased energy use by roughly the same amount as a 11-20% increase in price.

[9] Another potential danger of not providing households with such information is heuristic decision-making, which has been studied in the energy conservation literature (Gillingham et al., 2009). When households do not have enough information about their energy use, they may rely on heuristics to determine their energy consumption, which a number of experiments found led to miscalculations of use and overconsumption (Kempton et al., 1992; Kempton and Montgomery, 1982).

Given this, we designed our experiment to explore the influence of social rank and peer information in the context of a water conservation mailer program.

2.1 Existing Theories on Rank and Response

A number of important theories from social science literature might explain how social rank and peer information affect behavior – with very different predictions. A brief discussion of these theories and their predictions is presented here.

2.1.1 Social Norms Theory

Social norms theory predicts that peer information, including social rank, motivates behavior change because it provides a social standard to follow. Most notably, the theory of social comparison processes presented in Festinger (1954) suggests that social comparison occurs when objective, non-social standards are unavailable. This could lead individuals to evaluate their opinions and abilities by comparing themselves to others - and to take action to reduce any found discrepancies. Furthermore, Festinger argues that individuals are most likely to compare themselves to, and more likely to reduce discrepancies when compared to, people who are similar to them.[10] Schultz et al. (2007) similarly state that social information can send the message that "being deviant is being above or below the norm."[11] Social norms theory then implies that providing individuals with rank information would cause their outcomes to compress towards the displayed social norm.

In recent years, there have been an increasing number of experimental tests of these theories. For example, Schultz et al. (2007) conducted a field study with several hundred households in San Marcos, California, using door hangers with aggregate-level social information on energy use to motivate energy reduction. They find that social information caused high energy use households to decrease their energy use, but encouraged low energy use households to increase energy use. On the one hand, this implies a desirable response to social information from low-performing individuals. However, it also predicts a detrimental response from high-performers, referred to as the "boomerang effect."

The boomerang effect hypothesis has its roots in the psychology of motivation. Proponents of the effect suggest that a favorable social comparison provides a license for high-performing people to behave worse (Clee and Wicklund, 1980). For example, telling individuals that most people in their workplace do not put in overtime sends the message that putting in overtime is unnecessary. The boomerang effect has been

[10] For a discussion on the specific variables relevant for comparison (e.g. expertise, similarity and previous agreement), see Suls et al. (2002).

[11] Schultz et al. (2007), p. 430

documented in some interventions focused on social rank, but there is a relative shortage of experimental evidence on the effect (Fischer, 2008). In this experiment, we explore how framing and context might influence the boomerang effect.

It is important to note that any observed compression towards the mean could be attributed to mean reversion, which can happen in any experimental setting in the absence of messaging with social norms (Kahneman, 2011). We must be careful when assessing interventions using social rank information to ensure that we do not confuse mean reversion with responses to social norms, particularly in the short run. Using a randomized experiment that varies whether subjects are exposed to specific social norms information helps us distinguish between these effects.

2.1.2 Motivation Effects

Academic literature on motivation and self-efficacy suggests a third possibility: that individual outcomes will widen away from the mean, as those who rank well among their peers will work harder to improve and those who rank poorly will "give up."[12] There is a rich body of research underpinning this prediction in the social sciences. Research with children has shown that our beliefs about our abilities are influenced by what others tell us.[13] If we feel – or are told – that we are good at an activity, we are more likely to engage in it, whereas we avoid activities for which we feel ill-equipped (Bandura, 1977).[14] Research in exercise and sport performance has shown that verbally reciting instruction messages that convey positive beliefs improves ensuing performance outcomes (Shelton and Mahoney, 1978). This suggests that individuals with positive beliefs about their ability may set higher goals for themselves and try harder to achieve them. This is commonly attributed to the view that "high efficacy" people view difficult or new tasks as challenges rather than threats (Bandura, 1994; Yim and Graham, 2007).

Individuals with low-efficacy (those who receive low-rankings), on the other hand, may quit once they learn of their poor rank (Hagger et al., 2002). Beshears et al. (2014) found that low-savings individuals were discouraged by information about peers' savings rates, which the authors attributed to the discouraging effects of upward social comparison. This response, now commonly referred to as the "what the hell effect," was also identified in dieters by Polivy and Herman (1985). The authors found that once a dieter exceeds their caloric intake goal for a single day, they proceed to eat significantly more calories than the goal. In short, they perceive their performance as a failure and respond by "binging" on food.

[12] For a summary of literature in this area, see the Pajares chapter in (Pajares, 1997).

[13] **?**

[14] Educational research has notably used this theory to suggest that teacher efficacy, or a teacher's belief in his or her ability to bring out the best in students, has powerful effects on student achievement, student motivation, and teacher behavior (Tschannen-Moran and Hoy, 2001).

The prediction of demotivated low-performers also finds support in loss aversion, which posits that a given loss affects the psyche more than an equivalent gain (Kahneman and Tversky, 1979). In this context, a poor performance relative to the social norm may be perceived as a "loss," and the worst-performing individuals may feel demotivated by the impossibility of catching up. The high emotional weight of losses may therefore bring "loss demotivation." Taken together, the what-the-hell effect, self-efficacy theories, and loss demotivation suggest that providing information on social rank may cause the spread of outcomes to widen, with on the top inspired to try harder and those at the bottom giving up.

Figure 1 provides a simple visual of the various predictions of behavioral response to rank and supporting theories outlined here.

Figure 1: Rank Response Predictions and Supporting Theories

Predicted Response to Rank	**Supporting Literature/Theories**
Top Performers Improve	Self-efficacy (Bandura, 1977, 1994) and Rank Competition (Garcia et al., 2006)
Top Performers Worsen	Social comparison (Festinger, 1954) and Social norms (Clee and Wicklund, 1980; Schultz et al., 2007)
Bottom Performers Improve	Social comparison (Festinger, 1954) and Social norms (Clee and Wicklund, 1980; Schultz et al., 2007)
Bottom Performers Worsen	Self-efficacy (Bandura, 1977, 1994), Demotivation (Hagger et al., 2002; Beshears et al., 2014; Polivy and Herman, 1985), "Loss Demotivation" (based on Kahneman and Tversky (1979))

2.2 Social Messaging Frames

Literature in psychology, behavioral economics, and marketing has reliably found that the framing of information can alter behavior (Winter, 2008). While existing work has not explored framing in the context of social rank specifically, there are related results that provide testable predictions. For example, Schultz et al. (2007) argue that the "boomerang effect" can be counteracted if an injunctive message, which conveys social approval or disproval, is included. The authors found that low energy users receiving an injunctive message maintained their low use rates, while those who did not suffered a boomerang effect.[15] Similarly, Allcott and Muillainathan (2010) attribute the lack of a boomerang effect in their experiment to the use of similar injunctive messaging.

[15]In Schultz et al. (2007), the injunctive message used was a "happy face" if they consumed less than the social average and a "sad face" if they consumed more than the social average.

In this experiment, we use both competitive and cooperative frames to explore the nuances of how social information influences behavior, building on past work. For example, in an experiment exploring the effect of negative and positive frames of cooperative messaging on behavior, Cialdini et al. (2006) found that positive framing of cooperative messaging encouraged people to cooperate. Conversely, negatively-framed cooperative messaging provided people with a justification for their own bad behavior. Other research suggests that cooperative frames can backfire when there is no widespread rule adherence, as individuals might use the cooperative setting to free ride (Tauber, 1972; Olson, 1965).

Framing social comparison as a competition, through status rewards such as medals, can also be powerful. Online and social media giants like Foursquare, Overstock, Yelp, and Wikipedia all use non-financial status rewards to motivate users. These rewards prime our competitive desire to obtain a higher social rank, and can serve as a form of personal affirmation that increases self-efficacy (Antin and Churchill, 2011). Furthermore, Garcia et al. (2006) argue that ranking may itself drive competitiveness, finding that individuals are most competitive when they or their competitors are highly ranked. They also argue that degree of competition between rivals "depends on their proximity to a meaningful standard."[16] Therefore, amongst top performers, rankings and competitive framing may mutually reinforce in a way that motivates positive behavior change.

Meanwhile, we might expect low-performing individuals to change their behavior to avoid "being last," particularly if they are close to the margin, due to "last place aversion" (Kuziemko et al., 2014). However, research on goal-setting and attainment suggests that the use of competition and rank can also be demotivating. For example, both Beenen et al. (2004) and Harding and Hsiaw (2014) suggest that individuals may do worse if they feel that their target goals are unachievable. Similarly, Little (2012) finds that competitive frames can demotivate individuals if they reinforce patterns of failure.

Overall, there is little consensus on the relative value of cooperative versus competitive framing (Little, 2012; Qin et al., 1995; Kohn, 1996; Julian and Perry, 1967). While this experiment will not resolve these debates, it contributes experimental evidence in a specific context that can inform further research on the use of cooperative and competitive frames to motivate behavior change.

2.3 Policy Context

Water use is a serious issue in many parts of California.[17] In a statement in April 2014, the Director of the California Department of Water Resources Mark Corwin said, "We're already seeing farmland fallowed

[16] **Garcia et al. 2006, p. 970**

[17] President Barack Obama declared in a February 2014 visit, "As anybody in this state could tell you, California's living through some of its driest years in a century. Right now, almost 99 percent of California is drier than normal - and the winter snowpack that provides much of your water far into the summer is much smaller than normal." (See Obama (2014))

and cities scrambling for water supplies. We can hope that conditions improve, but time is running out and conservation is the only tool we have against nature's whim."[18] The ongoing crisis in California brought President Barack Obama to the state as well. In public remarks in Los Banos in February 2014, President Obama stated, "everybody, from farmers to industry to residential areas [... is] going to have to start rethinking how we approach water for decades to come."[19]

This is not a problem unique to California. At present, roughly one billion people worldwide lack access to safe drinking water. Increasingly, experts warn that this persistent shortage of water resources will have ramifications not only for human health and the environment, but for political stability and national security.[20] Based on the notion that small efforts at reduction can have major impacts, water conservation efforts have focused on the overuse of water at the household level. California Governor Jerry Brown has used this approach to address the current water crisis in the state, saying, "every day this drought goes on we are going to have to tighten the screws on what people are doing."[21]

Some argue that directly restricting water use or raising prices may be the answer, but these two options have flaws. Government-imposed restrictions on water use may make individuals less enthusiastic about conserving water, undermining long-run behavior change (Lynne et al., 1995; Watson et al., 1999). Furthermore, a report by the Pioneer Institute concludes that such programs can also be expensive, especially considering that empirical evidence regarding their aggregate effects is mixed (Olmstead and Stavins, 2007). Price-based approaches and subsidies are similarly underwhelming. Hunt Allcott argues that they are politically infeasible, "in practice a large drain on increasingly-limited public funds," and difficult to evaluate (Allcott, 2011, 1082). The problem of price inelasticity discussed earlier also means that the impact of price-based regulation would be limited. This suggests a possible role for behavioral interventions in water conservation policy.

3 Experiment Overview

3.1 Motivating Literature for Experiment Design

With this field experiment, we seek to understand the mechanisms of social information effects, and how framing affects behavior change. The basic design features of the experiment draw and build on existing studies on social information, including in the water conservation context (Petersen et al., 2007; Davis, 2011;

[18] Thomas and Carlson (2014)
[19] Obama (2014)
[20] Community (2012)
[21] Nagourney and Lovett (2014)

Allcott, 2011; Ferraro and Price, 2013; Allcott and Rogers, 2014). We extend existing work by framing peer information using both cooperative and competitive frames in some of our treatment groups. These frames were chosen to prime different elements of social information, which we argue might influence individuals differently. For example, cooperative frames motivate behavior change by inducing norm compliance, whereas competitive frames motivate behavior change by making an action seem valuable (Fehr and Gintis, 2007; **?**).

Additionally, households received multiple mailers in this experiment, which is uncommon in the literature. Most existing experimental work on water conservation has used single mailers sent to households (Ferraro et al., 2011; Ferraro and Miranda, 2013). Furthermore, studies that do use multiple mailers in other contexts tend not to address the effect of additional mailers (Karlan and Zinman, 2009; Bertrand et al., 2010). One exception is Allcott and Rogers (2014), which finds that households responded to repeated mailer treatments over two years, with some decay in effects when mailers ceased. In this paper, we assess both the short- and long-term impacts of the treatment during the experimental period.

3.2 Experimental Design

3.2.1 Partners

My research partner for this field experiment was a firm based in California that works directly with public water utilities to promote more efficient water use by California homeowners.[22] The firm sends a personalized mailer, called a Home Water Report (HWR), to households every two months. The HWRs are transmitted either electronically or through traditional mail, and incorporate messages designed to engage customers and reduce water use.[23] Through the utilities, the firm tracks water use and customer engagement over time. The public utility partner in the study was a local water provider that serves a subset of homes in the greater San Francisco Bay area. The utility provided the water use data for analysis of the experiment's impact.

3.2.2 Subjects

We conducted the field experiment in Castro Valley, a town of 60,000 residents in Alameda Country, California, roughly 15 miles southeast of Oakland. Subjects in the study were residents of 5,180 single-family households in the C2A pressure zone in Castro Valley, who receive water through the public utility.[24] The visuals in Appendix 1 show the specific location of both Castro Valley and of the C2A pressure zone within

[22]The firm will remain anonymous in this paper.

[23]Approximately 10% of customers receive the Home Water Report by email, with the rest receiving paper mailers.

[24]A "pressure zone" is a geographical area defined by the public utility based on the area's elevation above sea level. A map of these pressure zones in Castro Valley is visible in the Appendix 1.

Castro Valley specifically. Prior to the start of this experiment, the firm was already working with roughly 4,000 households in the other pressure zones in Castro Valley. This study specifically targeted households in the C2A pressure zone, who were being added to the existing base of Castro Valley customers. This is critical - all those receiving mailers in the study received the Home Water Report for the first time.

3.2.3 Study Design

The 5,180 households in the experiment were first subdivided into 20 "cohorts" based on two categorical variables: 1) outdoor irrigable area; and 2) the number of occupants in the household. There were four possible irrigable area sizes for a household (small, medium, large, and extra large), with irrigable area computed by the firm using real estate data on lot size and home footprint from DataQuick. There were five possible household occupant "buckets" (1, 2, 3, 4, and 5+). Therefore, with four possible "irrigable area" values and five possible "occupant" values, there were 20 cohorts of households in the experiment. Chart 1 in Appendix 3 outlines the number of households in each of the cohorts.

Every household was then individually assigned a random subset of four households in their cohort, referred to as their "water group." Using the cohorts ensured that homes were only paired with homes with roughly the same water needs. Importantly, a water group was assigned for all households in the experiment, including those receiving the control mailer. This allowed us to use the control group directly to analyze the effects of ranking, group performance, and other characteristics of the peer comparison. Specifically, we can consider what control households "would have" received as a peer comparison, had they been assigned to receive one.

Finally, once households had been assigned a water group from within their cohort, households were randomly assigned to one of the four experimental mailer groups in the study - a control mailer group (the "In-Sample Control" group) and three treatment mailer groups (the "Rank," "Team," and "Competitive Rank" treatment groups). The treatment and control groups are outlined in detail in section 3.2.4. Each of these groups received up to four Home Water Reports over the course of the experiment, but with different information in the treatment area, as outlined below.[25] This mailer was delivered to each household in the experiment every two months, by postal mail or email. Households in each experimental group got the same version of the mailer each time (in other words, a household assigned to the "Rank" treatment group received up to four "Rank" treatment mailers). A sample HWR is included in Appendix 2, with the "treatment area" labeled.

[25]Some homes did not receive all four mailers because of logistical issues or asynchronous timing of water delivery and water readings.

