Centralized vs Decentralized Demand Response: Evidence from a Field Experiment*

Megan Bailey† David P. Brown‡ Blake Shaffer§
Frank A. Wolak¶

March 10, 2023

Draft prepared for NBER EEE Spring Conference
Please do not cite or circulate

Abstract

We conduct a field experiment with residential electricity customers to evaluate the effectiveness of centralized (utility-initiated) vs decentralized (customer-initiated) demand response. Participants receive dynamic incentives to reduce electricity use during randomized peak events. Treatment groups differ in the ease with which they can respond to events in terms of a) provision of technology to remotely control devices in their home, including hot water heaters, baseboard thermostats, and electric vehicle chargers, and b) whether the default response is to reduce consumption during an event (centralized) or requires customer-initiated action to do so (decentralized). We find centrally-initiated households reduce consumption by 26% on average during 3-hour demand response events, whereas customer-initiated households reduce by only 4% despite both groups receiving the same financial compensation for the same percentage demand reduction. Having to take an action, one as small as pushing a button on an app versus not having to do so, results in a 6-fold difference in response. We find an additional "manual" decentralized group, one with the same incentives but without remote control technology installed, indistinguishable in their consumption reduction (5%) to the decentralized group with technology. These findings speak to the importance of reducing the effort and cognitive burden on residential customers to unlock flexible electricity demand.

*We are grateful to our partner electric utility company for sponsoring and managing the field experiment on which this study is based. We thank attendees at the 2022 Berkeley Energy Camp, 2022 Electricity Camp in the Rockies, 2022 Canadian Resource and Environmental Economics annual conference, and 2023 Berkeley POWER conference for useful feedback. This work is supported by excellent research assistance from Mallika Sharma. We acknowledge funding from our utility partner, the University of Calgary, the Canada First Research Excellence Fund as part of the University of Alberta's Future Energy Systems research initiative and the University of Calgary's Global Research Initiative, and the Social Sciences and Humanities Research Council Insight Development Grant. This experiment is registered in the AEA RCT Registry under ID #AEARCTR-0008680. All errors are our own.

†Haskayne School of Business and the School of Public Policy, University of Calgary megan.bailey@ucalgary.ca

‡Department of Economics, University of Alberta dpbrown@ualberta.ca

§Department of Economics and the School of Public Policy, University of Calgary blake.shaffer@ucalgary.ca

¶Department of Economics, Stanford University and NBER wolak@zia.stanford.edu

1 Introduction

A fundamental challenge for electricity markets is the need to balance supply and demand at every instant despite limited storability and inelastic demand. Historically, this has been achieved by forecasting demand and adjusting (dispatching) supply. Going forward, however, the growth of intermittent renewable generation and new sources of flexible demand (such as electric vehicles) may lead to the prior adage being flipped upside down, with grid operators forecasting supply and dispatching demand.

Economists have long recognized the benefits of price-responsive demand to improve the efficiency of electricity markets. However, electricity consumers are generally inattentive to both their electricity price and usage (perhaps rationally so), and even when made aware, the task of acting on this information can be onerous relative to the relatively small gains on individual decisions. Put simply, given the opportunity cost of time and effort, electricity consumers may be reluctant to participate in demand response programs that amount to "picking up pennies" in a series of irregular and relatively low stakes opportunities.

In this paper, we examine the role centrally-initiated demand response can play in overcoming barriers to unlocking flexible demand from residential electricity customers. Partnering with a large electric utility, we recruited approximately 1700 participants to a demand response field experiment, whereby households receive occasional "peak event" notifications and can earn money by reducing their electricity consumption during the 3 hour window of each event.

The treatment groups differ in the technology they are provided with to assist in responding to the event and whether consumption reductions are initiated by the household (decentralized) or by the utility (centralized). Our first treatment group, the **Manual group**, receives rewards for consumption reductions during peak events but is not supplied with any enabling technology to minimize the effort required to do so. They must, as the name suggests, manually reduce consumption among their home appliances. Rewards range from \$1 to \$6 Canadian dollars per 3-hour event, depending on the level of consumption reduction achieved and whether the event is specified as a high or regular rewards event type.[1] The second treatment group, the **Tech group**, receives the same incentives but is also equipped with app-

[1] The rewards are designed to mimic the cost savings of demand reduction during peak market conditions when matching supply and demand is challenging. This corresponds to approximately \$1.18 to \$2.35 per kWh as compensation for consumption reductions for the average household in our sample.

enabled load controllers on devices our Utility partner installs in their home. The controllable devices include (i) baseboard thermostats, (ii) electric hot water heaters, and (iii) level 2 electric vehicle (EV) chargers. The Tech group, like the Manual one, is part of our *decentralized* demand response—they must actively respond to a peak event notification by taking an action. In the case of the Tech group, however, the effort required is less than the Manual group. Finally, our third treatment group, the **Central group**, mirrors the Tech group in receiving incentives and having load control technology installed in their home, but with one crucial difference: their load controllers' default response to a peak event notification is to have their devices automatically reduce their consumption.[2] Households in the Central group can opt-out of consumption reductions during an event but, in contrast to the Tech group, they must actively make an effort to *not* reduce consumption by pushing a button on their app.

We estimate the causal effect of receiving peak event incentives on electricity consumption for each group by comparing household consumption during event and non-event periods. We randomize events for each household at the household-day-time of day-incentive level; that is, each household receives their own unique randomized schedule of peak events over the course of the experiment. Doing so allows for identification both across and within treatment groups and also ensures events are not correlated with other drivers of household electricity consumption. In total, we analyze approximately 10 million hourly household-level consumption observations with over 15,000 household-event days.

The results are stark. We find the Central group reduces consumption by an average of 26% during events as compared to only 4% for the Tech group. This difference speaks to the importance of minimizing the costs of taking action in settings where both inattention is rife and the incentives per event are relatively small. Having to take an action, even as small as having to push a button on an app, to "opt-in" to a demand response event versus being defaulted to opt-in and having to push a button to "opt-out" results in a six-fold difference in effect. Perhaps more surprisingly, we see no meaningful difference between the Tech group's performance (-4%) and that of the Manual group's (-5%). This emphasizes that technology alone is not sufficient to overcome barriers to action; switching the default is imperative.

One possible trade-off for the superior performance of the Central group is that

[2]There are a number of recently developed programs that include utility-managed load-control for various appliances including hot water heaters (Wattersaver, 2023), thermostats (PG&E, 2023), electric vehicles (DTE Energy, 2022), and solar-plus-storage systems (Spector, 2020).

households may be less inclined to accept this centrally-managed program compared to alternatives. Despite our expectations, we find a relatively modest difference in the take-up rates of the Central and Tech group offers (42% and 48%, respectively). This surprising result indicates that consumers were not deterred by the idea of the electric utility managing the consumption of their devices.

The difference in responsiveness to incentives between the Central and Tech groups remains in our intention-to-treat (ITT) estimates (13% vs 3%), which encompass both take-up rates and per-event electricity consumption responses by program/group. These estimates indicate that centralized management of household device electricity use during peak events is the clear program winner (of those considered here) for utility companies considering demand-side measures.

A novel feature of our experiment is the inclusion of EV chargers and electric hot water heaters.[3] These two devices represent large shares of household electricity demand. We see stark differences between the event treatment effect estimates by homes with various controllable devices across the Central and Tech groups. For example, we find households in the Central group with hot water heaters reduce their demand by 24% during events, as compared to 1% for similar households in the Tech group. This difference across groups for households with controllable hot water heaters highlights that there may be potentially undiscovered sources of electricity demand flexibility in large-consuming household appliances. Tech group households with EV chargers seem to reduce demand during events more (by 7.5%) than those with hot water heaters (1%). We suspect this may be driven by consumers' greater understanding of, or familiarity with, using EV chargers for altering household electricity demand, compared to other devices such as hot water heaters. However, utility load control is still valuable for increasing demand flexibility during events for homes with EV chargers, given the difference in treatment effects we see between the Central and Tech groups with EV chargers (14% and 7.5%).

The importance of flexible electricity demand was on display in California in Summer of 2022, where statewide appeals for conservation narrowly avoided rolling blackouts (Balaraman, 2022). This event, as well as the February 2021 blackout in Texas during a cold snap (King et al., 2021), illustrates the urgency of the problem. A critical policy question is whether greater consumer demand flexibility could be cheaper to facilitate than costly system upgrades to accommodate changing patterns

[3] Much of the literature on automated electricity demand response has focused on smart thermostats. See, for example, Bollinger and Hartmann (2020); Blonz et al. (2021); Brandon et al. (2022).

in electricity demand and supply. The behavioral question addressed in this paper is how best to overcome impediments to otherwise beneficial demand response.

Our paper builds on several strands of the literature. First, we add to the rich set of empirical research estimating household responsiveness to time-varying pricing in electricity.[4] Our experimental design is most similar to the critical peak pricing (CPP) strand of this literature. The results from our Manual group with no load control technology nor automation is inline with those observed in prior CPP studies. Our main estimate of a 5% reduction in the Manual group falls within the range of past findings.

Second, our paper contributes to a growing strand of literature that explores automation options for consumers to overcome barriers to demand response. There is evidence that automation of smart thermostats can assist in facilitating short-run demand responsiveness when combined with pricing (Harding and Lamarche, 2016; Bollinger and Hartmann, 2020; Blonz et al., 2021). However, consumers may override important settings with such technology, reducing the anticipated benefits (Brandon et al., 2022). Somewhat consistently, we find that our Tech group performs no better than the Manual group in altering electricity consumption in response to events. That is, given the ability to remotely control large appliances as well as automate some aspects of their electricity usage (such as thermostat settings and turning back on EV chargers and water heaters after events), they fare no better than consumers who require a more manual action to change electricity consumption. In our setting, smart assistive technology is not resolving demand-side failures in price coordination.

Third, our paper relates to Fowlie et al. (2021) who look at opt-in vs opt-out default effects at the extensive margin of selecting time-varying electricity pricing plans. Our paper complements this work by focussing on default effects at the intensive margin of the consumption response decision during peak events. Our key contribution beyond the existing literature is the finding of significantly greater responsiveness when consumption reductions are made the default, or passive, action in response to demand response events. Requiring customers to actively take action—even with the provision of technology that makes the associated cost as minimal as remote control with a mobile phone app—is no match for the power of demand response that is managed on the consumer's behalf.

Our analysis proceeds as follows. Section 2 describes our experimental design. In Sections 3 and 4, we describe the data we obtain through the experiment and

[4] See Faruqui and Sergici (2010) and Yan et al. (2018) for surveys of this literature.

provide descriptive statistics of our sample. We then lay out our empirical framework, including our estimation strategies and robustness checks, in Section 5. In Section 6, we present our results, and in Section 7, we conclude.

2 Experimental Design

In this section, we provide an overview of our experimental design. We then summarize our various experimental groups and event categories. We describe our sample selection criteria and recruitment strategy, and we detail how we randomized households into the different experimental groups.

2.1 Overview

We explore how centralized decision-making can overcome barriers to customers expressing demand preferences in electricity markets via a large-scale field experiment. We partnered with a large regulated Canadian electric utility (hereafter referred to as the "Utility") to randomly offer customers one of several programs. One of these programs involves the Utility remotely controlling customers' select devices during "peak events", times in which customers are offered compensation for reducing electricity demand. Households that were offered and accepted each program subsequently comprise each of our treatment groups, and our sample consists of all treatment and control groups.[5]

Notably, we recover causal treatment effects, the average effect of an event household consumption by treatment group, by randomizing events at the household-day-time of day-event "type" level. That is, average treatment effects are recovered by randomizing a treatment (events) *within* treatments groups. Households within each treatment group receive a random, independent schedule of events. In our main specification, identification of our parameters of interest that give us the average household electricity consumption response to events, by group, relies on comparing event-time consumption from households that randomly receive events to non-event consumption from all households that do not receive events at a particular time, conditional on controls. With this design, we do not need to be concerned about selection affecting our identification.

[5] We consider participants in each group as similar to those that would accept many utilities' offers for "demand-side management" programs. Indeed, they give us more conservative estimates of consumer responsiveness to these types of programs, as utilities often select and recruit customers for these programs that they anticipate will be the most responsive.

2.2 Treatment Groups

Table 1 summarizes our experimental groups. The *Central* group is our centralized decision-making group. This group received load controller equipment that is controlled by the Utility to turn off the select devices during "peak events". This equipment was available for several major electricity-consuming devices: electric baseboard heaters, electric hot water tanks, and/or level 2 electric vehicle chargers. This equipment allows for enhanced control of the devices via the Utility's App, allowing households to turn on and off the devices remotely. Peak events (hereafter "events") are times when the Utility offers consumers compensation for reducing electricity consumption, which are framed as "rewards" or "earnings" in our experiment.

Our *Tech* group also received load controller equipment. However, households in this group differ from the Central group in that the Utility does not control the load controller devices during events. Both the Central and Tech groups can remotely control the devices on which they have load controllers. However, while the Utility controls the Central groups' devices on their behalf, the Tech group must take action to turn off or adjust their devices.[6]

The *Manual* group earns financial rewards for demand reductions during events, but does not have load controller equipment. If there is an asymmetric response to events between the Manual and Tech groups, this captures the impact of the load controller technology in facilitating demand reductions.

All participants in the Central, Tech, and Manual groups receive financial compensation for reducing electricity consumption during peak events. The events are randomized at the day-household-event time-event type level.[7] All participants in these groups plus a fourth group, our *Info* group, receive real-time household-level electricity usage information via the App.[8] The Info group is not exposed to events (so they do not receive rewards for reducing electricity consumption) or load controller equipment. This allows us to observe and estimate the impact of events independently from the impact of providing consumers with real-time price information.

[6] In general, the Tech group must respond to event offers in-the-moment by adjusting their devices; however, they do have the capability to pre-set their thermostats and make settings that ensure any device turns back on after events. As described below, all experimental groups receive the first notifications regarding events 21 hours in advance.

[7] As described below in Section 2.3, event times were constrained to two periods in the day in which electricity demand is typically high. Event types reflect two magnitudes of compensation households could receive for demand reduction, during evening events only.

[8] Households that received load-controllers also receive real-time information about electricity consumption from these specific devices on the App.

Finally, we have a *Control* group, a group of Utility customers whose electricity consumption we just observe. This group has never been contacted after they installed the utility's App, and they can only view their electricity usage at a one-day lag through the same App.

Table 1. Summary of Treatment Groups

Groups	DR Control	Control Tech	Price Incentive	Usage Info
Central	Utility	✓	✓	✓
Tech	Household	✓	✓	✓
Manual	Household		✓	✓
Info	Household			✓
Control	Household			

Notes. DR Control represents whether demand response to events is controlled entirely by the household (decentralized) or by the Utility for the load-controlled devices (centralized). Control Tech denotes whether the household has load controller equipment installed. Price Incentive reflects if households receive peak events and rewards for reduced demand during events. Usage Info denotes whether households receive real-time household-level consumption information. ✓indicates categories that are applicable to each group.

2.3 Events

We use peak events to evaluate households' responses to financial incentives, and determine if this response differs across the treatment groups. There are two event times and two event types. Events occur either in the morning (7 AM - 10 AM) or evening (5 PM - 8 PM). We consider two event types that have different reward levels: a "normal" and a "high" peak event. The high peak event receives elevated compensation for larger demand reductions. We designed high peak events for several reasons: (1) Incentives for consumers to reduce electricity consumption, even during peak times, are relatively low on the margin. The peak and high peak event incentive structures capture the range of peak-time pricing that consumers would face if they had real-time pricing. The peak price schedule allows us to observe behavior under common peak pricing, and the high peak schedule gives us a view of consumer behavior under high peak pricing incentives that may occur when electricity demand approaches the maximum supply available. (2) We hypothesized that while the Central group would not need relatively large per-event incentives to make consumption changes, participants in other groups might. (3) Infrequent, large rewards may keep consumers engaged in a peak pricing rewards program. All morning events are normal peak events. Evening events are permitted to be either normal or high-peak events.

Households receive financial rewards for demand reductions relative to their household-specific baseline consumption. Baselines reflect the average consumption during the relevant event time window of the last five weekdays prior to the event. We designed this rolling baseline to evade customer efforts to "game" their baseline by over-consuming during times when peak events occur. Customers do not know how the baseline is calculated.[9]

We consider three reward levels that depend on the reduction in consumption relative to the household's baseline. During normal peak events, households receive \$1 for a 10% reduction, \$2 for a 30% reduction, or \$3 for a 50% reduction. High peak evening events elevate the incentives to \$1 for 10%, \$3 for 30%, or \$6 for 50% reductions. These incentives translate to incentive payments ranging from approximately \$1.18 to \$2.35 per kWh for the average household.[10] These incentives are in the range of wholesale price caps that are used to limit scarcity pricing in a number of jurisdictions in North America.[11]

For households that receive events, we randomly allocate events to households across all weekdays, excluding holidays. Consequently, households are unable to predict when they will be exposed to a peak event. In addition, events are not correlated with other drivers of household electricity consumption. The two event times and types are randomly assigned to these event slots throughout the month. In each month, households experience one to five events, with an average of three events. There is an average of one normal evening event, one high peak evening event, and one morning event per month. This schedule was altered in the summer months of July and August when the likelihood of peak events is low in Canada. During these months, households experience no "high peak" events. Events started on February 22, 2022, and continue for 18 months.

This randomization approach allows us to estimate treatment effects by household, using households' own consumption during non-event times as a control. In

[9] The baseline calculation was adjusted on March 3, 2022 (after 7 treatment days) from the average of the highest 3 of the past 5 weekdays to the average of the last 5 weekdays. This change was made to mitigate the impact of outliers on the baseline calculation.

[10] The average household consumes 1.7 kWh in each hour in our sample. A 10%, 30%, and 50% reduction translates to a 0.51, 1.53, and 2.55 kWh reduction over the three-hour event, respectively. Consequently, for a 1 kWh reduction in consumption, we are compensating households on the lower bound $\frac{3}{2.55} = 1.18$ per kWh (i.e., \$3 for a 2.55 kWh reduction over three hours) for a 50% reduction during a normal peak event to a higher bound of $\frac{6}{2.55} = 2.35$ for a 50% reduction during a high peak event. The other percent reductions lie between these two cases.

[11] Examples include the wholesale price cap of \$1.00/kWh in Alberta (Brown and Olmstead, 2017), \$3.50/kWh in the Mid Continent Independent System Operator that operates in the Midwest United States (IRC, 2017), and \$5/kWh in Texas (Smith, 2022).

contrast to other common experimental approaches in the literature estimating consumer demand responses to electricity prices, we have unique identifying variation: Households do not all receive the events at the same time.

Households that are exposed to events receive event notifications 21 and 2 hours prior to the event. These event notifications provide information on the time of the event and the financial incentives for the different demand reduction levels. When consumers receive the 21-hr notifications, they are also able to see event details in the App itself. See Appendix B.1 for examples of the notification and in-App event messages.

Households' rewards for consumption reduction during events are displayed in the App at a two-to-three day lag. The App also provides households with a summary of their total rewards to date. See Appendix B.2 for details.

2.4 Sample Selection, Recruitment, and Assignment

2.4.1 Phase I Recruitment: Onto the App

The study sample was drawn from the population of residential customers in the Utility's service territory in and near a large metropolitan city. We employed a two-step recruitment strategy. First, starting in August 2021, the Utility invited households to join a home electricity consumption management App operated by a third-party company, in partnership with the Utility. The App provides households with household-level hourly consumption posted at 11 AM the following morning. The App can be coupled with other devices to provide more detailed information on real-time usage and device control. Households were recruited to the App using several marketing strategies, including advertisements on the Utility's website, social media posts, the Utility's newsletter, website notifications when users logged into their Utility accounts, and emails sent to approximately 306,000 residential households.[12]

[12] Emails were sent only to customers who had indicated prior to the recruitment that they could be contacted via email. One portion of the App Recruitment campaign targeted a set of customers in an energy-savings program consisting of regular communication from the Utility regarding energy-saving tips and incentives. We consider these customers sophisticated users. Approximately 64% of the households in our final sample (all groups combined) consists of these sophisticated users (639 out of 1005). The share of these users across groups is fairly even: 70% of each of the Central and Tech groups, 63% of the Manual group, and 57% of the Info and Control groups (combined). We conduct a robustness check when estimating Equation 1 that interacts our coefficients of interest with sophisticated user status. We find no evidence that sophisticated users in the Central or Tech groups have a different response to events than others. Further, we find no evidence that the Manual group participants who are sophisticated users reduce demand more during events than non-sophisticated users. We are not concerned that this set of customers is driving our main treatment effects.

The recruitment onto the Utility's App provided us with a pool of 9,020 households to draw from. When households signed up to join the App, they were required to answer a six question survey. The survey asked households about their motivation for joining the App and whether the household rents or owns their home. It asked about devices eligible for load control in our experiment, including whether the household has an electric hot water tank, an electric vehicle (EV), and electric baseboard heaters as the primary heat source. Households with EVs were asked what type of charger (level 1 or 2) they use. It also asked whether households have air conditioning, a major source of demand flexibility.

We applied several selection criteria to this pool of households. Customers had to be in and near a large metropolitan city in the province for which it was feasible for Utility-partnered electricians to install load control equipment, as needed. Only homeowners were permitted to participate. Condos and apartments were removed, leaving primarily single-family homes, duplexes, and row homes as eligible. Households must have at least one month of historical consumption data as of September 2021, and the customers must have at least one controllable electric device. The set of controllable electric devices include a level 2 electric vehicle charger, electric baseboard heaters used as the primary heat source, and an electric hot water heater tank. This left us with a sample of 1,661 potential households that we used for our randomized assignment to experimental groups.

2.4.2 Phase II Recruitment: Group Offers

We randomized eligible households into our treatment and control groups.[13] Starting in October 2021, we sent group-specific emails to households inviting them to join a new "Trial" program. These emails provided details about the specific experiences households would face in the group to which they were being invited, including a summary of the expected rewards they could earn over the course of the Trial, equipment they would receive and its estimated value, and future peak events. Households

[13] Specifically, we used a randomization procedure designed to balance important observable characteristics over groups. We first used the machine learning algorithm "kmeans" to group households based on observable characteristics. These included cumulative household electricity consumption (in kWh) and load factor by season (Fall, Spring, Winter, and Summer), variables that indicate if a household has an electric vehicle, electric baseboard heating, or air conditioning, and census data on median household income. Load factor is average electricity consumption divided by maximum consumption over a specific time period; it is a way to capture the relative utilization rate of consumption at the household-level. We then randomized group assignment with the condition that households within cluster were balanced across groups.

were also randomly offered a small sign-on incentive of the amounts \$10 or \$20, or no incentive. Recall that all households faced a yes/no decision regarding whether they would accept our group-specific offer. The Control group that receives no equipment, price incentives, nor real-time usage information (recall Table 1) received no further communication beyond joining the App in the first phase of recruitment.

Households had to actively accept the invitation to join the relevant experimental group. After selecting to join, households were mailed a device called the "Hub" that facilitates monitoring real-time energy usage via the App. Households in the Central and Tech groups were contacted by installers to install the load controller equipment.

This two-phase recruitment process occurred over the months August 2021 - February 2022. The second phase of recruitment occurred in fives waves starting in October 2021. As additional households joined the App, we collected the survey responses, identified eligible households, randomized households into groups, and sent the second-phase recruitment emails. This process was used to facilitate the time required to schedule and install the load controllers, as well as to achieve the targeted sample size.

3 Data and Validation of Randomization

3.1 Data Description

Our analysis uses hourly household-level consumption (in kWh) for all households in our experiment. Specifically, we have hourly electricity consumption data starting on October 1, 2020. We complement this household-level data with device-level consumption data on devices that receive load controllers during our trial. These data will allow us to identify changes in consumption behavior during events and evaluate if there are specific changes to the use of particular devices.

We have information on a number of household characteristics that were provided through survey responses as a necessary condition to enter the first phase of our recruitment process. In addition, the Utility provided supplementary household information including the type of household (e.g., single-family/duplex, row home) and whether the household is enrolled in any other Utility programs.

We complement the detailed household-level data with information from the 2016 Canadian Census (Statistics Canada, 2021). We are provided a household's Census Dissemination Area (CDA) identifier; the CDA is the most granular geographical unit for which all Census information is provided publicly. We collect hourly weather information to control for environmental factors that impact electricity consumption, including temperature and humidity at three stations that are geographically representative of the households located in our study.[14] These data were accessed at Environment and Climate Change Canada.

3.2 Validation of Randomization

We evaluate if there are differences in pre-treatment observable characteristics across our various groups to assess the quality of our randomization. Table 2 provides summary statistics by group for a number of variables, including those used in the clustering procedure during randomization (recall Section 2.4). The sample presented in this Table represents all 1,661 households invited to participate in the experiment. For all variables, we report the results of a one-way ANOVA test to evaluate if there are statistical differences in means across the groups.[15]

Table 2 demonstrates that the majority of households in our sample have electric hot water heating and use baseboard heating as the primary heat source, while electric vehicles are less common representing approximately 30% of households. The majority of households are single-family homes or duplexes, with the remainder being primarily row homes. The households consume considerably more electricity during the winter, with the lowest consumption arising during the summer. This demonstrates the potential for larger opportunities for load shifting during these months.

4 Descriptive Results

In this section, we present information on program adoption rates by group. We provide initial descriptive evidence that households reduce their electricity consumption during peak events and demonstrate that this response differs across our treatment groups. We also use detailed device-level data to illustrate how households adjust their load controlled devices during peak events. These descriptive results will be supplemented by a formal econometric model below.

[14] We match the households in our sample with their closest weather station.

[15] The seasonal cumulative consumption and load factor data only contain households with a full year's worth of historical consumption. We computed analogous statistics using only data from September 2021, the month in which all households have complete pre-treatment consumption data. We find no evidence of statistically significant differences in means across the groups using this data.

4.1 Program Adoption and Comparability Between Groups

Table 3 summarizes acceptance rates for each group offer. Acceptance rates among all groups were high. The acceptance rate for the Central offer, 42%, was quite high and only marginally statistically different than the acceptance rate for the Tech offer (48%).[16] The acceptance rates of the Central and Tech groups were lower than the others due to in part to the need for load controllers to be successfully installed in households that accepted these offers.[17]

We take the similarity among final acceptance rates between the Central and Tech groups as the first set of evidence that we can confidentially compare our estimated treatment effects between these groups. While the Manual group had a higher final acceptance rate, concerns that the Manual group systematically differs from the other two groups are mitigated based on a comparison of observable characteristics across groups in our final groups.

Table A1 in the Appendix compares average pre-treatment household characteristics by group for households that were in the final treatment groups. Consistent with Table 2, we observe limited statistically significant differences in these characteristics across groups. One difference across groups is in the proportion of households that live in single-family homes/duplexes. There is a larger proportion of households in this building type in the Manual group than other groups, in particular. We observed differential attrition across groups between households' initial acceptance of group offers and final participation in each group program that was driven by technical challenges of participants connecting the smart Hubs, which transmit real-time meter data to participants' App, to their meters. Participants in row homes, the other house type, could not connect Hubs to meters at a higher rate than participants in detached homes or duplexes. The reason for this is the nearness of one's unit to the meter area in a row home setup, which is likely random in relation to a household's electricity consumption decisions. We also observe a difference across groups in the proportion of households that have EVs, but this difference is only marginally statistically significant. Overall, these results suggest that the balance on observables that arose as a result of the initial randomization remains in the final treated sample.