A few things are worth noting about this setup. First, each of the households was linked with unique water use outcomes – each water meter is associated with a single household unit. There is no complication from shared housing units with a single water bill. Second, all groups discussed so far received the Home Water Reports, which contained information about water use above and beyond what was randomly assigned to the treatment area. This information, which included a "WaterScore" driven by overall data on mean water use in the town, would likely have had an effect on water use independent of the experiment treatments. This is true for both the treatment mailer groups and the control mailer group. While this can be controlled for in the analysis to some extent, data was also collected from a neighboring town within the public utility's coverage area to serve as an additional control group (the "Out-of-Sample Control" group). This data enables us to assess the efficacy of the mailers compared to the counterfactual of not receiving any mailers. Third, each individual household's water group was unique - just because household A was assigned a "water group" consisting of households B, C, D, and E, this did not (and in fact rarely) meant that household A appeared in B, C, D, or E's water group, due to the randomization at the individual level. Fourth, note that each household's water group consisted of homes in the same cohort but not necessarily in the same treatment group. So while the rest of a given household's water group are certain to be in the same "number of occupants" category and the same "irrigable area" category, the rest of the group was unlikely to be receiving the same "treatment" as that household.

3.2.4 Treatments and Controls for In-Sample Households

Households in the sample were randomly assigned to receive one of four different mailer versions, which differed in the information displayed in the "Treatment Area" of the mailer, as labeled in Appendix 2. Figure 1 below displays the control and three treatment mailer versions.

Figure 1: Control and Treatment Mailer Versions

Households assigned to the In-Sample Control group received the standard HWR, with a "Got water questions?" insert in the treatment area. Note that no information about the water group was transmitted to households in the In-Sample Control, nor was the household made aware that any comparison water group had been created.

Households in the Rank treatment group received an HWR with a ranked comparison in the treatment area. This treatment provided a simple social comparison for informational purposes, with neutral framing.

Households in the Competitive Rank treatment group received an HWR with a competitively-framed ranked comparison in the treatment area. This treatment provided the same social comparison as in the Rank treatment, but with a competitive frame (using "Go for the Win!" messaging and a ribbon icon), to assess the effect of priming a competitive instinct to encourage behavior change.

Households in the Team treatment group received an HWR with a cooperatively framed "Team Challenge" in the treatment area. Note that the Team treatment did not include information on social rank or peer comparison within the water group, but emphasized the group as a collective (and provided a comparison of the household's water group with other water groups). As a result, subjects in this treatment did not know the precise water use of the other homes in their water group, though they would have been able to deduce how their personal water use differed from that of their team by comparing the team average to their own usage (available elsewhere in the HWR).

Finally, since the In-Sample Control group shown above received mailers, it was important to have a second control group in the experiment that was not sent any mailers. For logistical and administrative reasons, it was not feasible to have this control group in the experimental location itself. Therefore, 2,880 households from the nearby Dingee pressure zone were used as an "Out-of-Sample Control" group for the experiment.[26]

3.2.5 Timeline

The experiment began in November 2012. The firm sent out the first mailers at the end of November, using October 2012 meter reads. The firm then sent three additional mailers, with the same treatment/control messaging, in January 2013 (based on December 2012 meter reads), March 2013 (based on February 2013 meter reads), and May 2013 (based on April 2013 meter reads). Households in the experiment that had meter reads ouside of the four key meter read months (October 2012, December 2012, February 2013, or April 2013) did not receive a experimental mailer in the month that followed their read.

3.3 Data and Baseline Characteristics

We collected the data for this study from the firm, who obtained it through their partnership with the public utility. Two types of data were collected. First, we collected water use data for the households in the experiment, for the periods before and during the experiment. Second, the firm provided us with data on the characteristics of the households in the study, which they obtained both from the public utility and from independent data sources including DataQuick.

3.3.1 Descriptive Data and Baseline Characteristics

We observed data from all 5,180 experimental households in Castro Valley, of which 4,265 received all four experimental mailers. Chart 2 in Appendix 3 outlines the number of households in each treatment and control group, and the number of households in each group that received all four mailers. In addition, we observed data for 2,880 additional households from the neighboring Dingee area to serve as the Out-of-Sample Control. Table 1 shows demographic information for both In-Sample and Out-of-Sample households. The In-Sample households' demographics are presented both broken up by treatment, and overall. Note that

[26] The Dingee pressure zone (code B5A) spreads across parts of the Berkeley/Piedmont/Oakland area. All homes in Dingee, with the exception of a small number of homes in the pressure zone that had received mailers as part of a prior firm pilot, were used as the Out-of-Sample control group.

the Out-of-Sample households differ from the In-Sample households in these measurables, which is expected since they are in different areas. The Out-of-Sample homes and properties are, on average, slightly larger and older than the In-Sample homes, and used more water than the In-Sample homes in the 2012 months prior to the experiment.

3.3.2 Pre-Treatment Water Use Trends

Meter read technicians from the water utility measured water use every two months (the key reads for this experiment were in December 2012, February 2013, April 2013, and June 2013). Most meters used CCF units for water use (1 CCF = 100 cubic feet of water = 748 gallons), and the CCF reads were converted into a "gallons per day" (GPD) measure by the public utility. The mean water use in the In-Sample area (as well as in the Out-of-Sample Control area) prior to the experiment, measured in GPD, is visible in Table 1. Additionally, Appendix 4 provides a graphic of the water use trends in the two areas prior to the experiment. Notice the key role that seasonality plays in water use; water use is higher in the summer than in the winter. Because of this, we used month fixed effects in certain specifications to control for seasonal trends.

Note that while mailers were sent on the same date for all households in each mailing cycle, households did not have meter reads on the same date. As a result, there is some variance in how many days a given household was treated by a single mailer. For example, some households would have received the mailer in the week before their next meter read while others received it in the week after their previous meter read. This is not an uncommon issue, having appeared in similar experiments using read-based mailers, including Allcott (2011). Successful randomization deals with this issue to some extent, in that there is no correlation between treatment and meter read cycle, meaning that on average each condition's households was "treated" by a given mailer cycle for the same fraction of the period post mailing.

3.3.3 Randomization Check

Some recent research questions the need for randomization checks in experiments (Mutz and Pemantle (2011), for example). However, in this instance randomization checks are warranted for two main reasons. First, the randomization process itself was conducted by the firm and not the researcher. Though the firm has a track record of experimentation and a strong background in randomization procedures, a check is needed to ensure that there was no systematic error in randomization. Second, some households were dropped after randomization but prior to study implementation. In particular, 355 households did not receive a mailer despite being assigned to one of the treatment or control groups, for logistical reasons (the subject moved from the property, the address was not verified, etc.).

To test the balance of the samples on these observed demographic characteristics, we run a regression of the various demographic characteristics (written as y_i below) on dummy variables for the three treatment groups (written as T_m with m ranging from 1-3 for the three treatment groups, below), omitting the In-Sample Control. We also compute f-test statistics to determine joint significance. The econometric model is as follows:

$$y_i = \beta_0 + \sum_{m=1}^{3} \beta_m (T_m)_i + \varepsilon$$

Table 2 presents the results from these regressions. Note that none of the f-statistics and associated p-values suggest joint significance for the coefficients, suggesting that randomization resulted in balanced treatment and control groups.

3.3.4 Handling Outliers

The primary outcome measure in a given water read cycle, GPD, had occasional extreme values. We remove outliers from the analysis on both the high and low ends. First, households occasionally register a GPD of zero for a given read period. This is generally because household members are either not at home during the read period, or because their water use is so low that it fails to register on the water meter. In all household water use data (8,060 households, including the 2,880 Out-of-Sample Control households), only 135 households had a zero GPD reading for at least one meter read period after October 2011 (the relevant period for the difference-in-differences analysis used in this paper), and 90 households in the In-Sample area had at least one zero reading during the experimental period (December 2012 - June 2013). To prevent these low values from influencing the results, we conduct our analyses without these zero GPD observations, depending on the periods being analyzed in a given specification.[27]

Second, there were a few meter reads that were far above normal values. One observation in particular was dramatically in excess of normal levels (over 120,000 GPD for a single household read - the median value of mean household water use in 2011 was 193.68 GPD). The utility identified these high reads as meter malfunctions or abnormalities. To deal with such outliers on the upper tail of the distribution, we use 5,000 GPD as a cutoff for a single meter read to define excluded outliers. In total, three households in the data had a read in excess of 5,000 GPD in a read period after October 2011, and two households had a read above this threshold during the experimental period.

[27] For any analysis that used only post-experiment, In-Sample group data (means comparisons across In-Sample groups and the place effects analysis, for example), only those GPD reads of zero after the experiment was initiated were excluded. However, for difference-in-differences analysis using both In-Sample and Out-of-Sample data from the year prior to treatment as a "pre" period, any household with a zero GPD reading in any single read after October 2011 was excluded.

4 Empirical Methods

In this experiment, subjects received multiple treatment mailers. The information in the treatment area of the mailers differed by treatment group, as did the information displayed in the non-treatment sections of the mailers. As a result, we must use a variety of econometric methods to analyze the experiment and its effects. Also, the distinction between the In-Sample and Out-of-Sample Control groups is important, as comparisons between the treatments and the two control groups require different interpretations. In this section, we outline the econometric strategies used for answering different research questions in the data.

4.1 Average Treatment Effects: Difference-in-Differences, Matching, and Regression

We estimate the average treatment effect in three ways. First, we use the Out-of-Sample Control group to determine overall mailer effects in the short and long run, using a difference-in-differences approach. Second, we test the robustness of the difference-in-differences results using nearest-neighbor matching. Finally, to assess the impact of the different treatment mailer versions on household water use behavior, we use regressions that compare means across treatments while controlling for relevant variables, then disaggregate the analysis using past water use.

4.1.1 Difference-in-Differences (In-Sample vs. Out-of-Sample)

The first question we seek to address is whether the HWR mailers, including the control mailer, influenced water use. To do this, we use a neighboring pressure zone as an Out-of-Sample control and perform two types of analysis. First, we use a difference-in-differences approach.[28] The "parallel trends" assumption needed for difference-in-differences analysis seems credible given the data from the pre-treatment period. The pre-experiment mean water use in the two areas is visible in Appendices 4 and 5, which provide visuals to justify the parallel trends assumptions.

As the image in Appendix 4 shows, one concern may be that the gap in mean water use between the In-Sample and Out-of-Sample areas was larger in the summer than in the winter. Therefore, when using the difference-in-differences approach, we looked at the matching month or months for each household in the year before and the year after the treatment to control for the seasonal water use differences between the areas. We first restrict the analysis to the first meter read after the initial mailer (December 2012) to the

[28] This is necessary because the Out-of-Sample and In-Sample areas differ in observable characteristics and water use patterns.

read from the same month a year prior to treatment (December 2011). Using this approach enables us to handle differential seasonality and attain something closer to the requisite "parallel trends," as is visible in chart (A) in Appendix 5. The econometric specification is as follows:

$$GPD_i = \beta_0 + [\sum_{m=1}^{4} \beta_m (T_m)_i * (Post)] + [\sum_{m=5}^{8} \beta_m (T_m)_i] + \beta_9 (Post) + \varepsilon$$

In this specification, "Post" is a dummy variable for whether or not the reading was from December 2012. The β_1,β_2,β_3, and β_4 coefficients serve as difference-in-difference estimators of the causal impact of the mailers on water use. Note there are four treatment dummies here (In-Sample Control, Rank, Competitive Rank, and Team), with the Out-of-Sample Control the omitted group. Furthermore, no household characteristic controls are used, as these variables did not vary within a household over time in the data.

We then repeat this approach using the mean water use in each household in all relevant pre- and post-experiment periods.[29] Specifically, we use the mean household water consumption in the December 2012, February 2013, April 2013, and June 2013 meter reads as the post-experiment water use outcome. We then collect data on the water use of these households in the matching months pre-experiment (from the December 2011, February 2012, April 2012, and June 2012 reads), and compute a mean for each household in the four pre-experiment reads. By doing this, we obtain a pre- and post-experiment mean that aggregates multiple meter reads.[30] We then run the above regression specification for a difference-in-differences estimator using the overall mean water use, to approximate an overall effect of the mailers. This approach helps us deal with the seasonality issue across the treated and control areas, and strengthens the justification for the parallel trends assumption (as visible for this specification in chart (B) in Appendix 5). Furthermore, it acknowledges the inherent issues with serial correlation in difference-in-differences estimators raised by Bertrand et al. (2004), improving our faith in the standard errors of our estimates.[31]

4.1.2 Matching Estimators (In-Sample vs. Out-of-Sample)

To check the robustness of the above difference-in-differences results, we compute average treatment effects using a matching framework. Specifically, we use nearest-neighbor matching to match individual households in the In-Sample area (who all received mailers) to households in the Out-of-Sample area (who did not receive mailers). The matching is based on the water use of the households in the four water reads prior to

[29] Recall that most households received all four treatment mailers.

[30] When all four reads were not available either pre- or post- experiment for a given household, the mean for any matching pre/post months that were available was used instead.

[31] The authors write, in reference to the serial correlation problem in many difference-in-differences estimation, "collapsing the data into pre- and post-periods produce consistent standard errors, even when the number of states is small." (see Bertrand et al. (2004))

the experiment, and three household characteristics that are plausibly linked to water use (home size, lot size, and exact matching to the number of occupants in the home). The outcome variables in this analysis are water use in the first period after mailer initiation (to estimate the short-run effect) and mean water use over the four periods following mailer initiation (to estimate the long-run effect).[32] When appropriate, these matching estimators use the bias adjustment procedure outlined in Abadie and Imbens (2002) and Abadie et al. (2004), which adjusts for differences in matches in finite samples. Also, both Euclidian distance and Mahalanobis distance are used as matching metrics, though in the paper we present the matching results using Mahalanobis distance only.[33]

4.1.3 Regression for Means Comparison (In-Sample)

To determine whether mailer version influenced water conservation behavior, we use regressions that compare mean water use across mailers after the initiation of the treatment, in two ways. First, we compare mean water use in the first period (from the meter read following receipt of the the first mailer) across treatment groups. Second, to provide an estimate of the long-run differences in water use across treatment mailers, we compare mean water use for all post-treatment periods across conditions.

The econometric specification for this comparison is a regression of water use by the household (measured in gallons per day) in the relevant periods on dummy variables for the three treatments, and is visible below:

$$GPD_i = \beta_0 + \sum_{m=1}^{3} \beta_m (T_m)_i + \varepsilon$$

We also run these regressions with controls for home characteristics (lot size, home square footage, and the number of bathrooms, bundled below as ϱ_i). This specification is shown below:

$$GPD_i = \beta_0 + \sum_{m=1}^{3} \beta_m (T_m)_i + \varrho_i + \varepsilon$$

Importantly, we disaggregate the analyses to determine whether the average treatment effect differs across conditions based on past water use. Specifically, we classify households as being “low” or “high” water users using data on water use in the pre-experiment reads in 2012. Low-use households are defined as those in the bottom third of water use within each irrigable area category, and high-use households are defined as those in the top third within each irrigable area category. By assessing water use within the irrigable area

[32] For this analysis, homes for whom one or more meter reads in the four pre and post periods was missing were excluded. This reduced the number of households in the analysis from 4,908 to 4,109 for the various in-sample treatment households, and from 1,668 to 1,426 for the out-of-sample control households.

[33] The match results based on Euclidian distance show similar effects, and will be made available through the online appendix.

classifications, we are able to control for the differences in water needs based on property size, which allows for large homes with efficient residents to still be classified as “low” users.

This diversity of approaches helps us address theories about the differential effect of framing on household response to peer information, capturing heterogeneities in treatment effects that a simple average treatment effect will miss. The critical specifications center on identifying if certain mailer versions were more or less effective for households with “high” or “low” water use, pre-experiment. This allows us to test the hypotheses around potential “boomerang” or “what-the-hell” effects in social comparison messaging.

4.2 Ranking Effects

In the Rank and Competitive Rank treatments, which both displayed social rank information relative to four peers, each possible rank position can be viewed as a distinct treatment. In other words, a “first place” Competitive Rank mailer may enduce a different response than a “last place” Competitive Rank mailer. This feature of the experimental setup allows us to explore and test theories about social rank and its influence on behavioral response. We do this using two different econometric strategies.

4.2.1 Restrict Focus to Last/First Place Mailers Only

First, we treat each mailer and the household’s water use in the ensuing period as a distinct treatment/outcome pair. This requires a model of behavioral response whereby a household’s behavior in the period following a mailer is a direct response to the content of that mailer and is independent of the content of previous mailers. From the perspective of maintaining randomization, this is not an issue with the first mailer and subsequent behavior. However, since we use multiple mailers per household in the analysis, this does threaten our identification by moving away from pure randomization. This is because the content of previous mailers may have influenced household response to subsequent mailers.

There are precedents for this approach to assessing the impact of multiple treatments in existing research. For example, Doherty and Adler (2014) argue that mailer effects in a political campaign context are short-lived. The authors suggest that individual level responses can be considered in the period immediately following a given mailer, as timing may be more important to outcomes than mailer quantity. Additionally, Allcott and Rogers (2014) find evidence of cycles of significant backsliding in the weeks immediately following social information mailer receipt, using data from Opower’s Home Energy Reports. A similar sort of backsliding here could lead to near-total decay of mailer effects by the end of a single period post-mailer.

One could also justify this approach, as Bertrand et al. (2010) do, using the behavioral concepts of "System I" and "System II" thinking (Stanovich and West, 2000). That is, the receipt of a given mailer (and the social information contained in the mailer) can have two effects. First, in the short run it can cause an intuitive, System I response in the household, whereby the specific information in the mailer has an immediate effect on behavior. Second, in the long run they may invoke a System II impact, whereby the mailer causes deliberative changes in behavior (such as replacing household fixtures). It is plausible that in our data, the short-run, System I response would be more visible than the long-run, System II impacts of any previous mailers.

While we feel this analytical technique is justifiable in this case, we attempt to control for potential biases from repeat mailer exposure by using fixed effects for the number of mailers seen prior to the read in question. These controls do not significantly change our results. We feel this justifies our identification assumption that the impact of a given mailer on behavior is primarily restricted to the period immediately following that mailer's receipt, and is independent of prior messaging.

To assess the rank effects econometrically, we look at all mailer/outcome pairs for households that finished in "last place" in that specific mailer, and regress household water use on the treatments. We only do this for the In-Sample Control, Rank, and Competitive Rank treatments, as the Team treatment does not have any visible ranking. Note that households in the In-Sample Control group are assigned to water groups, but information about their group is never displayed to them. Consequently, "last place" homes in the two treatment groups can be compared to would-be "last place" homes in the In-Sample Control group. The In-Sample Control group is the omitted group in the regression, leaving regression coefficients that represent treatment effects of Rank and Competitive Rank mailers for "last place" homes. We then repeat this analysis for "first place" households. Since there are multiple observations for each household in the sample in this analysis, we cluster standard errors at the household level. The specification is shown below.

$$GPD_{ijk} = \beta_0 + \beta_1(T_{Rank})_i + \beta_2(T_{CompRank})_i + \beta_3(MailerGPD_{ijk}) + \varrho_i + \delta_j + \gamma_{ijk} + \rho_k + \varepsilon_{ijk}$$

Note that the specification includes controls for household water use displayed (in gallons per day) in the mailer ($MailerGPD_{ijk}$), household demographics (lot size, home size, and bathrooms, captured by ϱ_i), month fixed effects (δ_j), WaterScore fixed effects (γ_{ijk}), and fixed effects for the number of mailers seen prior to the observation mailer (ρ_k).

4.2.2 Rank Effects Amongst Middle Third of Water Users

A second approach to evaluating rank effects is to restrict attention to homes in the middle third of water users pre-experiment, whose water use was around the mean given their irrigable area. Due to the random assignment of water groups, some of these households had a water group composed of relatively low or high water users. Therefore, there is variation in rank position amongst these homes that is not a function of their actual water use behavior. We exploit this variation and run the following specification, which uses data from all mailers received by individuals in the middle third of water use and includes interaction effects for rank position (1st, 2nd, 3rd, 4th, or 5th) and treatment (Control, Rank, Competitive Rank):

$$GPD_{ijk} = \beta_0 + [\textstyle\sum_{m=1}^{3}(\sum_{n=1}^{5} \beta_{m,n}(Position_n)_{ijk} * (T_m)_i)] + \beta_{15}(MailerGPD_{ijk}) + \varrho_i + \delta_j + \gamma_{ijk} + \rho_k + \varepsilon$$

The 14 interaction terms will reveal whether or not there is a differential impact of rank position based on mailer version, which will help us identify any evidence for a "last place effect" or "first place effect" in the Rank and Competitive Rank treatments (note that all In-Sample Control households are unaware of their social rank position and therefore serve as an effective control group). Again, this specification controls for household water use displayed in the mailer ($MailerGPD_{ijk}$), lot size, home size, and bathrooms (ϱ_i), month fixed effects (δ_j), WaterScore fixed effects (γ_{ijk}), and fixed effects for the number of mailers seen prior to the observation mailer (ρ_k). Standard errors are again clustered at the household level.

5 Results

5.1 Overall Mailer Effects: Difference-in-Differences and Matching

We begin by using the Out-of-Sample Control group to assess the impact of the four experimental mailers (the In-Sample Control, Rank, Team, or Competitive Rank treatments) on water use. The short-run results, which focus on the read following the first mailer (December 2012) compared to the same month in the previous year, suggest that the experimental mailers did not have a significant effect on water use in the short run. These results are visible in Table 3, in both linear and log forms, and in the tables below. We also compute matching estimators as described in section 4.1.2. The matching results suggest a greater short-run decrease in water use than the difference-in-differences estimators (particularly when matching was done using only one match). However, when we use four matches instead of one, the effect sizes are smaller. Taken together, we conclude from these results that the mailers had minimal impact in the short run.

For convenience, we present the short-run ATE estimates from the difference-in-differences and matching analyses, in both level and log form, in Figure 2 below.

Figure 2: Short-Run ATE Estimates (for the first period post-mailer initiation, December 2012)

	Difference-in-Differences	Matching (1 match)	Matching (4 matches, bias adjusted)
In-Sample Control Mailer	1.65 GPD (0.81%)	-21.52 GPD *** (-6.09%) ***	-8.49 GPD * (1.41%)
"Rank" Mailer	4.24 GPD (1.36%)	-19.84 GPD *** (-4.80%) **	-7.41 GPD * (1.55%)
"Team" Mailer	4.26 GPD (0.24%)	-21.22 GPD *** (-6.38%) ***	-6.92 GPD (1.60%)
"Competitive Rank" Mailer	2.11 GPD (0.41%)	-22.62 GPD *** (-6.46%) ***	-15.45 GPD *** (-0.55%)

The results for the entire experimental period, however, suggest that the different mailers reduced water use by 13-17 gallons per day, or around 3%. The reduction in water use in the linear regression was statistically significant at the 95% level for the In-Sample Control (16.17 GPD) and Rank (16.28 GPD) mailers, and at the 90% level for the Team mailer (14.62 GPD) and the Competitive Rank mailer (13.44 GPD). However, none of the log specifications are significant (though the estimates are all in the 3-5% range). Table 4 shows these results. To verify the difference-in-differences estimates, we compute average treatment effects using matching as described in section 4.1.2. We present the overall ATE estimates from the difference-in-differences and matching analyses, in both level and log forms, in Figure 3 below.

Figure 3: Overall ATE Estimates (Estimates for all post-mailer initiation reads, December 2012-June 2013)

	Difference-in-Differences	Matching (1 match)	Matching (4 matches, bias adjusted)
In-Sample Control Mailer	-16.17 GPD ** (-3.36%)	-27.87 GPD *** (-7.84%) ***	-12.09 GPD *** (0.58%)
"Rank" Mailer	-16.28 GPD ** (-3.35%)	-31.48 GPD *** (-8.08%) ***	-17.87 GPD *** (-0.66%)
"Team" Mailer	-14.62 GPD * (-4.59%)	-23.61 GPD *** (-6.00%) ***	-9.83 GPD *** (1.61%)
"Competitive Rank" Mailer	-13.44 GPD * (-3.23%)	-24.14 GPD *** (-6.70%) ***	-9.56 GPD *** (1.03%)

These results suggest that the initial mailer with social information failed to spur significant behavior change in the short run, but over time the mailers did influence behavior. It is possible that the effects observed here are artifacts of a breakdown in the parallel trends assumption underlying the analysis. However, given the past data on the trends in water use in the In-Sample and Out-of-Sample areas and our approach of aggregating means in the pre- and post-periods, we are relatively confident that the observed patterns capture genuine conservation efforts on the aggregate in response to the initial mailers.

The most interesting result here is that the Competitive Rank mailer was the least effective across specifications. This hints at possible underlying differences in how individuals responded to the mailers, particularly the Competitive Rank mailer. We explore explanations for these differential impacts in the next section.

5.2 Across Mailer Differences: Means Comparisons with Regression

Having found evidence that the mailers do influence behavior, we now set aside the Out-of-Sample Control, and look only at the households that did receive mailers, to explore differences in behavioral response across mailer versions.

5.2.1 Aggregate Means Comparison Regressions

Tables 5 and 6, along with the visuals in Appendix 6, present the results of the regressions of water usage on treatments for all in-sample groups. The goal of this analysis is to estimate the average treatment effect of each version of the mailer relative to the control mailer, which featured no social information except for the "WaterScore" on the left hand side of the mailer. Table 5 presents regressions using water use in the first period following experiment initiation as the outcome variable, and Table 6 presents the regressions using mean water use in all periods following experiment initiation as the outcome variable. Appendix 6 provides charts displaying the average treatment effect estimates from these analyses.

The results show no strong evidence of a differential impact across mailers in the short run or overall. The only statistically significant result (at the 90% level) is that the Competitive Rank mailer performed worst, increasing water use by 8.19 GPD relative to the control mailer overall (Table 6, model (2)). The similarity in effect across mailer versions is not surprising, since this analysis treats receipt of any version of a given mailer as part of the same treatment, whether or not you performed well or poorly in the displayed peer comparison. In other words, a household receiving a Competitive Rank mailer and finding themselves in "first place" in the ranking is, in this analysis, grouped with a household receiving a Competitive Rank mailer and finding themselves in "last place." It is likely that these two types of households will have very different responses to the Competitive Rank mailer. More analysis is therefore needed (and follows in section 5.3).

However, this is still a significant result from a policy perspective. No mailer version is inherently better than the other, in terms of its average effects. This suggests that a policymaker seeking to make a blanket decision on which form of messaging to use in a water mailer across a large population cannot expect one type of messaging to work better. Instead, a targeted approach that considers disaggregated treatment effects may prove most effective in reducing overall water consumption.

5.2.2 Disaggregation by Past Water Use

We next repeat the analysis, but disaggregate based on a key, visible covariate - past water use. By doing so we can determine if certain mailer versions were more effective for households with high or low water use habits.[34] Tables 7-14 show the output from these regressions, while Appendices 7A, 7B, 8A, and 8B provide visuals, as noted below.

The results suggest that the Rank treatment increases water use amongst households with low water use prior to the experiment, relative to the control mailer. The effect is mostly from the impact of the mailers in the first post-mailer period, where the Rank treatment increased water use by 12.02 GPD relative to the control mailer, statistically significant at the 95% level (Table 7, model 3, and Appendix 7A, chart (i)). It is instructive to compare the Rank and Competitive Rank mailers directly as well, since the only difference between these mailer versions was the framing around social rank information. When compared directly with

[34] "High"/"low" users were defined as being in the top/bottom third of water users within their irrigable area category in the pre-experiment period in 2012. See section 4.1.3 for more.

the Rank mailer, the Competitive Rank mailer is associated with 14.22 GPD lower household water use in the first period - 10.3% less than the Rank mailer (Table 8, models 2 and 4, and Appendix 7A, chart (ii)). When looking at the mean water use during all periods following the initiation of the experiment, however, the detrimental effect of the Rank mailer relative to the control mailer is smaller and not statistically significant (3.85 gallons per day more than the control, around 3.0% higher water use when a log-level regression is used, visible in Table 9, models 2 and 4, and Appendix 7B, chart (i)). However, the difference between the Rank and Competitive Rank mailers remains significant at the 90% level, with the Competitive Rank mailer associated with 7.15 GPD lower household water use than the Rank mailer over the entire experimental period (visible in Table 10, model 2, and Appendix 7B, chart (ii)).

This is a notable result - there is evidence of a "boomerang effect" for low-water-use households from rank information, but one that is counteracted by a competitive frame. Note that this effect comes solely from the peer comparison, and not the other social information on the mailer, which is controlled for in the regression. One possible explanation for this finding is that the Rank treatment's neutral messaging does not provide sufficient incentive for efficient households to continue conservation efforts. The Competitive Rank treatment mailer provided peer comparison and social rank as well, but did so with an added competitive motivation, which arguably prevents the boomerang effect observed for households receiving the Rank mailer.

Meanwhile for households with high levels of water use pre-treatment, the treatment effects are different, and less statistically compelling. As Tables 11 and 12 demonstrate, the mailers were similarly effective in the short run. While the Team treatment performed slightly worse than the other two treatments, this difference was not statistically significant (Table 11, models 3 and 6, and Appendix 8A, chart (i)). When we look at the mean water use in all periods following the first mailer, in Tables 13 and 14, we see that the Competitive Rank mailer performs worse than the other mailers, increasing mean water use by 13.28 GPD relative to control mailer (Table 13, model 2, and Appendix 8B, chart (i)) and by 15.89 GPD relative to the Rank mailer (Table 14, model 2, and Appendix 8B, chart (ii)). The second result is statistically significant at the 10% level, and equates to roughly 4.0% higher water use than the Rank mailer when using the log-level specification (Table 14, model 4).

This is suggestive of a competitive framing effect for high water use households that is the exact opposite of that for low water use households - while competitive framing of rank information had a positive effect on low water use households (preventing a boomerang effect), it seemed to increase water use in high water use households. This could be because it is demotivating to perform poorly in a competitive comparison with your peers. Note that the higher water use relative to the control group was not observed in the Rank treatment (Table 13, model 2) - the competitive framing seems to be the key element.

5.3 Ranking Effects

In assessing the effect of specific rankings, we focus on first and last place in particular. We begin by restricting analysis to the following mailers and subsequent outcomes: 1) households receiving first/last place rank messaging in the Rank and Competitive Rank treatments; and 2) households receiving the Control mailer who "would have" ranked in first/last had they been shown a displayed rank. We use regressions to estimate the effect of displayed "first" and "last" place messaging on behavior following mailer receipt using data from all experimental mailers, and we cluster standard errors at the household level. The full results on ranking effects are in Tables 15 and 16.

The main result in this analysis is the existence of a persistent and strong "last place effect" for households in the Competitive Rank treatment (Table 15). Specifically, households ranked last in the Competitive Rank treatment show higher post-mailer water use than households in the In-Sample Control who would have been in last place had they seen their position (17.85 GPD more water use than the In-Sample Control mailer, visible in Table 15, model 4). This effect is also significant in comparison to the last-placed individuals in the Rank treatment.[35]

The results suggest that priming a sense of competition makes social rank information demotivating for people who perform worst in the displayed rank. This is especially interesting because the Competitive Rank treatment does not seek to prime negative thoughts about poor performance in the household - it actually encourages a last place household to improve in an effort to attain 4th place.