Finally, during the invitations to join each group, we randomized upfront incentives. While we find a higher rate of initial acceptance with higher upfront incentive

[16] A difference in means test between these two values yielded a p-value of 0.072.

[17] We observed unsuccessful installation at households that initially accepted these offers due to, for example, households never responding to subsequent inquiries to receive and install equipment or households not being in compliance with local electrical codes.

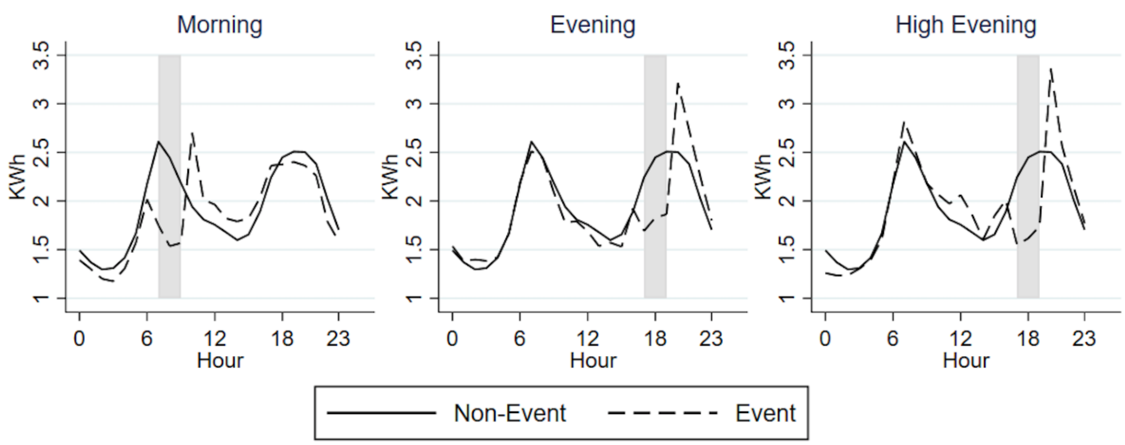
payments, the differences are small and not statistically significantly different.[18]

4.2 Descriptive Consumption Patterns

Figures 1 – 3 provide average hourly household-level consumption for the Central, Tech, and Manual groups for March 2022 during non-event and event days. The shaded regions reflect the relevant event hours. These figures are illustrative of the broader patterns we observe throughout our sample.

Figure 1 demonstrates that the Central group has a large reduction in average consumption during events regardless of the event type. After each event, we observe a large spike in consumption. This "snap-back" is consistent with the devices turning on immediately after the event (e.g., to reheat the water tank and/or home).[19]

Figure 1. Average Household Consumption - Central (March 2022)



Figures 2 and 3 show that the Tech and Manual groups exhibit substantially smaller responses to events. Despite the fact that the Tech group has access to the same equipment as the Central households, its average consumption changes by a considerably smaller amount during events. The Manual group shows a relatively limited change in its average consumption outside of the high evening events for this month. Neither group experience the same snap-back effect immediately following the events. Taken together, these results suggest that the Central group has a considerably larger response to each event type.

[18]Households that received a \$0, \$10, and \$20 upfront incentive accepted the initial invitation with a 63%, 67%, and 68% probability, respectively.

[19]The observed snap-back could be mitigated by the Utility by staggering the beginning and/or end of the load controlled event.

Figure 2. Average Household Consumption - Tech (March 2022)

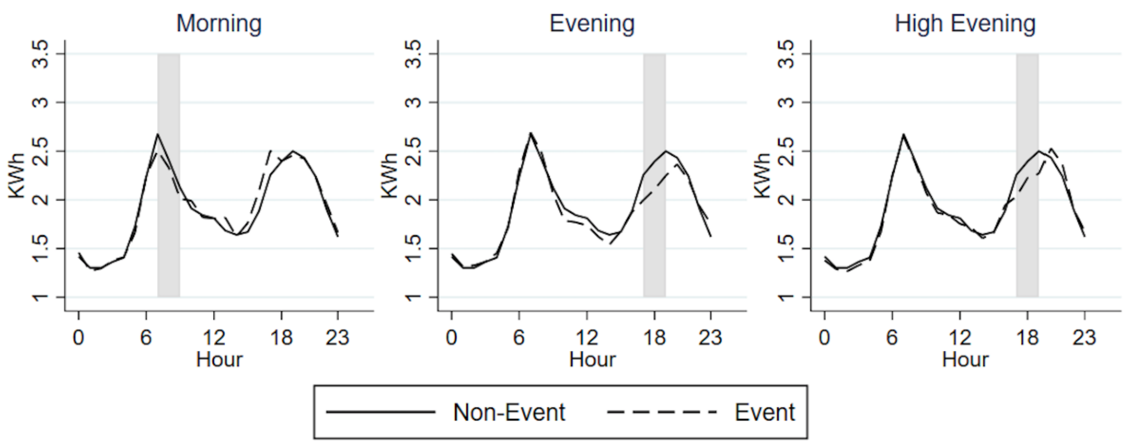
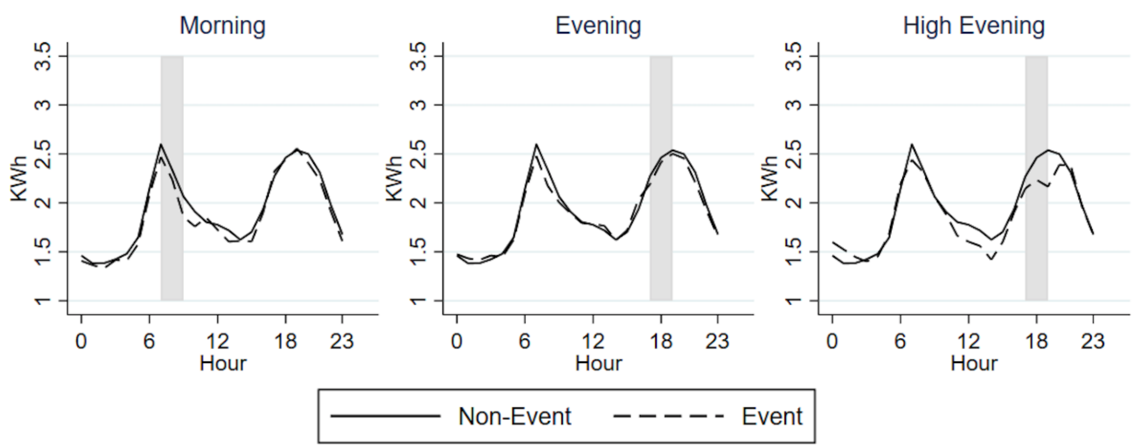
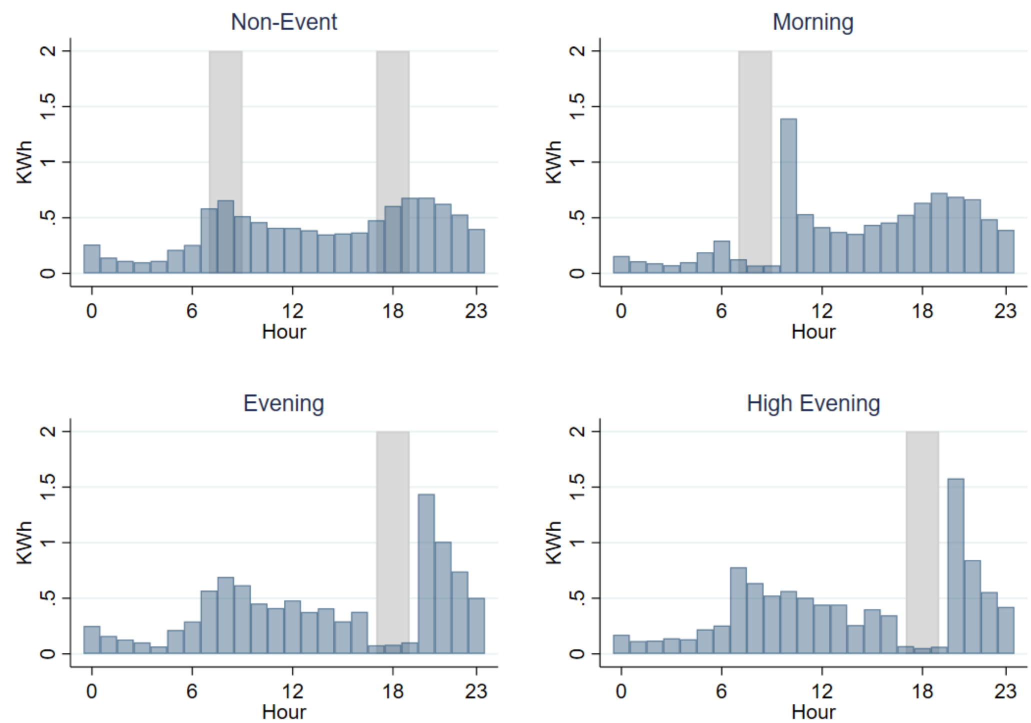


Figure 3. Average Household Consumption - Manual (March 2022)



We use the device-level data for devices that have load controllers to begin to evaluate if households in the Central group allows control of devices and if households in the Central and Tech groups use the devices during events. Figures 4 and 5 present the average hourly consumption of hot water heaters on non-event days and on the three different event day types for the Central and Tech groups in March 2022. We observe a large reduction in average hot water use during the event windows for the Central group. Consistent with the household-level results above, there is a sizable snap-back after the event. Alternatively, Figure 5 provides limited evidence that households in the Tech group are using the hot water heater load controllers during events. It is important to note that for the Tech group, households can adjust their hot water usage during an event by pushing a button in the Utility's App.

Figure 4. Average Hot Water Consumption - Central (March 2022)

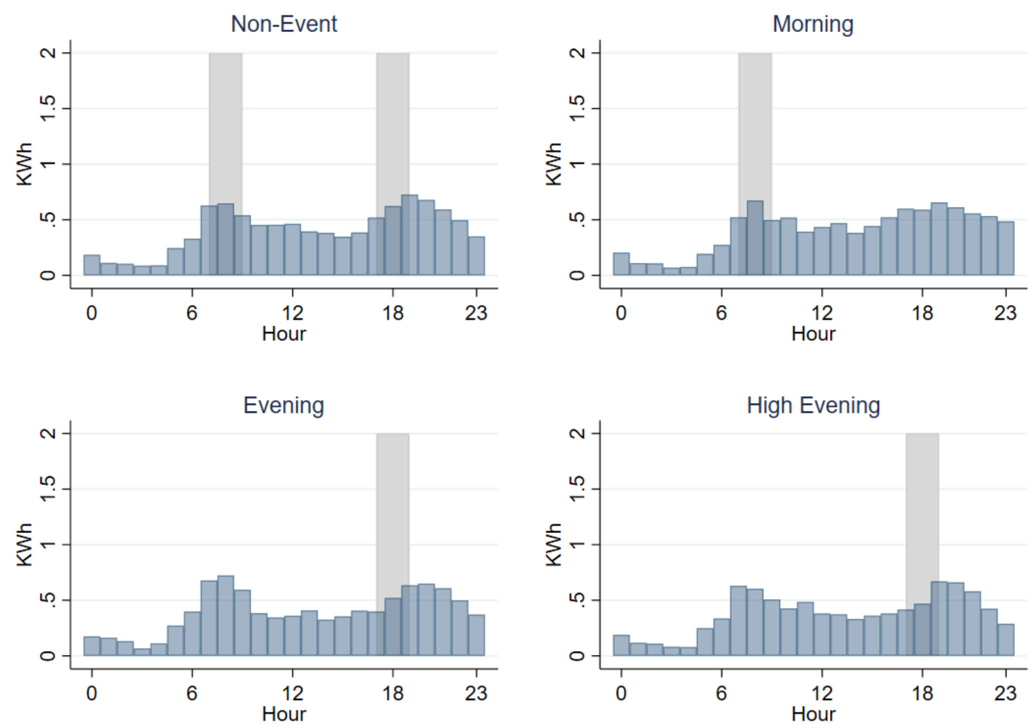


Figures A1 – A4 in the Appendix provide analogous figures for the load controllers installed on electric baseboard heaters and level 2 EV chargers.[20] We continue to observe a distinct reduction in consumption during events for the Central group for both the baseboard heater(s) and level 2 EV chargers.[21] In particular, average consumption from level 2 EV chargers decreases to essentially zero during events. Alternatively, for households in the Tech group, we observe relatively limited differences in consumption patterns on baseboard heaters during events. In contrast, there is some evidence that these households are adjusting their level 2 EV consumption during events. This provides suggestive evidence that these households may be using the EV load controllers.

[20]For the level 2 EV chargers, we present the average hourly EV consumption (i.e., charging) across all months in 2022. We take this approach because of the relatively small number of households that have level 2 EV chargers in our sample. Focusing on one month can lead to higher variability in charging patterns because there is (typically) only one day of each event type for each household.

[21]For the baseboard heaters, the consumption does not decrease to zero during the events. This is driven in part by the fact that the thermostats do not completely turn off these devices during events, but the temperature set points are reduced.

Figure 5. Average Hot Water Consumption - Tech (March 2022)



These descriptive results provide initial evidence that the Central group experiences large reductions in consumption during events. This contrasts with households in the Tech and Manual groups which have considerably smaller adjustments in hourly consumption patterns. The device-level data provide suggestive evidence that the Tech group does not use the full capability of the load controllers to achieve larger demand reductions during events, with the exception of the use of the level 2 EV load controllers. We undertake a formal empirical analysis to quantify these effects and control for potentially confounding factors.

5 Empirical Framework

5.1 Treatment effects by group

We begin by estimating the following specification to identify the average treatment effect of events on electricity consumption, by group, for the population of customers that participate in our experiment.