The evidence for a comparable "first place effect" is not as compelling. As model 4 in Table 16 shows, the visible "first place effect" in a simple specification without controls disappears with the inclusion of controls.

For robustness, we use the second approach outlined in 4.2.2, restricting analysis to only those homes in the middle third of water users. We estimate the effect of rank here by interacting treatment and rank to determine if there was a differential response to rank position by treatment. Table 17 presents the results of this regression, and Appendix 9 provides a visual depiction of the coefficients on the interaction terms by treatment and position (because it was necessary to omit a coefficient, Control households in 3rd position were omitted to generate the images). While not statistically significant, the clear message from this analysis is that the Rank and Competitive Rank treatments seem to consistently drive up water use for households in last place.

When these results are coupled with the earlier results showing that the Competitive Rank mailer performed worst on aggregate, a clear story emerges. The competitive framing discourages high water users, particularly

[35] The Rank treatment seems to cause a smaller "last place effect" of its own relative to the control when a log-level specification is used, however.

those individuals who find themselves in "last place" in a displayed rank. These individuals perform worse than they would have had they not seen the rank information. Simultaneously, the competitive frame has a smaller positive impact on low water users. However, the detrimental effects of the competitive frame (on the high water users) outweigh any positive impacts (on the low water users), meaning that on aggregate the Competitive Rank mailer performs worst of all mailer versions used.

6 Discussion and Conclusions

Our experiment provides insights on some important underlying drivers of behavioral response to social information and peer comparison. First, the experiment replicates existing work on social information by showing that mailers using this information can reduce water use (by roughly 13-17 gallons per day in this study). Overall, the experiment finds that the different frames used in the mailers (neutral, competitive, and cooperative) had minimally different effects on water use, with the competitively-framed mailer performing marginally worse. However, this aggregate comparison of mailers masks more interesting results on the underlying mechanisms of rank and response.

The most robust results come from the disaggregated analysis of treatment effects, and the analysis of specific rank effects. The analysis shows that the display of a neutrally-framed peer comparison with four similar homes caused a "boomerang effect" in water-efficient households, increasing the households' water use relative to the control. Interestingly, this boomerang effect was eliminated by the inclusion of a competitive frame. This result was bolstered by weak evidence that receiving a "first place" ranking in a competitively-framed mailer led to greater water conservation for a household. Though this "first place" result was not statistically significant, the results together are supportive of a conclusion that high acheivers thrive (or, at least, do not struggle) in competitive settings, and may need competition to avoid boomerang effects from explicit rank information.

However, the competitive framing of rank information had large demotivational effects on water-inefficient households. These households responded poorly to the competitively-framed rank information, more than offsetting the beneficial effects of the competitive framing on high achievers. Furthermore, it appears that rank effects play a significant role here as well. The results show that households who finish in "last place" in a competitively-framed rank comparison were demotivated, and increased their water use relative to both the control and the neutrally-framed rank information. Interestingly, this implies that this demotivation effect is primarily driven by the competitive frame (rather than by the low ranking).

These results have direct implications for the competing theories related to rank and response. This experiment finds that in competitive settings, the theories on motivation and self-efficacy seem more consistent with observed behavior, with top performers holding steady while poor performers worsen. This supports the existence of a "what-the-hell effect" (Polivy and Herman, 1985), and also is consistent with the oppositional reactions to peer information found by Beshears et al. (2014) in the savings context. Furthermore, the finding that rank information offsets the boomerang effect for top performers is consistent with Garcia et al. (2006). However, without the competitive frame, simple rank information seems to encourage a behavioral response that is more in line with social norms and social comparison theories, with "boomerang effects" for top performers the clearest result. This provides some structure to existing theories on rank and response, and suggests that the framing of peer comparisons and social rank is key to their success or failure.

The implication of these findings for public policymakers and "nudgers" seeking to promote conservation behavior is mixed. On the one hand, there seem to be benefits from social information overall. The mailers did influence behavior on the aggregate. However, the experiment also reveals some potential pitfalls of social information, namely that it can demotivate poor performers in a way that has detrimental effects on their water use. This leads to important follow up questions. What types of social information are best to motivate those who are performing poorly? Why does a competitive frame prevent backsliding for top performers and to what extent is this context-dependent? Further research is needed to better understand the observed effects and what forms of social messaging are needed to negate those effects.

On the whole, experiments of this form offer a compelling way to develop public policy for water conservation. Exploring messaging and developing innovative ways to convey information to individuals and households can increase the salience of water use, cost, and environmental impact, and lead to changes in aggregate water use outcomes. Follow up research could extend this work in a number of ways. First, the "last place effect" outlined here could be tested in a randomized setting with a larger sample size. Second, studies could explore increasing the salience of water cost and how to frame messaging around water bills to increase its significance to households. Third, future research needs to explore how low-cost messaging can be used to promote major household behavior change. In some ways, water conservation suffers from an "energy paradox" as outlined in Jaffe and Stavins (1994).[36] That is, undertaking repairs that reduce water use may have significant long-run financial benefits for households, but the upfront cost and mental effort may prevent households from pursuing them. While behavioral nudges using social information might provide cost-effective and environmentally-significant savings, further work should explore the use of behavioral

[36] Note: Allcott and Taubinsky extended this "energy paradox" into electricity conservation as the "lightbulb paradox" in Allcott and Taubinsky (2014).

science to influence household decision making on the big changes, like the replacement of water-hungry appliances, which could reduce water use significantly in the long run.

References

Abadie, Alberto, David Drukker, Jane Leber Herr, and Guido W. Imbens (2004) "Implementing matching estimators for average treatment effects in Stata," *The Stata Journal*, Vol. 4, pp. 240–311.

Abadie, Alberto and Guido W. Imbens (2002) "Simple and bias-corrected matching estimators,"Technical report, Department of Economics, University of California, Berkeley.

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

Allcott, Hunt and Sendhil Muillainathan (2010) "Behavioral Science and Energy Policy," *Science*.

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

Allcott, Hunt and Dmitry Taubinsky (2014) "The Lightbulb Paradox: Evidence from Two Randomized Experiments," *Working Paper*.

Antin, Judd and Elizabeth F. Churchill (2011) "Badges in Social Media: A Social Psychological Perspective," *CHI 2011 Gamification Workshop Proceedings*.

Bandura, A. (1977) "Self-efficacy: Toward a unifying theory of behavioral change," *Psychological Review*, pp. 191–215.

——— (1994) "Self-efficacy," in *Encyclopedia of human behavior*, New York: Academic Press.

Beenen, Gerard, Kimberly Ling, Xiaoqing Wang, Klarissa Chang, Dan Frankowski, Paul Resnick, and Robert E. Kraut (2004) "Using social psychology to motivate contributions to online communities," *CSCW Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pp. 212–221.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004) "How Much Should We Trust Differences-in-Differrences Estimates?" *The Quarterly Journal of Economics*, Vol. 119, pp. 249–275.

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman (2010) "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment," *The Quarterly Journal of Economics*, Vol. 129, pp. 263–306.

Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, and Katherine L. Milkman (2014) "The Effect of Providing Peer Information on Retirement Savings Decisions," *Journal of Finance, Forthcoming*.

Cialdini, Robert B., Linda J. Demaine, Brad J. Sagarin, Daniel W. Barrett, Kelton Rhoads, and Patricia L. Winter (2006) "Managing social norms for persuasive impact," *Social influence*, Vol. 1, pp. 3–15.

Clapp, J.D., J.E Lange, C. Russell, A Shillington, and R. Voas (2003) "A failed norms social marketing campaign," *Journal of Studies on Alcohol*, pp. 409–414.

Clee, Mona A. and Robert A. Wicklund (1980) "Consumer behavior and psychological reactance," *Journal of Consumer Research*, pp. 389–405.

Community, Intelligence (2012) "Global Water Security," February, URL: http://www.dni.gov/files/documents/Newsroom/Press%20Releases/ICA_Global%20Water%20Security.pdf.

Davis, Matt (2011) "Behavior and Energy Savings Evidence from a Series of Experimental Interventions," *Environmental Defense Fund*.

Dickerson, C. A., R. Thibodeau, E. Aronson, and D. Miller (1992) "Using Cognitive Dissonance to Encourage Water Conservation," *Journal of Applied Social Psychology*, Vol. 22, pp. 841–854.

Doherty, David and E. Scott Adler (2014) "The Persuasive Effects of Partisan Campaign Mailers," *The Political Research Quarterly*, Vol. 67, pp. 562–573.

Espey, M., J. Espey, and W.D. Shaw (1997) "Price Elasticity of Residential Demand for Water: A Meta-Analysis," *Water Resources Research*, Vol. 33, pp. 1369–1374.

Fehr, Ernst and Herbert Gintis (2007) "Human Motivation and Social Cooperation: Experimental and Analytical Foundations," *Annual Review of Sociology*, Vol. 33, pp. 43–64.

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

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

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

Festinger, L. (1954) "A Theory of Social Comparison Processes," *Human Relations*, Vol. 7, pp. 117–140.

Fischer, Corinna (2008) "Feedback on household electricity consumption: a tool for saving energy?" *Energy efficiency*, Vol. 1, pp. 79–104.

Garcia, Stephen, Avishalom Tor, and Richard Gonzalez (2006) "Ranks and Rivals: a Theory of Competition," *Personality and Social Psychology Bulletin*, Vol. 32, pp. 970–82.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer (2008) *Social Pressure and Voter Turnout: Evidence from a Large-scale Field Experiment*: American Political Science Review.

Gillingham, Kenneth, Richard G. Newell, and Karen Palmer (2009) "Energy Efficiency Economics and Policy," *Resources for the*.

Granfield, R. (2005) "Alcohol use in college: Limitations on the transformation of social norms," *Addiction Research & Theory*, Vol. 13, pp. 281–292.

Grisevicius, Vladas, Robert B. Cialdini, and Noah J. Goldstein (2008) "Social Norms: An Underestimated and Underemployed Lever for Managing Climate Change," *RESEARCH*.

Hagger, M. S., N. L. Chatzisarantis, and S. J. Biddle (2002) "A meta-analytic review of the theories of reasoned action and planned behavior in physical activity: Predictive validity and the contribution of additional variables.," *Journal of sport and exercise psychology*.

Harding, Matthew and Alice Hsiaw (2014) "Goal Setting and Energy Conservation," *Journal of Economic Behavior and Organization*, Vol. 107, pp. 209–227.

Jaffe, Adam and Robert Stavins (1994) "The Energy Paradox and the Diffusion of Conservation Technology," *Resource and Energy Economics*, Vol. 16, pp. 91–122.

Jessoe, Katrina and David Rapson (2014) "Knowledge Is (Less) Power: Experimental Evidence from Residential Energy Use," *American Economic Review*, pp. 1417–438.

John, Leslie K. and Michael I. Norton (2013) "Converging to the Lowest Common Denominator in Physical Health," *Health Psychology*, Vol. 32, pp. 1023–1028.

Julian, James W. and Franklyn A. Perry (1967) "Cooperation Contrasted with Intra-Group and Inter-Group Competition," *Sociometry*, pp. 79–90.

Kahneman, Daniel (2011) *Thinking Fast and Slow*, Chap. Regression to the Mean: Farrar, Straus and Giroux.

Kahneman, Daniel and Amos Tversky (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, pp. 263–291.

Karlan, Dean and Jonathan Zinman (2009) "Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment," *Econometrica*, Vol. 77, pp. 1993–2008.

Kast, Felipe, Stephan Meier, and Dina Pomeranz (2012) "Under-Savers Anonymous: Evidence on Self-Help Groups and Peer Pressure as a Savings Commitment Device," *NBER Working Paper*.

Kempton, W, D Feuermann, and A McGarity (1992) "I Always Turn It on "Super": User Decisions About When and How to Operate Room Air Conditioners," *Energy and Buildings*, pp. 177–191.

Kempton, W and L Montgomery (1982) "Folk Quantification of Energy," *Energy*, pp. 817–27.

Kohn, Alfie (1996) *No Contest: The Case Against Competition*, Chap. The "Number One" Obsession: Houghton Mifflin.

Kuziemko, Ilyana, Ryan W. Buell, Taly Reich, and Michael I. Norton (2014) ""Last-place Aversion": Evidence and Redistributive Implications," *The Quarterly journal of economics*, Vol. 129, pp. 105–149.

Lee, Juneseok and Stephanie A. Tanverakul (2015) "Price elasticity of residential water demand in California," *Journal of Water Supply*, Vol. 64, pp. 211–218.

Little, Donna E. (2012) "A Crucible of Competition and Cooperation: Where Do the Concepts Fit in Recreation Activity Delivery?" *University of Waikato Research Commons*, pp. 141–154.

Lynne, Gary D., C. Franklin Casey, Alan Hodges, and Mohammed Rahmani (1995) "Conservation Technology Adoption Decisions and the Theory of Planned Behavior," *Journal of Economic Psychology*.

Mayer, Peter W. and William B. DeOreo (1999) "Mean Per Capita Daily Water Use," in *Residential End Uses of Water*: AWWA Research Foundation and American Water Works Association, p. 87.

Mutz, Diana and Robin Pemantle (2011) "The Perils of Randomization Checks in the Analysis of Experiments," *Paper presented at the Annual Meetings of the Society for Political Methodology*.

Nagourney, Adam and Ian Lovett (2014) "Severe Drought Has U.S. West Fearing Worst," *The New York Times*, URL: http://www.nytimes.com/2014/02/02/us/severe-drought-has-us-west-fearing-worst.html?_r=0.

Obama, President Barack (2014) "Remarks by the President on the California Drought," February, URL: http://www.whitehouse.gov/the-press-office/2014/02/14/remarks-president-california-drought.

Olmstead, Sheila M. and Robert N. Stavins (2007) "Managing Water Demand Price vs. Non-Price Conservation Programs," *A Pioneer Institute for Public Policy Research*, URL: http://www.hks.harvard.edu/fs/rstavins/Monographs_&_Reports/Pioneer_Olmstead_Stavins_Water.pdf.

Olson, Mancur (1965) *The Logic of Collective Action: Public Goods and the Theory of Groups*, Cambridge, MA.

Pajares, Frank (1997) *Advances in motivation and achievement*, Chap. Current Directions in Self-efficacy Research, Greenwich, CT: JAI Press.

Petersen, John E., Vladislav Shunturov, Kathryn Janda, Gavin Platt, and Kate Weinberger (2007) "Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives," *International Journal of Sustainability in Higher Education*, Vol. 8, pp. 16–33.

Polivy, J. and C.P. Herman (1985) "Dieting and binging: A causal analysis," *American Psychologist*, pp. 193–201.

Qin, Zhining, David W. Johnson, and Roger T. Johnson (1995) "Cooperative Versus Competitive Efforts and Problem Solving," *Review of educational Research*, Vol. 65, pp. 129–143.

Schultz, Wesley P., Jessica M. Nolan, Noah J. Goldstein Robert B. Cialdini and, and Vladas Griskevicius (2007) "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science*.

Shelton, T. O. and M.J. Mahoney (1978) "The content and effect of psyching-up strategies in weight lifters," *Cognitive Therapy and Research*, Vol. 2, pp. 275–284.

Stanovich, Keith and Richard West (2000) "Individual Differences in Reasoning: Implications for the Rationality Debate," *Behavioral and Brain Sciences*, pp. 645–665.

Suls, J., R. Martin, and L. Wheeler (2002) "Social comparison: Why, with Whom, and With What Effect?" *Current Directions in Psychological Science*, Vol. 11, pp. 159–163.

Tauber, Edwards M. (1972) *Why Do People Shop?*, Chap. Why Do People Cooperate?, pp. 11–26: Journal of Marketing.

Thomas, Ted and Doug Carlson (2014) "Despite Storms, Snowpack Still Far Below Normal: Drought Retains Grip as Summer Approaches,"Technical report, California Department of Water Resources, News For Immediate Release.

Tran, Ahn and Richard Zeckhauser (2009) "Rank as Incentive," *HKS Faculty Research Working Paper Series*.

Tschannen-Moran, Megan and Anita Woolfolk Hoy (2001) "Teacher efficacy: capturing an elusive construct," *Teaching and Teacher Education*, Vol. 17, pp. 783–805.

United States Environmental Protection Agency, Office Of Ground Water and Drinking Water (2009) "Water On Tap: What You Need To Know," December.

WaterSense (2015) "Fix a Leak Week,"Technical report, Environmental Protection Agency, Accessed: April 2015.

Watson, Raymond K., Margot H. Murphy, Frank E. Kilfoyle, and Susan M. Moore (1999) "An Opportunistic Experiment in Community Water Conservation," *Population and Environment*.

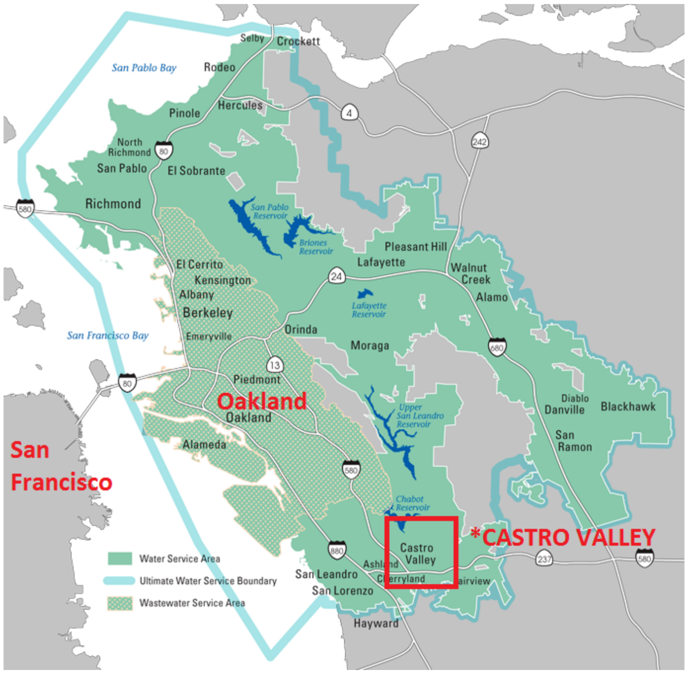
Wechsler, Henry, Toben F. Nelson, Jae Eun Lee, Mark Seibring, Catherine Lewis, and Richard Keeling (2003) "Perception and Reality: A National Evaluation of Social Norms Marketing Interventions to Reduce College Students' Heavy Alcohol Use," *Journal of Studies on Alcohol*.

Winter, P.L. (2008) "Science Notes: Park signs and visitor behavior: A research summary," *Park Science*, pp. 34–35.

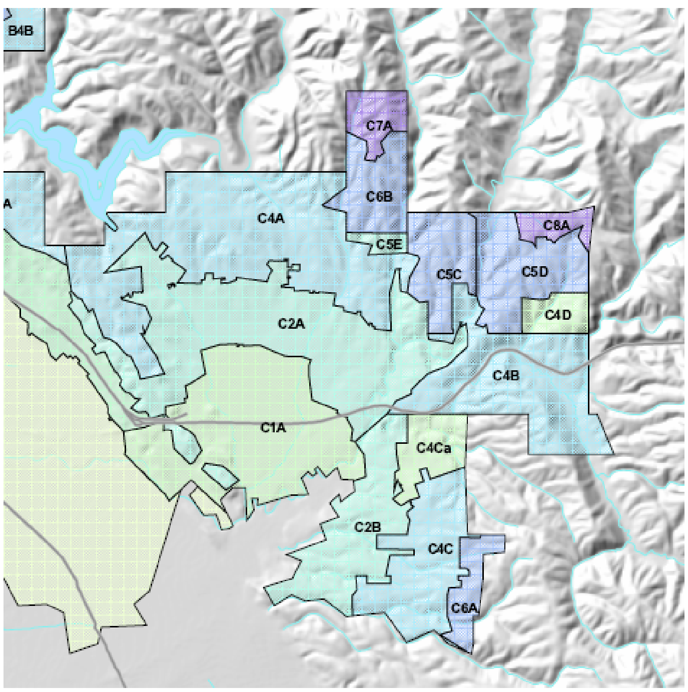
Yim, Jeffrey and T.C. Graham (2007) "Using games to increase exercise motivation," *Proceedings of the 2007 conference on Future Play*, pp. 166–173.

Appendix 1

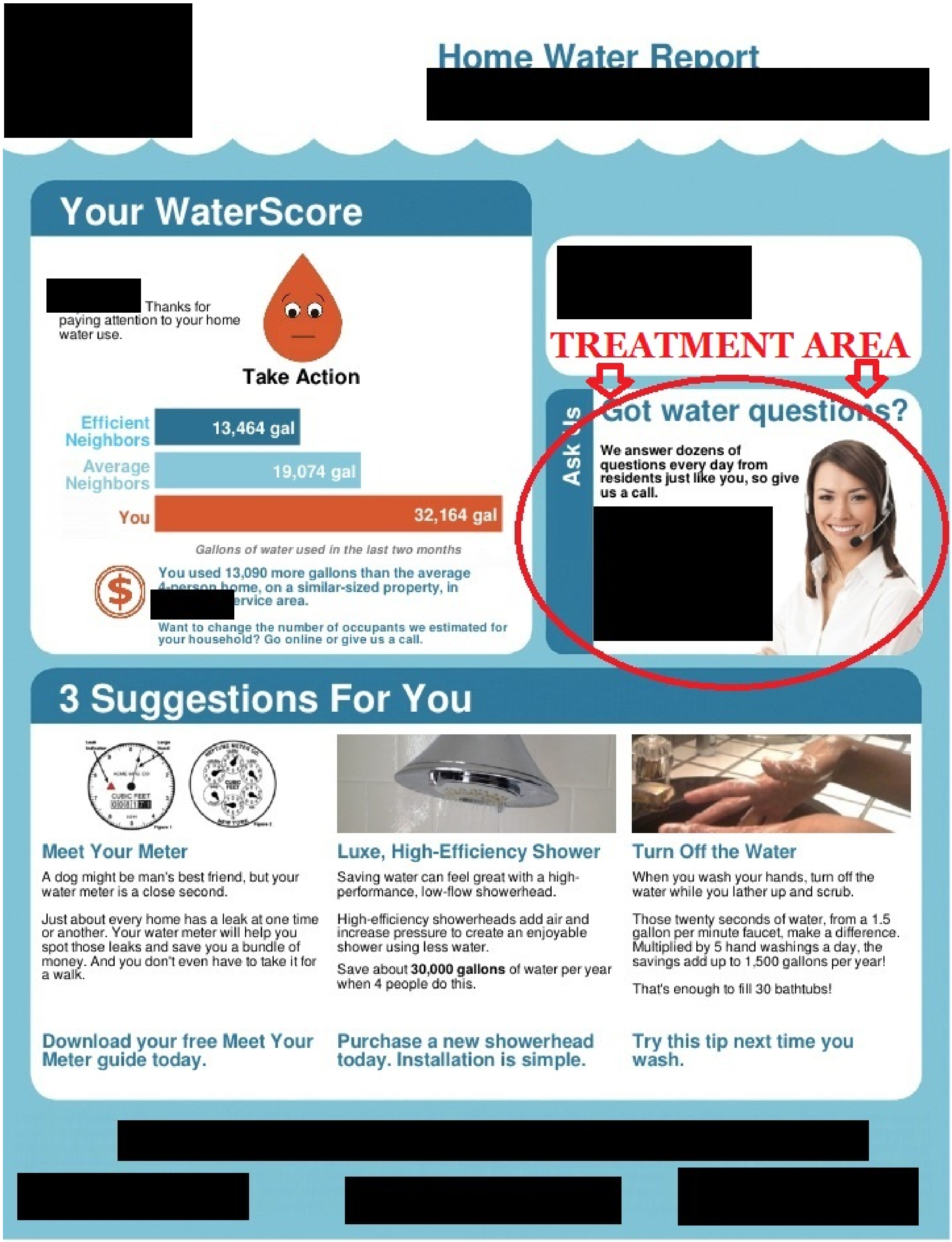
i. Experiment Location



ii. Pressure Zones



Appendix 2


[redacted]
Home Water Report
[redacted]
Your WaterScore
[redacted] Thanks for paying attention to your home water use.
Take Action
Efficient Neighbors 13,464 gal
Average Neighbors 19,074 gal
You 32,164 gal
Gallons of water used in the last two months
You used 13,090 more gallons than the average [redacted] home, on a similar-sized property, in [redacted]ervice area.
Want to change the number of occupants we estimated for your household? Go online or give us a call.
[redacted]
TREATMENT AREA
Ask Us
Got water questions?
We answer dozens of questions every day from residents just like you, so give us a call.
[redacted]
3 Suggestions For You
Meet Your Meter
A dog might be man's best friend, but your water meter is a close second.
Just about every home has a leak at one time or another. Your water meter will help you spot those leaks and save you a bundle of money. And you don't even have to take it for a walk.
Download your free Meet Your Meter guide today.
Luxe, High-Efficiency Shower
Saving water can feel great with a high-performance, low-flow showerhead.
High-efficiency showerheads add air and increase pressure to create an enjoyable shower using less water.
Save about 30,000 gallons of water per year when 4 people do this.
Purchase a new showerhead today. Installation is simple.
Turn Off the Water
When you wash your hands, turn off the water while you lather up and scrub.
Those twenty seconds of water, from a 1.5 gallon per minute faucet, make a difference. Multiplied by 5 hand washings a day, the savings add up to 1,500 gallons per year!
That's enough to fill 30 bathtubs!
Try this tip next time you wash.
[redacted]

Appendix 3

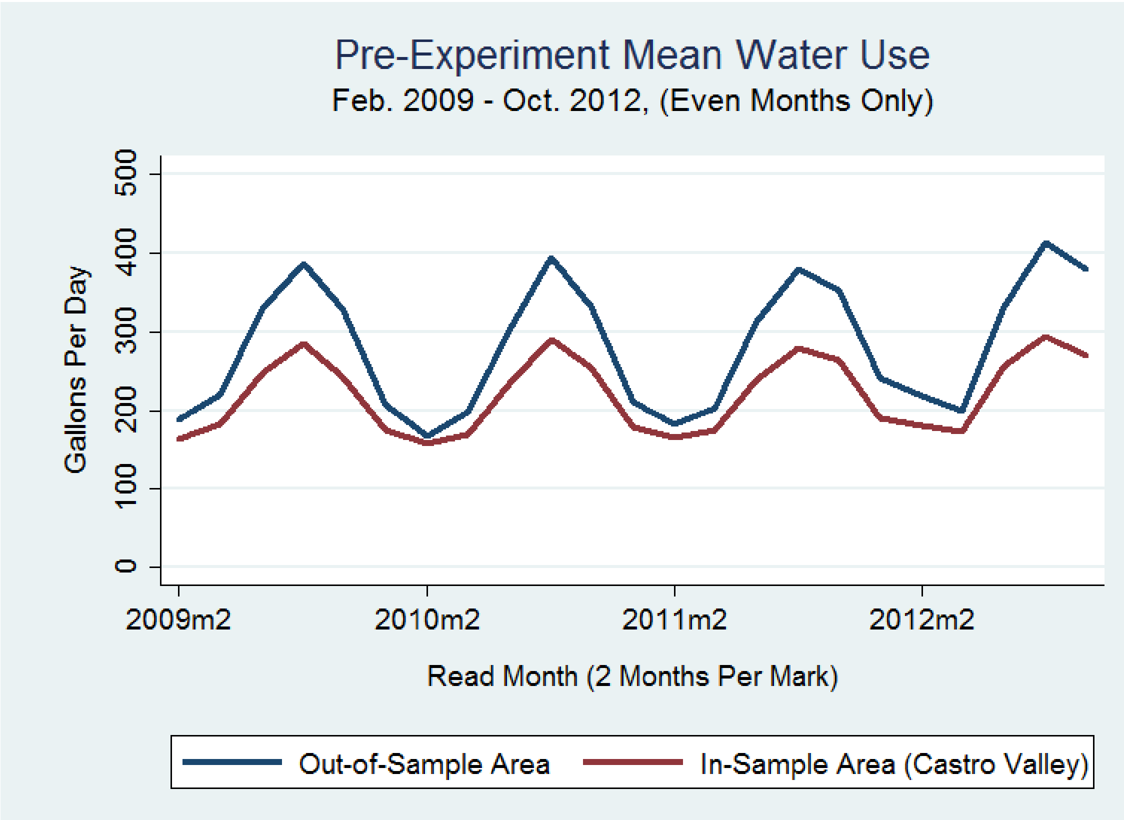
Chart 1:

	1 Occupant	2 Occupants	3 Occupants	4 Occupants	5+ Occupants
Small	350	867	1,381	577	255
Medium	131	264	605	324	131
Large	21	31	81	36	16
Extra Large	12	22	37	17	13

Chart 2:

	Number of Households (total)	Number of Households receiving all four treatment mailers
Control Group	1,308	1,091
Treatment #1: Rank	1,288	1,050
Treatment #2: Team	1,284	1,056
Treatment #3: Competitive Rank	1,300	1,068

Appendix 4



Pre-Experiment Mean Water Use
Feb. 2009 - Oct. 2012, (Even Months Only)
Gallons Per Day
0
100
200
300
400
500
2009m2
2010m2
2011m2
2012m2
Read Month (2 Months Per Mark)
Out-of-Sample Area
In-Sample Area (Castro Valley)

Appendix 5

Chart (A)