We model the effect of peak events on a household i's consumption c_{it} (in log kWh) in hour t using the following model:

$$ln(c_{it}) \;=\; \alpha + \beta\, D_i\, E_{it} + \theta_i + \gamma_t + \delta X_t + \varepsilon_{it} \tag{1}$$

where α is a constant, D_i is a vector of treatment group dummies that each equal one if household i is in the Central, Tech, or Manual groups and zero otherwise. E_{it} is the household-specific event indicator that equals one if the household is (randomly) assigned an event in hour t. We use the log of household electricity consumption on the left hand side to normalize the right-skewed variable.[22] We include θ_i, household fixed effects, which control for time-invariant household characteristics. We also include γ_t, a vector of time fixed effects that includes hour, day-of-week, and year-month, which control for time-varying factors that impact consumption. In our main specification, we include household consumption data on all groups (experimental groups and control groups).

Consumer responsiveness to events (especially with thermostat settings) may vary in local weather conditions. To capture this variation, we include X_t, a vector of hourly weather controls that include the relative humidity and cooling degrees and heating degrees above and below 65° F (18.33° C). Since consumer responsiveness may vary in weather conditions in a nonlinear way, we include a flexible functional form with a polynomial up to the third degree for each weather-related covariate. ε_{it} is the error term. We cluster standard errors at the household-level.

Our parameters of interest are β, which measure the change in household-level electricity consumption during peak events for each of the Central, Tech, and Manual groups. Identification of our parameters of interest relies on comparing event-time consumption from households that randomly receive events to non-event consumption from all households (treatment and control households) that do not receive events at a particular time, conditional on average consumption in each hour of day, average household-level consumption over all hours, and other controls. The identifying assumption behind Equation (1) recovering a causal effect of events on consumption is that events are not correlated with other drivers of household electricity consumption. This is met via our randomization of events.[23] As discussed in Section 2.1,

[22] Our results are robust to functional form; we observe similar results with a linear-linear specification.

[23] One may be concerned, for the sake of external validity, that non-event consumption among treatment group households is not a good control for event-time behavior, if participation in a program with events leads people to alter non-event time behavior. To some extent, non-event consumption among treated households changing after people enter such a program is expected and part of the overall program effect we want to capture; in particular, people shifting consumption out

our randomization of events within treatment groups allows us to causally identify treatment effects without employing common selection-correction techniques, such as estimation of local average treatment effects (LATE) parameters.

We report the average marginal effect of an event on households' electricity consumption by group, which is a group-specific function of $\widehat{\beta}$, $f(\widehat{\beta})$. Because of our log-linear specification, $f(\widehat{\beta})$ is a semi-elasticity. We transform this function to report the percentage change in hourly consumption during an event via $100 \times (exp(f(\widehat{\beta})) - 1)$.

We include data from before the events started as well as after (from September 1, 2021 through October 31, 2022), which gives more precision to our estimates. (Recall events started on February 22, 2022.)

5.2 Treatment effects by group: Heterogeneous treatment effects by event type and installed devices

We consider additional specifications to test for the presence of heterogeneous treatment effects across multiple dimensions. First, we consider a specification that interacts our treatment group indicator D_i by a vector of Event Type$_{it}$ indicator variables for each of the morning, evening, and high evening event types. This allows us to evaluate if there are asymmetric responses to events both by the treatment group and event-type. In particular, we anticipate a large response during the high evening events, which offer consumers greater incentives for demand reductions.

Second, we focus on the Central and Tech groups to investigate how the responses to events vary by the installed load controller devices. For these groups, we run separate regressions that interact the event indicator E_{it} with a vector of indicator variables that are device-specific and equal one if the household has a load controller installed on a given device and zero otherwise.[24]

of event windows may lead to shifting consumption to other times. Additionally, we imagine that our program, including the frequency of events, mimics the experience households may have in coming years due to the variety of factors impacting electricity markets. Event-level estimates are net of consumption shifting to other hours, which is ultimately of interest to a utility concerned about event-time consumption only. Additionally, we took several measures customers to prevent customers from "gaming" their baseline to earn rewards which may result in over-estimates of treatment effects: We use only recent data in the baseline and did not communicate the baseline calculation method to consumers. (Note that the random timing of events mitigates gaming as well.) Ultimately, however, we run a version of Equation (1) without the Info and Control groups and obtain similar results. This suggests that non-event consumption of the treatment groups during the months after events started is not appreciably different than that of households who do not receive events at all.

[24]Note that these indicators do not precisely capture consumption reductions due to load controllers on devices; they capture household consumption reductions from households during events with particular installed devices. However, we observe a low rate of opt-outs among the Central

5.3 Intention-to-treat

We estimate a regression specification that allows us to estimate an effect similar to an intention-to-treat (ITT) estimate. A classic ITT estimate is a treatment effect that includes all households (or other units) that were invited to participate in an experiment, yielding an "overall program" effect estimate (Athey and Imbens, 2017). In our setting, this is of particular interest, as we anticipated a lower take-up rate of the Central program offer but, conditional on take-up, large household-level reductions in electricity demand during events. In contrast, we anticipated a larger take-up of the Tech program offer with smaller household-level reductions in electricity consumption during events. The net result of these extensive- and intensive-margin effects determines the ultimate event-level demand flexibility from the pool of customers that were offered each program. This is of particular interest to electric utility companies and balancing authorities which are weighing options for demand flexibility in the changing electricity landscape with system upgrade costs.

Unlike many experimental settings, our treatment (events) are randomized within experimental groups and are not assigned globally by group. Therefore, unlike standard ITT specifications, we cannot have a binary treatment indicator that turns on for all households assigned to a specific experimental group, regardless of whether they accept that program. Because we have periodic events that are randomly assigned to only those households that chose to participate in our trial, we create an analogous environment in our setting for all households invited to each group. To do this, we assign households that were randomized to receive the Central, Tech, and Manual group offers but did not accept the offer a distribution of randomized events that is the same as the households that participated. This creates a new variable $\widehat{E}_{it}$ that includes the observed events for households in our experiment and a synthetic distribution of events for households that were invited to the Central, Tech, or Manual groups but did not accept our offer.

We specify the following estimating equation:

group households: Households opt-out from about 1.8% of device-event observations. Opt-outs, when done, are for thermostats about 88% of the time and for hot water heaters about 9% of the time. Additionally, households in our sample can have a combination of one or more of the eligible load-controlled devices. We run additional specifications where we interact the event indicator with each possible load controller device combination. This allows us to evaluate if there are potential "synergies" associated with having different device combinations. We did not find evidence of such synergies, indicating that the marginal effects we report from this specification can be interpreted as the marginal impact of a household with a given load controller receiving an event on consumption.

$$ln(c_{it}) \; = \; \alpha + \omega \, I_i \, \widehat{E}_{it} + \theta_i + \gamma_t + \delta X_t + \varepsilon_{it} \tag{2}$$

where I_i represents a vector of indicator variables that equal one if household i was invited to participate in the Central, Tech, or Manual groups and zero otherwise. $\widehat{E}_{it}$ is the household-specific event indicator that equals one if household i experiences an event in hour t (or is assigned a synthetic event). This regression is estimated on the full sample of household hourly consumption c_{it}, including the households that did not participate in our experiment. Similar to our main specification in (1), we include household fixed effects, a vector of time fixed effects, and X_t hourly weather controls. We cluster standard errors at the household-level.

5.4 Removing Info and Control Groups

To estimate our parameters of interest, Equation (1) compares event-time consumption to non-event time consumption (conditional on controls). Non-event time consumption includes that from treatment groups when they do no experience events as well as that of the Info and Control groups (who never receive events). One may be concerned that non-event time consumption of the treatment groups may be altered once they start to receive events. Notably, we see no appreciable difference in non-event time consumption between the treatment and control groups when plotting average, hourly non-event time consumption by month for each group. Still, we estimate a version of Equation (1) without the Info and Control groups included. Put another way, only the Central, Tech, and Manual groups are included, and the estimation of treatment effects is from comparing the event to non-event time consumption using treatment group data only. If the results of this estimation strategy are similar to those obtained from our main estimation, it implies that the "control" behavior of treatment group participants during non-event times is similar enough to the control groups' to not be of concern.

6 Empirical Results

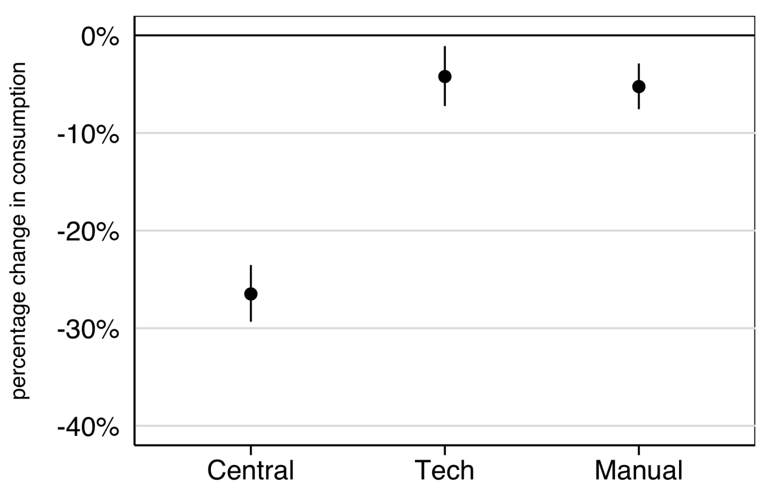
In this section, we present the results of our econometric model. We begin by presenting the average treatment effect for the population of households that participate in our experiment. We then describe heterogeneous treatment effects by event type and installed household devices as well as our ITT results. Finally, we present the results of when we remove the Info and Control households from the sample.

6.1 Treatment effects by group

Figure 6 provides the estimated average response to events by group as a percentage change in household-level consumption. (See Appendix Table A2 for more detail.)

We observe an average 26% reduction in consumption during events for the Central group. In contrast, the Tech and Manual group have an approximate 4% and 5% reduction in demand, respectively. All of these effects are statistically significantly different from zero. Despite the fact that the Tech group has the same equipment as the Central group, it has a statistically significantly lower response to events. In particular, the average response for the Tech and Manual group are not statistically significantly different. These results are consistent with the descriptive data analysis presented in Section 4.2 that suggest that households in the Tech group were not broadly using the load controller equipment to the same extent as the Central group.

Figure 6. Average Treatment Effect of Participants by Group



Notes: The reported results are group-specific marginal effects calculated from estimating $\widehat{\beta}$ in (1). We adjust marginal effects $f(\widehat{\beta})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\beta})) - 1)$. Vertical lines indicate 95% confidence intervals. Standard errors are clustered at the household level.

6.2 Treatment effects by group: Heterogeneous treatment effects by event type and installed devices

Figure 7 presents the estimated response to events allowing for differential responses by event type and group. (See Appendix Table A3 for detailed results.) For the Central group, we see a large demand reduction for all event types, with an approximate 26% reduction during both morning and evening events and a 28% average reduction during high evening events.

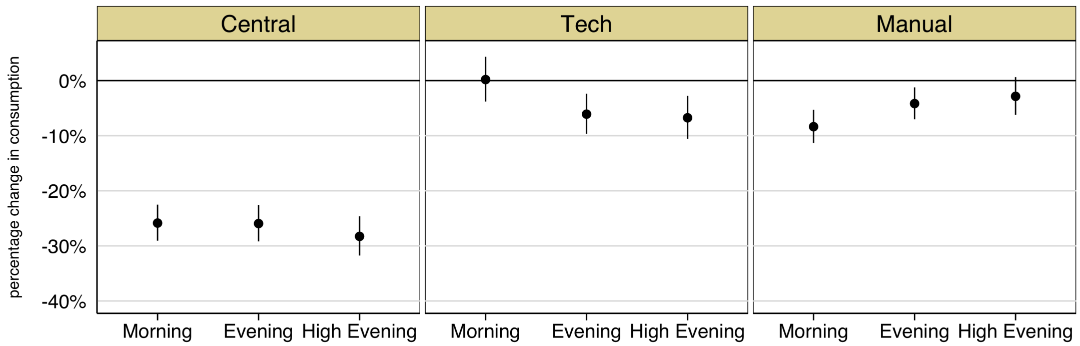
The Tech and Manual groups exhibit different patterns of event-time consumption changes across event type, compared to the Central group. The Tech group has a response to morning events that are not statistically significantly different from zero. This differs (statistically significantly) from the Manual group, which has an average estimated reduction of 8% during the morning events. This is a counter-intuitive result, as the Tech group has all the same information, incentives, and abilities as the Manual group in making electricity consumption reductions during events, with the added ability to remotely control thermostats, EV chargers, and hot water heaters on which they have load controllers installed. The evening and high evening Tech and Manual group effects are not significantly different from each other, when compared within each event type. This suggests that the Tech group's tech is not facilitating a response to the price incentive.

During the evening and high evening events, the Tech group reduces its demand by 6% and 7%, while the Manual group has 4% and 3% estimated reductions during these event types, respectively. These effects are statistically different from zero, except for the Manual group's behavior during high evening events. The Tech and Manual group do not reduce consumption more during high evening than regular evening events, which suggests the increased financial incentives for reduced electricity consumption during these times is not enough for participants in these groups to make greater reductions in usage.

Figure 8 presents the results of the device-specific treatment effects for the Central and Tech groups. We observe no responses to events from households in the Tech group that have load controllers on their hot water heater or thermostat. These results suggest that these households are not using the load controller technology on these devices, during events. We do observe a 7% reduction in consumption among Tech group households with EV load controllers installed. Though this effect is not statistically significant, we observe a subset of high-performing Tech group households with EV chargers that do appear to be reducing consumption during events, but there is high variability in behavior in this group.