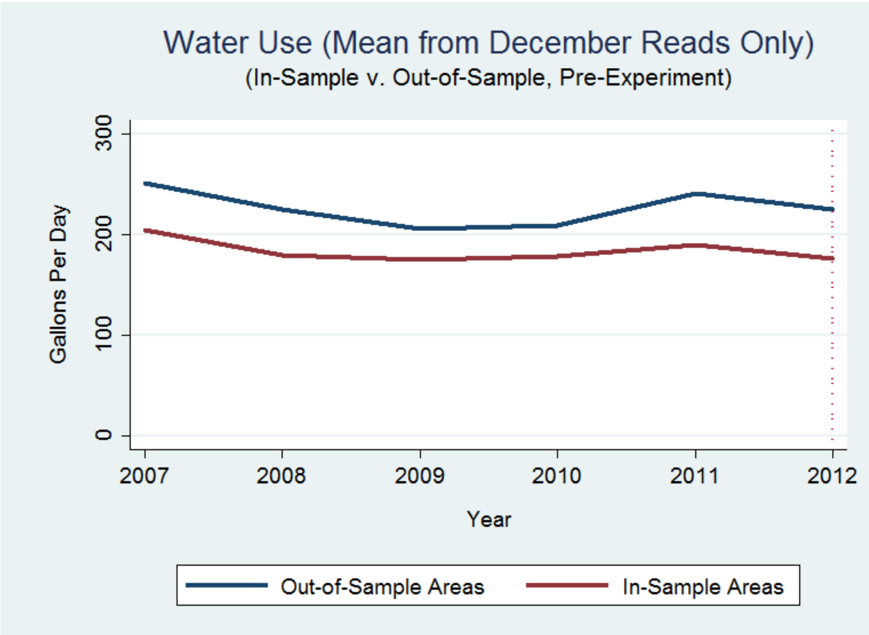
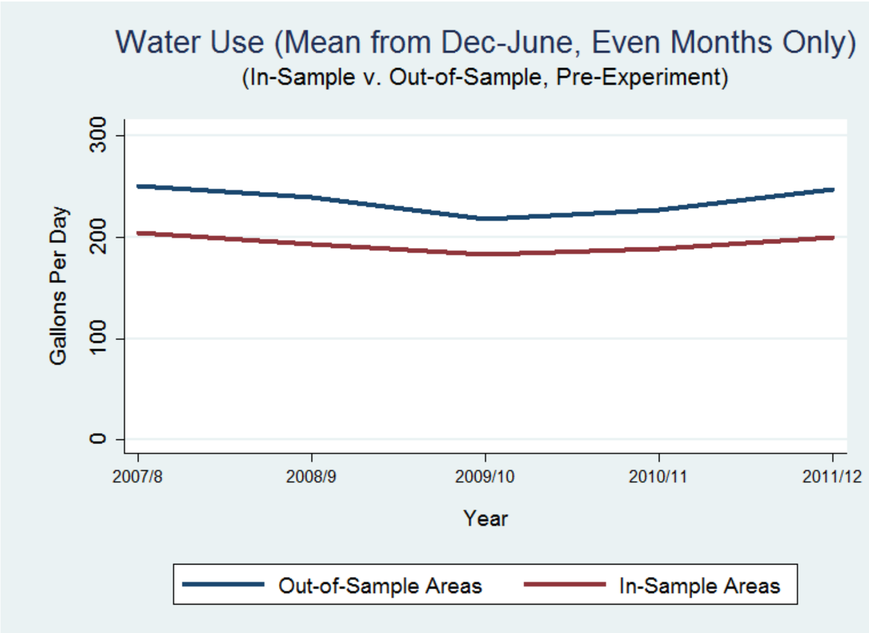
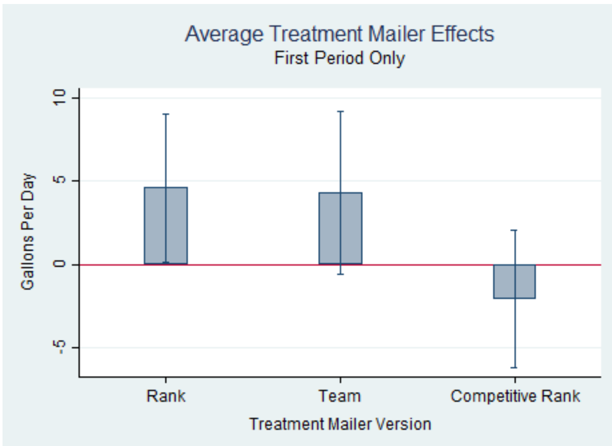


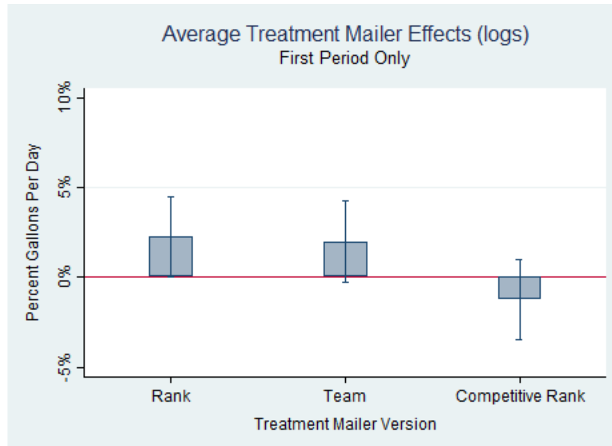
Chart (B)



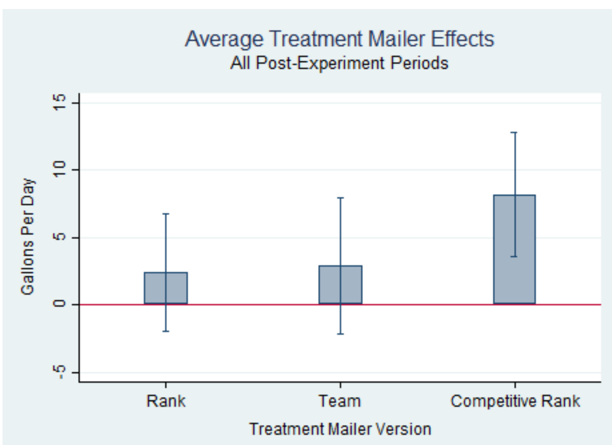
Appendix 6: Average Treatment Effects (Treatment Mailers Relative to the Control Mailer)

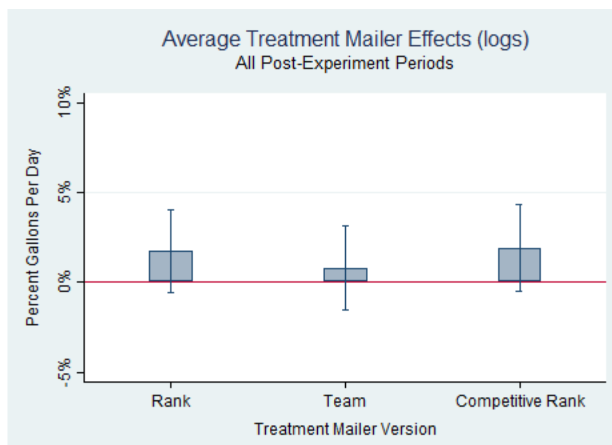
A) First Post-Mailer-Initiation Period Only (Dec. 2012) - Standard Error Marked





B) Mean of all Post-Experiment Periods - Standard Error Marked

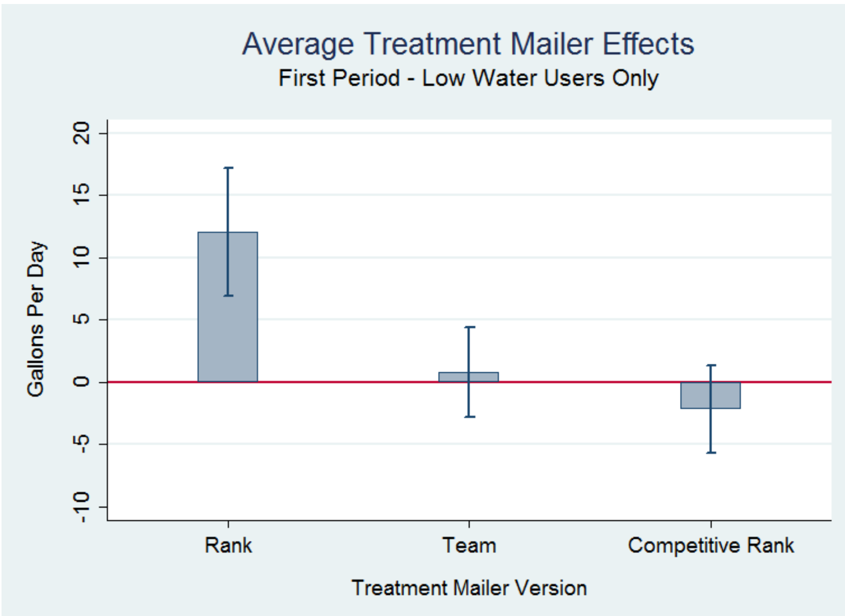




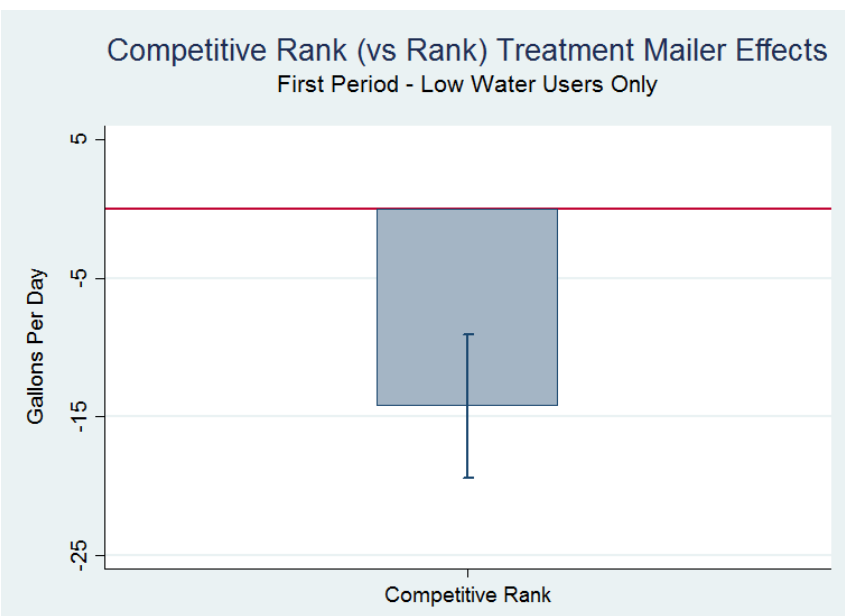
Appendix 7A: Average Treatment Effects (Treatment Mailers Relative to the Control Mailer) - Low Water Users

First Post-Mailer-Initiation Period Only (Dec. 2012)

i. All Treatments - Standard Error Marked



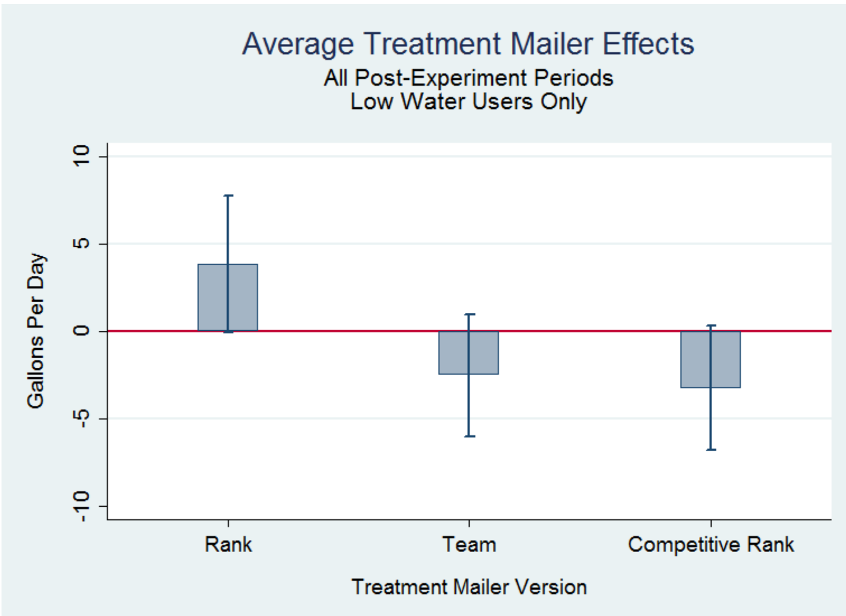
ii. Competitive Rank and Rank Treatments Only - Standard Error Marked



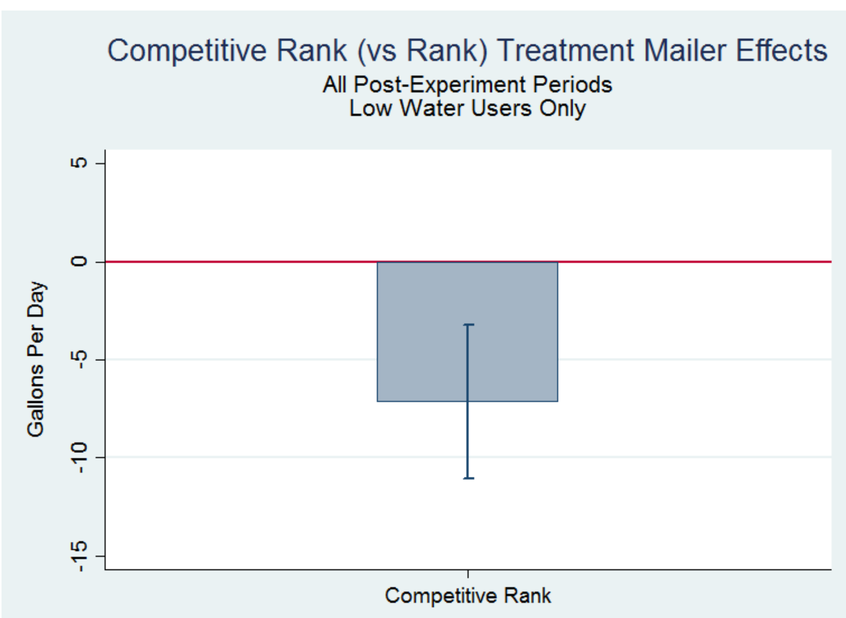
Appendix 7B: Average Treatment Effects (Treatment Mailers Relative to the Control Mailer) - Low Water Users

All Post-Experiment Periods (Mean Gallons Per Day)

i. All Treatments - Standard Error Marked



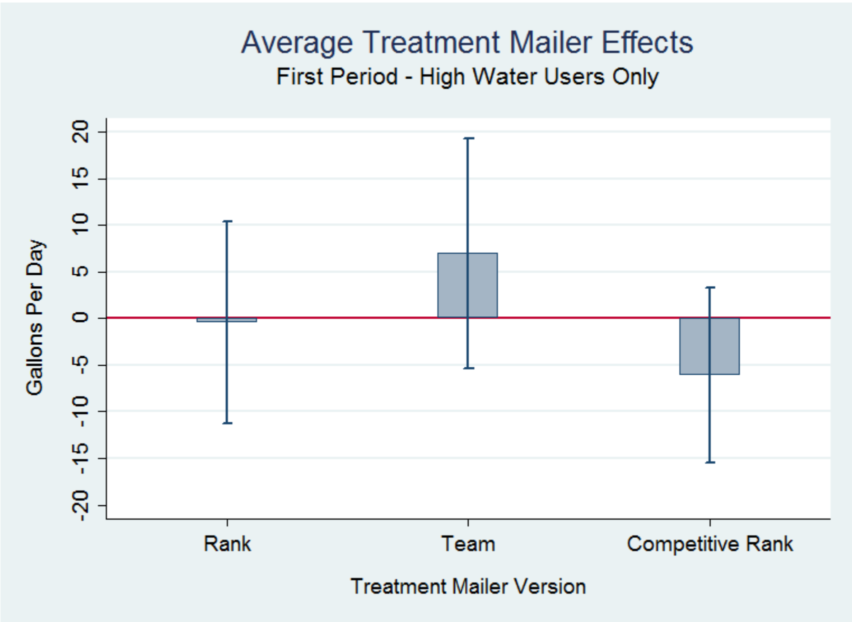
ii. Competitive Rank and Rank Treatments Only - Standard Error Marked



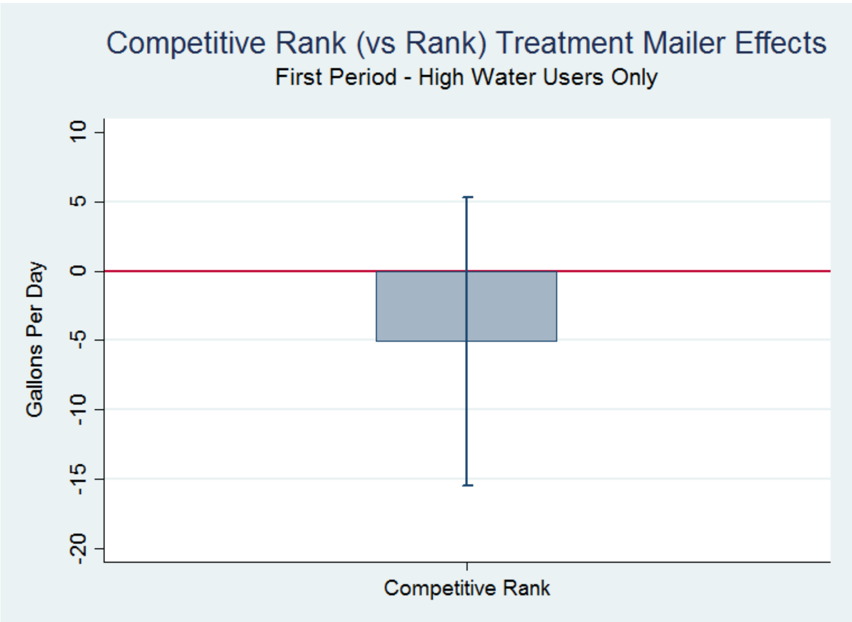
Appendix 8A: Average Treatment Effects (Treatment Mailers Relative to the Control Mailer) - High Water Users