Figure 8 shows that the households in the Central group respond to events with statistically significant reductions in consumption across all device types. The largest effect arises from households with load-controlled hot water heaters, resulting in an

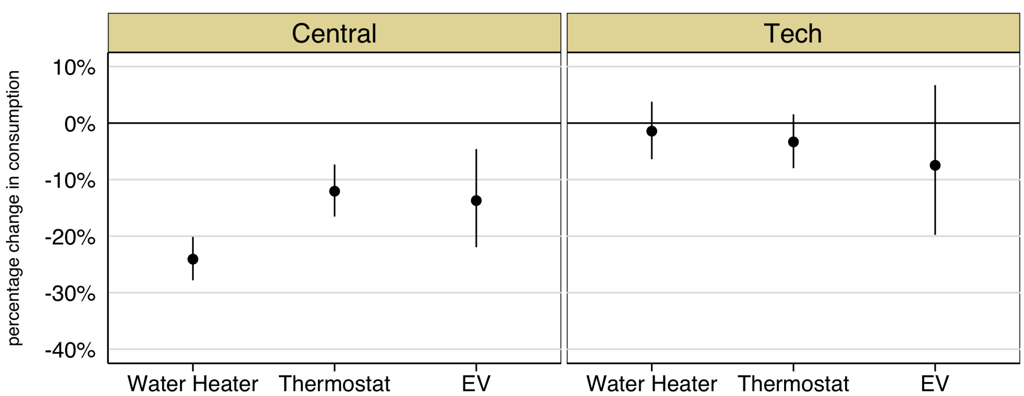
Figure 7. Average Treatment Effect of Participants by Group and Event Type



Notes: The reported results are group- and event type-specific marginal effects calculated from estimating $\widehat{\beta}$ in (1), adjusted to allow for event-type interactions with the group indicator variables D_i. We adjust marginal effects $f(\widehat{\beta})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\beta})) - 1)$. Vertical lines indicate 95% confidence intervals. Standard errors are clustered at the household level.

estimated reduction in consumption of 24%. Households with thermostats and EVs have reductions equal to about 12% and 14%, respectively. The results in Figure 8 are consistent with the descriptive results presented in Section 4.2.

Figure 8. Average Treatment Effect of Participants by Group and Technology

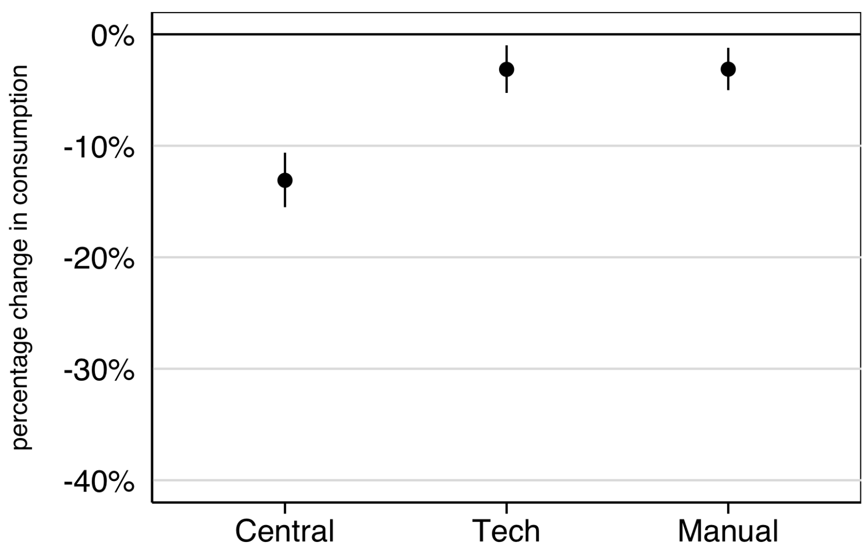


Notes. The reported results are group- and household-type specific marginal effects, obtained from estimating specification (1) run separately for the Central and Tech groups and interacting an indicator for whether a household has each device type with the event indicator E_{it}. We adjust marginal effects $f(\widehat{\beta})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\beta})) - 1)$. Vertical lines indicate 95% confidence intervals. Standard errors are clustered at the household level.

6.3 Intention-to-treat

Overall, we find statistically significant event-time reductions in electricity consumption among households in the Central, Tech, and Manual groups, when including households who declined the offer to participate. Figure 9 shows group-specific effects: On average, households in the Central invitee group have reduced consumption by about 13% during events. (See Tables A5 and A6 for details.) The Tech and Manual invitee groups have reduced consumption during events by about 3% each. These effects are about half as large as the treatment effects estimated from participants only, as one might anticipate from our take-up rates of 42% and 48%, respectively, for the Central and Tech groups. The Central group response is significantly different from the responses of the Tech and Manual groups, as in the effects estimated from participants. The Tech and Manual group responses do not significantly differ from each other.

Figure 9. ITT Results by Group

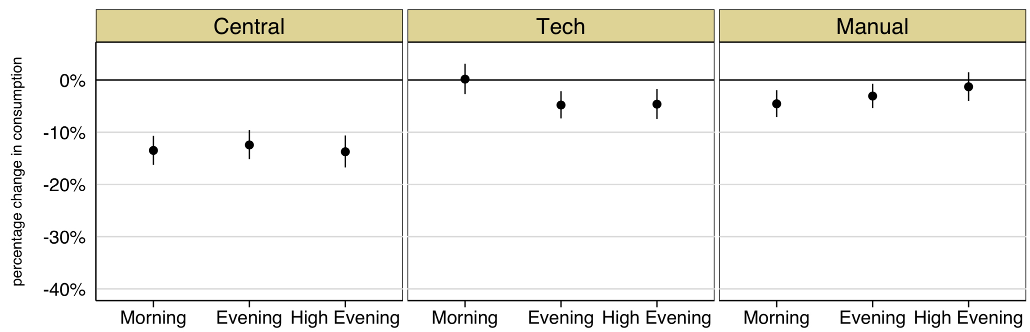


Notes. The reported results reflect group-specific marginal effects computed from estimates of $\widehat{\omega}$ from specification (2). We adjust marginal effects $f(\widehat{\omega})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\omega})) - 1)$. Vertical lines indicate 95% confidence intervals. Standard errors are clustered at the household level.

Figure 10 shows ITT results by group and event type. As with the treatment effect estimates from participants, effects are statistically significant, except for the Tech group during morning events, and the Manual group during high evening events.

The results continue to show the qualitative conclusions and pattern in our main specification, but with reduced magnitudes.

Figure 10. ITT Results by Group and Event Type



Notes. The reported results reflect group- and event-type specific marginal effects computed from estimates of $\widehat{\omega}$ from a version of specification (2) that included event-type interactions with the event indicator $\widehat{E}_{it}$. We adjust marginal effects $f(\widehat{\omega})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\omega})) - 1)$. Vertical lines indicate 95% confidence intervals. Standard errors are clustered at the household level.

6.4 Removing Info and Control Groups

Tables A7 and A8 show the results of estimating Equation (1) with group and group-by-event type indicators, respectively, and calculating marginal effects without the Info and Control groups included in the data. These tables are comparable to Tables A2 and A3. The results are remarkably similar, suggesting that non-event type behavior between these groups (which never receive events) and the Central, Tech and Manual treatment groups are similar. This indicates that using non-event consumption from treatment groups as a control in our main estimation is valid.[25]

[25] We also run a robustness check using the opposite approach: We run independent regressions of Equation (1) with group and group-by-event type indicators using data of only one treatment group (such as the Central group) at a time as well as the Info and Control group data. These regressions identify the coefficients of interest by comparing the sole treatment group's event-time behavior to that of their non-event time behavior and that of the Info and Control groups', conditional on controls. Therefore, treatment effects are estimated using fewer non-event time behavior from treatment groups as in our main estimation. We obtain very similar results. In short, our results are robust to the specific comparison group we use; the non-event-time comparison group has little effect on the main results.

7 Conclusion

We explore the potential for centralized decision-making to resolve consumers' ability to respond to incentives in a relatively small stakes yet complex market environment. Retail electricity markets are one such context, and we conduct a novel, large-scale field experiment in partnership with an electricity Utility company. We find several important results.

First, though we did not expect consumers to be as comfortable with utility-controlled electricity consumption as a consumer-controlled arrangement, we find only a marginally statistically significant difference in the take-up rate of such a utility-controlled demand response program with a comparable program that offers consumers the same consumption reduction incentive structure, technology, and device-specific electricity consumption information but permitting self-directed remote control of device consumption. This surprising result indicates that consumers may recognize the potential that centralized decision-making offers in reducing costs of action.

Second, we find that the Central group in our experiment reduces electricity consumption by a large 26% (on average) when incentivized to do so with "peak events" from their utility company. In contrast, the Tech group only makes, on average, about a 4% reduction in consumption during events. We find that the Tech group responds to events on par with our Manual group, which receives the same incentives as the Central and Tech groups but does not have the ability to remotely control device electricity consumption or view device-specific consumption data. This striking result indicates that the tech adopted by the Tech group provides little, if any, resolution of consumers' perhaps rational inattention to dynamic electricity prices.

The difference in responsiveness to incentives between the Central and Tech groups remains in our intention-to-treat (ITT) estimates (13% vs. 3%), indicating that utility-controlled management of household device electricity use during peak events is the clear program winner for utility companies looking to address potential peak market conditions with demand-side measures. It demonstrates that, when considering program take-up rates and per-event electricity consumption responses by program/group, centralized decision-making in the winner.

We also uncover an important device-specific results: We observe the Central group households with hot water heater load controllers making large (24%) per-event reductions in electricity usage during events, on average. (The Tech group households with hot water heater load controllers, in contrast, do not reduce consumption during events.) We suspect there is “untapped potential” in consumer electricity consumption flexibility that is embedded in similar devices, those with which consumers can be flexible in usage but about which there is little consumer familiarity of electricity usage flexibility. Centralized decision-making may overcome a variety of consumer barriers in shifting usage of these devices, including informational barriers and fears over adjusting their usage.

We interpret the large difference in the response to incentives by the Central and Tech group as indicating that in-the-moment actions needed to respond to the incentives involve large costs for consumers (relative to the benefits of taking action), and centralized decision-making can resolve these costs. In particular, consumers in these groups face a difference in the default electricity consumption of the devices they have with load-controllers: The Central group, by default, has the utility managing these devices' consumption during events (though they can opt-out), and the Tech group, by default, has their normal usage of these devices that they must take action to alter during events. We suspect that “taking the load off” of customers' shoulders with decision-making in settings with large costs of action relative to gains can lead to new choice outcomes, and centralized control is one policy option that accomplishes this in electricity markets.

Table 2. Comparison of Means by Group

	Central	Tech	Manual	Info	Control	ANOVA
Cumul. kWh						
Winter	5,279	5,268	5,442	4,859	5,265	1.29
	(2,694)	(3,032)	(3,076)	(2,748)	(2,950)	
Spring	3,760	3,773	3,818	3,503	3,712	0.87
	(1,924)	(2,112)	(1,911)	(2,116)	(1,974)	
Summer	2,845	2,836	2,708	2,614	2,729	0.77
	(1,742)	(1,872)	(1,539)	(1,861)	(1,710)	
Fall	3,633	3,670	3,700	3,458	3,623	0.60
	(1,663)	(1,945)	(1,974)	(1,796)	(1,860)	
Load Factor						
Winter	24.66	24.98	25.41	24.73	24.67	0.40
	(8.20)	(8.15)	(8.80)	(8.29)	(8.63)	
Spring	19.52	20.12	20.01	19.28	19.91	0.62
	(7.25)	(6.97)	(6.70)	(7.73)	(7.41)	
Summer	16.82	16.55	16.73	16.12	16.32	0.38
	(7.89)	(6.30)	(5.93)	(8.11)	(8.29)	
Fall	18.56	18.90	19.34	18.42	19.06	0.97
	(5.89)	(6.23)	(6.00)	(6.48)	(6.50)	
Electric Vehicle	0.27	0.27	0.27	0.33	0.27	0.99
	(0.44)	(0.45)	(0.45)	(0.47)	(0.45)	
Baseboard Heating	0.61	0.64	0.61	0.63	0.63	0.17
	(0.49)	(0.48)	(0.49)	(0.48)	(0.48)	
Air Conditioning	0.52	0.51	0.50	0.51	0.54	0.17
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	
Electric Hot Water	0.70	0.66	0.70	0.66	0.72	1.06
	(0.46)	(0.47)	(0.46)	(0.47)	(0.45)	
House/Duplex	0.77	0.76	0.81	0.78	0.84	1.61
	(0.42)	(0.43)	(0.39)	(0.42)	(0.37)	
Median Income	86,376	88,291	85,931	87,470	85,948	0.87
	(19,503)	(22,227)	(19,255)	(21,574)	(21,541)	
Households	423	382	409	259	188	

Notes. This table compares pre-treatment average values across the five different groups. Parentheses contain the standard deviations. Cumul. kWh and Load Factor represent the cumulative household-level consumption and load factor by season. The seasonal cumulative consumption and load factor data only contain households with a full year's worth of historical consumption. (See text.) Electric Vehicle, Baseboard Heating, Air Conditioning, and Electric Hot Water are indicator variables denoting the presence of each device. House/Duplex is an indicator variable that equals one if the home type is a single-family home or duplex and zero otherwise. ANOVA reports the F-statistic from one-way ANOVA tests for differences in means across groups. Statistical significance: $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Table 3. Program Acceptance by Group

	Central	Tech	Manual	Info	Control
Invited	423	382	409	259	188
Accepted	177	184	242	177	188
Pct. Accepted	(42%)	(48%)	(59%)	(68%)	(100%)

Notes. "Invited" reflects the number of households invited to participate in the experiment, by group. "Accepted" is the number of households that accepted our offer and made it through equipment installation process (as applicable, by group). "Pct. Accepted" displays acceptance rates relative to the number of households invited.

References

Athey, Susan and Guido W Imbens, "The Econometrics of Randomized Experiments," in "Handbook of Economic Field Experiments," Vol. 1, Elsevier, 2017, pp. 73–140.

Balaraman, Kavya, "The Power Grid Faced Heat Waves, Record Demand and Tight Conditions in 2022. What Happens Next?," *Utility Dive*, 2022.