First Post-Mailer-Initiation Period Only (December 2012)

i. All Treatments - Standard Error Marked



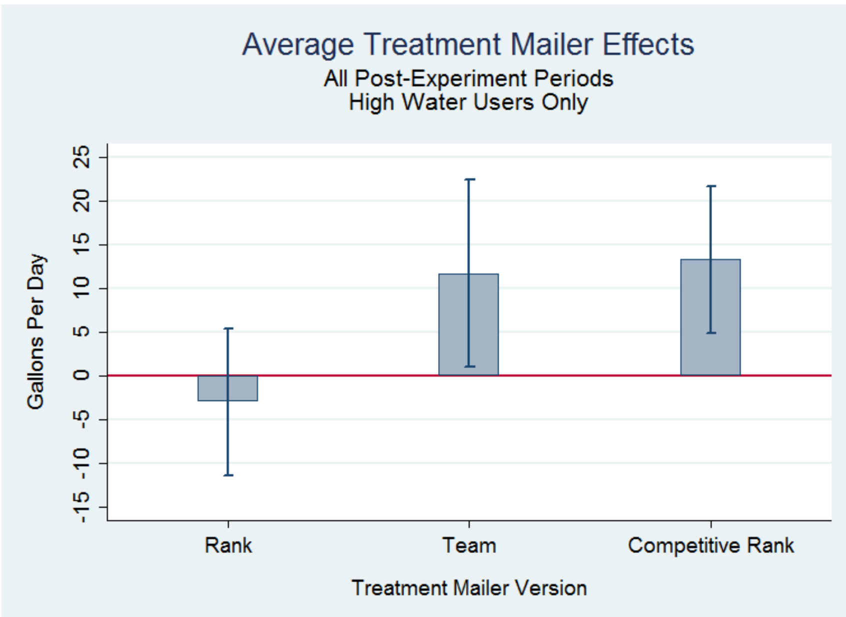
ii. Competitive Rank and Rank Treatments Only - Standard Error Marked



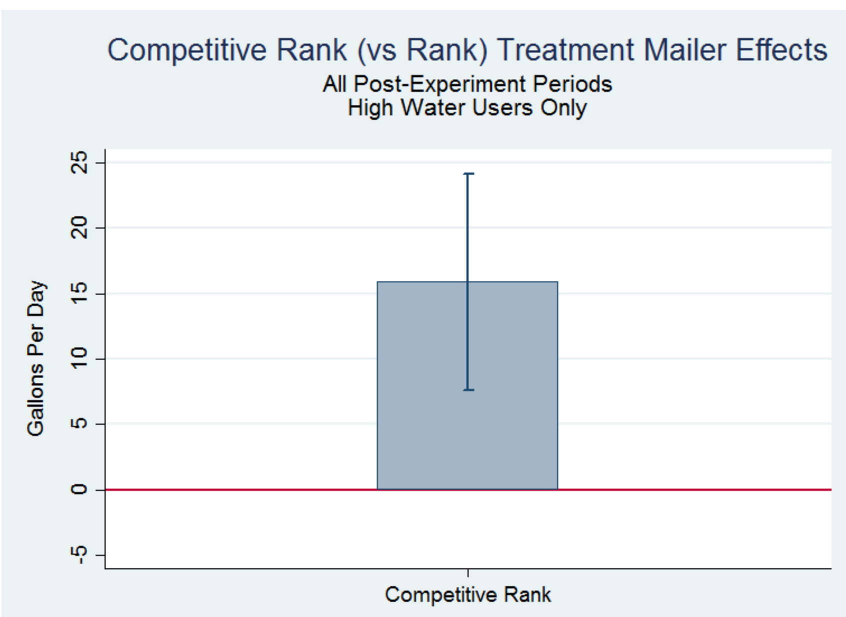
Appendix 8B: Average Treatment Effects (Treatment Mailers Relative to the Control Mailer) - High Water Users

Mean of all Post-Experiment Periods

i. All Treatments - Standard Error Marked



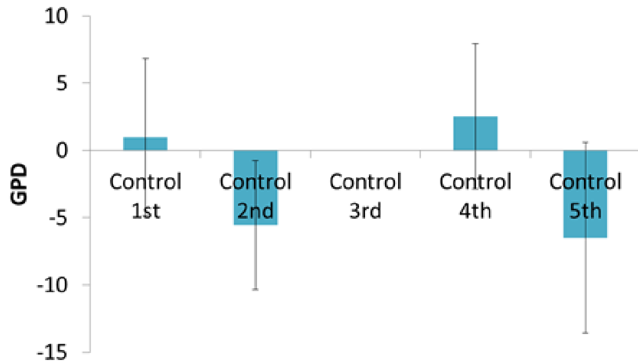
ii. Competitive Rank and Rank Treatments Only - Standard Error Marked



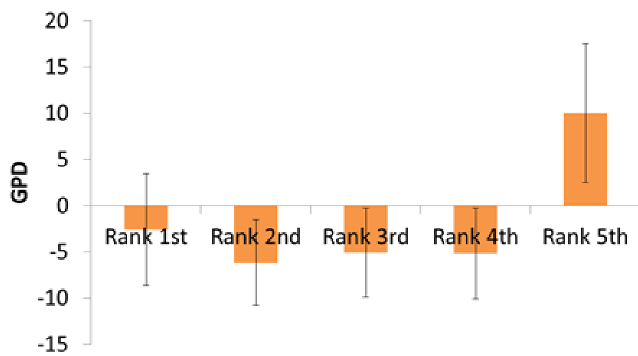
Appendix 9: Coefficients from Interactions of Treatment and Rank Position (Table 17)

Middle-Third (Pre-Experiment) of Water Users Only; Control 3rd Place Omitted; Clustered Standard Errors Marked with Bars

i. Control



ii. Rank



iii. Competitive Rank

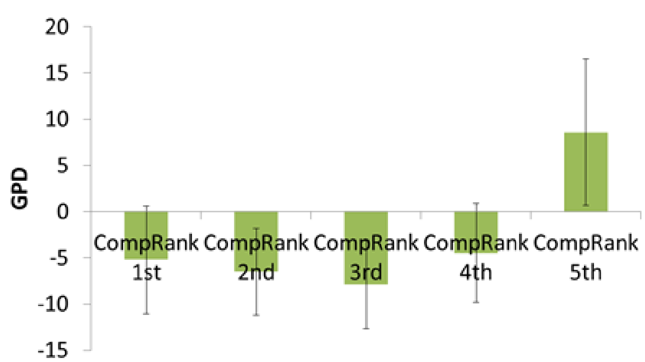


Table 1: Demographic Variables by Treatment Group

	Control Mailer	Rank Mailer	Team Mailer	Comp. Rank Mailer	Out-of-Sample Control
Home Size (sqft.)	1650.5	1627.5	1628.0	1622.2	1976.2
	(561.7)	(533.4)	(544.0)	(538.3)	(932.2)
Lot Size (sqft.)	7503.7	7320.8	7657.0	7376.4	8655.2
	(4390.9)	(3615.7)	(4821.9)	(5175.4)	(9397.3)
Year Home Built	1958.0	1957.5	1957.4	1957.4	1954.3
	(13.81)	(13.46)	(13.24)	(13.05)	(27.71)
Number of Bedrooms	3.129	3.110	3.160	3.109	3.181
	(0.725)	(0.735)	(0.759)	(0.757)	(0.968)
Number of Bathrooms	2.071	2.055	2.049	2.022	2.311
	(0.827)	(0.859)	(0.832)	(0.833)	(1.066)
Mean Water Use (Pre-Exp 2012)	229.7	231.0	230.5	236.1	304.0
	(129.6)	(136.9)	(141.3)	(136.9)	(264.5)
N	1308	1288	1284	1300	2880

Means, with standard deviations in parentheses

Table 2: Randomization Checks

	(1) Home Size	(2) Lot Size	(3) Year Built	(4) Bedrooms	(5) Bathrooms	(6) Pre-Treat Mean Water Use
Rank	-25.26	-203.9	-0.565	-0.0185	-0.0147	0.859
	(23.04)	(169.4)	(0.575)	(0.0309)	(0.0354)	(5.264)
CompetitiveRank	-29.98	-120.7	-0.651	-0.0172	-0.0475	5.912
	(23.09)	(202.6)	(0.564)	(0.0312)	(0.0346)	(5.253)
Team	-19.68	163.6	-0.564	0.0352	-0.0135	0.853
	(23.36)	(196.3)	(0.572)	(0.0315)	(0.0349)	(5.374)
Observations	4485	4485	4485	4445	4518	5090
R^2	0.000	0.001	0.000	0.001	0.000	0.000

Standard errors in parentheses

This table presents the regressions of various household characteristics on dummies for the three treatment groups (Rank, Competitive Rank, and Team), as a randomization check using in-sample data.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Difference-in-Differences: Dec 2011 (Pre-Treat) vs. Dec 2012 (Post-Treat)

	(1) GPD	(2) Log GPD
Control*Post	1.650	0.00808
	(8.539)	(0.0355)
Rank*Post	4.237	0.0136
	(8.567)	(0.0357)
Team*Post	4.261	0.00241
	(8.823)	(0.0352)
CompRank*Post	2.107	0.00407
	(8.514)	(0.0353)
Control	-50.20***	-0.207***
	(6.004)	(0.0251)
Rank	-52.49***	-0.216***
	(5.940)	(0.0251)
Team	-51.70***	-0.204***
	(5.873)	(0.0247)
CompRank	-48.53***	-0.190***
	(5.907)	(0.0249)
Post	-17.41**	-0.0887***
	(6.970)	(0.0241)
N	12658	12658
R^2	0.024	0.022

Standard errors in parentheses

This table presents difference-in-differences regression results comparing the water use of the In-Sample and Out-of-Sample groups. The pre-period is the Dec 2011 read and the post-period is the Dec 2012 read, the period after the first experimental mailer was sent. Regressions without controls are presented, in both level-level and log-level forms.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Difference-in-Differences: Pre-Treat Mean vs. Post-Treat Mean

	(1) Mean GPD	(2) Log Mean GPD
Control*Post	-16.17**	-0.0336
	(7.958)	(0.0324)
Rank*Post	-16.28**	-0.0335
	(7.939)	(0.0323)
Team*Post	-14.62*	-0.0459
	(8.539)	(0.0320)
CompRank*Post	-13.44*	-0.0323
	(8.103)	(0.0325)
Control	-46.55***	-0.172***
	(5.348)	(0.0229)
Rank	-47.32***	-0.175***
	(5.355)	(0.0228)
Team	-47.08***	-0.166***
	(5.378)	(0.0224)
CompRank	-43.45***	-0.157***
	(5.397)	(0.0227)
Post	14.58**	0.0369*
	(6.489)	(0.0223)
N	13164	13164
R^2	0.027	0.018

Standard errors in parentheses

This table presents difference-in-differences regression results comparing the water use of the In-Sample and Out-of-Sample groups. The pre-period water use measure is the mean household water use in the Dec 2011, Feb 2012, Apr 2012, and Jun 2012 reads and the post-period is the mean household water use in the Dec 2012, Feb 2013, Apr 2013, and Jun 2013 reads. Regressions without controls are presented, in both level-level and log-level forms.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Means Comparison: Water Use in the First Post-Mailer Period (In-Sample Only)

	(1) GPD	(2) GPD	(3) GPD	(4) Log GPD	(5) Log GPD	(6) Log GPD
Rank	0.774	5.270	4.598	0.000105	0.0279	0.0227
	(4.767)	(5.027)	(4.438)	(0.0256)	(0.0267)	(0.0227)
Team	0.427	2.743	4.306	-0.00547	0.00991	0.0198
	(5.199)	(5.448)	(4.893)	(0.0255)	(0.0268)	(0.0227)
CompetitiveRank	1.751	2.148	-2.051	0.00595	0.00732	-0.0123
	(4.719)	(4.825)	(4.109)	(0.0255)	(0.0267)	(0.0223)
Lot Size		0.00427***	0.00489***		0.0000111***	0.0000156***
		(0.000874)	(0.000717)		(0.00000248)	(0.00000230)
Num Bathrooms		5.512*	2.576		0.0581***	0.0385**
		(3.109)	(2.676)		(0.0184)	(0.0151)
Home Size (SqFt)		0.0269***	0.0229***		0.000120***	0.0000979***
		(0.00562)	(0.00477)		(0.0000303)	(0.0000247)
Constant	175.5***	86.97***	74.15***	4.980***	4.572***	4.553***
	(3.266)	(8.422)	(7.262)	(0.0180)	(0.0366)	(0.0320)
Observations	5041	4440	4414	5039	4439	4414
R^2	0.000	0.054	0.237	0.000	0.038	0.306
Read Month Fixed Effects	No	No	Yes	No	No	Yes
WaterScore Fixed Effects	No	No	Yes	No	No	Yes

Standard errors in parentheses

This table first presents three regressions (1-3) that compare water use in the first period following mailer initiation in the in-sample groups only. The following three regressions (4-6) present the same results using a log-level specification. The controls used are for household characteristics, the WaterScore that households observed on the first mailer, and Read Month.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Means Comparison: Water Use in All Periods (In-Sample Only)

	(1) Mean GPD	(2) Mean GPD	(3) Log Mean GPD	(4) Log Mean GPD
Rank	-0.740	2.398	-0.00413	0.0172
	(4.356)	(4.427)	(0.0227)	(0.0234)
Team	1.391	2.894	-0.00426	0.00799
	(5.063)	(5.105)	(0.0228)	(0.0237)
CompetitiveRank	6.194	8.194*	0.0107	0.0189
	(4.616)	(4.663)	(0.0234)	(0.0243)
Lot Size		0.00628***		0.0000150***
		(0.00163)		(0.00000254)
Num Bathrooms		7.061**		0.0487***
		(3.127)		(0.0156)
Home Size (SqFt)		0.0285***		0.000135***
		(0.00541)		(0.0000253)
Observations	5041	4440	5039	4439
R^2	0.000	0.100	0.000	0.057

Standard errors in parentheses

This table first presents a simple means comparison of mean water use in all periods following mailer initiation in the in-sample groups only. The subsequent regression controls for household characteristics. The next two regression repeat the analysis, using a log-level specification.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Low Water Users in the First Post-Mailer Period (In-Sample Only)