Blonz, Joshua, Karen Palmer, Casey J Wichman, Derek C Wietelman et al., "Smart Thermostats, Automation, and Time-Varying Prices," 2021.

Bollinger, Bryan K and Wesley R Hartmann, "Information vs. Automation and Implications for Dynamic Pricing," *Management Science*, 2020, *66* (1), 290–314.

Brandon, Alec, Christopher M Clapp, John A List, Robert D Metcalfe, and Michael Price, "The Human Perils of Scaling Smart Technologies: Evidence from Field Experiments," Technical Report, National Bureau of Economic Research 2022.

Brown, David P and Derek EH Olmstead, "Measuring Market Power and the Efficiency of Alberta's Restructured Electricity Market: An Energy-Only Market Design," *Canadian Journal of Economics*, 2017, *50* (3), 838–870.

DTE Energy, "DTE Energy Smart Charge Program," 2022.

Faruqui, Ahmad and Sanem Sergici, "Household Response to Dynamic Pricing of Electricity: A Survey of 15 Experiments," *Journal of regulatory Economics*, 2010, *38* (2), 193–225.

Fowlie, Meredith, Catherine Wolfram, Patrick Baylis, C Anna Spurlock, Annika Todd-Blick, and Peter Cappers, "Default effects and follow-on behaviour: Evidence from an electricity pricing program," *The Review of Economic Studies*, 2021, *88* (6), 2886–2934.

Harding, Matthew and Carlos Lamarche, "Empowering Consumers Through Data and Smart Technology: Experimental Evidence on the Consequences of Time-of-Use Electricity Pricing Policies," *Journal of Policy Analysis and Management*, 2016, *35* (4), 906–931.

IRC, "Market Design Executive Summary: 2017 IRC Markets Committee," Technical Report, ISO/RTO Council 2017.

King, Carey W, Josh D Rhodes, Jay Zarnikau, Ning Lin, Erhan Kutanoglu, Benjamin Leibowicz, Dev Niyogi, Varun Rai, Surya Santoso, David Spence et al., "The Timeline and Events of the February 2021 Texas Electric Grid Blackouts," *The University of Texas Energy Institute*, 2021, *2.*

PG&E, "SmartAC Switches," 2023. https://www.pge.com/en_US/residential/save-energy-money/savings-solutions-and-rebates/smart-ac/smartac-switch.page. Accessed: 2-27-2023.

Smith, Jordan, "Texas Regulators Lower Electric Price Cap to $5,000 per Megawatt Hour," 2022. https://www.chooseenergy.com/news/article/texas-regulators-lower-electric-price-cap/.

Spector, Julia, "Portland General Electric Jumps Into Virtual Power Plant Business," 2020. https://www.greentechmedia.com/articles/read/portland-general-electric-launches-525-home-battery-fleet-virtual-power-plant.

Statistics Canada, "Census of Population. Data Products, 2016 Census," 2021. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/index-eng.cfm, Accessed: 2021-07-15.

Wattersaver, "WatterSaver Program," 2023. https://www.watter-saver.com/. Accessed: 2-27-2023.

Yan, Xing, Yusuf Ozturk, Zechun Hu, and Yonghua Song, "A Review on Price-Driven Residential Demand Response," *Renewable and Sustainable Energy Reviews*, 2018, *96*, 411–419.

A Supplementary Tables and Figures

Table A1 shows observables across groups for the final set of households in each group.

Table A1. Comparison of Means by Group - Final Accepted Households

	Central	Tech	Manual	Info	Control	ANOVA
Cumul. kWh						
Winter	5,507	5,302	5,422	5,037	5,265	0.54
	(2,706)	(2,737)	(3,240)	(2,768)	(2,950)	
Spring	3,900	3,739	3,797	3,642	3,712	0.34
	(1,934)	(1,791)	(1,939)	(2,159)	(1,974)	
Summer	2,851	2,672	2,766	2,702	2,729	0.22
	(1,869)	(1,759)	(1,547)	(1,849)	(1,710)	
Fall	3,754	3,550	3,677	3,547	3,623	0.32
	(1,733)	(1,659)	(1,992)	(1,788)	(1,860)	
Load Factor						
Winter	24.62	25.56	24.93	24.93	24.67	0.26
	(8.68)	(8.21)	(9.04)	(7.76)	(8.63)	
Spring	19.33	20.48	19.80	19.59	19.91	0.52
	(7.43)	(6.30)	(6.45)	(7.05)	(7.41)	
Summer	16.33	16.87	16.95	16.61	16.32	0.25
	(8.54)	(6.02)	(5.95)	(7.80)	(8.29)	
Fall	18.17	18.97	19.11	18.53	19.06	0.61
	6.27	5.78	6.29	6.11	6.50	
Electric Vehicle	0.25	0.21	0.30	0.34	0.27	2.20*
	(0.43)	(0.41)	(0.46)	(0.47)	(0.45)	
Baseboard Heating	0.68	0.70	0.60	0.59	0.63	1.83
	(0.47)	(0.46)	(0.49)	(0.49)	(0.48)	
Air Conditioning	0.46	0.46	0.51	0.51	0.54	0.99
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	
Electric Hot Water	0.75	0.74	0.68	0.65	0.72	1.66
	(0.43)	(0.44)	(0.47)	(0.48)	(0.45)	
House/Duplex	0.82	0.77	0.89	0.84	0.84	2.89**
	(0.39)	(0.42)	(0.32)	(0.37)	(0.37)	
Median Income	84,978	88,274	86,718	89,504	85,948	1.41
	(19,647)	(20,432)	(19,494)	(21,079)	(21,541)	
Households	177	184	242	177	188	

Notes. This table compares pre-treatment average values across the five different groups for households that were in our final treatment groups. Parentheses contain the standard deviations. Cumul. kWh and Load Factor represents the cumulative household-level consumption and load factor by season. Electric Vehicle, Baseboard Heating, Air Conditioning, and Electric Hot Water are indicator variables denoting the presence of each device. House/Duplex is a indicator variable if the home type is a single-family home or duplex. ANOVA reports the F-statistic from one-way ANOVA tests for differences in means across groups. Statistical Significance $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Figure A1. Average Baseboard Heater Consumption - Central (March 2022)

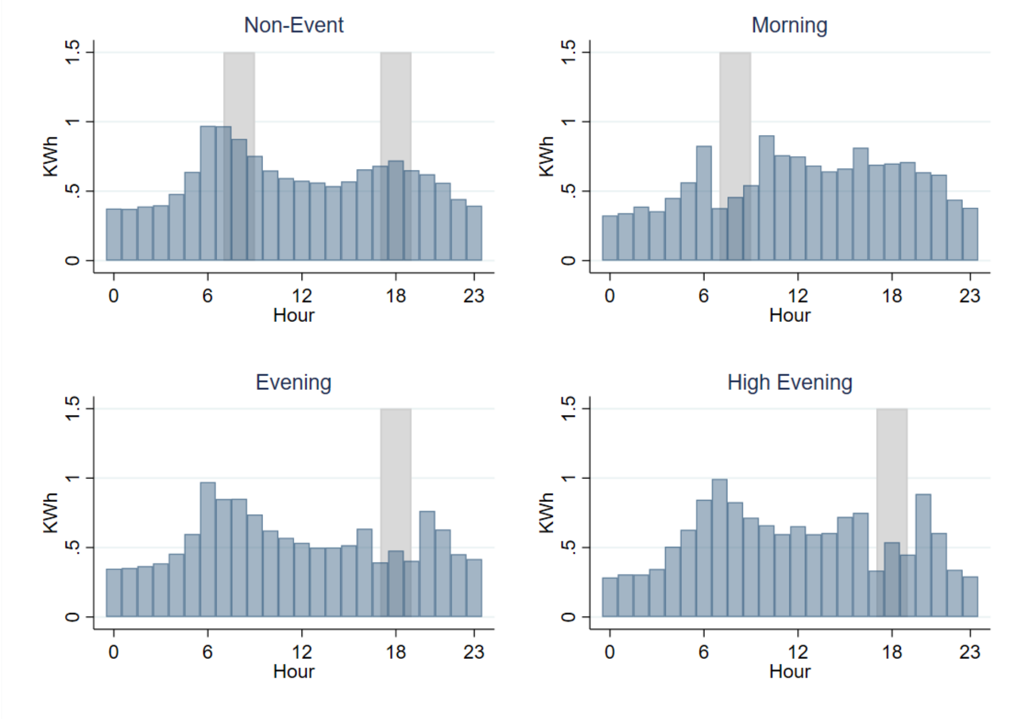


Figure A2. Average Baseboard Heater Consumption - Tech (March 2022)

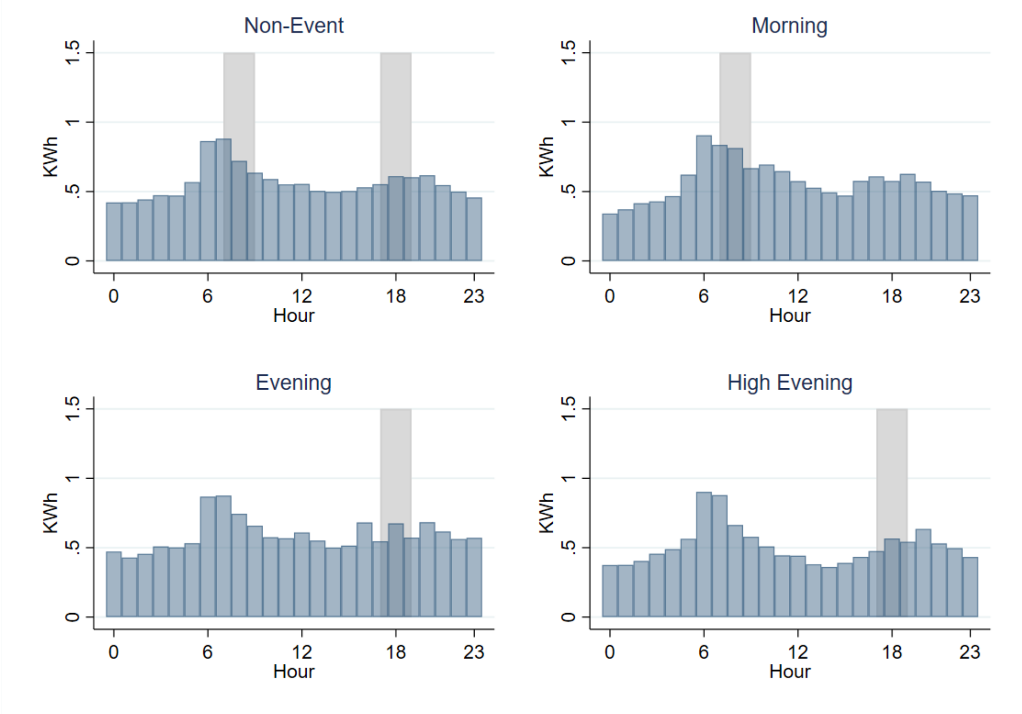


Figure A3. Average Level 2 EV Consumption - Central (Year 2022)

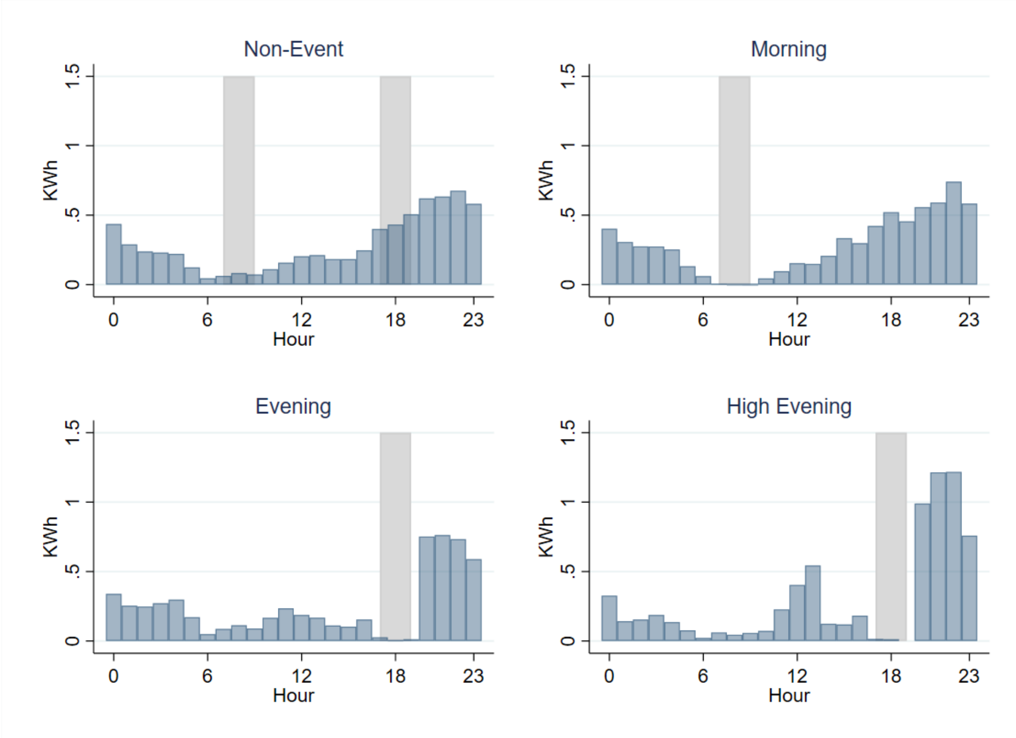


Figure A4. Average Level 2 EV Consumption - Tech (Year 2022)