	(1) GPD	(2) GPD	(3) GPD	(4) Log GPD	(5) Log GPD	(6) Log GPD
Rank	9.213**	12.18**	12.02**	0.0485	0.0776*	0.0780*
	(4.655)	(5.139)	(5.165)	(0.0404)	(0.0428)	(0.0427)
Team	-0.423	-0.0609	0.815	0.00199	0.00479	0.0168
	(3.391)	(3.607)	(3.612)	(0.0384)	(0.0411)	(0.0408)
CompetitiveRank	-1.730	-2.461	-2.208	-0.0251	-0.0339	-0.0251
	(3.437)	(3.528)	(3.534)	(0.0389)	(0.0411)	(0.0408)
Lot Size		0.000740	0.00104*		0.00000494	0.00000710
		(0.000616)	(0.000619)		(0.00000497)	(0.00000499)
Num Bathrooms		4.083*	4.027**		0.0546*	0.0543**
		(2.091)	(2.044)		(0.0280)	(0.0274)
Home Size (SqFt)		0.00105	0.00129		0.00000373	0.00000865
		(0.00365)	(0.00361)		(0.0000500)	(0.0000492)
Constant	95.42***	80.75***	86.62***	4.425***	4.279***	4.364***
	(2.418)	(5.537)	(5.654)	(0.0274)	(0.0600)	(0.0595)
Observations	1651	1446	1434	1649	1445	1434
R^2	0.005	0.015	0.036	0.002	0.013	0.034
Read Month Fixed Effects	No	No	Yes	No	No	Yes
WaterScore Fixed Effects	No	No	Yes	No	No	Yes

Standard errors in parentheses

This table first presents three level regressions (1-3), with different controls, providing means comparisons of water use in the first period following mailer initiation amongst households who were low water users in advance of the experiment. This means that they were in the lowest third of water users in their irrigable area category in the 2012 months that preceded the experiment. The next three regressions (4-6) provide the same results in log-level form. The controls used are for household characteristics, the WaterScore that households observed on the mailer, and the meter read month.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Low Water Users in the First Post-Mailer Period (Rank and CompRank Only)

	(1) GPD	(2) GPD	(3) Log GPD	(4) Log GPD
CompetitiveRank	-14.64***	-14.22***	-0.111**	-0.103**
	(5.216)	(5.211)	(0.0436)	(0.0435)
Lot Size	0.000329	0.000643	0.00000291	0.00000532
	(0.000592)	(0.000588)	(0.00000607)	(0.00000603)
Num Bathrooms	1.834	1.941	0.0295	0.0321
	(2.583)	(2.548)	(0.0321)	(0.0315)
Home Size (SqFt)	0.00625	0.00606	0.0000892*	0.0000855*
	(0.00504)	(0.00502)	(0.0000522)	(0.0000516)
Constant	92.22***	95.13***	4.289***	4.352***
	(9.324)	(9.002)	(0.0849)	(0.0850)
Observations	716	708	716	708
R^2	0.015	0.044	0.020	0.036
Read Month Fixed Effects	No	Yes	No	Yes
WaterScore Fixed Effects	No	Yes	No	Yes

Standard errors in parentheses

This table first presents two regressions (1-2) providing a simple means comparison of water use in the first period following mailer initiation amongst households who were low water users in advance of the experiment. This means that they were in the lowest third of water users in their irrigable area category in the 2012 months that preceded the experiment. The subsequent regressions (3-4) provide the same result in log-level form. The controls used are for household characteristics, the WaterScore that households observed on the mailer, and the meter read month.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Low Water Users in All Post-Mailer Periods (In-Sample Only)

	(1) meanGPDall	(2) meanGPDall	(3) lnmeanGPDall	(4) lnmeanGPDall
Rank	2.166	3.850	0.0110	0.0298
	(3.613)	(3.925)	(0.0335)	(0.0357)
Team	-2.658	-2.532	-0.0150	-0.0121
	(3.286)	(3.526)	(0.0325)	(0.0349)
CompetitiveRank	-2.974	-3.255	-0.0362	-0.0400
	(3.400)	(3.604)	(0.0340)	(0.0362)
Lot Size		0.000834**		0.00000676**
		(0.000358)		(0.00000343)
Num Bathrooms		3.402*		0.0338
		(1.932)		(0.0210)
Home Size (SqFt)		0.00194		0.0000279
		(0.00327)		(0.0000364)
Observations	1651	1446	1649	1445
R^2	0.002	0.013	0.001	0.012

Standard errors in parentheses

This table first presents two level regressions (1-2) of water use in all periods following mailer initiation on the mailer treatments for households who were low water users in advance of the experiment (this means that they were in the lowest third of water users in their irrigable area category in the 2012 months that preceded the experiment. The two regressions that follow (3-4) provide the same results in log-level form. Regressions 2 and 4 use demographic controls.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Low Water Users in All Post-Mailer Periods (Rank and Competitive Rank Only)

	(1) meanGPDall	(2) meanGPDall	(3) lnmeanGPDall	(4) lnmeanGPDall
CompetitiveRank	-5.140	-7.150*	-0.0472	-0.0697*
	(3.536)	(3.913)	(0.0343)	(0.0373)
Lot Size		0.000719		0.00000489
		(0.000571)		(0.00000551)
Num Bathrooms		0.943		0.0199
		(2.387)		(0.0255)
Home Size (SqFt)		0.00600		0.0000793*
		(0.00465)		(0.0000439)
Observations	827	716	827	716
R^2	0.003	0.012	0.002	0.016

Standard errors in parentheses

This table presents both log (1-2) and level (3-4) regressions of mean water use in all periods following mailer initiation amongst households who were low water users in advance of the experiment (this means that they were in the lowest third of water users in their irrigable area category in the 2012 months that preceded the experiment.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: High Water Users in the First Post-Mailer Period (In-Sample Only)

	(1) GPD	(2) GPD	(3) GPD	(4) Log GPD	(5) Log GPD	(6) Log GPD
Rank	-5.248	-2.275	-0.453	-0.0269	-0.0175	-0.0122
	(10.11)	(11.04)	(10.83)	(0.0327)	(0.0347)	(0.0340)
Team	1.731	4.287	6.950	-0.0105	0.000372	0.00947
	(12.19)	(12.80)	(12.44)	(0.0324)	(0.0337)	(0.0332)
CompetitiveRank	-4.986	-4.691	-6.061	-0.0158	-0.0134	-0.0130
	(9.650)	(10.01)	(9.432)	(0.0306)	(0.0323)	(0.0316)
Lot Size		0.00637***	0.00611***		0.0000127***	0.0000122***
		(0.00115)	(0.00105)		(0.00000222)	(0.00000213)
Num Bathrooms		-10.41	-8.805		-0.0187	-0.0138
		(6.762)	(6.743)		(0.0202)	(0.0200)
Home Size (SqFt)		0.0306***	0.0319***		0.0000927***	0.000102***
		(0.0108)	(0.0104)		(0.0000305)	(0.0000291)
Constant	267.8***	187.6***	147.0***	5.482***	5.262***	5.105***
	(6.845)	(14.41)	(21.10)	(0.0224)	(0.0422)	(0.0505)
Observations	1692	1459	1454	1692	1459	1454
R^2	0.000	0.067	0.078	0.000	0.045	0.068
Read Month Fixed Effects	No	No	Yes	No	No	Yes
WaterScore Fixed Effects	No	No	Yes	No	No	Yes

Standard errors in parentheses

This table first presents three level regressions (1-3), with different controls, providing means comparisons of water use in the first period following mailer initiation amongst households who were high water users in advance of the experiment. This means that they were in the highest third of water users in their irrigable area category in the 2012 months that preceded the experiment. The next three regressions (4-6) provide the same results in log-level form. The controls used are for household characteristics, the WaterScore that households observed on the mailer, and the meter read month.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: High Water Users in the First Post-Mailer Period (Rank and CompRank Only)

	(1) GPD	(2) GPD	(3) Log GPD	(4) Log GPD
CompetitiveRank	-2.709	-5.111	0.00350	-0.000215
	(10.78)	(10.45)	(0.0335)	(0.0330)
Lot Size	0.00502***	0.00501***	0.00000963***	0.00000956***
	(0.00100)	(0.000985)	(0.00000194)	(0.00000200)
Num Bathrooms	-17.46	-16.06	-0.0318	-0.0328
	(10.68)	(10.02)	(0.0283)	(0.0273)
Home Size (SqFt)	0.0479***	0.0467***	0.000127***	0.000136***
	(0.0179)	(0.0165)	(0.0000452)	(0.0000412)
Constant	181.3***	139.3***	5.238***	5.127***
	(17.24)	(19.64)	(0.0534)	(0.0600)
Observations	746	745	746	745
R^2	0.079	0.093	0.044	0.063
Read Month Fixed Effects	No	Yes	No	Yes
WaterScore Fixed Effects	No	Yes	No	Yes

Standard errors in parentheses

This table first presents two regressions (1-2) providing a simple means comparison of water use in the first period following mailer initiation amongst households who were high water users in advance of the experiment. This means that they were in the highest third of water users in their irrigable area category in the 2012 months that preceded the experiment. The subsequent regressions (3-4) provide the same result in log-level form. The controls used are for household characteristics, the WaterScore that households observed on the mailer, and the meter read month.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: High Water Users in All Post-Mailer Periods (In-Sample Only)

	(1) meanGPDall	(2) meanGPDall	(3) lnmeanGPDall	(4) lnmeanGPDall
Rank	-1.126	-2.999	-0.00473	-0.00565
	(8.211)	(8.447)	(0.0251)	(0.0255)
Team	11.31	11.69	0.0196	0.0203
	(11.25)	(10.78)	(0.0256)	(0.0261)
CompetitiveRank	9.934	13.28	0.0224	0.0344
	(8.513)	(8.479)	(0.0254)	(0.0258)
Lot Size		0.00955***		0.0000171***
		(0.00309)		(0.00000383)
Num Bathrooms		-8.389		-0.0277*
		(5.749)		(0.0149)
Home Size (SqFt)		0.0290***		0.000102***
		(0.00972)		(0.0000235)
Observations	1692	1459	1692	1459
R^2	0.002	0.172	0.001	0.108

Standard errors in parentheses

This table first presents two level regressions (1-2) of water use in all periods following mailer initiation on the mailer treatments for households who were high water users in advance of the experiment (this means that they were in the highest third of water users in their irrigable area category in the 2012 months that preceded the experiment. The two regressions that follow (3-4) provide the same results in log-level form. Regressions 2 and 4 use demographic controls.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: High Water Users in All Post-Mailer Periods (Rank and Competitive Rank Only)

	(1) meanGPDall	(2) meanGPDall	(3) lnmeanGPDall	(4) lnmeanGPDall
CompetitiveRank	11.06	15.89*	0.0271	0.0399
	(8.455)	(8.310)	(0.0256)	(0.0260)
Lot Size		0.00650***		0.0000140***
		(0.000914)		(0.00000378)
Num Bathrooms		-14.76*		-0.0338
		(7.967)		(0.0218)
Home Size (SqFt)		0.0490***		0.000137***
		(0.0128)		(0.0000347)
Observations	875	746	875	746
R^2	0.002	0.177	0.001	0.112

Standard errors in parentheses

This table presents both log and level regressions of mean water use in all periods following mailer initiation amongst households who were high water users in advance of the experiment (this means that they were in the highest third of water users in their irrigable area category in the 2012 months that preceded the experiment.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Last Place Effect: Using All Post-Treatment Periods

	(1) GPD	(2) GPD	(3) GPD	(4) GPD
Rank	9.297	2.552	1.014	1.347
	(10.48)	(7.385)	(6.009)	(6.087)
Competitive Rank	29.88***	21.00***	18.94***	17.85***
	(10.92)	(7.441)	(5.957)	(5.957)
prev_GPD		0.448***	0.608***	0.588***
		(0.0444)	(0.0457)	(0.0509)
Lot Size		0.00305***	0.00211***	0.00226***
		(0.000570)	(0.000525)	(0.000540)
Num Bathrooms		-2.511	1.888	1.623
		(5.594)	(4.729)	(4.664)
Home Size (SqFt)		0.0269***	0.0156**	0.0168**
		(0.00799)	(0.00669)	(0.00662)
Constant	286.7***	77.84***	-71.64***	-86.57***
	(7.634)	(12.23)	(16.41)	(13.87)
Observations	3161	2731	2731	2727
R^2	0.005	0.298	0.494	0.497
Read Month Fixed Effects	No	No	Yes	Yes
Mailers Seen Fixed Effects	No	No	Yes	Yes
WaterScore Fixed Effects	No	No	No	Yes

Clustered standard errors in parentheses.

This table presents regression results comparing the effect of the Neutral and Competitive treatments for 'last place' households on subsequent water use. The omitted group here is households in the In-Sample Control who 'would have' been in last place in their groups had they seen a ranking in their mailer. Standard errors were clustered at the household level.

Table 16: First Place Effect: Using All Post-Treatment Periods

	(1) GPD	(2) GPD	(3) GPD	(4) GPD
Rank	-10.16**	1.269	1.154	2.578
	(4.670)	(3.609)	(3.340)	(3.391)
Competitive Rank	-12.69***	0.0177	-0.0711	1.008
	(4.597)	(3.645)	(3.463)	(3.485)
prev_GPD		0.731***	0.810***	0.892***
		(0.0366)	(0.0354)	(0.0419)
Lot Size		0.00116**	0.00105**	0.00101**
		(0.000504)	(0.000457)	(0.000452)
Num Bathrooms		1.350	2.288	2.207
		(2.450)	(2.244)	(2.214)
Home Size (SqFt)		0.0104**	0.00878*	0.00746
		(0.00502)	(0.00466)	(0.00459)
Constant	128.1***	21.85***	-20.76***	-36.83***
	(3.064)	(7.359)	(7.793)	(8.675)
Observations	2948	2649	2649	2649
R^2	0.004	0.217	0.341	0.344
Read Month Fixed Effects	No	No	Yes	Yes
Mailers Seen Fixed Effects	No	No	Yes	Yes
WaterScore Fixed Effects	No	No	No	Yes

Clustered standard errors in parentheses.

This table presents regression results comparing the effect of the Neutral and Competitive treatments for 'first place' households on subsequent water use. The omitted group here is households in the In-Sample Control who 'would have' been in first place in their groups had they seen a ranking in their mailer. Standard errors were clustered at the household level.

Table 17: Differential Impact of Rank by Treatment (Control 3rd Place Omitted)

	(1) GPD	(2) GPD	(3) GPD
Control 1st Place	-7.053	-0.920	0.988
	(6.292)	(5.685)	(5.848)
Control 2nd Place	-3.765	-5.892	-5.542
	(5.117)	(4.750)	(4.788)
Control 4th Place	2.543	2.158	2.542
	(5.941)	(5.379)	(5.393)
Control 5th Place	-3.339	-7.189	-6.488
	(7.787)	(7.076)	(7.098)
Rank 1st Place	-8.378	-4.129	-2.587
	(6.549)	(5.934)	(6.034)
Rank 2nd Place	-10.47**	-6.774	-6.156
	(5.078)	(4.565)	(4.590)
Rank 3rd Place	-7.628	-5.335	-5.090
	(5.175)	(4.748)	(4.775)
Rank 4th Place	-2.882	-5.641	-5.209
	(5.450)	(4.853)	(4.890)
Rank 5th Place	12.09	9.293	10.00
	(8.202)	(7.472)	(7.511)
Comp. Rank 1st Place	-10.25*	-7.027	-5.194
	(6.044)	(5.682)	(5.842)
Comp. Rank 2nd Place	-9.333*	-7.082	-6.495
	(4.996)	(4.639)	(4.666)
Comp. Rank 3rd Place	-10.55**	-8.157*	-7.878*
	(5.131)	(4.751)	(4.767)
Comp. Rank 4th Place	-3.106	-4.824	-4.464
	(5.929)	(5.323)	(5.359)
Comp. Rank 5th Place	10.90	7.997	8.618
	(8.828)	(7.852)	(7.917)
Constant	134.4***	34.77***	36.30***
	(5.724)	(7.502)	(7.719)
Observations	4553	4121	4121
R^2	0.101	0.398	0.398
Read Month Fixed Effects	No	Yes	Yes
Demographic and Mailer Number Fixed Effects	No	Yes	Yes
WaterScore Fixed Effects	No	No	Yes

Clustered standard errors in parentheses.

Regressions include only the middle third of water users, based on pre-experiment 2012 water use.

Regressions omit Control 3rd Place households.