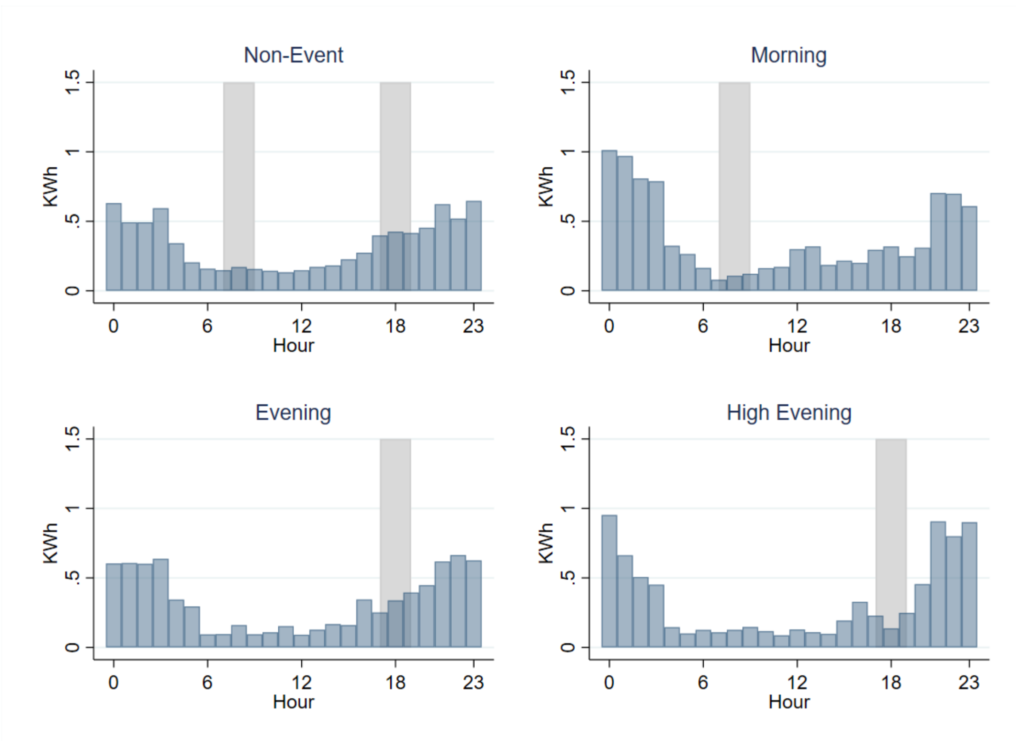


Table A2. Regression Results: Event - Group Interactions

	Marginal Effect	Standard Error	P-Value	Marginal Effect (%)
Central	-0.3078	0.0202	0.0000	-26.4903
Tech	-0.0431	0.0164	0.0085	-4.2192
Manual	-0.0539	0.0126	0.0000	-5.2511
Adj. R^2	0.4751			
N	11,716,985			

Notes. The reported results are group-specific marginal effects calculated from estimating $\widehat{\beta}$ in (1), with their associated standard errors and p-values to indicate statistical significance. Standard errors are clustered at the household level. We adjust marginal effects $f(\widehat{\beta})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\beta})) - 1)$. Adj. R^2 reflects the adjusted R-squared value and N denotes the number of observations.

Table A3. Regression Results: Event Type - Group Interactions

		Marginal Effect	Standard Error	P-Value	Marginal Effect (%)
Central	Morning	-0.2992	0.0225	0.0000	-25.8593
	Evening	-0.3004	0.0228	0.0000	-25.9477
	High Evening	-0.3323	0.0254	0.0000	-28.2749
Tech	Morning	0.0018	0.0207	0.9293	0.1837
	Evening	-0.0628	0.0198	0.0015	-6.0901
	High Evening	-0.0699	0.0213	0.0011	-6.7507
Manual	Morning	-0.0874	0.0168	0.0000	-8.3655
	Evening	-0.0426	0.0154	0.0058	-4.1714
	High Evening	-0.0290	0.0180	0.1072	-2.8536
Adj. R^2	0.4751				
N	11,716,985				

Notes. The reported results are group- and event type-specific marginal effects calculated from estimating $\widehat{\beta}$ in (1), adjusted to allow for event-type interactions with the group indicator variables D_i. The associated standard errors and p-values are reported to indicate statistical significance. Standard errors are clustered at the household level. We adjust marginal effects $f(\widehat{\beta})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\beta})) - 1)$. Adj. R^2 reflects the adjusted R-squared value and N denotes the number of observations.

Table A4. Regression Results: Event Type - Group - Device Interactions

		Marginal Effect	Standard Error	P-Value	Marginal Effect (%)
Central	Thermostat	-0.1284	0.0267	0.0000	-12.0499
	EV	-0.1474	0.0512	0.0042	-13.7053
	Water Heater	-0.2753	0.0258	0.0000	-24.0670
Adj. R^2	0.4727				
N	6,511,846				
		Marginal Effect	Standard Error	P-Value	Marginal Effect (%)
Tech	Thermostat	-0.0338	0.0251	0.1791	-3.3243
	EV	-0.0776	0.0728	0.2866	-7.4674
	Water Heater	-0.0144	0.0263	0.5844	-1.4288
Adj. R^2	0.4765				
N	6,629,249				

Notes. The reported results are group by event by device marginal effects calculated from estimating $\widehat{\beta}$ in (1), adjusted to allow for device-type interactions with the group indicator variables D_i. The regressions are run separately for the Central and Tech groups compared to the Info and Control groups. The associated standard errors and p-values are reported to indicate statistical significance. Standard errors are clustered at the household level. We adjust marginal effects $f(\widehat{\beta})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\beta})) - 1)$. Adj. R^2 reflects the adjusted R-squared value and N denotes the number of observations.

Table A5. Regression Results: Event - Group Interactions - Intention-to-Treat

	Marginal Effect	Standard Error	P-Value	Marginal Effect (%)
Central	-0.1404	0.0144	0.0000	-13.0973
Tech	-0.0319	0.0112	0.0046	-3.1442
Manual	-0.0318	0.0100	0.0015	-3.1288
Adj. R^2	0.4760			
N	17,750,713			

Notes. The reported results are group-specific marginal effects computed from estimates of $\widehat{\omega}$ from specification (2), as well as their associated standard errors and p-values to indicate statistical significance. Standard errors are clustered at the household level. We adjust marginal effects $f(\widehat{\omega})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\omega})) - 1)$. Adj. R^2 reflects the adjusted R-squared value and N denotes the number of observations.

Table A6. Regression Results: Event Type - Group Interactions - Intention-to-Treat

		Marginal Effect	Standard Error	P-Value	Marginal Effect (%)
Central	Morning	-0.1449	0.0164	0.0000	-13.4849
	Evening	-0.1328	0.0162	0.0000	-12.4375
	High Evening	-0.1478	0.0181	0.0000	-13.7418
Tech	Morning	0.0017	0.0148	0.9100	0.1674
	Evening	-0.0491	0.0140	0.0005	-4.7936
	High Evening	-0.0473	0.0153	0.0021	-4.6245
Manual	Morning	-0.0466	0.0137	0.0007	-4.5552
	Evening	-0.0312	0.0123	0.0111	-3.0729
	High Evening	-0.0130	0.0141	0.3566	-1.2948
Adj. R^2	0.4760				
N	17,750,713				

Notes. The reported results reflect group- and event-type specific marginal effects computed from estimates of $\widehat{\omega}$ from a version of specification (2) that included event-type interactions with the event indicator $\widehat{E}_{it}$. The associated standard errors and p-values are reported to indicate statistical significance. Standard errors are clustered at the household level. We adjust marginal effects $f(\widehat{\omega})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\omega})) - 1)$. Adj. R^2 reflects the adjusted R-squared value and N denotes the number of observations.

Table A7. Regression Results: Event - Group Interactions - Removing Info and Control

	Marginal Effect	Standard Error	P-Value	Marginal Effect (%)
Central	-0.3133	0.0201	0.0000	-26.8999
Tech	-0.0487	0.0160	0.0025	-4.7576
Manual	-0.0596	0.0122	0.0000	-5.7871
Adj. R^2	0.4764			
N	7,344,863			

Notes. The reported results are group-specific marginal effects calculated from estimating $\widehat{\beta}$ in (1), with their associated standard errors and p-values to indicate statistical significance. Standard errors are clustered at the household level. We adjust marginal effects $f(\widehat{\omega})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\omega})) - 1)$. Adj. R^2 reflects the adjusted R-squared value and N denotes the number of observations.

Table A8. Regression Results: Event Type - Group Interactions - Removing Info and Control

		Marginal Effect	Standard Error	P-Value	Marginal Effect (%)
Central	Morning	-0.2993	0.0225	0.0000	-25.8673
	Evening	-0.3092	0.0225	0.0000	-26.5985
	High Evening	-0.3398	0.0253	0.0000	-28.8069
Tech	Morning	0.0017	0.0198	0.9304	0.1728
	Evening	-0.0718	0.0194	0.0002	-6.9261
	High Evening	-0.0775	0.0209	0.0002	-7.4577
Manual	Morning	-0.0876	0.0160	0.0000	-8.3861
	Evening	-0.0514	0.0149	0.0006	-5.0144
	High Evening	-0.0366	0.0173	0.0344	-3.5975
Adj. R^2	0.4764				
N	7,344,863				

Notes. The reported results are group- and event type-specific marginal effects calculated from estimating $\widehat{\beta}$ in (1), adjusted to allow for event-type interactions with the group indicator variables D_i. The associated standard errors and p-values are reported to indicate statistical significance. Standard errors are clustered at the household level. We adjust marginal effects $f(\widehat{\omega})$ to be a percentage change in consumption using the transformation $100 \times (exp(f(\widehat{\omega})) - 1)$. Adj. R^2 reflects the adjusted R-squared value and N denotes the number of observations.

B Treatment details

B.1 Group-specific event notifications

Each treatment group experienced event notifications tailored to their treatment. We describe these below. Each group received a notification 21 hours before an event as well as 2 hours before the event. As a function of the App company's load control system, the Central group was sent an additional, generic notification at 2 hours before the event.

All participants are shown a short notification, according to their device and in-app notification settings. If participants touch and press the notification, they are shown the long notification specific to their group, featured below, with event incentive details.)

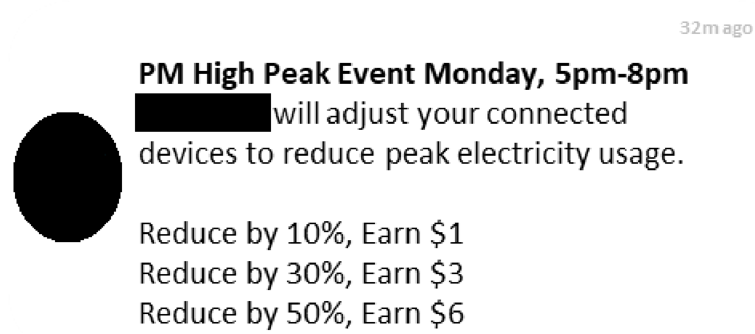


Figure C1. Long Notification for Central group

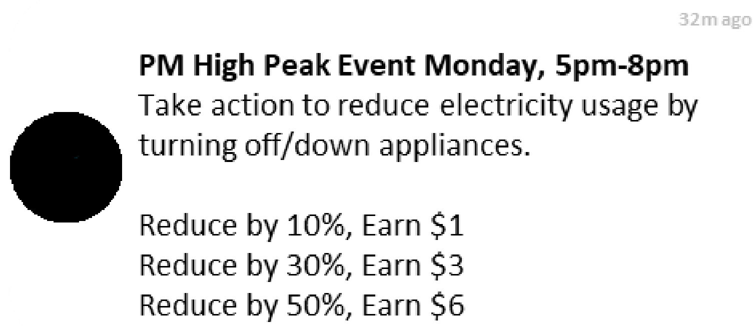


Figure C2. Long Notification for Tech group

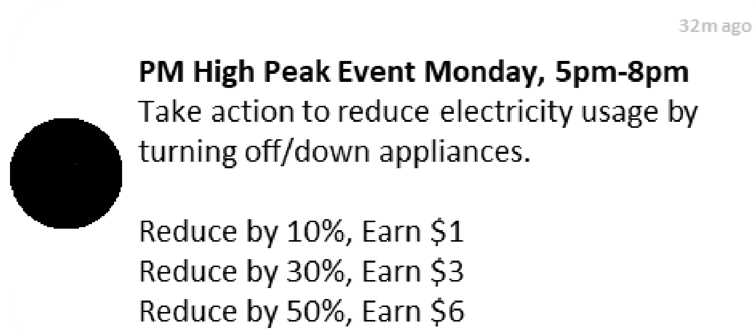


Figure C3. Long Notification for Manual group

Note that all group participants in the three groups below are able to locate event details in the "Advisor" tab of the App, a centralized location for information from the App company, once they receive an event notification. The "Learn More" button at the bottom right of this information card takes participants to the "FAQs" section of the group-specific experiment website.

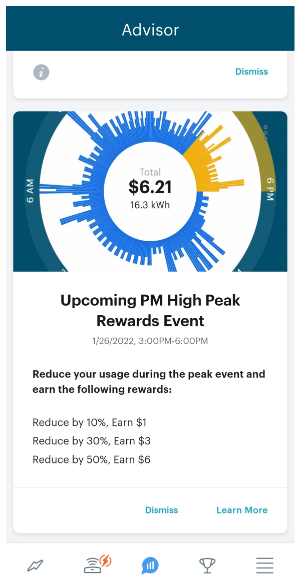


Figure C4. Event info in App

B.2 Treatment group-specific app functionality

Each group in our experiment had an App experience and functionality that differed according to their group assignment. We detail that here and walk through how participants in each group could have responded to peak events, given the options in the App.

B.2.1 Central Group

The Central group participants receive 21-hr and 2-hr notifications regarding upcoming events, as described in Appendix B.1. These notifications allow them to see the timing of the event and the magnitude of rewards for electricity consumption reductions. They also remind participants that their devices with load controllers would be altered by the Utility to reduce consumption, unless they opted-out of the event.

There are several ways that Central group participants can opt-out of events. Before an event starts, they can push an "Opt-out" button in the "My Devices" tab of the App. (This tab is a central App location that allows App users to remotely control devices that have load controllers and see the individual electricity consumption of those devices.) This button removes the participant from the event globally by removing all of their load-controlled devices from the event.

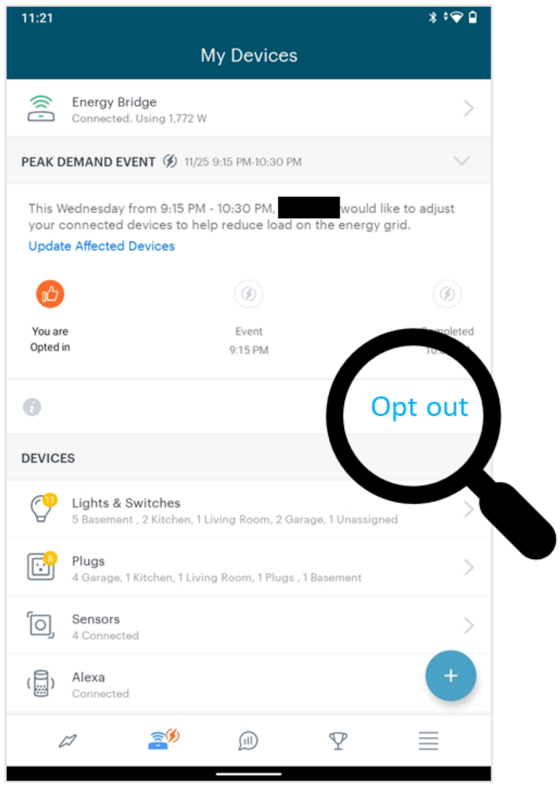


Figure C5. Central group Opt-out functionality

If they do not opt-out in this way, they see a series of screens in the "My Devices" tab. These indicate the progression of the event to the participant and signal when their devices' electricity consumption is being controlled by the utility. The first screen has an orange icon above the text "You are Opted in", as shown below in Figure C6. This occurs before an event starts. When the event is underway and participants' devices are being controlled, they see the icon above "Event" turn orange. The icon above "Completed" turns orange after an event is completed.

During an event, participants can cancel Utility device control in a device-specific way. For EV chargers and hot water heaters, they can remotely opt-out their device from being controlled, or they can physically turn off the load controller at the device

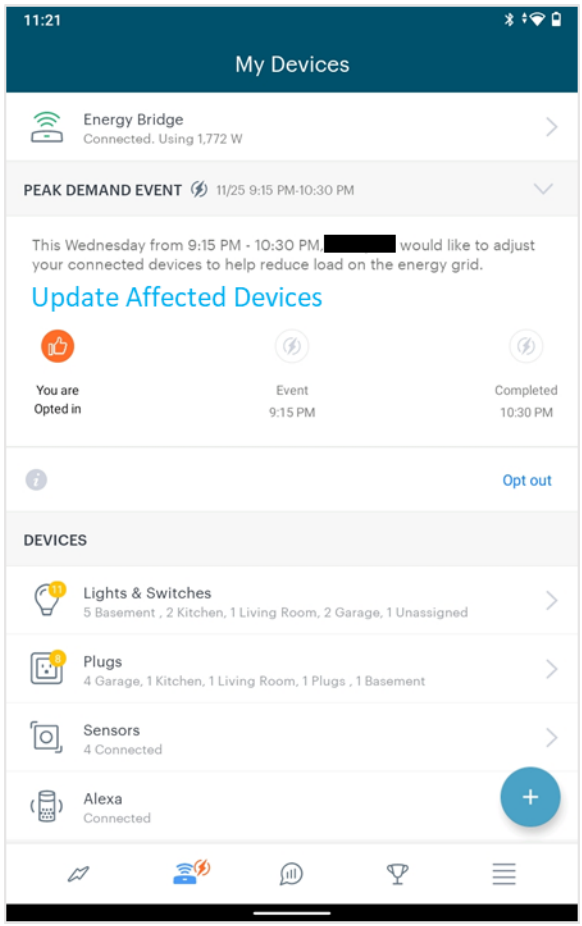


Figure C6. Central group event experience itself. For thermostats, participants can opt-out of load control by adjusting them physically or remotely through the App, during an event.

Note that the Central group has remote and manual control of all devices with load controllers, just like the Tech group. Central group households can also change anything else in the house to alter their electricity consumption during events.

B.2.2 Tech Group

The Tech group participants receive 21-hr and 2-hr notifications regarding upcoming events, as described in Appendix B.1. These notifications allow them to see the timing of the event and the magnitude of rewards for electricity consumption reductions. They also remind participants that they need to "take action" to make consumption changes to receive the rewards offered.

The Tech group can remotely control any device that has an installed load controller through the App. Tech group participants can "opt-in" these devices, in ways that differ by device. For EV chargers and hot water heaters, they can turn them off via two clicks from the My Devices section of the App. (See Figure C7 below for the instructions sent to participants that explain these actions.) Tech group participants cannot make a schedule to turn off these devices before events start and must turn them off before or during events to reduce consumption this way. (They must also remember to turn them on unless they set up a turn-on schedule.)

For thermostats, the Tech group can set up schedule for their thermostat set-point before events, using the App. They can also adjust their thermostats remotely during events with the App.

B.2.3 Manual Group

The Manual group participants receive 21-hr and 2-hr notifications regarding upcoming events, as described in Appendix B.1. These notifications allow them to see the timing of the event and the magnitude of rewards for electricity consumption reductions. They also remind participants that they need to "take action" to make consumption changes to receive the rewards offered.

Manual group participants do not load controllers given to them as part of this experiment or Utility control of any devices. They therefore only observe these notifications as well their aggregate, real-time household consumption through the App. If Manual group participants install their own smart home devices, they may be able to link them to the smart electricity consumption technology ecosystem used in this

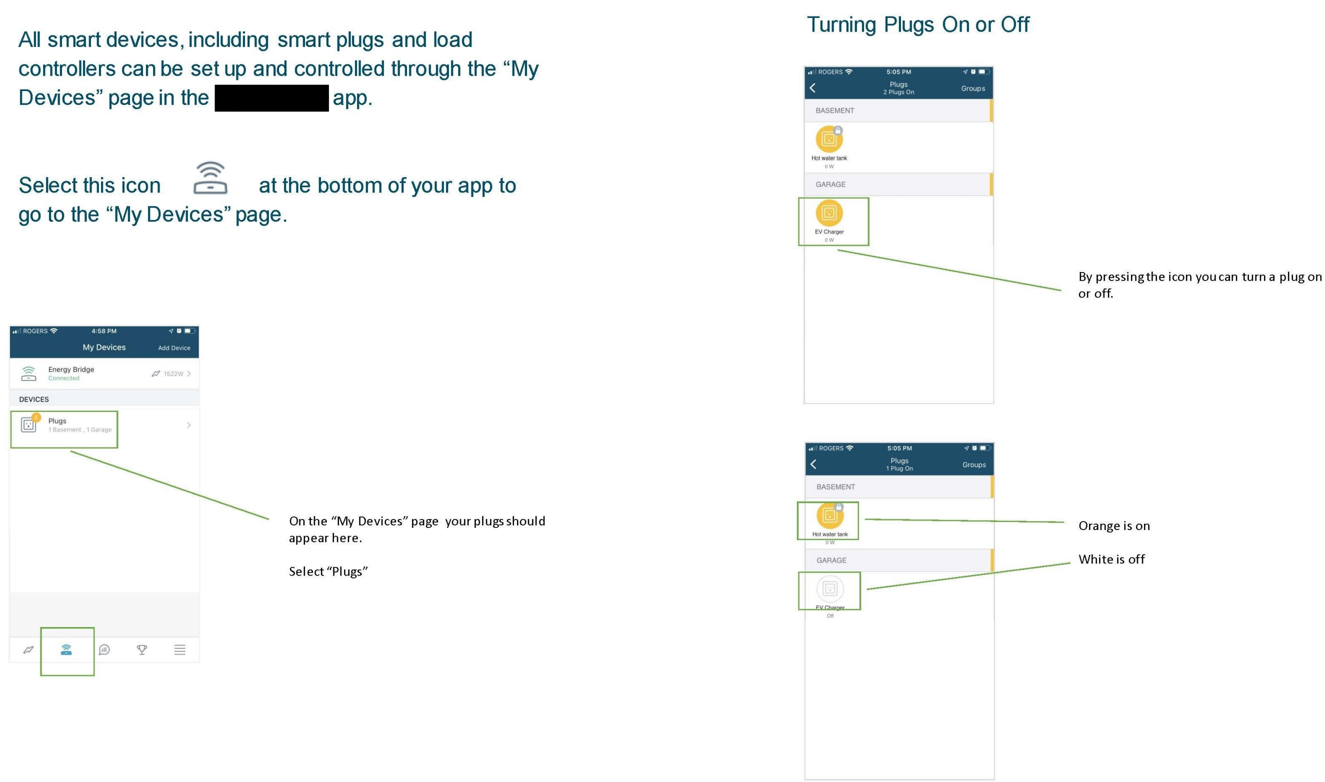


Figure C7. Controller guide for Tech group experiment. If so, they may have the capabilities of the Tech group to observe the real-time consumption of those devices/devices individually and adjust them remotely through the App. (So far we only see two households in the Manual Group that have done this, with smart thermostats.)

B.2.4 Central, Tech, and Manual groups

After each event, all three of the Central, Tech, and Manual groups receive a result on their performance, as depicted below. This appears in the "Advisor" tab of the App, a central location for information from the App company. This result card reminds participants of the event type (reward magnitudes being "high" or not) and the day and time of the event. It shows the incremental reward the participant earned from the event as well as their cumulative rewards throughout the entire experiment, including the reward from the prior event. The text below the reward for the last event is variable and depends on whether a participant met one of the reward tiers. The rewards screen with one of these text options is shown below in Figure C8.

From this rewards screen, participants can select "Event History" and see their recent history of event rewards, as shown in Figure C9.

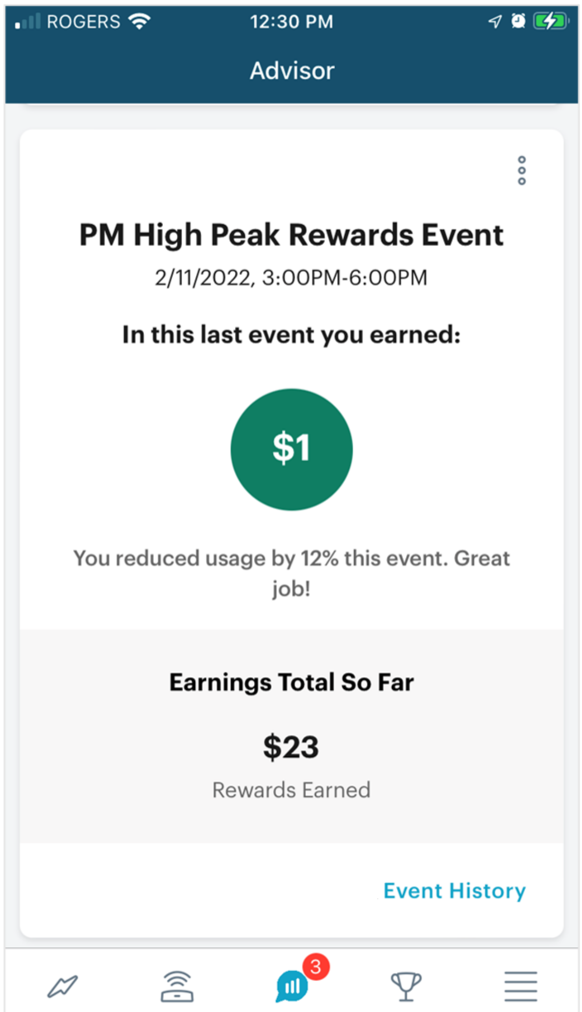


Figure C8. Rewards screen

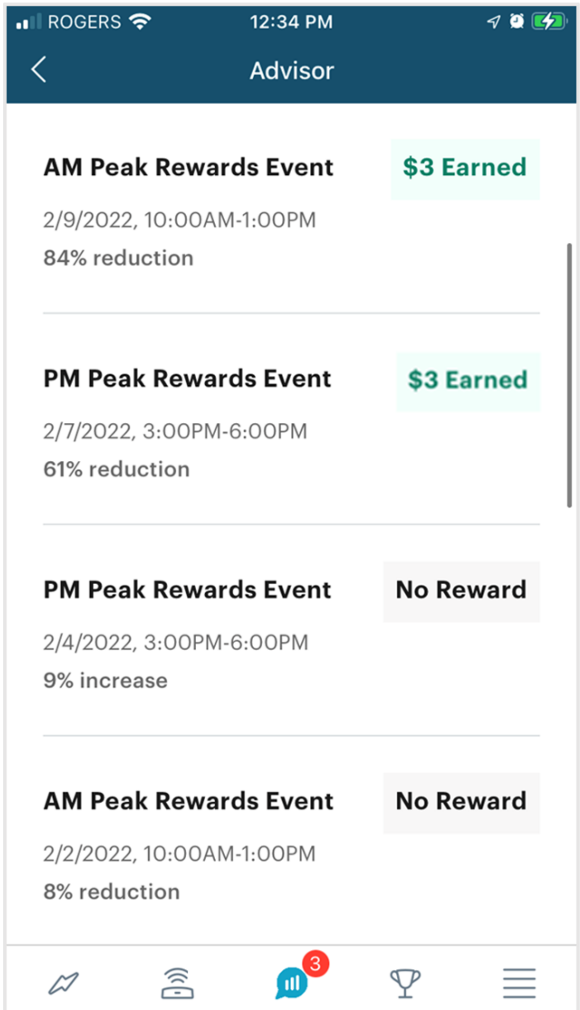


Figure C9. Event history